Assessing differential responders and mean changes in muscle size, strength, and the cross-over effect to two distinct resistance training protocols

Scott J. Dankel¹, Zachary W. Bell², Robert W. Spitz², Vickie Wong², Ricardo B. Viana², Raksha N. Chatakondi², Samuel L. Buckner⁴, Matthew B. Jessee⁵, Kevin T. Mattocks⁶, J. Grant Mouser⁷, Takashi Abe², Jeremy P. Loenneke²

¹ Department of Health and Exercise Science, Exercise Physiology Laboratory, Rowan University, Glassboro, NJ.
² Department of Health, Exercise Science, & Recreation Management, Kevser Ermin Applied Physiology Laboratory, The University of Mississippi, University, MS.
³ Faculty of Physical Education and Dance, Federal University of Goiás, Goiânia.
⁴ USF Muscle Laboratory, Exercise Science Program, University of South Florida, Tampa, FL.
⁵ Department of Health, Exercise Science, & Recreation Management, The University of Mississippi, University, MS.
⁶ Department of Exercise Science, Lindenwood University - Belleville, Belleville, IL.
⁷ Department of Kinesiology and Health Promotion, Troy University, Troy, AL.

Running head: Differential responders and random error

Corresponding Author
Jeremy P. Loenneke, PhD
Kevser Ermin Applied Physiology Laboratory
Department of Health, Exercise Science, and Recreation Management
The University of Mississippi
231 Turner Center
University, MS 38677
E-Mail: jplouene@olemiss.edu
Phone: 662-915-5567
Fax: 662-915-5525
Abstract

Purpose: To determine differences in two distinct resistance training protocols and whether true variability be detected after accounting for random error. Methods: Individuals (n=151) were randomly assigned to one of three groups: (1) a traditional exercise group performing four sets to failure; (2) a one-repetition maximum (1RM) performing a 1RM test; and (3) a time-matched non-exercise control group. Both exercise groups performed 18 sessions of elbow flexion exercise over six weeks. Results: While both training groups increased 1RM strength similarly (~2.4kg), true variability was only present in the traditional exercise group (true variability = 1.80kg). Only the 1RM group increased untrained arm 1RM strength (1.5kg), while only the traditional group increased ultrasound measured muscle thickness (~0.23cm). Despite these mean increases, no true variability was present for untrained arm strength, or muscle hypertrophy in either training group. Conclusion: These findings demonstrate the importance of taking into consideration the magnitude of random error when classifying differential responders, as many studies may be classifying high and low responders as those who have the greatest amount of random error present. Additionally, our mean results demonstrate that strength is largely driven by task specificity, and the cross-over effect of strength may be load dependent.

Novelty Bullets:

- Many studies examining differential responders to exercise do not account for random error.
- True variability was present in 1RM strength gains, but the variability in muscle hypertrophy and isokinetic strength changes could not be distinguished from random error.
- The cross-over effect of strength may differ based on the protocol employed.
Keywords: 1RM training; individual responders; personalized training; randomized controlled trial; strength training; variability; weight lifting

Introduction
It is commonly believed that if a group of individuals all perform the exact same exercise intervention, some individuals will respond more favorably than others (Bouchard and Rankinen 2001; Hubal et al. 2005). This idea has led to millions of dollars being spent on genome-wide association studies attempting to link common genetic mutations with differences in exercise outcomes (Visscher et al. 2012). As it relates to variability in exercise responses, these studies have been largely unsuccessful likely due to two factors: (1) difficulties in detecting people as being differential responders and (2) poor reliability of gene expression measurements (Islam et al. 2019). To date, there is little evidence to support that there are individual differences with respect to changes in muscle size and strength that occur in response to resistance exercise. The strongest support comes from data pooled within several studies which used various measurement techniques (Ahtiainen et al. 2016); however, this may be subject to inflated variability given differential sensitivities in the ability of measurement tools to detect changes from baseline (Dankel et al. 2018a).

Although it appears there is support that individuals respond differently to the same exercise intervention, it is important to question how variable the response would look even if there were no exercise intervention. This can be termed random error, and consists of both measurement error and random biological variability (Atkinson and Batterham 2015). Thus, we cannot be certain individuals respond differently to exercise unless we have a method to quantify how much of the variability present is simply random error and how much can truly be attributed to the exercise intervention itself. Arguably the best approach to quantify random error would be to perform the exact same training intervention multiple times following a sufficient wash-out
period (Hecksteden et al. 2015). This would allow for a confidence interval to be created around each individual’s mean response, but this assumes that a similar response will be observed during subsequent training periods. For some variables, such as muscle strength, the wash-out period would likely span years (Fatouros et al. 2005), making this approach infeasible. A more feasible approach involves examining the level of variability that is present in a time-matched control group not performing the exercise intervention (Atkinson and Batterham 2015). This allows for the quantification of random error present even without an intervention. Thus, any magnitude of variability that exceeds this random error could be attributed to the exercise intervention itself.

Previous research has suggested that the level of variability in adaptations to exercise may differ depending on the intensity (Bonafiglía et al. 2016) or volume (Damas et al. 2018) of exercise performed. Given the difficulty of deciphering whether these differences were the result of random error, we sought to examine whether the magnitude of true variability (if present at all) would depend on the type of intervention employed. Therefore, the purpose of this study was to determine if there is individual variability in response to two distinct resistance exercise protocols even after accounting for random error quantified by a time-matched control group. These specific protocols have been employed by our laboratory previously (Dankel et al. 2017; Mattocks et al. 2017) and have both been shown to produce similar results for increasing muscle strength while also showing differences in muscle growth.

**Materials and Methods**

*Participants*
A total of 158 individuals (60 males and 98 females) were recruited for the study. To be eligible, individuals must have been between the ages of 18-35 years, could not be a regular tobacco user, and could not have performed any structured resistance exercise within the previous 6 months. All individuals provided written informed consent for this study which was approved by the University’s Institutional Review Board (Protocol # 18-027).

Study Design

Individuals visited the laboratory and had the following pre-test measurements taken in the order listed: (1) height and body mass, (2) muscle thickness, (3) one-repetition maximum (1RM) strength, and (4) maximal isokinetic strength. The post-testing visit was identical to the pre-test visit (height and body mass measurements excluded) and was performed between 2-5 days after the final training session. To account for the potential influence of circadian rhythms, the pre-test and post-test visits were performed around the same time of day (i.e. within 2 hours). At the end of the pre-visit, the first 12 males and 12 females were assigned to one of three groups using a random number generator. This was done so the researchers would not be able to predict the group assignment for all subsequent individuals who were randomly assigned to a group based on quartiles of dominant arm 1RM strength using a covariate adaptive randomization program that was supplied by the authors of a previous paper (Kang et al. 2008). The three groups included: (1) a control group, (2) a traditional exercise group, and (3) a 1RM training group. All of the training and testing procedures for all individuals were completed at the same location (i.e. it was not a multi-site study).

Training Protocols
All exercise training involved dumbbell elbow flexion exercise performed on the dominant arm only. The control group performed only the pre-test and post-test visits spanning the same six week period as the training groups. Both training groups performed elbow flexion exercise three times per week for six weeks (18 total sessions) with all sessions supervised by a research investigator. The traditional exercise group performed four sets of exercises to volitional failure at a load which individuals could achieve between 8-12 repetitions. Ninety seconds of rest were allotted between sets. The 1RM group performed up to five heavy single repetitions during each of the visits (Mattocks et al. 2017). The exercise was terminated if all 5 repetitions were successfully completed, or one of the repetitions was unsuccessful. The load started at around 80-85% 1RM and was progressively increased in an attempt to meet or exceed their previously established 1RM. Ninety seconds of rest were allotted between repetition attempts. All exercises for both groups were performed while the individual’s back and heels were against a wall to ensure strict form. Individuals in all groups were instructed to continue their normal diets and activity patterns, and were instructed not to begin performing any resistance exercise outside of the intervention.

Muscle Thickness

Ultrasound images (Logiq e, General Electric, Fairfield, CT) of muscle thickness were taken at the anterior aspect of the individual’s upper arm at each 50, 60 and 70% of the distance between the acromion process and lateral epicondyle. The probe (Logiq e, L4-12t probe, General Electric, Fairfield, CT) was coated with gel and held transversely against the skin (Dankel et al. 2017). Images were taken in duplicate and saved for later analysis in a blinded fashion with the researcher unaware of the group assignment for all ultrasound images. This was performed on
both the trained and untrained arms. All of the ultrasound images were taken by the same
ultrasound technician. To ensure that similar measurement sites were used for the pre-test and
post-test sessions, the arm length and distance from the lateral epicondyle were recorded for each
of the measurement sites during the pre-test visit and used for reference during the post-test visit.
The probe angle was always placed perpendicular to the measurement site, and we have
previously shown that the magnitude of unintentional probe tilt expected between measurements
would be unlikely to have a meaningful impact on muscle thickness measurements (i.e. even an
8° probe tilt resulted in a coefficient of variation of 1.5% on repeated measurements) (Dankel et
al. 2018b).

Isotonic (One-Repetition Maximum) Strength

Maximum unilateral concentric strength (the heaviest weight that can be lifted one time) of the
individual’s arms during elbow flexion was measured with dumbbells. Each individual
completed the same protocol on both arms. The weight was progressively increased to find the
maximum weight that could be lifted through the full range of motion. All 1RM attempts were
separated by 90 seconds of rest and were performed with the individual’s back and heels against
a wall to ensure strict form. All 1RMs were measured to the nearest 0.2 kg and were usually
obtained in around 5-7 attempts. In a randomized fashion, the 1RM was completed in its entirety
in one arm, before measuring 1RM strength in the contralateral arm. The instructions and verbal
encouragement provided to the participant were matched as closely as possible for the pre-test
and post-test sessions.

Isokinetic Strength
Individuals were seated on a dynamometer (Biodex Medical Systems, Shirley, New York, USA) with the seat and lever arm adjusted appropriately and the settings recorded and standardized for all future tests. After weighing the individuals forearm by having them hold onto the handle with their arm limp to correct for gravity, three consecutive isokinetic contractions at 60°/s were performed. Following 60 seconds rest, another three contractions at the same speed were performed. Individuals were provided with verbal and visual feedback during all contractions. The highest torque value produced was recorded as the peak torque value for analysis. This was performed on both arms.

**Statistical Analyses**

Statistical analyses were performed using the Bayes Factor package in R version 3.5.3 (RStudio, Inc., Boston, MA, 2019) and JASP version 0.9.2 (JASP Team, Amsterdam, The Netherlands, 2019). A Bayesian ANCOVA was used to test for mean changes across groups while adjusting for pre-values. All analyses were computed using uninformed priors ($r=0.5$ for fixed effects and $r=0.354$ for the covariate) as suggested previously (Wagenmakers et al. 2018). To assess whether the variability in response to each exercise intervention differed from that of the control group, Levene’s tests were computed to compare variances. Bayes factors ($BF_{10}$) were used to provide support for the null ($BF_{10} \leq 0.33$) or alternate ($BF_{10} \geq 3$) hypotheses. A $BF_{10}$ in the range of 0.34 to 2.99 represents ambiguity.

If there was support for the alternative hypothesis, indicating an added variability that could be attributed to the exercise intervention, we quantified how many individuals were distinguishable from random error using methods previously discussed by our laboratory (Dankel and Loenneke
Briefly, error was quantified by multiplying the change score standard deviation of the control group by 1.96 to encompass 95% of the responses in the absence of an exercise intervention. Then to quantify individuals as being differential responders in the intervention group, we added and subtracted this degree of random error (1.96 × the control group change score standard deviation) from the mean intervention response to detail the number of individuals that could be identified as responding differently from the mean. Only those individuals exceeding this level of random error above or below the mean response were classified as differential responders. An additional analysis (post-hoc) was performed using the methods of Swinton et al. (2018) to predict the proportion of response to exercise (i.e. the proportion of individuals who exceeded the smallest worthwhile change of 0.2 baseline standard deviation units). Confidence intervals (CI) for these estimates were obtained using a bootstrapping procedure which was provided as supplementary material from the authors of the aforementioned paper (Swinton et al. 2017).

Results

Descriptive statistics and exercise adherence

A total of 151 of the 158 individuals (96%) completed the study. There were 3 males (2 in the traditional exercise group and 1 in the control group) and 4 females (3 in the traditional exercise group and 1 in the control group) that dropped out of the study for unrelated reasons. Except for one individual in the 1RM training group who missed two exercise sessions, all individuals completed all training sessions. There were no injuries or adverse events that occurred from either training or testing. Descriptive statistics are shown in Table 1. By design, the traditional exercise group completed more repetitions and a greater volume of exercise.
Mean changes in 1RM and isokinetic strength

Changes in 1RM strength are shown in Figure 1A. The results of the ANCOVA demonstrated a difference in 1RM strength change across groups ($BF_{10} = 405,502$) where strength increased in both the 1RM ($BF_{10} = 4.791e+6$) and traditional ($BF_{10} = 11,915$) exercise groups when compared to the control group. The strength increases in the 1RM and traditional exercise groups were equivalent ($BF_{10} = 0.21$). With respect to strength changes in the untrained arm, the ANCOVA demonstrated a difference across groups ($BF_{10} = 73.5$). Follow-up comparisons demonstrated the 1RM training group increased untrained arm 1RM strength more so than both the control ($BF_{10} = 271$) and traditional ($BF_{10} = 3.00$) exercise groups. The traditional exercise group did not increase untrained arm 1RM strength when compared to the control group ($BF_{10} = 0.530$).

Changes in isokinetic strength are shown in Figure 1B. The initial ANCOVA demonstrated there were no differences in the changes in isokinetic strength of either the trained ($BF_{10} = 0.18$) or untrained ($BF_{10} = 0.21$) arm when made relative to the control group. The estimated proportion of individuals exceeding the smallest worthwhile change for 1RM strength of the trained arm was 79% (95% CI: 67 – 95) in the traditional exercise group and 99% (95% CI: 78 – 100) in the 1RM training group. With respect to 1RM strength of the untrained arm it was estimated that 91% (95% CI: 66 – 100) of individuals in the 1RM training group and 41% (95% CI: 10 – 64) of individuals in the traditional exercise group exceeded the smallest worthwhile change.

Mean changes in muscle size

Changes in muscle size of the trained arm are shown in Figure 1C. Each of the three measurement sites demonstrated a difference in muscle size across groups (all $BF_{10} \geq 332$). All
sites yielded the same conclusion in that muscle growth of the trained arm was only present in the traditional exercise group when compared to that of the control group (50% proximal site $BF_{10} = 3,733$, 60% mid site $BF_{10} = 7,184$, 70% distal site $BF_{10} = 224$). The increase in muscle size of the traditional exercise group also exceeded that of the 1RM group (50% proximal site $BF_{10} = 62.0$, 60% mid site $BF_{10} = 26.0$, 70% distal site $BF_{10} = 5.66$). Changes in muscle size of the untrained arm are shown in Figure 1D. There were no differences in muscle size for the untrained arm in either the 1RM or traditional exercise groups when compared to the control group. Specifically, the $BF_{10}$ for the ANCOVA provided support for the null hypothesis at each of the three sites as follows: proximal 50% site $BF_{10} = 0.070$, 60% mid site $BF_{10} = 0.067$, and 70% distal site $BF_{10} = 0.086$. The estimated proportion of individuals in the traditional exercise group exceeding the smallest worthwhile change for muscle thickness was 80% (95% CI: 60 – 99), 83% (95% CI: 64 - 99) and 79% (95% CI: 59 - 99) for the proximal, mid, and distal sites, respectively.

Variability in muscle size and strength

All variability statistics and results of the Levene’s test are shown in Table 2. None of the variability in muscle growth of either exercise group differed from the control group, and this was true for both the trained and untrained arm. A histogram is shown in Figure 2 detailing a similar degree of variability for the change in muscle size of the trained arm across each of the groups. When examining differences in variability for 1RM strength of the trained arm, there was no difference between the variability in the 1RM and control group, but there was between the traditional exercise and control groups. No differences were present in the variability for 1RM strength of the untrained arm for either the 1RM or traditional exercise groups. With
respect to isokinetic strength, there were no differences in either group for variability of either
the trained or untrained arms. A histogram of the changes in 1RM strength of the trained arm is
shown in Figure 3. As there was true variability in the change in 1RM strength of the traditional
exercise group, we demonstrate a method by which only those confidently responding differently
from the mean response can be analyzed (Figure 4). Notably, after teasing out random error, 10
of the 48 individuals (20.8%) could be identified as responding differently from the mean
response (6 high responders and 4 low responders).

Discussion

Main findings

The main findings of this study were as follows: (1) except for the change in 1RM strength in the
trained arm of the traditional exercise group, none of the variability present could be
distinguished from random error, (2) increases in muscle size were only present in the trained
arm of the traditional exercise group, (3) similar increases in 1RM strength of the trained arm
were observed for both the training groups, and (4) the cross-over effect of strength to the
untrained arm was only present in the 1RM group.

Differential responders to exercise

The idea that we were not able to distinguish differential responders from random error in all
variables except for 1RM strength of the traditional exercise group brings into question the
findings of previous resistance training studies not including control groups (Chmelo et al. 2015;
Gentil et al. 2017). It is entirely plausible that studies using traditional approaches of classifying
differential responders as those more or less than one standard deviation from the mean (Erskine
et al. 2010; Ahtiainen et al. 2016), those in the top or bottom percentiles (Davidsen et al. 2011; Morton et al. 2018) of responses, or cluster analyses (Bamman et al. 2007; Kim et al. 2007; Petrella et al. 2008; Thalacker-Mercer et al. 2009, 2013; Roberts et al. 2018; Haun et al. 2019)(Bamman et al. 2007; Kim et al. 2007; Petrella et al. 2008; Thalacker-Mercer et al. 2009, 2013; Roberts et al. 2018; Haun et al. 2019), are simply analyzing random error as opposed to true differences in response to exercise. The importance of quantifying the degree of random error that would have been present in the absence of an exercise intervention becomes particularly apparent in the inability to replicate various findings. For example, studies have suggested that individuals who incorporate more myonuclei with training (Petrella et al. 2008) or have higher baseline androgen receptor content (Morton et al. 2018) gain more muscle mass, but these findings have not been able to be replicated (Mobley et al. 2018; Haun et al. 2019). This may be due to individuals being categorized based on random error as opposed to true differences in the exercise response. This is supported in that the entirety of the variability in muscle growth in the present study could not be distinguished from random error. To minimize the number of individuals misclassified as high and low responders, only those individuals that exceed random error should be analyzed (Figure 4). While the present study pertains to resistance exercise, studies showing large degrees of variability in response to aerobic exercise (Bouchard and Rankinen 2001) also appear to be largely driven by random error (Hopkins 2015). While our results would suggest otherwise, we cannot be certain that previous studies were not analyzing true variability with respect to the changes in muscle size occurring with resistance exercise. The idea that associations were found between different exercise responses and physiologic variables may provide some indication that true variability was
present within these studies. As such, more research is necessary to examine whether there is true variability present in response to resistance exercise, and if so, if this variability can be explained by physiological differences across individuals.

The reason as to why there was variability in the change in 1RM strength in the traditional exercise group, but not the 1RM group is unknown, but may illustrate that the magnitude of variability present may depend on the exercise protocol employed. Specifically, it would appear as though the variability in 1RM strength present in the traditional exercise group resulted from differences in how well the skill of performing 8-12 repetitions transferred to the skill of performing a 1RM. This is supported by the idea that the 1RM group, that trained by performing the 1RM test, did not have any true variability in the change in 1RM strength (i.e. not statistically different from control group variability). Thus, performing the 1RM test may provide more consistent increases in 1RM strength relative to performing a more traditional resistance exercise protocol.

*Mean differences in training protocols*

Our mean results corroborate the findings of previous studies within our laboratory examining changes in muscle size and strength (Dankel et al. 2017; Mattocks et al. 2017). Muscle growth was only present in the trained arm of the traditional exercise group, indicating that a sufficient volume of exercise is needed to induce muscle growth. This also demonstrates that, despite drastically lower exercise volume and the absence of muscle hypertrophy, performing the 1RM strength test results in similar 1RM strength increases compared to traditional exercise. Large increases in 1RM strength have been documented previously in response to performing the 1RM
test (12-22% increases) (Ploutz-Snyder and Giamis 2001). Interestingly, the cross-over effect (i.e. an increase in 1RM strength of the untrained arm) was only present in the 1RM group when compared to the control and traditional exercise groups. Our finding that there was not a significant cross-over effect in the traditional exercise group corroborates the results of a previous meta-analysis demonstrating that most individual studies do not observe a significant effect when compared to a control group (Carroll et al. 2006). The idea that there was a cross-over effect in the 1RM group may indicate that the cross-over effect is more sensitive to the principle of specificity and the load being lifted, given that there were no differences in strength changes of the trained arm across both training groups.

Conclusion

The primary finding of the present study was that only the variability in 1RM strength of the traditional exercise group could be distinguished from random error. Thus, categorizing individuals into groups based on any other variables would simply be separating these individuals based off random error as opposed to true differences in the exercise response. When there is a greater variability in the exercise group compared to that of a control group, we provide a method to only examine those individuals that can be confidently classified as differential responders. Our mean results across protocols demonstrate that 1RM strength gains can be achieved by lifting heavy loads (i.e. a 1RM – 10RM) and these strength gains are not dependent upon increasing muscle mass or performing a large volume of exercise. The cross-over effect of strength may be more apparent when very heavy loads are lifted.
Acknowledgements: None

Conflict of Interest: The authors declare that they have no conflict of interest

Funding: The preparation of this manuscript was funded by the School of Applied Sciences at The University of Mississippi
References


Bayesian inference for psychology. Part II: Example applications with JASP. Psychon. Bull.
Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Control (n=51)</th>
<th>IRM (n=52)</th>
<th>Traditional (n=48)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males (n)</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Age (years)</td>
<td>21 (3)</td>
<td>20 (1)</td>
<td>21 (2)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>168 (9)</td>
<td>169 (13)</td>
<td>169 (9)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>72.2 (15.7)</td>
<td>74.6 (22.0)</td>
<td>72.6 (17.2)</td>
</tr>
<tr>
<td><strong>Baseline muscle thickness of the trained arm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximal 50% site (cm)</td>
<td>2.66 (0.57)</td>
<td>2.72 (0.57)</td>
<td>2.79 (0.61)</td>
</tr>
<tr>
<td>Mid 60% site (cm)</td>
<td>2.88 (0.56)</td>
<td>2.92 (0.56)</td>
<td>2.99 (0.57)</td>
</tr>
<tr>
<td>Distal 70% site (cm)</td>
<td>3.22 (0.57)</td>
<td>3.27 (0.60)</td>
<td>3.32 (0.61)</td>
</tr>
<tr>
<td><strong>Baseline muscle thickness of the untrained arm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximal 50% site (cm)</td>
<td>2.61 (0.60)</td>
<td>2.65 (0.54)</td>
<td>2.71 (0.66)</td>
</tr>
<tr>
<td>Mid 60% site (cm)</td>
<td>2.78 (0.58)</td>
<td>2.83 (0.55)</td>
<td>2.89 (0.62)</td>
</tr>
<tr>
<td>Distal 70% site (cm)</td>
<td>3.10 (0.60)</td>
<td>3.18 (0.59)</td>
<td>3.22 (0.61)</td>
</tr>
<tr>
<td><strong>Baseline strength of the trained arm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1RM strength (kg)</td>
<td>13.5 (4.6)</td>
<td>13.3 (4.5)</td>
<td>13.6 (5.2)</td>
</tr>
<tr>
<td>Isokinetic strength (Nm)</td>
<td>35.4 (12.7)</td>
<td>36.9 (14.2)</td>
<td>36.6 (12.4)</td>
</tr>
<tr>
<td><strong>Baseline strength of the untrained arm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1RM strength (kg)</td>
<td>12.9 (4.4)</td>
<td>12.7 (4.5)</td>
<td>12.9 (4.9)</td>
</tr>
<tr>
<td>Isokinetic strength (Nm)</td>
<td>32.4 (12.6)</td>
<td>34.9 (14.2)</td>
<td>34.5 (12.2)</td>
</tr>
<tr>
<td><strong>Training Completed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total repetitions</td>
<td>0 (0)</td>
<td>76 (6)</td>
<td>732 (72)</td>
</tr>
<tr>
<td>Repetitions completed per session</td>
<td>0 (0)</td>
<td>4.2 (0.3)</td>
<td>40.6 (4.0)</td>
</tr>
<tr>
<td>Total exercise volume (kg)</td>
<td>0 (0)</td>
<td>1,008 (390)</td>
<td>7,244 (2,150)</td>
</tr>
<tr>
<td>Exercise volume completed per session (kg)</td>
<td>0 (0)</td>
<td>56 (21)</td>
<td>402 (119)</td>
</tr>
</tbody>
</table>

All results are expressed as mean (standard deviation).
Table 2. Variability Statistics

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>1RM</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trained arm muscle size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% proximal site</td>
<td>0.21 (7.9%)</td>
<td>0.20 (7.0%)</td>
<td>0.25 (7.2%)</td>
</tr>
<tr>
<td>60% mid site</td>
<td>0.20 (7.9%)</td>
<td>0.19 (6.8%)</td>
<td>0.25 (6.0%)</td>
</tr>
<tr>
<td>70% distal site</td>
<td>0.23 (9.4%)</td>
<td>0.18 (8.2%)</td>
<td>0.29 (8.5%)</td>
</tr>
<tr>
<td><strong>Untrained arm muscle size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% proximal site</td>
<td>0.23 (8.0%)</td>
<td>0.20 (7.9%)</td>
<td>0.16 (6.7%)</td>
</tr>
<tr>
<td>60% mid site</td>
<td>0.21 (7.4%)</td>
<td>0.19 (6.8%)</td>
<td>0.16 (6.1%)</td>
</tr>
<tr>
<td>70% distal site</td>
<td>0.21 (6.5%)</td>
<td>0.17 (5.6%)</td>
<td>0.18 (5.9%)</td>
</tr>
<tr>
<td><strong>Trained arm strength</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1RM strength</td>
<td>1.46 (10.3%)</td>
<td>1.54 (12.4%)</td>
<td>2.32 (21.8%)</td>
</tr>
<tr>
<td>Isokinetic strength</td>
<td>3.22 (9.2%)</td>
<td>3.35 (10.5%)</td>
<td>4.67 (11.7%)</td>
</tr>
<tr>
<td><strong>Untrained arm strength</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1RM strength</td>
<td>1.31 (12.3%)</td>
<td>1.38 (12.6%)</td>
<td>1.41 (12.3%)</td>
</tr>
<tr>
<td>Isokinetic strength</td>
<td>3.56 (12.5%)</td>
<td>4.01 (10.9%)</td>
<td>4.80 (14.0%)</td>
</tr>
</tbody>
</table>

Bayes Factors Comparing Variability with the Control Group

<table>
<thead>
<tr>
<th></th>
<th>1RM</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trained arm muscle size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% proximal site</td>
<td>0.232</td>
<td>0.247</td>
</tr>
<tr>
<td>60% mid site</td>
<td>0.247</td>
<td>1.00</td>
</tr>
<tr>
<td>70% distal site</td>
<td>0.473</td>
<td>0.526</td>
</tr>
<tr>
<td><strong>Untrained arm muscle size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% proximal site</td>
<td>0.235</td>
<td>0.290</td>
</tr>
<tr>
<td>60% mid site</td>
<td>0.212</td>
<td>0.387</td>
</tr>
<tr>
<td>70% distal site</td>
<td>0.258</td>
<td>0.241</td>
</tr>
<tr>
<td><strong>Trained arm strength</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1RM strength</td>
<td>0.643</td>
<td>5.381</td>
</tr>
<tr>
<td>Isokinetic strength</td>
<td>0.255</td>
<td>0.959</td>
</tr>
<tr>
<td><strong>Untrained arm strength</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1RM strength</td>
<td>0.273</td>
<td>0.219</td>
</tr>
<tr>
<td>Isokinetic strength</td>
<td>0.219</td>
<td>0.667</td>
</tr>
</tbody>
</table>

The top half of the table shows the standard deviations of the change from pre to post. The standard deviation was divided by the mean value at baseline to provide an indication of relative variability in parentheses. On the bottom half of the table are Bayes factors for the Levene’s test comparing the variance within each group to that of the control group. These can be interpreted as odds ratios comparing the support for the alternative hypothesis divided by support for the null hypothesis such that a Bayes factor $\leq 0.33$ provides support for the null hypothesis and a Bayes factor $\geq 3$ provides support for the alternative hypothesis.
FIGURE LEGEND

**Figure 1.** Mean changes in strength and muscle size across groups.

A) Changes in 1RM strength. B) Changes in isokinetic strength. C) Changes in muscle thickness of the trained arm. D) Changes in muscle size of the untrained arm. All results are expressed as median changes and 95% credible intervals. * different from the control group, # different from the traditional exercise group, & different from the untrained arm within group. A difference was identified if the Bayes factor was greater than or equal to 3, indicating the alternative hypothesis was three times as likely as the null hypothesis.

**Figure 2.** Histogram detailing the changes in muscle size across groups.

Each individual contributed three measurements to the histogram, one for each of the three sites of the arm measured. Therefore, there are 153 measurement in the traditional group (51 × 3), 156 measurements in the 1RM group (52 × 3) and 144 measurements in the traditional exercise group (48 × 3). The results of the Levene’s test demonstrated that neither of the training groups was more variable than the control group for any of the three sites measured. Most previous studies do not include a control group and conclude that the variability occurs in response to exercise, but we show that the same degree of variability would have been present even if individuals did not exercise. Specific variability statistics are shown in Table 2.

**Figure 3.** Histogram detailing the changes in 1RM strength across groups.

The results of the Levene’s test demonstrated the variability in the traditional exercise group exceeded that of both the control and 1RM groups. There was no difference in the variability of the 1RM and control groups. Specific variability statistics are shown in Table 2.
**Figure 4.** Classifying individuals exceeding random error.

The bars placed over the individual data points indicate the mean and interval of random error around the mean. The range of random error was quantified as $1.96 \times$ the control group SD on the change in 1RM strength from pre to post. Using the variability reported in Table 2, the random error would be calculated as $2.86 \times 1.46$. This degree of random error (2.86) was then added and subtracted to the mean response of each exercise group. As the Levene’s test demonstrated the variability in the 1RM group did not differ from the control group, we would not recommend employing this method for the 1RM group, although researchers may wish to if there is deemed to be a meaningful degree of true variability. The Levene’s test demonstrated there was a difference in variance between the traditional training group and the control group and thus we would recommend using this method to eliminate noise in classifying individuals into responder categories. Only those individuals above ($n=6$) and below ($n=4$) the error from the mean can be identified as high and low responders, respectively. The rest of the variability cannot be distinguished from random error and we would not recommend categorizing these individuals as differential responders.