Reliability of meta-analyses to evaluate resistance training programmes

Antonio Arruda, Daniel Souza, James Steele, James Fisher, Jürgen Giessing & Paulo Gentil


To link to this article: http://dx.doi.org/10.1080/02640414.2016.1243799

Published online: 04 Nov 2016.
Reliability of meta-analyses to evaluate resistance training programmes

We read with great interest the study by Schoenfeld, Ogborn, and Krieger (2016) and would, first of all, like to congratulate the authors for the well-written paper and valuable effort to treat the topic in an objective and straightforward manner. However, to improve our understanding of the topic, we offer some comments based on theoretical evidence and our practical experience.

A common criticism of meta-analyses is that they usually combine studies that have important methodological differences and, consequently, the summary effect can be largely influenced by these differences across studies (Field, 2015). For this reason, an important stage before the decision to use a meta-analysis involves consideration of variables that might explain variations in calculated effect sizes. However, the power of meta-analyses to test moderators depends on the number of studies available and the sample sizes used in studies.

Resistance training outcomes are influenced by many variables that interact with each other (Paoli, 2012). Hence, attempts to estimate the impact of 1 variable when others are not controlled has a high risk of bias. Although the authors made an effort to control for confounding variables by using regression models, these did not include other resistance training variables, and the small number of studies may also have reduced the power to detect interactions. This seems to be the case when analysing different body parts. Of the 5 studies that have analysed both upper and lower body muscle size, 4 of them (Bottaro, Veloso, Wagner, & Gentil, 2011; McBride, Blaak, & Trippllet-McBride, 2003; Ostrowski, Wilson, Weatherby, Murphy, & Lyttle, 1997; Radaelli et al., 2014) showed different patterns of muscle hypertrophy for upper and lower body muscles, with upper body, but not lower body muscles, generally responding better to higher volumes of resistance training. Only the study of Rønnestad et al. (2007) showed similar behaviour among different body parts. However, Rønnestad et al. (2007) evaluated the trapezius muscle, while the other studies evaluated additional arm muscles (biceps and/or triceps brachii). Therefore, including studies that measured legs and arm muscles in the same analyses could be misleading.

Another possible caveat concerns analysis of the number of sets per exercise, instead of the number of sets in which the muscle/s was/were involved. For example, the participants in the study by Correa et al. (2015) performed both leg press and knee extensions. Although the number of weekly sets per exercise was 3 or 9, the number of sets involving knee extensions were 6 or 18, respectively. With regard to arm muscles, multijoint exercises (e.g., lat pull downs) promote the same increases in muscle size as single-joint exercises (e.g., biceps curls) (Gentil, Soares, & Bottaro, 2015). This suggests that upper body multijoint exercises should be included when counting the number of sets for arm muscles. This was not performed in the studies of Correa et al. (2015), Ostrowski et al. (1997), Radaelli et al. (2015) Radaelli et al. (2014) Radaelli et al. (2015), and Ribeiro et al. (2015), and does not appear to have been accounted for in the analysis of Schoenfeld et al.

The poor control and inadequate reporting of effort is another point of concern. Most people that advocate training at low volume, suggest that exercises should be performed to momentary concentric failure. Training to momentary failure may be critical during low-volume resistance training (Giessing et al., 2014), making it important to control for intensity of effort in the studies analysed (Fisher & Smith, 2012; Steele, 2014). The inadequate reporting of training effort is evident in the study of Radaelli et al. (2015) who stated that participants performed 8–12 repetitions to concentric failure, but in addition, stated that loads were increased only when participants could perform more than 12 repetitions in all 3 sets. This seems implausible. Having reached true momentary failure in the first set, with a rest of only 90–120 s, there would be either a decrease in repetitions or it would be necessary to reduce the load used in subsequent sets, thus making it near impossible to keep participants within the 8–12 repetition range (Willardson & Burkett, 2005, 2006). Furthermore, increasing load for 1 exercise (e.g., leg press) would almost certainly reduce repetitions on subsequent exercises using similar muscle groups (e.g., leg extension and leg curl). Whether participants in this study did indeed train to momentary failure as suggested is unclear from the description in the methods offered by the authors. While, sensitivity analyses did not reduce the effect size estimated to result in statistical non-significance, we think it is important to note that the effect size and confidence intervals generated from the study by Radaelli et al. (2015) is the only to suggest convincingly that higher volumes are more beneficial.

Inspection of the methods used in included studies included suggests it is unlikely that participants reached momentary concentric failure in most cases. For example, Sooneste, Tanimoto, Kakigi, Saga, and Katamoto (2013) reported that the participants performed sets at 80% of 1 RM reaching momentary failure or until 10 RM was completed. The study of Cannon and Marino (2010) involved knee flexion and extension at 50% of 1 RM during week 1 and 75% of the 1 RM for weeks 2–10, and the participants were instructed to perform either 1 or 3 sets of 10 repetitions. Considering previous reports of number of repetitions performed at different percentages of 1 RM (Hoeger, Hopkins, Barette, & Hale, 1990), it seems likely that many participants in both studies did not reach momentary failure in the initial sets. Support for this is that participants performed the same number of repetitions in sets 1 and 2 in the study of Sooneste et al. (2013).

Based on these observations, it is possible that in many of these studies, groups that performed single sets per exercise...
were training at submaximal efforts while the multiple sets groups, though also using submaximal efforts, were accumulating fatigue from set to set and thus were at a closer proximity to momentary failure in later sets. This has been seen in previous studies with volume-matched training to failure and not to failure using rating of perceived exertion scales. So, in addition to comparing volume, these studies might be comparing groups that trained at differing intensities of effort, which can produce different outcomes (Gießing et al., 2014).

We acknowledge that it is not the responsibility of Schoenfeld et al. to “police” studies in this way, however, inclusion of such studies serves only to reduce meta-analyses outcomes to a series of numerical values with limited real-world validity. A wider question on use of meta-analyses arises: Is it possible to pool data from studies using disparate methods to generate an overall conclusion on the effects of manipulation of a variable in an intervention when that variable also interacts with the manipulation of other variables? Considering the above points, it is impossible that higher volumes are more beneficial when intensity of effort is not maximal (i.e., repetitions are not performed to momentary failure). However, with the poor reporting of set endpoints in many resistance training studies, it is often difficult to know whether 2 interventions have appropriately controlled for this variable. Indeed, many meta-analyses of resistance training variables pool studies that have compared them under varying conditions. From a practical perspective, one is left wondering whether higher or lower set volumes are better when performing repetitions to momentary failure, under heavy and/or light load conditions, with long or short repetitions, durations, etc. The role of interactions among other resistance training variables and set volume was apparently not considered in the meta-analysis of Schoenfeld et al. Indeed, even if they were to have included meta-regressions for these, some have argued that even with this type of consideration, a meta-analysis is not the right tool to tease apart aspects of interventions that work and those that do not (Field, 2015).

With these considerations in mind, we opine that fundamental inadequacies in the primary studies could have been carried over to the meta-analysis. Additionally, benefits of increasing training volume could have arisen as compensation for a low intensity of effort in training. The influence of intensity of effort could be one of the reasons why most studies have shown that lower body muscles respond better to higher-volume training, since participants trained to, or closer to, momentary failure more frequently in upper body than lower body exercises (Gentil & Bottaro, 2010).

We also note inclusion of studies with participants of different training history in the same analysis. The already well trained have reduced muscle hypertrophy response to training (Ahtiainen, Pakarinne, Alen, Kraemer, & Hakkinen, 2003). Other issues that could have influenced the results, and were not considered, were the sex of participants and their age. The dissimilarities of participants, methods, muscle groups, and types of exercise among the studies analysed brings into question if it is really possible to suggest the existence of a dose–response relationship. We wonder whether increasing number of sets leads to greater muscle gains or whether the conclusion drawn by Schoenfeld et al. was “contaminated” by methodological differences in the studies included. This seems to be an important issue, especially because if we analyse the studies that used more than 2 training groups, for most there is no clear sign of graded results with increasing number of sets. For example, in the study of Ostrowski et al. (1997) increase in triceps muscle thickness was 2.7%, 4.6%, and 4.7%, whereas in rectus femoris muscle thickness was 6.7%, 5.0%, and 13.1% for 1, 2, and 4 sets per exercise, respectively. Radaelli et al. (2015) reported a graded response in effects of 1, 3, and 5 sets per exercise for elbow flexor but not for elbow extensors muscle thickness. This means that the meta-analysis presents 2 studies in a total of 4 comparisons of more than 2 volumes, where only 1 analysis showed a graded response for increasing the number of sets. Additionally, it is important to note that Ostrowski et al. (1997) suggested a possibility of over-training as the number of sets increased, because of negative alterations in testosterone/cortisol ratio. Therefore, one should be cautious before adopting “the more, the better” as an approach to choosing the optimal number of sets.

The practical applications suggested by Schoenfeld et al. suggest a greater response to higher volumes, but do not consider how this information might be applied in a training programme. Consider that if 10+ sets per muscle group/week produce the greatest adaptation, can a person self-select how they divide these exercises throughout the week (i.e., the necessary volume performed in only 1 workout, or across 2, 3, 4, or 5 progressively lower volume workouts)? The results of this meta-analysis could be considered as supportive of practical recommendations from proponents of low-volume resistance training approaches (Fisher, Steele, & Smith, 2013) if one were to perform ~3 workouts a week consisting of ~4 exercises per muscle group, where a single set of each exercise is performed.

In conclusion, considering the large number of variables involved in resistance training and the methodological inconsistencies in the current literature, it seems impossible to make comparisons of different studies or include different studies in the same analysis. For a meta-analysis to be valid, a large amount of data on homogeneous subgroups should accumulate for topics where there is strong consensus about which variables have theoretical importance, and this does not seem to be the case for resistance training studies. Because of this, the generalisation of meta-analyses should be viewed with caution until we have a large number of studies providing adequate control of variables. Rather than prematurely perform meta-analyses on differing resistance training variables, which are all hindered by the inherent limitations of meta-analyses (Shapiro, 1994) including low study numbers and study heterogeneity (Field, 2015), and serve only to reduce the complexity of resistance training variables to a single statistic, greater value can be obtained by designing and conducting studies of larger and homogenous samples that can adequately address the topics considered. Otherwise, we can be comparing oranges with apples or, worse, we can be assuming that oranges and apples are the same.

**Disclosure statement**

No potential conflict of interest was reported by the authors.
References


Steele, J. (2014). Intensity; in-ten-si-ty; noun. 1. Often used ambiguously within resistance training. 2. Is it time to drop the term altogether?. British Journal of Sports Medicine, 48(22), 1586–1588. doi:10.1136/bjsports-2012-092127


Antonio Arruda
Human Performance Research Laboratory (LAPEDH), University of Pernambuco, Petrolina, Brazil
arrudabq@hotmail.com

Daniel Souza
Federal University of Rio Grande do Norte, Post Graduate Program in Physical Education, Natal, Brazil

James Steele and James Fisher
Sport Science Laboratory/Centre for Health, Exercise & Sport Science, Southampton Solent University, Southampton, UK

Jürgen Giessing
Institute of Sport Science, University of Koblenz-Landau, Landau, German

Paulo Gentil
College of Physical Education, Federal University of Goias, Goiania, Brazil

JOURNAL OF SPORTS SCIENCES

Accepted 28 September 2016