Medicina Sportiva
Med Sport 16 (3): 122-130, 2012
DOI: 10.5604/17342260.1011393
ICID: 1011393

CRITICAL REVIEW OF A META-ANALYSIS FOR THE EFFECT OF SINGLE AND MULTIPLE SETS OF RESISTANCE TRAINING ON STRENGTH GAINS

Ralph N. Carpinelli

Human Performance Laboratory, Adelphi University, Garden City, New York, USA

Abstract

A meta-analysis is a controversial statistical procedure that combines the data from several independent studies in an attempt to produce an estimate for the effectiveness of a specific intervention. The validity of a meta-analysis, also known by its critics as *numerological abracadabra*, is dependent on the arbitrarily defined criteria and discrimination of the analyst, and more importantly, on the quality of the inclusive studies. The focus of this commentary is a meta-analysis by Krieger, who claimed that multiple sets of each exercise are superior to a single set of each exercise for increasing muscular strength. Some examples of the inclusive studies that he awarded the highest quality scores are shown in this Critical Review to be very poor quality studies and not acceptable for inclusion in a meta-analysis.

Key words: *meta-analysis, sets, strength gains*

Meta-Analysis

The statistical process of a meta-analysis implies that theoretical and empirical science should be done by two different sets of people with different disciplinary abilities; that is, empirical research is performed by scientists and clinicians, but the interpretation of this research is performed by statisticians who decide what inferences should be drawn from the evidence [1]. The inclusion or exclusion of the studies in a meta-analysis is entirely based on the discrimination, opinions, and potential inherent bias of the statistician conducting the meta-analysis. The analyst may lack the appropriate training and experience to draw the correct inferences from the empirical research [1]. All the inclusive studies will have a profound effect on the results of the meta-analysis because the data from those studies may be the consequence of methodological flaws or fraud [2]. At best, the validity of a meta-analysis can only be as good as the quality of the inclusive studies; that is, a meta-analysis inherits all the flaws of the inclusive studies in addition to the inherent problems with peer review, editorial decisions and publication of those studies and the meta-analysis. If bias is present within the inclusive studies, the meta-analysis will reinforce that bias. A bad meta-analysis – like any bad research – may be useless or harmful, and bad research is more common than good research [3].

Krieger's Meta-Analysis

In a recent review entitled Determining Appropriate Set Volume for Resistance Exercise [4], Krieger stated that the primary way to manipulate the volume of training is to increase or decrease the number of sets for each exercise. In his next sentence he claimed: "*Thus, the number of sets can have a strong impact on the morphological and performance-based outcomes of a resistance training program*" (p. 30). Krieger did not cite any resistance training studies to support that statement, which may have been an early indication that readers were not going to get an impartial analysis of the topic.

In Krieger's review [4], he summarized and cited his own meta-analysis entitled Single Versus Multiple Sets of Resistance Training: A Meta-regression [5]. Although there are subtle differences between a meta-analysis and a meta-regression, Krieger never explained these differences in his mega-regression and several times in his review [4] he referred to his *hierarchical, random-effects meta-regression* simply as a meta-analysis. Therefore to maintain consistency with Krieger's own description, this Critical Review will refer to his meta-regression as a meta-analysis.

Krieger [5] set his own inclusive criteria arbitrarily for the meta-analysis: resistance training studies that involved at least one major muscle group with a minimum duration of four weeks, single and multiple sets with other equivalent training variables, pre-training and post-training 1RM (maximum resistance used for one dynamic repetition), and sufficient data published in the English language to determine frequency of training and calculate effect sizes for healthy participants at least 19 years of age. An effect size is a number that represents how many standard deviations the groups differ in outcomes such as strength gains.

Krieger [5] used the sum of two 0-10 scale-based scores to rate the quality for each of his 14 inclusive resistance training studies in his meta-analysis. His assessment of those studies resulted in quality scores that ranged from 9-15. A few of Krieger's highest scoring inclusive studies are examined in this Critical Review and readers can decide on the quality of those studies and consequently the credibility of his meta-analysis and its conclusions.

Examples of Krieger's Highest Quality Scored Inclusive Studies

Rhea and Colleagues

Krieger [5] awarded a resistance training study by Rhea and colleagues [6] the highest quality score. Although many of the flaws in the study by Rhea and colleagues have been previously exposed in detail [7], some of those criticisms are presented here because Krieger gave that study the highest quality rating (score = 15). Rhea and colleagues recruited 16 young males (mean age ~21 years) who were classified as trainees with at least two years of resistance training experience prior to the investigation. The participants were randomly assigned to perform either one set or three sets of leg press and bench press exercises three times a week for 12 weeks. There was no control group. The range of repetitions varied for each of the three weekly sessions (8-10RM, 6-8RM and 4-6RM, sessions 1, 2 and 3, respectively). There was no control for repetition duration during the training or the 1RM assessments.

Baseline assessment of the 1RM revealed no significant difference in 1RM bench press between groups but the 1RM leg press was 19% greater (although not statistically significant) in the 1-set group [6]. Both the 1-set and 3-set groups significantly increased bench press and leg press 1RM at the end of the study. There was no significant difference between groups in the bench press strength gains (1RM). The increase in 1RM leg press was significantly greater in the 3-set group compared with the 1-set group. However, perhaps because the 1-set group had a 19% greater 1RM baseline strength compared with the 3-set group, both groups attained similar leg press strength levels at the end of the study (337.2 kg and 343.5 kg, 1-set and 3-set groups, respectively). Although Rhea and colleagues compared the 1RM between groups at baseline, they did not report any statistical comparison of the strength levels between groups at the completion of the study.

Rhea and colleagues [6] stated that their data would be analyzed with a repeated measures analysis of variance and Tukey's post hoc tests to determine any significant differences between groups. They did not mention any calculation of effect sizes in their Statistical Analyses section. Perhaps when Rhea and colleagues failed to show a significant difference between groups for the increase in 1RM bench press with their designated statistical analysis, they chose to calculate – apparently incorrectly – effect sizes (see reference #7 for an in-depth critique of their effect size calculations). Nevertheless, by manipulating their statistical procedures they reported a significant difference between groups and claimed an effect size of 2.3 for the bench press. The effect size for the leg press was 6.5, which represents an unprecedented 6.5 standard deviations between the means of the 1-set and 3-set groups. The post-training standard deviation for the bench press and leg press was 3-4 times greater than the pre-training standard deviation in both groups and the magnitude of that change raises doubt concerning the ability to accurately estimate a true effect size. Rhea and colleagues did not report confidence intervals for their effect sizes.

Rhea and colleagues [6] reported that the effect size for the bench press was 2.3, which was almost three times greater than what statisticians designate as a *large* (0.8) effect size [8]. The effect size for the leg press (6.5) was more than eight times greater than a *large* effect size. Rhea [9] later proposed elevating the scale to interpret effect sizes and suggested that 1.5 should be considered a *large* effect size in resistance training studies with experienced trainees. Even if his unsupported arbitrary suggestion was valid – and he failed to present any rationale that was supported by logic or data – the extraordinary effect sizes (2.3 and 6.5) reported in their study [6] are unlike anything published in the scientific literature.

Rhea and colleagues [6] failed to speculate how their relatively strong experienced trainees in the 3-set group showed an increase in 1RM leg press from ~226 kg (~497 pounds) to ~344 kg (~757 pounds) and 1RM bench press from ~67 kg (~147 pounds) to ~86 kg (~189 pounds) in 12 weeks. Neither the 3-set nor the 1-set group showed any significant change in lean body mass, body fat, chest circumference or thigh circumference.

As previously noted, Krieger awarded this highly questionable and flawed study by Rhea and colleagues [6] the highest quality score in his meta-analysis [5], which questions the quality of his other inclusive lower scoring studies (see sections below).

Kremmler and Colleagues

Krieger [5] awarded a study by Kremmler and colleagues [10] the second highest quality rating (score = 14). Kremmler and colleagues reported the results of 50 females (mean age ~57 years) who were part of an ongoing osteoporosis prevention study. After 18 months of multiple set resistance training, the participants were assigned to one of two groups to continue their resistance training in a crossover design program. The purpose was to compare the effects of single versus multiple sets of each exercise. All the trainees performed five upper body and six lower body machine exercises once a week in what they designated as *Session 1*. The 1RM was assessed and reported for the leg press, bench press, rowing, and leg (hip) adduction exercises. There was no control for repetition duration during the 1RM assessments. Group 1 (n = 29) exercised using a multiple set periodized protocol with 65-90% 1RM for 12 weeks, switched to a low intensity program (2 sets of 20 repetitions with 50-55% 1RM) for five weeks, and then another 12 weeks of a single set periodized protocol with 65-90% 1RM. Group 2 (n =21) completed the same training program in the reverse order beginning with 12 weeks of the single set program, switching to five weeks of low intensity training, and then completing another 12 weeks of multiple set training. Kremmler and colleagues did not encourage any of the trainees to perform the maximal number of repetitions in any set for any exercise during the periodized single set, multiple set or low intensity periods. Consequently, the level of effort – and perhaps motor unit recruitment – may have been significantly different among the trainees at various times during the study. The intensity of effort is the predominant factor that determines motor unit recruitment [11].

There was a minimal although statistically significant increase in 1RM that ranged from 3-5% in both groups when multiple sets were employed; approximately 5%, 4%, 3%, and 4% for the leg press, bench press, rowing, and hip adduction, respectively [10]. The miniscule increase in 1RM after the 60-70 minute sessions of multiple set periodized training is extraordinarily small, even for previously trained subjects. There was a significant decrease (1-2%) from baseline 1RM following the single set protocol in both groups. It is unclear and not discussed by Kremmler and colleagues why the trainees in Group 2, who trained with multiple sets for 18 months prior to the crossover study, were unable to maintain their strength gains with the 1-set protocol.

In his meta-analysis, Krieger [5] committed several errors in reporting the study by Kremmler and colleagues [10]. Although these may be arguably minor errors, they question the accuracy of performing a complex statistical procedure such as a meta-analysis.

- Krieger's Table 1 (p. 1894) shows that Kremmler and colleagues compared 1-set and 2-set training. However, the participants actually performed 2-4 sets of each exercise during the multiple set protocol and two sets of each exercise only during the five weeks of intermediate low intensity training.
- Unlike the classification of the other 13 studies in Krieger's meta-analysis, he failed to report the training status of the participants in the study by Kremmler and colleagues.
- Krieger reported the effect sizes for the leg press, bench press and rowing exercises but did not report effect sizes for the hip adduction exercise, even though similar data were available from the study.
- Krieger reported the frequency of training as one session per week. The participants actually trained in two supervised sessions per week with different exercises that involved muscle groups used in the 1RM evaluations (see paragraph below for details).

Perhaps the most important aspect of this so-called comparative study of single versus multiple sets by Kremmler and colleagues [10] is that all of the trainees simultaneously participated in a second weekly supervised training session (*Session 2*). Those 60-70 minute sessions consisted of 2-4 sets of 10RM dumbbell bench presses, unilateral dumbbell rowing, squats and power cleans. These exercises involved muscles used in the performance and assessment of the 1RM leg press, bench press, rowing and hip adduction. They were performed weekly throughout the 29-week study by both groups. In addition to *Session 2*, all the trainees performed two weekly home exercise sessions that included selected isometric and elastic belt exercises. Kremmler and colleagues referred to these other sessions in the study included in Krieger's meta-analysis [5] and Kremmler and colleagues cited two of their previous publications [12-13] that described those sessions in detail. Krieger failed to report all these potential confounding variables.

The unexplained minimal changes in strength, the careless reporting of the data by Krieger [5], and his failure to note all the aforementioned potential confounding variables, question how Krieger awarded this study by Kremmler and colleagues [10] the second highest quality score of 14 in his meta-analysis.

Kraemer

Krieger [5] awarded an *experiment* by Kraemer [14] the third highest quality rating (score = 13). The data from this so-called *experiment #2* were resurrected from a database that Kraemer had accumulated as a coach approximately 15 years prior to publication. He matched and randomly assigned 40 young (mean age ~20 years) resistance trained (~2 years) Division I male American football players, which was noted incorrectly by Krieger as male and female participants, into one of two training groups. There was no control group. The single set group performed one set for each of five upper body and five lower body exercises; the multiple set group performed three circuits of these exercises. Both groups exercised with an 8-12RM load three times a week for 10 weeks. The single set group was given forced repetitions (assistance for additional reps) at the end of each set (failure) but the multiple set group did not receive forced repetitions. The 1RM leg press and bench press was assessed prior to and after the 10-week study. There was no control for repetition duration during the training or the 1RM assessments.

Kraemer claimed that there was 100% compliance for all 40 participants [14]. Both groups had a significant increase in 1RM bench press and leg press. The increase in the multiple set group was significantly greater than the single set group for both exercises. Compared with the 1-set group, the 1RM increase for the 3-set group was approximately three times greater for the bench press and approximately seven times greater for the leg press. These are unprecedented differences in strength gains for experienced trainees with a starting strength of ~144 kg (~317 pounds) in the bench press and ~175 kg (~387 pounds) in the leg press. Kraemer speculated that the difference in strength gains may be attributed to greater hormonal responses in the multi-set group. However, he did not measure hormonal responses.

Krieger [5] noted that *experiment #2* [14] was an unsupervised program. In the same article, Kraemer gave 115 football players an anonymous questionnaire to ask about their compliance to single set protocols (*experiment #5*). The response was that 89% of the players reported using additional multiple set programs at home or at health clubs because they wanted to supplement the single set protocol prescribed by the strength coach and perform additional exercises as well as multiple sets. If this were true for any of the trainees in Kraemer's *experiment #2*, it makes the reported differences between the single set and multiple set groups even more questionable.

It is critically important that the studies included in a meta-analysis have high methodological quality and ideally should be free from bias [15]. Kraemer [14] stated in his Introduction that for him the answers to important training questions were initially determined through his role as a coach. That statement revealed a potential bias prior to the resurrection, analysis and publication of his series of *experiments*. The scientific method first requires researchers to formulate a hypothesis and a subsequent null hypothesis, test that hypothesis, and then draw conclusions based on the confirmation or rejection of the null hypothesis.

Krieger's 1RM Inclusive Criterion

One of the inclusive criteria established by Krieger [5] was that a study must have reported the pre-training and post-training 1RM. Perhaps Krieger mistakenly believed that the 1RM is the only way to assess strength gains. That belief is not supported by the scientific literature [16]. There are at least 19 studies [17-35] that investigated various health related aspects of total body resistance training (~12 exercises) in males and females, and reported the pre-training and post-training 3RM as a result of training with one set for each upper body exercise and two sets for each lower body exercise with 5-15 repetitions for each set. The difference in the number of sets was within the whole study group (within-subject comparison) rather than comparing different groups, which would have been confounded by genetic variation among the participants. All the studies were supervised, well controlled and a sufficient duration of 4-6 months. The average strength gain in the 19 studies was ~42% for the 1-set training (upper body) and ~39% for the 2-set training (lower body). The question is whether Krieger was unaware of these studies, which were all published prior to his meta-analysis, he did not include them because of his arbitrarily predetermined 1RM criterion, or simply because 1-set and 2-set training produced similar strength gains.

Most of Krieger's [5] inclusive studies compared 1-set with 3-set training and he claimed that 2-3 sets were significantly better than one set for strength gains. If there really is a dose-response relationship between the number of sets and strength gains – as many multiple set proponents believe – then two sets should be better than one set and three sets should be better than two sets. Although Krieger had a section entitled Dose-Response Model, he failed to report on this potential relationship.

Numerological Abracadabra

On the first page in each issue of *Journal of Strength and Conditioning Research,* the Editorial Mission Statement claims that the National Strength and Conditioning Association attempts to "*...bridge the gap from the scientific laboratory to the field practitioner.*" Krieger's [5] lengthy Statistical Analyses section (p. 1892, 1895-6) published in that journal appears to be what Shapiro [36] has described as *numerological abracadabra*, rather than meaningful information that trainers and trainees could understand and apply to resistance training. Shapiro has noted: *"Perhaps this technique* [meta-analysis] *will succumb to its own absurdity; but if not, the next step will be the meta-analysis of meta-analyses, in which the meta-analyst will be totally divorced from reality, and totally surrounded by numbers without context"* (p. 229).

Publication Bias

Reviewers, editors and publishers are inclined to reject studies that show no significant difference between specific training protocols [37-39] such as the effect of single versus multiple sets on strength gains. Statisticians call this *the file drawer effect* or *publication bias*; that is, the editors cherry-pick the studies for publication that report a significant difference between protocols. The result of this *file drawer effect* (studies not published) should be that after a complete search for all the published research on a specific topic, the majority of the published research should be skewed toward studies that reported a statistically significant advantage of one training protocol over another [39]. However, on the specific topic of the effect of single versus multiple sets on strength gains, the majority of published studies – even considering the potential *file drawer effect* – reported no significant difference between protocols [see references 7, 40-42 for reviews of those studies].

Journal editors should require that the meta-analyst provide valid reasons – in a comprehensible language – for judging the quality of each inclusive study. Readers could then decide if they agree or disagree with the analyst's judgments regarding inclusion and exclusion [36]. Krieger [5] did not state which studies he rejected – or the reasons he rejected them – as not being acceptable for his meta-analysis. Consequently readers are unable to judge the reasons for his inclusion or exclusion of studies.

Blind Studies

One argument from the proponents of multiple sets is that the majority of studies recruited previously untrained participants and that multiple sets elicit superior outcomes in experienced trainees. Unfortunately, there are no double blind – or even truly blind (see section below) – resistance training studies on single versus multiple sets in any demographic of previously untrained or resistance trained participants. The researchers who are assessing pre-training and post-training strength can be blinded to the individual trainee's protocol but it may be extremely difficult to blind the trainees to an intervention such as single versus multiple sets. Consequently, if the scientific method was applied properly to the resistance training research, there would be no evidence for the superiority of multiple sets for strength gains. Science places the entire burden of proof on those who claim that multiple sets produce significantly greater strength gains than a single set of each exercise. Although there is no evidence to support the superiority of a single set protocol for strength gains, no one in the scientific literature has made that claim.

Marshall and Colleagues

In an interesting relevant study published after Krieger's meta-analysis [5], researchers attempted to blind the participants from the intent of the variation in training protocol. Marshall and colleagues [43] recruited 43 young males (mean age ~28 years) with an average 6.6 years of resistance training experience who could perform a barbell squat with at least 130% of their body mass. Prior to randomization, there was a 2-week washout from previous training. The washout consisted of four sets of 6-12RM for upper and lower body exercises. Using a 3-way split routine, each session involved six primary exercises that were performed three times during the 2-week washout. The barbell squat was not performed during the washout.

After the 2-week washout, the trainees were randomly assigned to either a 1-set, 4-set or 8-set squat exercise protocol [43]. The same upper body protocol as during the 2-week washout period was continued in an attempt to blind the participants from the manipulation of the lower body training variable: 1, 4 or 8 sets of barbell squats, which was the only lower body exercise for the next six weeks. The trainees performed the squat exercise two times a week with 80% of their baseline 1RM (after the 2-week washout period) and the 1RM was assessed again after three weeks in an attempt to maintain 80% 1RM. Prior to their *work sets*, they performed a warm-up consisting of 10 body mass squats, 10 repetitions with 50% 1RM, one repetition with 60% 1RM and one repetition with 70% 1RM. The multiple set groups rested three minutes between sets.

The participants in the three groups trained with 80% 1RM to volitional exhaustion; that is, the trainees were encouraged to give a maximal effort on each set and they were supervised by experienced exercise scientists [43]. The number of completed repetitions in the 1st set averaged 10.9, 9.0 and 8.2, for the 1-set, 4-set and 8-set groups, respectively. There was a significant difference in the number of repetitions for the 1stset between the 1-set and 8-set groups. In another publication of that study [44], the authors speculated that this inverse relationship between the number of repetitions on the 1st set and the volume of exercise (1, 4 or 8 sets) may be caused by a slightly greater effort – and thus a greater stimulus – in the 1-set group because that protocol did not require subsequent maximal efforts in additional sets. The authors stated that the trainees who followed the higher volume protocols may have held back in the early sets in order to perform better in subsequent sets. The average number of repetitions for all the sets was 10.9, 7.7 and 7.0 for the 1-set, 4-set and 8-set groups, which was significantly greater in the 1-set group compared with the 4-set and 8-set groups [43]. Eleven of the 43 trainees dropped out of the study; seven during the washout period and four during the 6-week primary training period.

All the remaining trainees in the three groups significantly increased their 1RM squat [43]. The 8-set group showed a significantly greater improvement (9.9%) than the 1-set group. However, there was no significant difference in strength gains between the 1-set group and the 4-set group, or between the 4-set group and the 8-set group. Compared with the 1-set group, the 8-set group performed approximately six times the volume of exercise (repetitions x sets x resistance). The repetition duration of each set began with a 2:1 eccentric:concentric ratio but changed as the sets approached volitional exhaustion (personal communication with Dr. Marshall, 12-21-11). The time of each session was not reported but with the designated warm-up, three minutes rest between sets, and assuming a total (concentric and eccentric) average repetition duration of at least four seconds, the

8-set group would have required approximately 40 minutes to complete their squat workout compared with about 12 minutes for the 1-set group.

Marshall and colleagues [43] stated that previous studies suggested that the responsiveness to resistance training is primarily determined by genetic factors. Based on the percent increase in strength after the six weeks of training, they subsequently sub-grouped their trainees as either low, medium or high responders. There was no significant difference in baseline strength after the 2-week washout between the 1-set, 4-set and 8-set groups. The high responders increased their 1RM squat by 29.4% compared with medium responders (14.3%) and low responders (2.6%), with no significant difference between responder groups for the average number of repetitions per set of squats. However, there were significant differences among responder groups for the increase in 1RM. The percent increase in strength for the high responders was more than 11 times greater than the low responders. Most importantly – 11 of the 13 low responders were from the 1-set and 4-set groups; that is, 80% (8 out of 10 trainees) of the 8-set group were medium or high responders. Marshall and colleagues noted that they did not know if the high responders would have achieved the same level of strength gains with a reduced training volume; that is, one set of squats instead of eight sets.

The primary author of this study [43] described the 8-set protocol as extremely hard and time consuming because of the excessive volume of exercise. In fact, the original plan was an 8-10-week loading phase but the trainees found the program so demanding that after 4-5 weeks the author was highly doubtful that they could complete the program. To minimize dropout, he reduced the primary training phase to six weeks (personal communication with Dr. Marshall, 12-21-11). It would be difficult to rationalize an 8-set protocol for just two compound upper body exercises. For example, performing eight sets to volitional exhaustion for the bench press and military press exercises would result in 16 high intensity sets for the triceps, which is a prime mover in both exercises. Consequently, an 8-set protocol for each exercise has no practical application to resistance training.

The authors stated that the results of their study supported multiple set resistance training (8 sets) in experienced individuals [43-44]. However, they also emphasized the possibility that inter-individual genetic variability could confound any attempt to draw conclusions regarding the appropriate number of sets [44].

When Marshall and colleagues [43] questioned the trainees at the end of the study, almost half (47%) thought that they knew what variable (the number of sets for the squat exercise) was manipulated in the study and seven (22%) knew exactly what was being manipulated. Therefore, their novel attempt to blind the subjects was not very successful.

There are a few important – often conflicting – points in the study by Marshall and colleagues [43] and the meta-analysis by Krieger [5]. Krieger claimed that 2-3 sets were significantly better than one set and there was no further benefit in performing more than 2-3 sets of each exercise. Marshall and colleagues reported no significant difference between the 1-set and 4-set groups, but that eight sets of squats produced a significantly greater strength gain compared with one set of squats. Neither discussed the extensive time required to perform multiple sets of each exercise compared with a single set. The 8-set protocol in the study by Marshall and colleagues would require more than six hours of training per session for 10 resistance exercises.

Genetics

Krieger [5] never mentioned the word *genetics* in his meta-analysis and its potential influence on the responsiveness to resistance training, which may be more important than the number of sets for each exercise. In contrast, Marshall and colleagues [43] understood the importance of genetics and reported the significant difference in strength gains among their sub-groups of low, medium and high responders (2.6%, 14.3% and 29.4%, respectively). A large genetically dependent range in strength gains for healthy young males and females to a specific training protocol is supported by other resistance training studies. For example, Hubal and colleagues [45] trained 585 males and females (mean age ~24 years) who followed an identical resistance training protocol for 12 weeks. The average strength gain (1RM biceps curl) was 54%, but ranged from 0%-250%. Thomis and colleagues [46] reported a significant 45.8% increase in 1RM biceps curl in 25 pairs of young (mean age ~22 years) male identical twins who participated in the same resistance training protocol for the elbow flexors three times a week for 10 weeks. These twins were an example of quintessentially matched groups. The variability in strength gains between the different pairs of twins was 3.5 times greater than the variability within the pairs of twins.

There is a scarcity of resistance training studies that emphasized the importance of genetics on strength gains and other chronic outcomes. Perhaps the reason for not reporting or discussing genetic variability is that the difference in strength gains resulting from a different number of sets or repetitions, repetition duration, periodization, inter-set rest intervals, frequency of training, exercise equipment, etc. are miniscule compared with the genetic influence. That revelation could minimize – if not eliminate – the need for extremely complex and time consuming protocols such as those recommended by the American College of Sports Medicine [47].

In a previous article on single versus multiple set resistance training [48], Krieger proclaimed: "*You can rule out the genetic factor when it comes to what gains to expect from single-set vs. multiple-set decisions*" (p. 15). He based that claim on a randomized crossover resistance training study by Humburg and colleagues [49]. They reported a small but significantly greater strength gain (~8%) for the two upper body exercises as a result of 3-set training compared with 1-set training. However, there was no significant difference in strength gains between 1-set and 3-set training for the two lower body exercises. The mean effect size for the 3-set protocol compared with the 1-set protocol for all four exercises was 0.23, which is considered a relatively small effect [8]. In contrast to the aforementioned interpretation of this study by Krieger, Humburg and colleagues concluded: "*The data of the present study revealed a large variation in individual adaptation to strength training with different volumes. Some individual subjects' 1RM progressed to a greater or equally large extent during the 1-set program compared with the 3-set program*" (p. 581). Although Krieger mistakenly dismissed the role of genetics in resistance training based entirely on the study by Humburg and colleagues, he gave that study the second lowest quality score (10) in his meta-analysis [5].

Krieger's reluctance to acknowledge or even mention the influence of genetics on strength gains [5] and his proclamation to rule out genetics as a significant contributing factor [48], is perhaps because of a firm – although unsubstantiated – belief in the superiority of multiple sets. Those strongly held beliefs may not be the optimal prelude to performing an unbiased meta-analysis.

Belief

The popularity of the meta-analysis was promoted by a statistician named Glass [50]. While earning his doctorate in statistics, he developed what he called a *major neurosis*. He truly believed that psychotherapy intervention helped him cope with his mental problems. However, the majority of scientific studies concluded that there was no significant difference in outcomes as a result of psychotherapy compared with a placebo treatment. Subsequently, Glass stated that he found this to be *personally threatening* and he admitted that he used a meta-analysis to confirm his belief in psychotherapy [51]. Those results were published in a psychology journal but Glass did not reveal his preconceived bias for the efficacy of psychotherapy in that meta-analysis [52].

People who have very strong beliefs will seek (consciously or sub-consciously) confirming evidence for support and ignore or misinterpret evidence that does not support their belief [53]. Prior to his meta-analysis [5], Krieger [48] claimed: "…*multiple sets are the way to go if you are looking to get the most out of your training* " (p. 17). Perhaps because of his strong belief in the superiority of multiple-set resistance training, which may be analogous to the strong belief in psychotherapy by Glass [51], he used a meta-analysis in an attempt to confirm that belief – especially when the preponderance of resistance training studies threatened that personal belief [see references 7, 40-42 for reviews of those studies].

Conclusions

Meta-analyses do not follow the rules of science. Eysenck [54] stated: "*A mass of reports – good, bad, and indifferent – are fed into the computer in the hope that people will cease caring about the quality of the material on which the conclusions are based*"(p. 517). The decision to include or exclude studies rests entirely on the supposedly unbiased discretion of the person performing the meta-analysis. Even a good meta-analysis of poorly designed or flawed studies results in bad statistics and misinformation. Combining poor quality data or overly biased data that do not make sense produces unreliable results [37]. This is also known in statistics as *garbage in, garbage out*. Charlton [1] noted: "*The notion that scientific interpretation can be reduced to statistical considerations, checklists and step-by-step flow diagrams applicable to any problem at any time would be laughable were it not becoming accepted practice in some circles. Inventories are not a substitute for substantive knowledge*" (p. 399).

Krieger [5] stated that he performed a sensitivity analysis by removing one study at a time from the meta-analysis in an attempt to identify the presence of studies that may have biased the analysis. He claimed: "…*given the robustness of the results to removal of individual studies, and the lack of evidence of publication bias, it is unlikely that significant bias was present*" (p. 1899). This Critical Review of Krieger's inclusive studies revealed that *numerological abracadabra* trumped *lack of evidence of publication bias*.

It is apparent from the previous discussion of the inclusive studies by Rhea and colleagues [6], Kremmler and colleagues [10], and Kraemer [14], which Krieger [5] awarded the highest quality scores (15, 14 and 13, respectively), that he did not critically analyze these studies before including them in his meta-analysis (Table). In Krieger's review [4] and meta-analysis [5] he concluded that for optimal strength gains, the overwhelming evidence showed that multiple sets of each exercise are superior to a single set. However, his conclusion lacked sufficient credible evidence for support.

Table. *Studies in Krieger's meta-analysis [5] that he rated the three highest quality scores*

Study	Score	Reasons to exclude these studies in a meta-analysis
Rhea et al. [6]	15	• No control group • Small sample size (n=8) • No control for repetition duration during the 1RM testing or training • No indication that the trainers or the assessors of the 1RM and body composition were blinded to the different training protocols • No statistical comparison between groups of post-training 1RM, which appeared to be similar • Unprecedented effect sizes unlike anything in the scientific literature (e.g., the leg press effect size of 6.5 is eight times greater than what is considered a *large* effect) • The 1RM standard deviation was 3-4 times greater from pre-training to post-training for both exercises in the two groups, which raises concerns about the ability to accurately estimate effect sizes • No confidence intervals reported with effect sizes • No explanation for the 52% strength gain in the leg press or a 33% gain in the bench press in relatively strong experienced trainees who showed no significant change in lean body mass
Kremmler et al. [10]	14	• Minimal (3-5%) strength gains after 12 weeks of training • Trainees not encouraged to perform any sets with a maximal effort • No control for repetition duration during 1RM testing or training • No indication that the trainers or the 1RM assessors were blinded to the different training protocols • Multiple potential confounding variables: additional supervised and unsupervised training sessions that involved similar muscles in the exercises that were assessed for 1RM
Kraemer [14]	13	• Resurrected data from at least 15 years prior to publication • No control group • Unprecedented 3-7 times difference in strength gains between groups in strong, previously trained (~2 years) Division I football players • Unsubstantiated speculation that the difference in strength gains may have been caused by greater hormonal responses, which were not measured • No control for repetition duration during 1RM testing or training • Forced repetitions in one group but not in the other group • No indication that the trainers or those who assessed the 1RM were blinded to the different training protocols • The author's claim that the answers to important training questions were determined in his role as a coach before he analyzed the data revealed a strong potential bias for a specific outcome

Acknowledgement

The author is extremely grateful to Arty Conliffe for his conceptual and editorial critique of this manuscript throughout its preparation.

References

1. Charlton BG. The uses and abuses of meta-analysis. *Fam Pract* 1996; 13: 397-401.
2. Klein DF. Flawed meta-analyses comparing psychotherapy with pharmacotherapy. *Am J Psychiatry* 2000; 157: 1204-11.
3. Shapiro D. Point/counterpoint: meta-analysis/shmeta-analysis. *Am J Epidemiol* 1994; 140: 771-8.
4. Krieger JW. Determining appropriate set volume for resistance exercise. *Strength Cond J* 2010; 32: 30-2.
5. Krieger JW. Single versus multiple sets of resistance exercise: a meta-regression. *J Strength Cond Res* 2009; 23: 1890-901.
6. Rhea MR, Alvar BA, Ball SD, et al. Three sets of weight training superior to 1 set with equal intensity for eliciting strength. *J Strength Cond Res* 2002; 16: 525-9.
7. Otto RM, Carpinelli RN. A critical analysis of the single versus multiple set debate. *J Exerc Physiol* 2006; 9: 32-57.
8. Rosenthal R, Rosnow RL, Rubin DB. *Contrasts and effect sizes in behavioral research*. Cambridge, UK: Cambridge University Press, 2000.
9. Rhea MR. Determining the magnitude of treatment effects in strength training research through the use of the effect size. *J Strength Cond Res* 2004; 18: 918-20.
10. Kremmler WK, Lauber D, Engelke K, et al. Effects of single- vs. multiple-set resistance training on maximal strength and body composition in trained postmenopausal women. *J Strength Cond Res* 2004; 18: 689-94.
11. Carpinelli RN. The size principle and a critical analysis of the unsubstantiated heavier-is-better recommendation for resistance training. *J Exerc Sci Fit* 2008; 6: 67-86.

12. Kremmler WK, Engelke K, Lauber D, et al. Exercise effects on fitness and bone mineral density in early postmenopausal women: 1-year EFOPS results. *Med Sci Sports Exerc* 2002; 34: 2115-23.
13. Kremmler WK, Lauber D, Weineck J, et al. Benefits of 2 years of intense exercise on bone density, physical fitness, and blood lipids in early postmenopausal osteopenic women. *Arch Intern Med* 2004; 164: 1084-91.
14. 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.
15. Egger M, Smith GD, Sterne JAC. Use and abuse of meta-analysis. *Clin Med* 2001; 1: 478-84.
16. Carpinelli RN. Assessment of one repetition maximum (1RM) and 1RM prediction equations: are they really necessary? *Med Sport* 2011; 15: 91-102.
17. Girouard CK, Hurley BF. Does strength training inhibit gains in range of motion from flexibility training in older adults? *Med Sci Sports Exerc* 1995; 27: 1444-9.
18. Hurley BF, Redmond RA, Pratley RE, et al. Effects of strength training on muscle hypertrophy and muscle cell distribution in older men. *Int J Sports Med* 1995; 16: 378-84.
19. Koffler KH, Menkes A, Redmond RA, et al. Strength training accelerates gastrointestinal transit in middle-aged and older men. *Med Sci Sports Exerc* 1992; 24: 415-9.
20. Lemmer JT, Ivey FM, Ryan AS, et al. Effect of strength training on resting metabolic rate and physical activity: age and gender comparisons. *Med Sci Sports Exerc* 2001; 33: 532-41.
21. Lott ME, Hurlbut DE, Ryan AS, et al. Gender differences in glucose and insulin response to strength training in 65- to 75-year-olds. *Clin Exerc Physiol* 2001; 3: 220-8.
22. Martel GF, Horblut DE, Lott ME, et al. Strength training normalizes resting blood pressure in 65- to 73-year-old men and women with high normal blood pressure. *J Am Geriat Soc* 1999; 47: 1215-21.
23. Menkes A, Mazel S, Redmond RA, et al. Strength training increases regional bone mineral density and bone remodeling in middle-aged and older men. *J Appl Physiol* 1993; 74: 2478-84.
24. Miller JP, Pratley RE, Goldberg AP, et al. Strength training increases insulin action in healthy 50- to 65-yr-old men. *J Appl Physiol* 1994; 77: 1122-7.
25. Nicklas BJ, Ryan AJ, Treuth MM, et al. Testosterone, growth hormone and IGF-I responses to acute and chronic resistive exercise in men aged 55-70 years. *Int J Sports Med* 1995; 16: 445-50.
26. Parker ND, Hunter GR, Treuth MS, et al. Effects of strength training on cardiovascular responses during a submaximal walk and a weight-loaded walking test in older females. *J Cardiopulm Rehabil* 1996; 16: 56-62.
27. Rhea PL, Ryan AS, Nicklas B, et al. Effects of strength training with and without weight loss on lipoprotein-lipid levels in postmenopausal women. *Clin Exerc Physiol* 1999; 1: 138-44.
28. Roth SM, Ivey FM, Martel GF, et al. Muscle size responses to strength training in young and older men and women. *J Am Geriatr Soc* 2001; 49: 1428-33.
29. Rubin MA, Miller JP, Ryan AS, et al. Acute and chronic resistive exercise increase urinary chromium excretion in men as measured with an enriched chromium stable isotope. *J Nutr* 1998; 128: 73-8.
30. Ryan AS, Treuth MS, Rubin MA, et al. Effects of strength training on bone mineral density: Hormonal and bone turnover relationships. *J Appl Physiol* 1994; 77: 1678-84.
31. Ryan AS, Pratley RE, Elahi D, et al. Resistive training increases fat-free mass and maintains RMR despite weight loss in postmenopausal women. *J Appl Physiol* 1995; 79: 818-23.
32. Ryan AS, Treuth MS, Hunter GR, et al. Resistance training maintains bone mineral density in postmenopausal women. *Calcif Tissue Int* 1998; 62: 295-9.
33. Ryan AS, Pratley RE, Elahi D, et al. Changes in plasma leptin and insulin action with resistive training in post-menopausal women. *Int J Obesity* 2000; 24: 27-32.
34. Ryan AS, Hurlbut DE, Lott ME, et al. Insulin action after resistive training in insulin resistant older men and women. *J Amer Ger Soc* 2001; 49: 247-53.
35. Treuth MS, Ryan AS, Pratley RE, et al. Effects of strength training on total and regional body composition in older men. *J Appl Physiol* 1994; 77: 614-20.
36. Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997; 50: 223-9.
37. Lau J, Ioannidis JPA, Schmidt CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997; 127: 820-6.
38. Egger M, Smith GD. Misleading meta-analysis. *BMJ* 1995; 310: 752-4.
39. Bailar JC III. The promise and problems of meta-analysis. *NEJM* 1997; 337: 559-61.
40. Carpinelli RN, Otto RM. Strength training: single versus multiple sets. *Sports Med* 1998; 26: 73-84.
41. Carpinelli RN, Otto, RM, Winett RA. A critical analysis of the ACSM position stand on resistance training: insufficient evidence to support recommended training protocols. *J Exerc Physiol* 2004; 7: 1-64.
42. Carpinelli RN. Berger in retrospect: effect of varied weight training programmes on strength. *Br J Sports Med* 2002; 36: 319-24.
43. Marshall PWM, McEwen M, Robbins DW. Strength and neuromuscular adaptation following one, four, and eight sets of high intensity resistance exercise in trained males. *Eur J Appl Physiol* 2011; 111: 3007-16.
44. Robbins DW, Marshall PWM, McEwen M. The effect of training volume on lower-body strength. *J Strength Cond Res* 2012; 26: 34-9.
45. Hubal MJ, Gordish-Dressman H, Thompson PD, et al. Variability in muscle size and strength gain after unilateral resistance training. *Med Sci Sports Exerc* 2005; 37: 964-72.
46. Thomis MAI, Beunen GP, Maes HH, et al. Strength training: importance of genetic factors. *Med Sci Sports Exerc* 1998; 30: 724-31.
47. Ratamess NA, Alvar BA, Evetoch [sic] TK, et al. Progression models in resistance training for healthy adults. *Med Sci Sports Exerc* 2009; 41: 687-708.
48. Krieger J. One vs. three. Another chapter in the debate about the number of sets. *J Pure Power* 2007; 2: 15-17.
49. Humburg H, Baars H, Schroder J, et al. 1-set vs. 3-set resistance training: a crossover study. *J Strength Cond Res* 2007; 21: 578-82.
50. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976; 10: 3-8.
51. Glass G. Meta-analysis at 25. (available at http://glass.ed.asu.edu/gene/papers/meta25.html).
52. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977; 32: 752-60.
53. Shermer M. The believing brain: why science is the only way out of belief-dependent realism *Sci Am* 2011; 305: 85.
54. Eysenck HJ. An exercise in mega-silliness. *Am Psychol* 1978; 33: 517.

Accepted: July 30, 2012
Published: September 25, 2012

Address for correspondence:
Ralph N. Carpinelli
P.O. Box 241,
Miller Place, NY 11764
USA
E-mail: ralphcarpinelli@optonline.net

Authors' contribution	B – Data Collection	D – Data Interpretation	F – Literature Search
A – Study Design	C – Statistical Analysis	E – Manuscript Preparation	G – Funds Collection