# Improbable data patterns in the work of Barbalho *et al.*

Andrew D. Vigotsky*[1], Greg Nuckols[2], James Fisher[3], James Heathers[4], James Krieger[5], Brad J. Schoenfeld[6], Jürgen Giessing[7], and James Steele[3]

[1]Northwestern University, [2]Stronger By Science LLC, [3]Southampton Solent University, [4]Northeastern University, [5]Weightology LLC, [6]Lehman College, [7]University of Koblenz-Landau

April 22, 2021

## Executive Summary

1. The studies by Barbalho *et al.* have extremely homogeneous baseline strength levels compared to the rest of the literature. In particular, we observed homogeneity up to ∼7.5 $z$-score units below what would be expected given the mean value. This homogeneity was not just extreme across one study or variable; rather, homogeneity was present across many studies, and many variables within each study. Simultaneous homogeneity across many variables is improbable. Finally, homogeneity was also present for variables that could not have been measured at baseline (muscle thickness and change scores). Therefore, biased sampling alone cannot explain this degree of homogeneity.

2. The effect sizes observed are both large and homogeneous. From a magnitude perspective, effect sizes for strength increases in the studies by Barbalho *et al.* were up to 13.5 $z$-score units greater than those in the rest of the resistance training literature. From a signal-to-noise perspective, multiple signal-to-noise effect sizes were undefined since the responses were perfectly homogeneous (i.e., standard deviation of change scores equal to zero). Excluding the perfectly homogeneous effects, the signal-to-noise effect sizes for strength increases reported by Barbalho *et al.* were up to 34 $z$-score units greater than those in the rest of the resistance training literature. While standardized effect sizes tend to scale with percent increases in strength in the literature, they do not in the studies by Barbalho *et al.*

3. The men's and women's volume studies are remarkably similar in terms of their observed effects and correlation structures. This is despite both studies being independent, and each study being randomized. These across-study consistencies yield $P < 1 \times 10^{-6}$ when we would in fact expect the null hypothesis to be true due to randomization. In addition, there is structure in raw data that is inconsistent with randomization (again, $P < 1 \times 10^{-6}$). Other patterns in the raw data, such as twice the number of even as odd numbers, were also noted—this holds even after removing the strength data.

4. In the single- vs. multi-joint vs. single+multi-joint studies, the effects observed in the multi-joint group nearly perfectly match those in the single+multi-joint group. This holds across studies.

5. Several patterns exist in the raw data, including "runs" of numbers and strength values for one exercise being exactly 8 kg more than those for another exercise (for the entire sample).

6. Squat strength increases in the recent squat versus hip thrust and single versus multi-joint papers are far beyond what would be expected for trained women of similar strength to those in the study. Even women who did not squat increased their squat strength at a rate of more than 2 $z$-score units above powerlifters who specifically train the movement. In those who did squat, $z$-scores of over 5 were observed.

7. In the elderly study, 98% of the sample lost weight from a resistance training intervention alone; no dietary intervention was implemented. This is in contrast to what is known about the role of exercise in weight loss and in contrast to other studies. This study also contained methodological inconsistencies, such as large imbalances in group size despite using block randomization.

8. We provide a statistical rationale for why the observed baseline homogeneities are not likely to stem from biased sampling; namely, because one would need to screen too many people.

---

*Corresponding Author: Andrew Vigotsky (avigotsky@gmail.com)

The first two authors, ADV and GN, were determined by their relative contributions. The rest of the authors are listed in alphabetical order by last name.

This work has not been peer-reviewed.

All authors have read and approved this version of the white paper for preprint.

Authors ADV (@avigotsky), GN (@GregNuckols), JF (@jpfisheruk), JH (@jamesheathers), JK (@YngvaiMalmsteve), BJS (@BradSchoenfeld), and JS (@JamesSteeleII) can be reached on Twitter.

This work can be cited as: Vigotsky AD, Nuckols G, Fisher J, Heathers J, Krieger J, Schoenfeld BJ, Giessing J, Steele J. (2020). Improbable data patterns in the work of Barbalho *et al.*, *SportRχiv*, doi:10.31236/osf.io/sg3wm

# 1 Data Anomalies

## 1.1 Statistical Properties Relative to Other Studies

### 1.1.1 Variances and Coefficients of Variation

We first became curious about the data in the studies authored by Barbalho when we consistently observed very tight SDs across nearly all measures and studies; SDs typically scale with mean values. Thus, we quantitatively addressed this observation using the reported strength measures in the studies by Barbalho *et al.* We have a database of 68 other studies [1–68], which was gathered as systematically as possible over the years for various articles (comparing periodized and non-periodized training, strength gains in male vs. female subjects, and analyzing the impact of frequency on strength gains). In these studies, SDs increase linearly as means increase, meaning CVs remain virtually unchanged, on average, as means increase (Figure 1a,b). However, the studies by Barbalho follow a different trend—the SDs are relatively constant across means, and thus, CVs decrease with increasing means (Figure 1a,b).

A more quantitative evaluation of the variances reported in the Barbalho studies reveals that, indeed, the variances are remarkably tight. We created a meta-regression based on the 68 studies; we used the resulting prediction interval to calculate $z$-scores to estimate how extreme Barbalho *et al.*'s variances are. We observe $z$-scores as low as $z \approx -7.5$, which is equivalent to a $P$-value of $3.2 \times 10^{-14}$ (Figure 1c). Examining Figure 1c, one can see that several of Barbalho *et al.*'s studies contain not just one, but many instances of extremely small variances relative to the rest of the literature. The degree of homogeneity is noteworthy.

### 1.1.2 Effect Sizes

The effects observed in the studies by Barbalho *et al.* are large from two perspectives: their magnitudes and consistency (signal-to-noise). These can be represented by Glass' $\Delta_{\text{pre}} = \bar{\delta}/\sigma_{\text{pre}}$ and Cohen's $d_z = \bar{\delta}/\sigma_{\delta}$, respectively, where $\bar{\delta}$ is the mean change score within a group, $\sigma_{\text{pre}}$ is the standard deviation of the baseline scores, and $\sigma_{\delta}$ is the standard deviation of change scores.

**Magnitude-based effect sizes.** Partially as a result of the small standard deviations, these studies also exhibit exceptionally large magnitude-based effect sizes, disproportionate to the actual changes in performance seen in the studies (Figure 2a,c). One other study had comparable effect sizes, also due to abnormally small standard deviations [62]. Within the rest of the studies analyzed, there was a strong ($r = 0.83$) linear relationship between percentage increases in strength measures and effect sizes ($\Delta_{\text{pre}}$), with many of the effect sizes in Barbalho's research strongly deviating from this trend (Figure 2e). 9 of the 10 effect sizes over $\Delta_{\text{pre}} = 10$ were found in Barbalho's studies, as well as 23 of the 34 effect sizes over $\Delta_{\text{pre}} = 5$. There were 16 effect sizes of $\Delta_{\text{pre}} > 5.0$ in Barbalho's studies from measures with strength increases below 28%. That pair of outcomes did not occur in any other study.

**Signal-to-noise effect sizes.** The effects reported by Barbalho *et al.* are also more consistent than those in literature (Figure 2b). We calculated and compared Cohen's $d_z$'s for the studies by Barbalho *et al.* and compared them to the literature using a random-effects meta-analysis with robust variance estimation. Of note, there were three outcomes in Barbalho *et al.* [74] for which the standard deviation of change scores was zero (i.e., perfectly homogeneous effects), meaning $d_z$ was undefined and could not be included. We observed $z$-scores as high as 34. Because Cohen's $d_z$ is dependent on the change scores, not necessarily baseline scores, effects this large/consistent cannot be wholly attributed to biased sampling. Like with Glass' $\Delta_{\text{pre}}$, Cohen's $d_z$ correlated with relative change scores in the rest of the literature ($r = 0.64$), with the effects reported by Barbalho *et al.* deviating from this trend (Figure 2f).

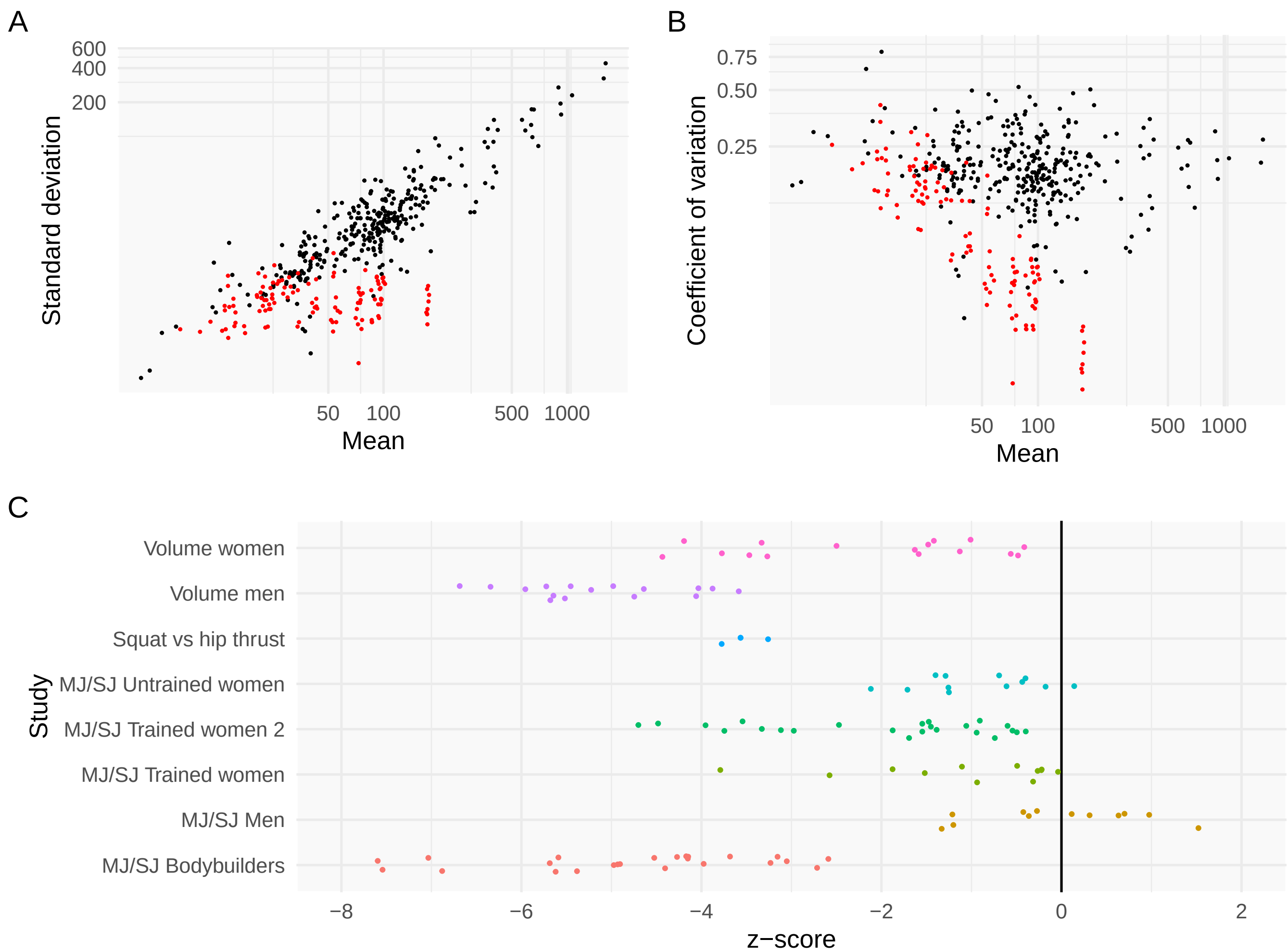



Figure 1: Studies by Barbalho *et al.* have tighter-than-expected baseline strength SDs which do not scale with mean values. (A) While much of the literature's SDs increase with mean values (black), Barbalho *et al.*'s SDs do not (red). As a result, (B) the CVs of Barbalho *et al.*'s studies decrease with increasing means, while much of the literature has a constant CV. (C) Results from a meta-regression with robust variance estimation reveal the degree to which the baseline homogeneity of strength in Barbalho *et al.*'s studies is surprising, with $z$-scores as low as $z \approx -7.5$, equivalent to $P = 3.2 \times 10^{-14}$. We used [1–68] as comparison studies, and the studies by Barbalho *et al.* are as follows:

1. Volume women [69]
2. Volume men [70]
3. Squat vs. hip thrust [71]
4. MJ/SJ Untrained women [72]
5. MJ/SJ Trained women 2 [73]
6. MJ/SJ Trained women [74]
7. MJ/SJ Men [75]
8. MJ/SJ Bodybuilders [76]

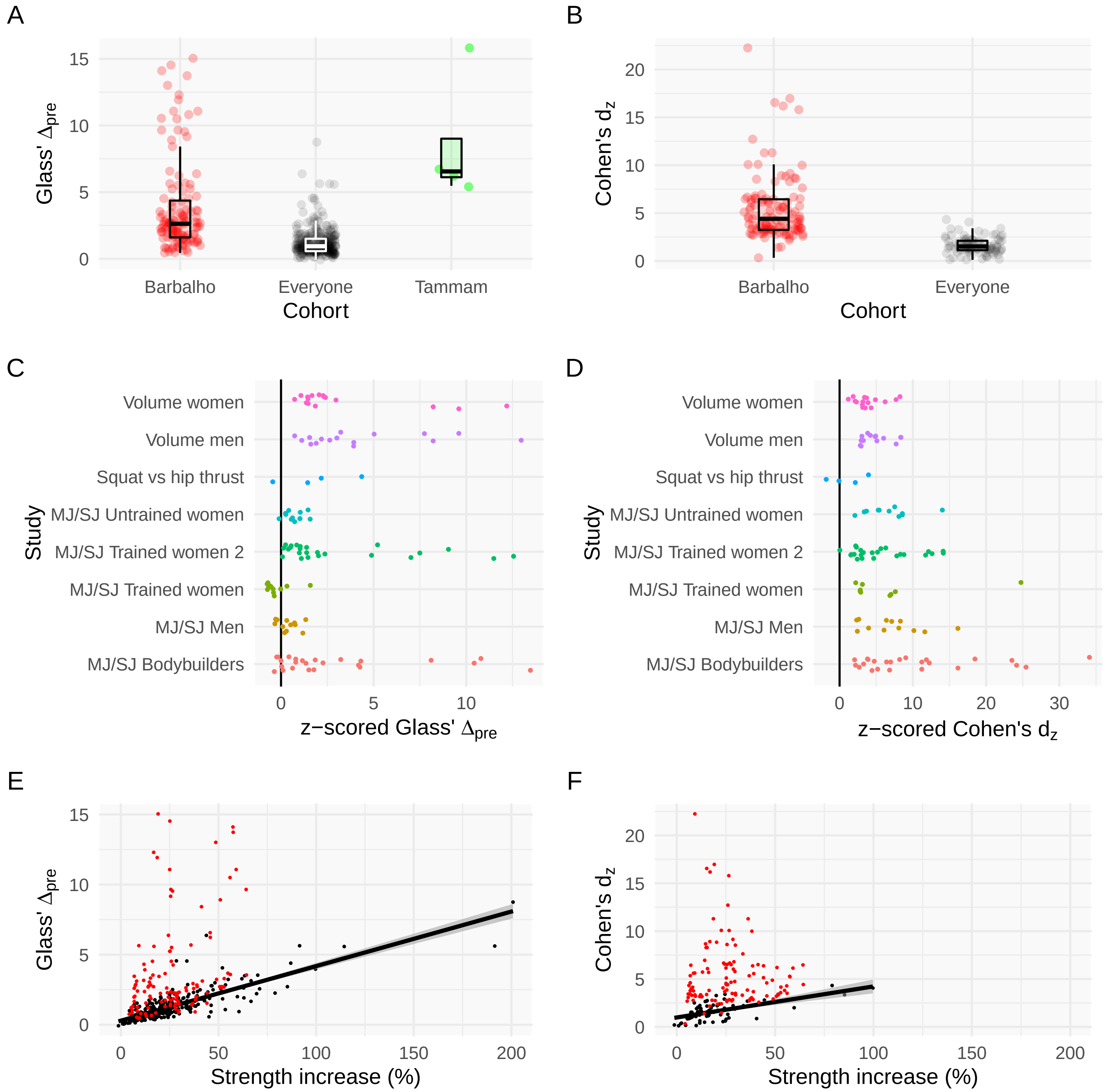



Figure 2: Studies by Barbalho *et al.* have larger effect sizes than the rest of the literature. (A) Magnitude-based effect sizes (Glass' $\Delta_{\text{pre}}$) observed in the studies by Barbalho *et al.* are often much higher than the average observed across the literature, with the exception of those from a single study [62]. (B) Signal-to-noise effect sizes (Cohen's $d_z$) are, again, greater in the studies by Barbalho *et al.* compared to the rest of the literature. (C) illustrates the magnitude-based effect sizes from the studies by Barbalho *et al.* $z$-scored relative to the rest of the literature. Note, these are crude estimates since we did not use a random-effects meta-analytic model to calculate the mean and SD of the literature values. Nevertheless, some $z$-scores are as high as $z = 13.5$, or $P = 1.3 \times 10^{-41}$. (D) illustrates the magnitude-based effect sizes from the studies by Barbalho *et al.* $z$-scored relative to the rest of the literature. In contrast to (C), the $z$-scores in (D) were calculated using robust variance estimation and random-effects meta-analysis. There were three outcomes in Barbalho *et al.* [74] for which the standard deviation of change scores was zero (i.e., perfectly homogeneous effects), meaning $d_z$ was undefined and could not be included. $z$-scores based on Cohen's $d_z$ are as high as 34, or $P = 4 \times 10^{-255}$; this is as unlikely as a fair coin landing on heads 845 times in a row. (E–F) Percent increases in strength correlate with (E) Glass' $\Delta_{\text{pre}}$ ($r = 0.83$) and (F) Cohen's $d_z$ ($r = 0.64$) in most studies, but the effects observed in the studies by Barbalho *et al.* deviate from this trend.

## 1.2 Volume Studies

Two of Barbalho's papers are methodologically parallel (with the exception of mid-point assessments), six-month volume studies, each with a separate groups of participants (one includes exclusively trained males, while the other includes exclusively trained females) [69, 70]. Despite being separate groups and studies, the data are strikingly similar in several ways. The results obtained by the corresponding male and female groups in both studies (e.g., male G5 change in squat 10RM vs. female G5 change in squat 10RM, male G15 change in biceps thickness vs. female G15 change in biceps thickness, etc.) have virtually identical raw effects, effect standard deviations, and standardized mean differences. Figure 3 displays these values for both the male study on the $x$-axis [70], and the corresponding effect sizes from the female study on the $y$-axis [69].

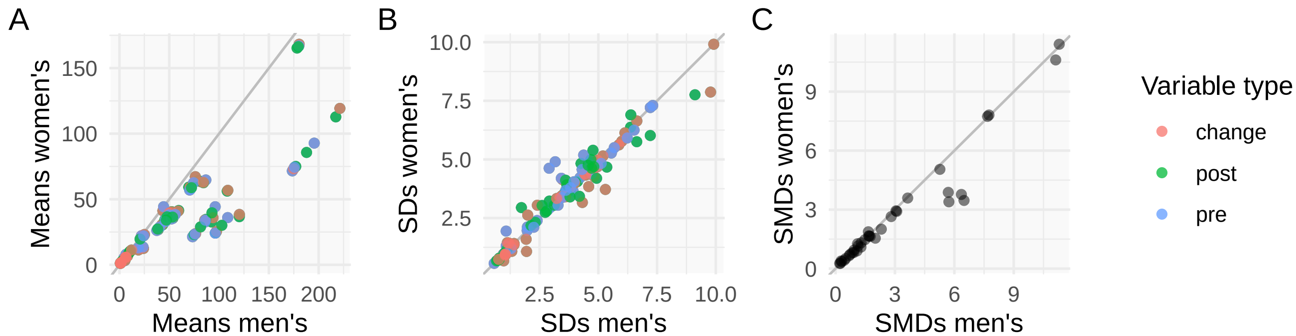
By looking at the raw data, we discovered that not only are the means, standard deviations, and effect sizes for the primary outcomes virtually identical, but so too are the correlations between pairs of individual variables. For example, if the correlation between two potentially unrelated variables is $r = 0.3$ in the G5 males, it will probably be very close to $r = 0.3$ in the G5 females. This holds for all correlations between two variables in corresponding groups, including variables where the correlation should be essentially random. Figure 4 shows color-coded heat maps (correlograms), where blue is a positive association, and red is negative. The two leftmost groups are G5 females and males. The next two are G10 females and males, etc. The mosaic pattern between each corresponding pair is virtually identical. The strength of the correlations between the correlation coefficients for corresponding groups in the two studies is $r > 0.8$ in all four cases. As a point of reference, G5 and G10 reported overall similar strength and hypertrophy results within both studies. However, the strength of the correlation between corresponding correlation coefficients in G5 vs. G10 in the male study is $r = 0.35$; for females, $r = 0.26$ (you can just compare the differences in patterns between the first and second mosaics in each row) (Figure 4). This strongly suggests an unexplained regularity between sources.

Because these correlograms include the effects of the intervention (i.e., change scores and post-intervention assessments), it is possible they are largely dominated by these columns. Thus, we also assessed the correlations of variables collected only at baseline, and the story is identical: unexplained regularities are present. Note that, despite each study being independently randomized, baseline correlations are strong between but not within studies (Figure 4).

The baseline scores have favorable theoretical properties in that, since there is a randomization scheme (i.e., groups are randomized at baseline), there is an easily calculable null distribution. This can be calculated by re-randomizing the groups and comparing the simulated baseline correlation matrices to the observed ones. We converted the correlations to Fisher's $z$, then used the sum of squared differences in Fisher's $z$'s (re-normalized to $z$-score units) between each of the correlations as a distance metric (analogous to a $\chi^2$ statistic). We performed this on each group individually and on the study as a whole (all groups together). On a group-by-group basis, using 100,000 permutations, the resulting one-sided $P$-values for G5, G10, G15, and G20, when comparing the similarity of the men's and women's correlation matrices, are $< 1 \times 10^{-5}$, 0.0006, $1 \times 10^{-5}$, and $7 \times 10^{-5}$, respectively (Figure 5). This indicates that correlation matrices between men and women for a given group are much more similar than we would expect for having randomized samples.

Next, we randomized all four groups (the entire study) at the same time rather than each group individually. This allowed us to calculate how extreme the observed similarity is across all groups at once. The process was similar: we re-randomized all individuals to one of four groups. None of the 1,000,000 simulations produced results more similar than what was observed in the real data (i.e., $P < 1 \times 10^{-6}$). Histograms of the observed distances (red) compared to the null distributions (grey) can be observed above (Figure 5). The consistency in distributions across studies is incredibly improbable.

When looking at the raw data from the men's and women's volume studies, it is apparent that there are twice as many even numbers as odd numbers. Distributions can be found in Figure 6. This relationship holds with and without the strength data which could conceivably be expected to consist primarily of even numbers, if the researcher primarily increased loads in increments of 2 kg when assessing strength. It is unclear how this could have happened, and its observation relative to




D

**Women**

TABLE 3. Groups characteristics.

| Group | Age | Height | Body Mass | Experience |
|---|---|---|---|---|
| G5 | 24.9 ± 1.97 | 165.3 ± 4.06 | 63.4 ± 4.14 | 3.3 ± 0.95 |
| G10 | 24.6 ± 1.17 | 168.2 ± 3.68 | 64.7 ± 4.90 | 3.2 ± 1.03 |
| G15 | 25.1 ± 1.20 | 167 ± 4.40 | 62.6 ± 4.67 | 3.6 ± 0.70 |
| G20 | 24.1 ± 1.20 | 166.4 ± 4.20 | 62.9 ± 3.84 | 3.5 ± 0.97 |

G5—5 sets per week per muscle group, G10—10 sets per week per muscle group, G15—15 sets per week per muscle group, G20—20 sets per week per muscle group.

**Men**

Table 3 Groups Characteristics

| Group | Age | Height | Body mass | Experience |
|---|---|---|---|---|
| G5 | 24.9 (1.97) | 178.3 (4.11) | 81.5 (3.47) | 5.3 (0.95) |
| G10 | 24.6 (1.17) | 180.4 (3.7) | 86.7 (3.16) | 5.2 (1.03) |
| G15 | 25 (1.22) | 180.1 (4.7) | 84.33 (5.36) | 5.6 (0.72) |
| G20 | 24.25 (1.28) | 179.6 (4.9) | 84.24 (4.59) | 5.5 (1.07) |

Abbreviations: G5, 5 sets per week per muscle group; G10, 10 sets per week per muscle group; G15, 15 sets per week per muscle group; G20, 20 sets per week per muscle group. Note: Values are represented mean (SD).

Figure 3: Relationships between data from the men's and women's volume studies. Diagonal line indicates the identity $y = x$. (A) Mean scores (including pre, post, and change scores) strongly align, and when they do not, there is structure, insofar as it "looks" as if points are simply shifted rather than randomly dispersed. (B) SDs strongly align, despite some differences in the means. (C) Standardized mean differences ($\frac{\mu_{\text{post}} - \mu_{\text{pre}}}{\sigma_{\text{pre}}}$) almost perfectly lie on the identity. (D) Example of the shift in means from the women's and men's volume studies; for each group, on average, the men have exactly two more years of training experience than the women. Women and Men tables are adapted from [69] and [70], respectively. NB, in A–C, the SDs and SMDs show almost perfect agreement: concordance correlation coefficients = 0.97 and 0.96, respectively.

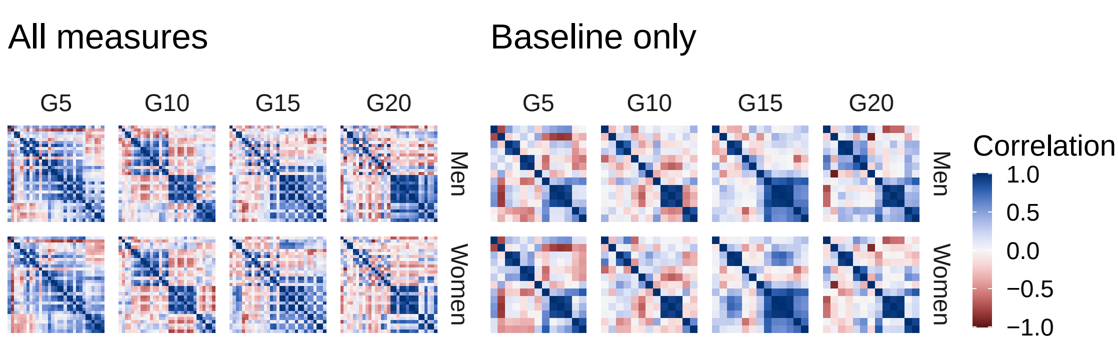



Figure 4: Correlations within and between trained women and men in the volume studies. (All measures) Includes pre, post, and change scores. Note the patterns are almost identical within-group/between-study, but not between-group/within-study. (Baseline only) Includes just pre-intervention scores and thus is unaffected by the effects of the intervention. Because there was a randomization process, we expect the differences between-group/within-study to have occurred by chance; however, it is extremely unlikely that these differences would be nearly identical between studies.

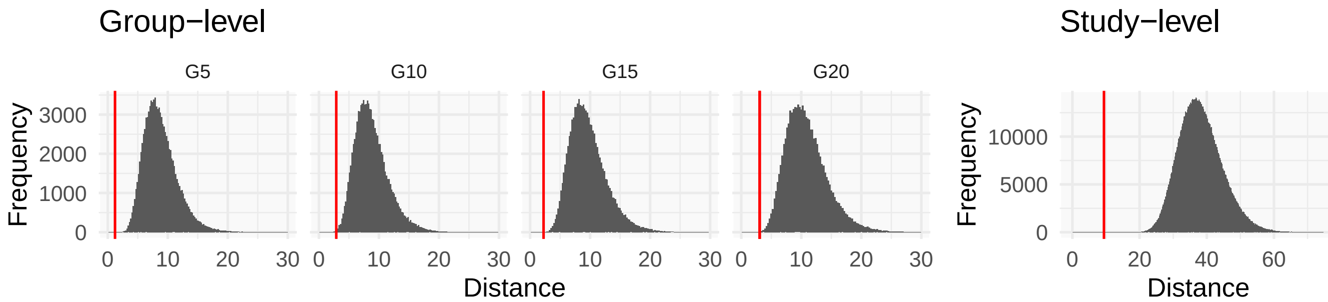



Figure 5: Permutation tests for the baseline similarity in covariances between the men's and women's studies. (Group-level) Results of permutation tests for the similarity in covariances between studies for individual groups. The red line indicates the observed similarity, while the distribution is a permutation distribution, or if one were to re-randomize participants. In other words, with randomization, we *expect* to see the red line fall within the plotted distribution. (Study-level) Results of a permutation test for the entire study at once. This is similar to (Group-level), but all individuals are assigned to groups at once and thus represents sampling without replacement. Note how far to the left the red lines are relative to the distributions. In the case of the study-level, the entire distribution, including its lower tail, is far from the red line (observed). This means the observed similarity is extremely unlikely under randomization.

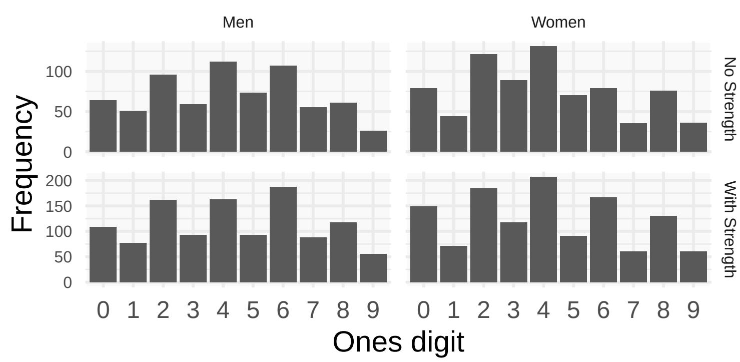



Figure 6: Double the number of even numbers as odd numbers in the trained men and women's volume studies.

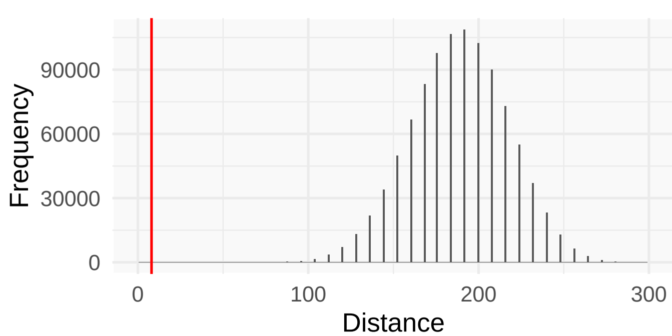



Figure 7: The thickness differences between the pectoralis major, triceps brachii, and biceps brachii were structured and ordered in a way that is highly unlikely to occur with randomization. With randomization, we would expect the red line (observed distance) to fall within the grey distribution; instead, the red line falls outside of the distribution, suggesting the data are improbably consistent.

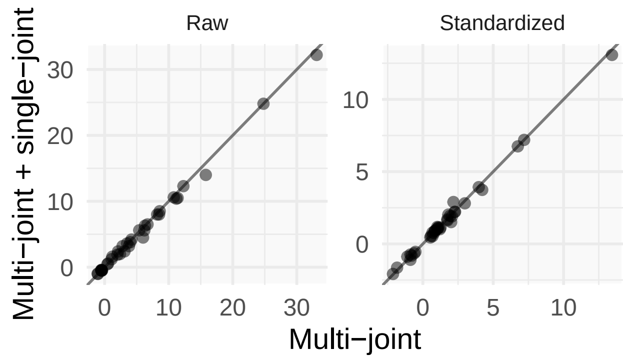



Figure 8: Within-study agreement of effects between multi-joint and multi-joint + single-joint outcomes. Both the raw (left) and standardized (right) effects across four multi-joint vs. single-joint studies display remarkable absolute agreement ($CCC > 0.99$ for both).

evenly distributed evens and odds is incredibly unlikely ($P \in \left[< 2.2 \times 10^{-16}, 1.3 \times 10^{-11}\right]$).

Finally, when looking at the raw data from the women's volume study [69], we noticed that the baseline muscle thicknesses of the pectoralis major, triceps brachii, and biceps brachii were strongly correlated. Upon closer examination, we noticed that the pairwise differences between pectoralis major, triceps brachii, and biceps brachii muscle thicknesses were nearly identical in G5 vs. G15 and G10 vs. G20. For example, subject 1 in G5 had an identical biceps minus triceps thickness as subject 1 in G15, and so on. To evaluate the extremeness of this observation, we performed yet another permutation test. Subjects were re-randomized to groups and ordered randomly within those groups; this was performed 1,000,000 times. We used the sum of squared differences between G5 and G15, and G10 and G20 as a measure of distance, and this took into account all three pairwise differences of the included muscle thicknesses (Figure 7). This permutation test showed that this observation was, indeed, very extreme and inconsistent with randomization—with a probability of occurrence of less than 1 in 1 million ($z = -6.26$, $P < 1 \times 10^{-6}$).

## 1.3 Single-joint versus multi-joint studies

### 1.3.1 Correlation of Effects

For all of the multi-joint vs. multi-joint plus single-joint studies [72, 74–76], corresponding groups also reported virtually identical results for every measure (Figure 8). Even if we assume the null is true, it would be fair to anticipate larger differences between groups simply due to sampling error (i.e., the small differences may fall in the lower tail of an $F$-distribution). The correlation between mean changes in corresponding groups in each study is $r > 0.99$. In the graph below, $x$-values are the change in the multi-joint only group for one measure, and $y$-values are the change in the MJ+SJ group for the same measure in the same study.

### 1.3.2 Patterns in Raw Data

In two of Barbalho's studies [72, 74] for which we had access to the raw data, there were patterns in the numbers. Specifically, the flexed arm circumference data were, in order, `0.8, 0.8, 0.8, 0.8, 0.8, 1.1, 1.1, 1.1, 1.1, 1.1` for group 1 and `1, 1, 1, 1, 1, 1.4, 1.4, 1.4, 1.4, 1.4` for group 2 in the first study, and `0.3, 0.3, 0.3, 0.5, 0.5, 0.5, 0.5, 0.5` for group 1 and `0.4, 0.4, 0.4, 0.4, 0.5, 0.5, 0.5, 0.5, 0.5` for group 2 in the second study. To the best of our knowledge, these data have not been sorted to produce this pattern (if that occurred, the subjects were re-numbered after the fact). Ignoring the probability of each group only having two values, and the probability of such small ranges in the data, simply attaining results with these characteristics ("runs" of one number, followed by "runs" of another number) is very unlikely, with probabilities of $\left(\frac{5!5!}{10!}\right)^2 = 1.6 \times 10^{-5}$ for the first study, and $\left(\frac{3!5!}{8!}\right)\left(\frac{4!5!}{9!}\right) = 1.4 \times 10^{-4}$ in the second study. The probability of obtaining data with these characteristics in both studies is

$$\left(\frac{5!5!}{10!}\right)^2 \left(\frac{3!5!}{8!}\right) \left(\frac{4!5!}{9!}\right) = 2.2 \times 10^{-9}.$$

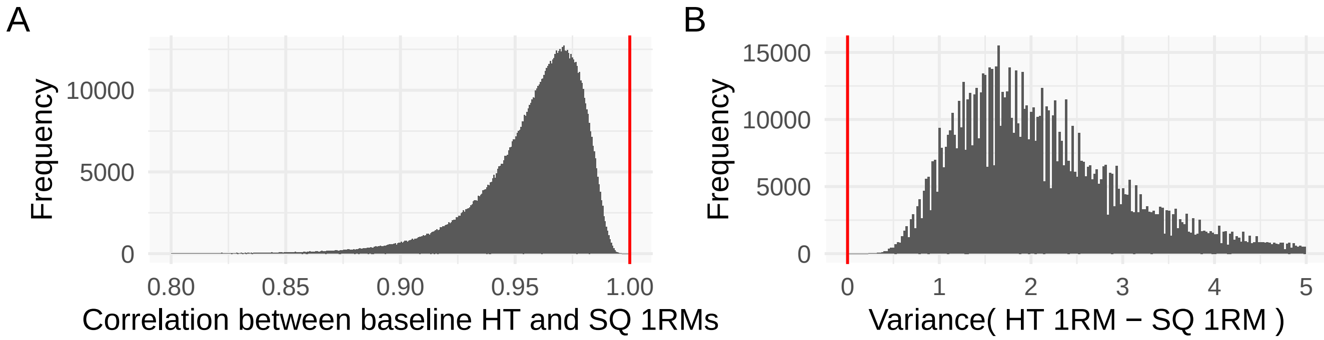



Figure 9: Simulations reveal that observing a perfect baseline difference between squat and hip thrust 1RMs is highly improbable. We simulated data based on [73], taking into account the test-retest reliability of the squat and hip thrust [71]. These simulations created a reference distribution, against which we could assess how extreme the observation of a perfect difference in baseline squat and hip thrust 1RMs is. After performing 1,000,000 simulations, we observed that (A) none of the simulations had a perfect correlation between baseline squat and hip thrust 1RMs, and similarly, (B) none of the simulations had a perfectly homogeneous difference between squat and hip thrust 1RMs. This indicates that the perfect baseline relationship observed by Barbalho *et al.* [73] has a $P$-value $< 1 \times 10^{-6}$.

When adding the probability of the "runs" being arranged low-to-high in all four groups, the probability drops to approximately 1 in ten billion:

$$\left(\frac{1}{2}\frac{5!5!}{10!}\right)^2 \left(\frac{3}{5}\frac{3!5!}{8!}\right)\left(\frac{4}{9}\frac{4!5!}{9!}\right) = 9.3 \times 10^{-11}.$$

#### 1.3.3 Baseline Squat and Hip Thrust Strength

In Barbalho *et al.*'s most recent paper [73], every lifter's baseline hip thrust 1RM was 8 kg more than their squat 1RM. This means that there was a baseline correlation of $r = 1$ between squat and hip thrust 1RMs. At face, this is unlikely because, among other reasons, measurement reliability would tend to prevent such a relationship from being observed. Specifically, we know from Spearman [77] that the correlations we observe are constrained by measurement precision,

$$r_{\text{obs}} = r_{\text{true}}\sqrt{r_{xx}r_{yy}},$$

where $r_{\text{obs}}$ is the observed correlation between two variables, $r_{\text{true}}$ is the true correlation between those two variables, and $r_{xx}$ and $r_{yy}$ are the test-retest correlations for the two variables being correlated.

Given the above, we aimed to quantify how unlikely it is that we would observe a perfect correlation between squat and hip thrust 1RMs, with the assumption that *true* squat and hip thrust 1RMs are perfectly correlated (hip thrust 1RM = squat 1RM + 8). To do so, we performed Monte Carlo simulations with the data simulated to be similar in nature to [73]. We incorporated the intraclass correlation coefficients (ICCs) for squat and hip thrust 1RMs reported by Barbalho *et al.* [71], along with their uncertainties. In these simulations, we also took into account that Barbalho *et al.* used loads that were increments of 1 kg.

The results of these simulations can be seen in Figure 9, and indicate that, after taking measurement error into account, the probability of observing the perfectly homogeneous baseline shift when one really exists is $P < 1 \times 10^{-6}$, meaning it is more surprising than a fair coin landing on heads 20 times in row. We note that the precision (and thus "smallness") of the $P$-value is constrained by the number of permutations performed, so this is a conservative estimate.

#### 1.3.4 Squat and Hip Thrust Change Scores

In addition to the perfectly homogeneous structure in the baseline scores, we also observe structure in the change scores. The differences between the hip thrust change scores and squat change scores have structure; they (a) are perfectly homogeneous (all = 4 kg) in the MJ+SJ group; (b) are perfectly bimodal (all are either 24 or 44 kg) in the SJ group; and (c) differ for each person in the MJ group. The distributions of differences in change scores can be seen in Figure 10A.

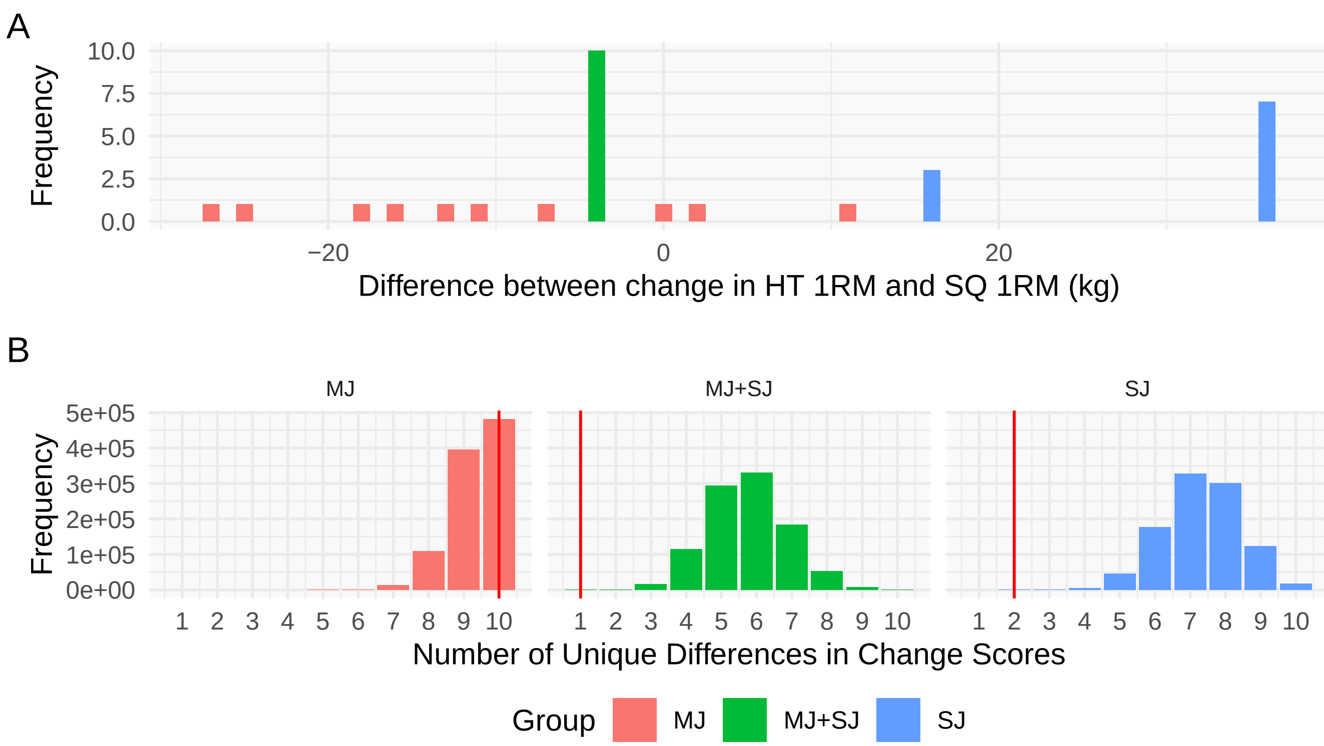



Figure 10: Simulations reveal that the observed structure in differences between change scores is highly unlikely. (A) There is a group-dependent structure in the difference between hip thrust 1RM and squat 1RM; MJ+SJ is perfectly homogeneous (all $= -4$ kg) and SJ falls into two groups (16 or 36 kg), while every individual in MJ has a different value. (B) The homogeneity in the MJ+SJ and SJ groups was highly improbable when taking measurement error into account (combined $P = 1.2 \times 10^{-11}$).

We would not expect to see such structure in the data, in part due to measurement error alone. Thus, we simulated more data to quantify the probability of observing data that looks like this. We will assume that the true differences are the ones observed; measurement error will increase variability. First, we investigated the within-group probabilities of observing $n \leq \{1, 2, 10\}$ unique differences in change scores. After taking measurement into account, we found that both the MJ+SJ and SJ distributions are highly unlikely (MJ+SJ, $P = 4 \times 10^{-6}$; SJ, $P = 3 \times 10^{-6}$) (Figure 10B). Unsurprisingly, the MJ group's heterogeneous distribution is unsurprising ($P = 1$). Second, we can look at the joint probabilities. The extreme findings in the MJ+SJ (perfectly homogeneous) and SJ (two or fewer outcomes) groups were not observed in any single simulation run (meaning $P < 1 \times 10^{-6}$); this is expected, as the product of the $P$-values suggests a combined $P = 1.2 \times 10^{-11}$, or about as surprising as a fair coin landing on head 36 times in a row.

#### 1.3.5 Distributions of Even and Odd Numbers in Muscle Thickness Data

The muscle thickness data in [73] have improbable distributions of even versus odd numbers (Figure 11). In particular, there are *no* odd-valued pre-intervention muscle thicknesses in any group or muscle ($^0/_{120}$, two-tailed $P = 2(0.5)^{120} = 1.5 \times 10^{-36}$ relative to an expected $^{50}/_{50}$ split of even and odd), while the post-intervention is roughly 40% odd ($^{47}/_{120}$, $P = 0.02$ relative to an expected $^{50}/_{50}$ split of even and odd), which evidences that odd numbers are possible. By comparing these proportions directly ($^0/_{120}$ vs. $^{47}/_{120}$), the pre-intervention distribution is still highly improbable ($P = 7.3 \times 10^{-14}$, or about as surprising as a fair coin landing on heads 43 times in a row).

#### 1.3.6 Distributions of Strength Numbers

The strength data in [73] also have improbable distributions (Figure 12). When looking at the distributions of ones digits in the pre- and post-intervention strength data, one can see there are spikes at 0, 3, 5, and 8 in the pre- but not post-intervention data. The difference between these distributions is marked ($P < 2.2 \times 10^{-6}$) and warrants explanation.

#### 1.3.7 Squat Strength Gains

In Barbalho *et al.*'s most recent paper [73], the magnitude and rate of squat strength gains is worth noting. In particular, all lifters—even those who performed only single-joint exercises—underwent appreciable strength changes. Given that the study was 24 weeks long, we calculated an average rate of squat strength increase for each subject ($\Delta$1RM/24 weeks), and we compared these rates to raw female powerlifters from the Open Powerlifting database (ages = 24–34; raw-only; tested *or* untested; and similar allometrically-scaled squat strength at their first meet (5–7 $\text{kg}^{\frac{1}{3}}$)). Subjects in the Barbalho *et al.* [73] study had rapid rates of squat strength increases – far superior to similarly skilled powerlifters (Figure 13). This was also the case for the squat group in the squat vs. hip thrust study [71] (Figure 13).

## 1.4 Elderly Study

Barbalho *et al.* [78] investigate the effect of exercise on, among other things, weight loss in elderly women. Strikingly, nearly all participants both lost weight and decreased their waist circumference. This finding is in contrast to other literature on exercise and weight loss without a dietary intervention [79]; for example, Ahtiainen *et al.* [2] only observed weight loss in 46% of participants. A test of weight loss proportions between Barbalho *et al.* [78] ($\frac{370}{376} = 0.98$) and Ahtiainen *et al.* [2] ($\frac{132}{285} = 0.46$) reveals drastic differences ($P < 2.2 \times 10^{-16}$).

Dr. Gentil responded to the aforementioned concerns about the elderly study on July 6, 2020, with the following:

> 1. They used a Ahtiainen et al. (Ahtiainen et al., 2016) to question our results. However, this study involved a heterogenous sample and only 36 older women, with no separate analysis for them. In fact, we were not able to find any graph or data regarding weight loss and waist circumference responsiveness nor specific information on the number of older women who lost weight in that study.
>
> 2. Weight loss is a multifaceted process and it is not possible to say that our results occurred exclusively due to the resistance

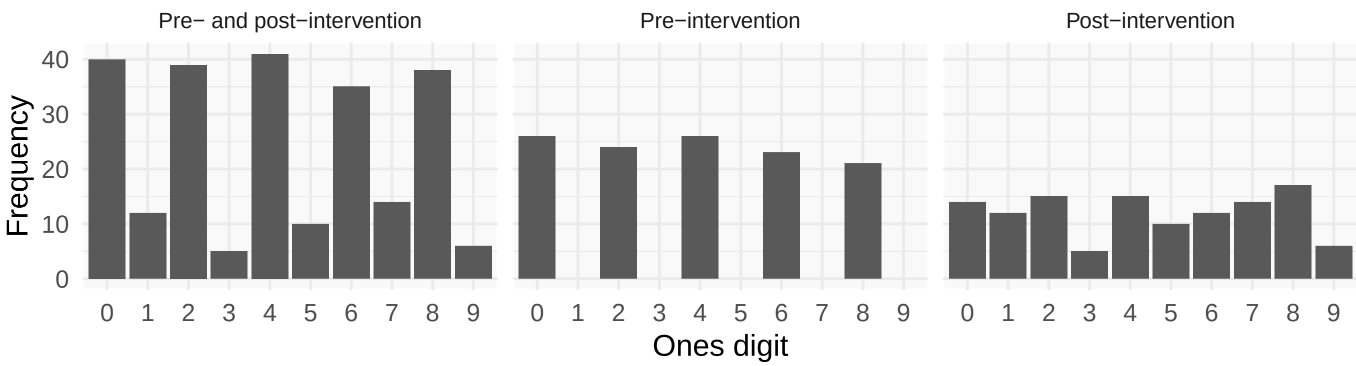



Figure 11: Even numbers dominate the distribution of muscle thicknesses because there are *no* odd values in the pre-intervention scores.

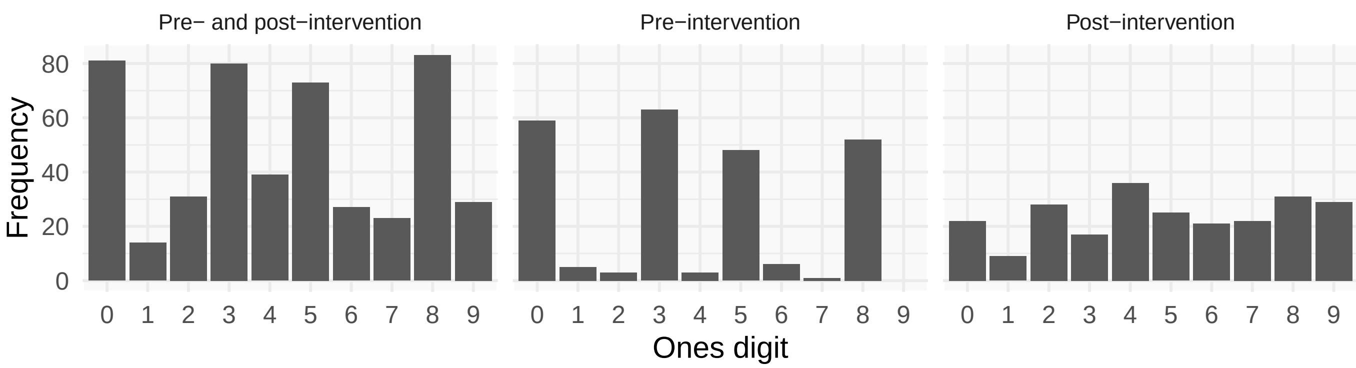



Figure 12: A select few numbers (0, 3, 5, and 8) dominate the pre- but not post-intervention strength measures. The pre- and post-intervention distributions of ones digits differ markedly from one another ($P < 2.2 \times 10^{-6}$).

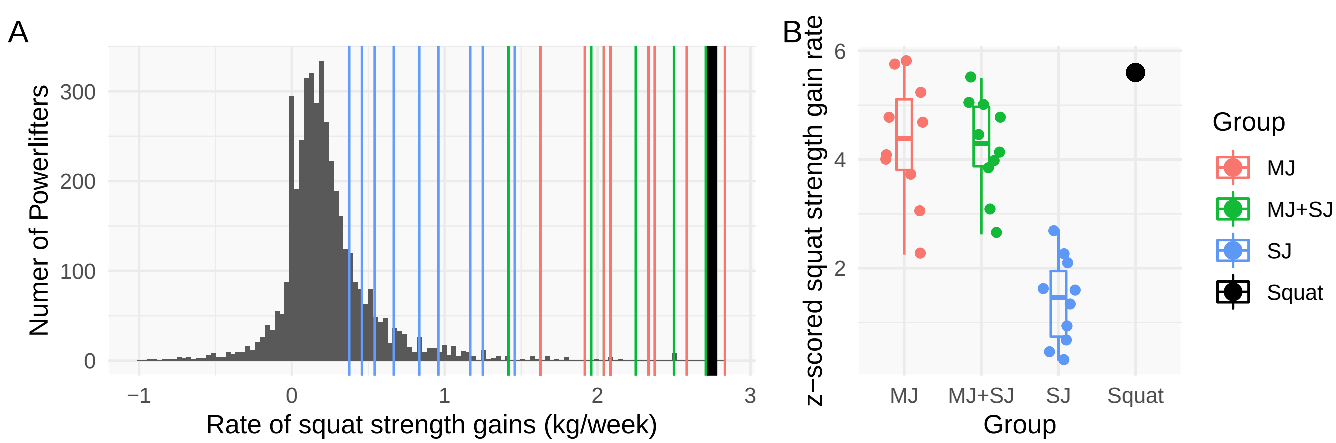



Figure 13: Participants from Barbalho *et al.* [71, 73] exhibit rapid increases in squat strength. (A) The distribution of the rate of squat gains (kg/week) in female powerlifters falls below most of the subjects in [73] (thin, colored bars) and the squat group average from [71] (thick, black bar). (B) This results in high $z$-scores for all three groups from [73] and the squat group average from [71], but especially so for those who performed multi-joint movements.

training protocols, which is recognized as a limitation: "One important limitation of the present study is the lack of nutritional control, which can influence in the results of anthropometric measures.". Therefore, it is possible that the participants changed their lifestyle during the study.

3. Weight loss has been shown to be inversely associated with strength gains in postmenopausal women (Bea et al., 2010) and our study showed a marked increase in muscle strength.

4. We try to explain our results stating that "the reductions in body mass and waist circumference found in the present study might be related to training intensity (i.e. training to momentary muscle failure), as reported in previous studies in which low-volume, high-intensity RT promoted positive changes in body composition in older people [43].".

5. We reported that the participants were closely supervised and the supervisors were oriented to encourage the participants to train with high efforts, which might have led to increased results and motivation to adopt positive lifestyle changes. As far as we know, these procedures have not been adopted in previous studies. Moreover, the study protocol used by Ahtiainen et al. (Ahtiainen et al., 2016) is not even described in the article.

Therefore, most of the concerns are already addressed in our article. Our results are completely comprehensible, and I have no reason to question the validity of our findings.

We are unsatisfied by Dr. Gentil's response for the following reasons (addressed in order):

1. We obtained the data directly from the authors of this study. That only 36 of them were "older women" does not substantially detract from our concerns; individuals, no matter their age, do not tend to lose weight with just a training intervention [80–84]. In fact, the literature suggests that younger individuals are more likely to lose weight on exercise-only interventions compared to older individuals [85], in turn rendering the Ahtiainen *et al.* [2] estimate a conservative one.

2. We certainly agree that weight loss is multifaceted, but it strains credulity that the consistency of weight loss would occur sans dietary or behavioral interventions for several reasons:

   - There is a massive body of literature demonstrating the behavioral changes—including dietary and lifestyle changes that result in weight loss—are extremely difficult to start and maintain [86]. In fact, behavioral interventions are necessary to improve adherence in exercise programs [87]. How a study without behavioral interventions could result in so much success—better success than studies with interventions—warrants explanation. The length of the study and consistency of the results adds to these improbabilities, in that longer studies are likely to result in poorer or more variable adherence.
   - Participants were explicitly asked not to change their diet. It would be strange for nearly every participant to improve their eating habits, to the extent of rendering weight loss, despite having been asked not to. Indeed, the Resist Diabetes trial, despite utilizing a similar resistance training protocol, did not find changes in weight in pre-diabetic participants aged 50-69 years across a 15-month study period [88]. This is despite secondary outcomes from that trial showing spontaneous reductions in dietary energy intake [89] and increases in non-resistance training aerobic physical activity [90]. Thus, it seems unlikely that spontaneous behavioral adaptations could explain the observed weight loss.
   - The energy expenditure from physical activity interventions alone is small. Estimates of energy expenditure for lower volume resistance training sessions range from around 50–150 kcal [91]. A conservative estimate of 150 kcal/session would yield 3600 kcal burned over the course of the

study. The lack of proportionality of weight loss to the exercise volume further suggests that the observed weight loss is not solely attributable to the exercise intervention.

- Given the above, the etiology and consistency of weight loss has not been explained. Vague, catch-all explanations are inadequate given that these results fly in the face of literature on the topic.

3. This is both orthogonal to our concerns and misleading. In fact, Bea *et al.* [92] exemplify our point; even after 6 years of exercise, on average, exercising participants gained (a negligible amount of) weight.

4. There does not exist a strong theoretical rationale as to why training to momentary muscular failure would substantially improve the probability of losing weight with resistance training alone.

   - Indeed, though there are data suggesting that, at a given work output, resistance training to momentary failure results in greater total energy expenditure; this amounts to ∼3 kcal difference [93]. Importantly, energy expenditure during resistance training is directly related to the amount of mechanical work performed [94]. Although performing a single set to momentary failure might increase mechanical work, across multiple sets, this does not appear to be the case [95]. Furthermore, if the reductions in body fat could be attributed to the work performed during the training sessions, one would anticipate that the subjects in the high volume group in the study would have lost approximately twice as much body fat as the subjects in the low volume group, which did not occur.

   - If we consider that, within each group, starting weights, height, age, and the amount of weight lost over the 12 week period (84 days) are relatively homogeneous, we can then use the National Institute of Diabetes and Digestive and Kidney Diseases model for predicting weight loss [96]. Specifically, we can estimate how much additional energy expenditure from the intervention alone would be required. The weight loss reported in the HV and LV groups would require an ∼70% and ∼78% increase in physical activity energy expenditure, respectively, assuming no dietary modifications in energy intake if weight loss were to be achieved over the 12 week period. This model considers metabolic compensations over time with weight loss. However, research also shows that behavioural compensation, such as that mentioned above, can range from +55% to +64%, which affects energy balance and thus weight loss in response to either dietary or exercise interventions [97]. Based on these assumptions, the required weekly net energy deficits (NIDDK model, NIDDK+55%, NIDDK+64%) from physical activity are estimated to be 1376.1 kcal, 2132.9 kcal, and 2256.8 kcal for the HV group, and 1528 kcal, 2368.4 kcal, and 2505.92 kcal for the LV group. If we consider the number of sets reported for either group in different weeks, we can estimate the energy expenditure that would be required to result from this. Data from one of our group's lab suggests negligible differences between different large muscle group exercises when performed to volitional failure [98]; therefore, we assume similar energy expenditure across exercises (though this likely makes our estimate more conservative as smaller muscle exercises included in the intervention are assumed to have a higher energy expenditure). The HV group ranged from 24 to 30 sets total per week; this would require each set to, on average, expend 45.9 kcal to 57.3 kcal, 71.1 kcal to 88.9 kcal, and 75.2 kcal to 94 kcal for each estimate, respectively, to achieve the weight loss reported. The LV group ranged from 12 to 18 sets total per week and thus would require sets to expend between 84.9 kcal to 127.3 kcal, 131.6 kcal to 197.4 kcal, and 139.2 kcal to 208.8 kcal for each estimate, respectively, to achieve the weight loss reported. It seems highly unlikely that this was achieved considering that our data have shown only an 118.9±22 kcals total energy expenditure when 4 exercises are performed for a single set to volitional failure. Moreover, other recent work has reported ∼25 kcal total energy expenditure per set of exercise performed to momentary failure [99]. Further, aside from set volume alone, the total absolute work (sets × reps × kg) performed is a strong predictor of energy expenditure [100, 101]; given the absolute loads being used by the participants in this study (given their low baseline strength values), it seems even more unlikely that they were able to achieve sufficient energy expenditure as a result of the resistance training intervention to produce the weight loss reported, even when considering the possibility of spontaneous behavioral modifications.

- Since physiological explanations do not seem to explain the colossal discordance between this study's findings and those in the literature, a more thorough explanation is warranted.

5. It seems unlikely that the intervention being supervised would have an appreciable effect on calories burned to the point of rendering the exercise routine itself a potent weight loss intervention. Indeed, in another study where older adults were provided with closely supervised, progressively implemented, high intensity of effort resistance training, there was a comparatively smaller weight loss over the intervention period (∼64-74% of that reported by Barbalho et al.) despite an intervention of twice the length (6 months) [102]. Further, although a smaller sample size ($n = 23$), 5 participants (∼21%) did not demonstrate weight loss. Positive lifestyle changes, on the other hand, are difficult to consistently implement. Given that the individuals were encouraged to not change their diet, such lifestyle changes would have to be independent of dietary changes. Thus, explanation is warranted regarding what lifestyle changes were encouraged and how those would render consistent weight loss across 370 elderly women.

In addition, the study employed block randomization. However, 217 participants were randomized to the high volume group and 203 participants were randomized to the low volume group. It is unclear how a 14-participant discrepancy could occur with block randomization.

We note that there are aspects of this study we find curious and are still looking into, such as the funding and resources necessary to complete this study considering its scale, in addition to some of the other measures/outcomes. We will update this white paper accordingly as additional information comes to light.

# 2 Arguments Against Extreme Homogeneity

It can be argued that the observed baseline homogeneity is a result of nonrandom (biased) sampling by the researchers, in that investigators purposely sampled individuals who had similar levels of strength, training experience, etc. While it is easy to sample a homogeneous sample conditional on one variable (e.g., squat strength), it is exponentially more difficult to sample conditional on more variables. This follows from the chain rule in probability—the population from which to sample becomes less dense for each variable on which you condition. Thus, although Barbalho *et al.* may have purposely recruited homogeneous samples, it seems tremendously difficult to have done so while matching on so many variables.

In addition to the low likelihood of matching on multiple dimensions, there is marked homogeneity for variables that were not assessed until after a participant was enrolled; namely, muscle thicknesses [69, 70], in addition to change scores (Figure 2b,d). This is incredibly unlikely given that this was not subject to explicitly biased sampling.

## 2.1 Example

Because our argument is fairly abstract, here, we further explain the theory behind it, and then we draw upon data from Open Powerlifting to demonstrate the appreciable effects of conditioning on multiple variables.

In the simplest case, wherein we are interested in the probability of both $A$ *and* $B$ occurring, chain rule in probability states $P(A \cap B) = P(A \mid B) \cdot P(B)$. Intuitively, if we are interested in both $A$ *and* $B$ occurring, then we know $A$ will only occur with $B$ a fraction of the time, and $B$ in general will only occur a fraction of the time. The means that the space from which to sample decreases for each variable we condition on (Figure 14). More tangibly, if a table has many fruits (berries, cherries, melons, apples, oranges, pears, etc.), looking for two properties simultaneously will quickly decrease the number of fruits that meet

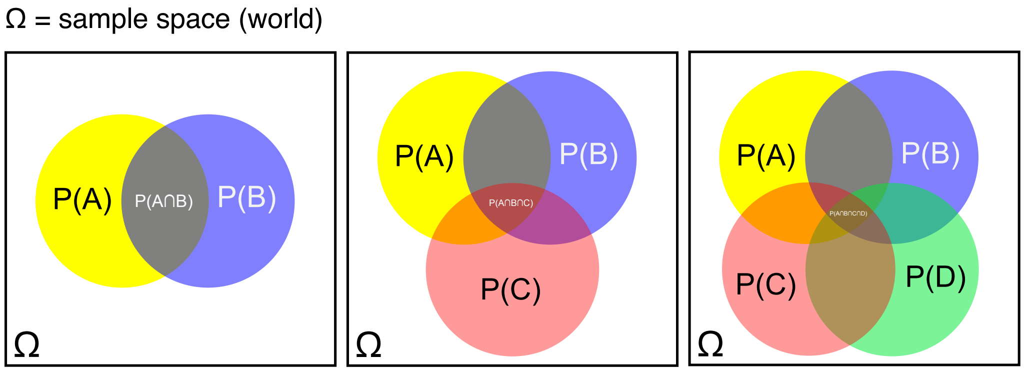



Figure 14: A visual explanation of the decrease in area from which to sample as you condition on additional variables. Note that the area of overlap gets smaller with each additional variable.

the criteria. For example, if I say I am looking for a red fruit, there are many options: cherries, strawberries, apples, watermelon, tomatoes, etc. However, if I say I am looking for something red that is also a berry, it seems I must be talking about strawberries. Alternatively, I can start with looking for berries and my options are plentiful; however, once I specify red, I get to the same answer. Thus, the more variables we condition on, the more unique or rare our event or state of interest becomes.

Now, say we were interested in sampling male powerlifters from the Open Powerlifting database. After cleaning the data (for duplicate lifters, missing data points, etc.), we have 71,037 data points with the information we need. Of these data points, suppose we are interested in raw lifters who compete in drug-tested federations and are between the ages of 20 and 34. In Figure 15, we see the effect of sequentially conditioning on raw, drug-tested, and age; with each additional variable we condition on, the number of lifters remaining decreases appreciably. From a logistical standpoint, it is much easier to condition on binary variables (e.g., we are only interested in raw, drug-tested lifters) than it is continuous variable, wherein we want our sample to look like a specific distribution. To emulate the biased sampling in the studies by Barbalho *et al.*, we will calculate the proportion of the "population" that can be used to generate new samples, each with tight SDs ($\sim$ 4 kg) and a specified mean for all three lifts.

To calculate the probability of finding an individual who can be used in the sample, we draw upon rejection sampling theory. In rejection sampling, we have two probability density functions (pdf), $f(x)$ and $g(x)$. $f(x)$ is our desired pdf, and $g(x)$ is the pdf from which we have to sample. More concretely, we wish to create a sample with the distribution $f(x)$ by taking a biased sample of $g(x)$. In rejection sampling, $M = \sup\left\{\frac{f(x)}{g(x)}\right\}$ is an optimal scaling factor, and $\frac{1}{M}$ is termed the acceptance probability. Another way of conceptualizing this is that $\frac{1}{M}$ is the proportion of individuals in $g(x)$ who can be sampled to form a distribution equal to $f(x)$. We applied this theory to the powerlifting data. We specified our distribution of interest to be the mean bench, squat, and deadlift 1RM, with an identical correlational structure to the original data, but with SDs of 5 kg for each lift. Note, 5 kg was chosen instead of 4 kg to be charitable, as 5 kg is on the higher end of the baseline SDs reported by Barbalho *et al.* Because powerlifting numbers tend to be discrete (multiples of 0.5 kg), we integrated around each mean to emulate the discretized distribution [103]:

$$p(\vec{\mu}) = \iiint_{\lfloor 2\vec{\mu} \rfloor/2}^{\lceil 2\vec{\mu} \rceil/2} f(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3,$$

where $f(\vec{x})$ is the trivariate normal density function, $\vec{\mu}$ is a vector of the mean one-repetition maximums of the three lifts, $\vec{x}$ is a vector of evaluated one-repetition maximums, and $p(\vec{\mu})$ is its discretized analogue (probability mass function) evaluated around the mean, with which we calculated $\frac{1}{M} = \frac{g(\vec{\mu})}{p(\vec{\mu})}$. Note, $\sup\left\{\frac{p(\vec{x})}{g(\vec{x})}\right\}$ is satisfied when $\vec{x} = \vec{\mu}$, meaning $M = \frac{p(\vec{\mu})}{g(\vec{\mu})}$. This approach produced nearly identical results (within 0.00001) to a more computationally costly grid approximation.

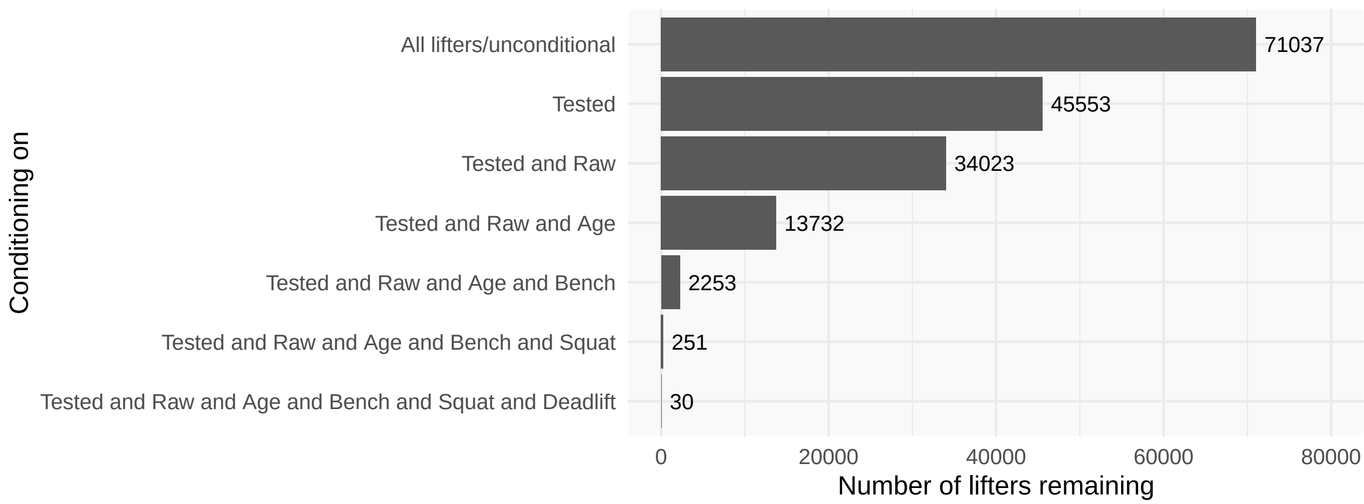



Figure 15: Sampling a homogeneous sample from the Open Powerlifting database. We are left with 30 of the original 71,037 lifters after conditioning on all variables.

By conditioning on each of these lifts, the number of lifters one can sample from decreases substantially; at the end, there are 30 lifters from the original 71,037 (Figure 15).

This principle is well established in probability; the more variables you condition on, the smaller your target population relative to the entire population. Here, we used strongly correlated lifts and thus our estimates are liberal; lower correlations between variables (e.g., 10RM triceps extensions and 10RM pull-downs rather than 1RM squat and 1RM deadlift) would result in even sparser populations from which to sample. By scaling the remaining lifters to the number needed for a 40-person study, the initial pool of lifters would need to contain 94,716 individuals; for context, as of 2020, Belém has a total population of 1.44 million. To actually recruit 40 subjects, all 94,716 would need to be screened and pre-tested, indicating that $\sim 2400$ subjects would need to be tested for each subject recruited. The numbers from this exercise suggest the homogeneity in the studies by Barbalho et al. is appreciable, especially for having recruited from a select few gyms. Finally, from a more applied perspective, not all of those who are eligible are willing to volunteer for studies or are able to (e.g., due to geographical restrictions). As a result, the lifters willing to participate would likely be even scarcer.

## 3 Conclusion

We noted several improbable observations present in studies published by Barbalho *et al.* These observations include improbably small SDs; large and consistent effects; consistent baseline structure following randomization; and effects that are inconsistent with other studies.

To be explicit, we have no evidence to suggest we understand the provenance of the data. We do not have any evidence beyond the fact that the data is unlikely to suggest how it became unlikely. Nevertheless, these improbable observations warrant explanation.

## 4 Acknowledgments

We would like to thank Aaron Caldwell and Kristin Sainani for their helpful feedback.

# 5 Appendix: Timeline 742

| | |
|---|---|
| February 11, 2020 | We first notified the senior author of these papers, Paulo Gentil, of our initial findings. |
| March 26, 2020 | The white paper was sent to Paulo Gentil to review. We asked for an explanation or rebuttal on or before April 10. Barbalho immediately requested a one-week extension, which we happily granted. |
| April 15, 2020 | The authors admitted that there were indeed "inconsistencies" in the data from Barbalho *et al.* [70]. The authors state that Barbalho *et al.* [70] was carried out, but the undergraduate student who was responsible for transferring the data from paper to Excel made errors in the process. |
| April 17, 2020 | The authors requested that Barbalho *et al.* [70] be retracted from *International Journal of Sports Physiology and Performance* for the aforementioned reasons. While one of the interrelated papers was retracted, our concerns with Barbalho *et al.* [69] remain. |
| June 10, 2020 | We contacted the journal editors with our concerns. *Experimental Gerontology*'s editor responded by working with Elsevier to contact Dr. Gentil directly. The remaining editors advised us to email Mr. Barbalho and Dr. Gentil, with the editors CC'd, to request an explanation. |
| June 22, 2020 | We emailed Mr. Barbalho and Dr. Gentil asking for an explanation. We gave them until July 13, 2020 @ 11:59 PM local time to respond. |
| July 6, 2020 | We received an email from Elsevier containing Dr. Gentil's response to our concerns regarding the *Experimental Gerontology* study. We are not satisfied by his explanations and have shared our concerns with Elsevier. |
| July 14, 2020 | Mr. Barbalho and Dr. Gentil did not respond to our concerns regarding the other studies; we requested retraction for these papers. 743 |
| July 28, 2020 | *Medicine & Science in Sports & Exercise* stated that, in accordance with COPE guidelines, they will be contacting the authors' institution. In the meantime, they will be publish an Expression of Concern. |
| August 13, 2020 | *European Journal of Sport Science* and Taylor & Francis requested a response and raw data from the authors. |
| September 1, 2020 | *International Journal of Sports Medicine* stated that they will not retract the articles at this time. We were invited to submit letters to the editor for [71] and [73]. We will respond to the editors and request the raw data for [71]. |
| September 4, 2020 | *Sports* and its publisher, MDPI, have contacted the authors' institution to open an investigation. On this day, we also followed up with the other journals. |
| October 1, 2020 | We responded to *International Journal of Sports Medicine* regarding their email from Sept 1. |
| October 15, 2020 | *Medicine & Science in Sports & Exercise* published an Expression of Concern regarding [69]. |
| March 16, 2021 | *Experimental Gerontology* stated that, after a University investigation, the journal will not take action regarding [78]. |
| March 17, 2021 | After requesting their names be removed from the author list, *International Journal of Sports Medicine* removed James Steele and James Fisher as coauthors to [73]. |
| April 1, 2021 | *Medicine & Science in Sports & Exercise*'s Editor-in-Chief retracted [69]. |
| April 22, 2021 | James Fisher and Jürgen Giessing, two of Barbalho's coauthors, were added as co-authors to the white paper. |

# References


[1] Sajad Ahmadizad, Shahab Ghorbani, Mansour Ghasemikaram, and Mehdi Bahmanzadeh. Effects of short-term nonperiodized, linear periodized and daily undulating periodized resistance training on plasma adiponectin, leptin and insulin resistance. *Clinical biochemistry*, 47(6):417–422, 2014.

[2] Juha P Ahtiainen, Simon Walker, Heikki Peltonen, Jarkko Holviala, Elina Sillanpää, Laura Karavirta, Janne Sallinen, Jussi Mikkola, Heli Valkeinen, Antti Mero, et al. Heterogeneity in resistance training-induced muscle strength and mass responses in men and women of different ages. *Age*, 38(1):10, 2016.

[3] Jytte M Apel, Ryan M Lacey, and Robert T Kell. A comparison of traditional and weekly undulating periodized strength training programs with total volume and intensity equated. *The Journal of Strength & Conditioning Research*, 25(3):694–703, 2011.

[4] Hamid Arazi, Abbas Asadi, H Aziri, A Asadi, H Arazi, and A Asadi. Effects of 8 weeks equal-volume resistance training with different workout frequency on maximal strength, endurance and body composition. *Int J Sports Sci Eng*, 5(2):112–8, 2011.

[5] Daniel Baker, Greg Wilson, Robert Carlyon, et al. Periodization: the effect on strength of manipulating volume and intensity. *J Strength Cond Res*, 8(4):235–42, 1994.

[6] Sandro Bartolomei, Jay R Hoffman, Franco Merni, and Jeffrey R Stout. A comparison of traditional and block periodized strength training programs in trained athletes. *The Journal of Strength & Conditioning Research*, 28(4):990–997, 2014.

[7] Sandro Bartolomei, Jeffrey R Stout, David H Fukuda, Jay R Hoffman, and Franco Merni. Block vs. weekly undulating periodized resistance training programs in women. *The Journal of Strength & Conditioning Research*, 29(10):2679–2687, 2015.

[8] Sandro Bartolomei, Jay R Hoffman, Jeffrey R Stout, Maddalena Zini, Claudio Stefanelli, and Franco Merni. Comparison of block versus weekly undulating periodization models on endocrine and strength changes in male athletes. *Kinesiology: International journal of fundamental and applied kinesiology*, 48(1):71–78, 2016.

[9] Felipe A Brigatto, Tiago V Braz, Thamires Cristina da Costa Zanini, Moises D Germano, Marcelo S Aoki, Brad J Schoenfeld, Paulo H Marchetti, and Charles R Lopes. Effect of resistance training frequency on neuromuscular performance and muscle morphology after 8 weeks in trained men. *The Journal of Strength & Conditioning Research*, 33(8):2104–2116, 2019.

[10] T. W. Buford, S. J. Rossi, D. B. Smith, and A. J. Warren. A comparison of periodization models during nine weeks with equated volume and intensity for strength. *J Strength Cond Res*, 21(4):1245–1250, Nov 2007.

[11] Darren G Candow and Darren G Burke. Effect of short-term equal-volume resistance training with different workout frequency on muscle mass and strength in untrained men and women. *Journal of strength and conditioning research*, 21(1):204, 2007.

[12] Tomas Carlsson, Lars Wedholm, Johnny Nilsson, and Magnus Carlsson. The effects of strength training versus ski-ergometer training on double-poling capacity of elite junior cross-country skiers. *European journal of applied physiology*, 117(8):1523–1532, 2017.

[13] EB Colliander and PA Tesch. Responses to eccentric and concentric resistance training in females and males. *Acta physiologica scandinavica*, 141(2):149–156, 1991.

[14] Ryan J Colquhoun, Christopher M Gai, Danielle Aguilar, Daniel Bove, Jeffrey Dolan, Andres Vargas, Kaylee Couvillion, Nathaniel DM Jenkins, and Bill I Campbell. Training volume, not frequency, indicative of maximal strength adaptations to resistance training. *The Journal of Strength & Conditioning Research*, 32(5):1207–1213, 2018.

[15] William L Daniels, James E Wright, Dan S Sharp, Dennis M Kowal, and Robert P Mello. The effect of two years training on aerobic power and muscle strength of male and female cadets. Technical report, ARMY RESEARCH INST OF ENVIRONMENTAL MEDICINE NATICK MA, 1980.

[16] C De Lima, DA Boullosa, AB Frollini, FF Donatto, RD Leite, PRG Gonelli, MIL Montebello, J Prestes, and MC Cesar. Linear and daily undulating resistance training periodizations have differential beneficial effects in young sedentary women. *International journal of sports medicine*, 33(09):723–727, 2012.

[17] Raphael Mendes Ritti Dias, Edilson Serpeloni Cyrino, Emanuel Péricles Salvador, Fábio Yuzo Nakamura, Fábio Luiz Cheche Pina, and Arli Ramos de Oliveira. Impact of an eight-week weight training program on the muscular strength of men and women. *Rev Bras Med Esp*, 11(4):213–216, 2005.

[18] Sandor Dorgo, Pradeep Edupuganti, Darla R Smith, and Melchor Ortiz. Comparison of lower body specific resistance training on the hamstring to quadriceps strength ratios in men and women. *Research quarterly for exercise and sport*, 83(2):143–151, 2012.

[19] Rodrigo Fernandez-Gonzalo, Tommy R Lundberg, Lucia Alvarez-Alvarez, and José A de Paz. Muscle damage responses and adaptations to eccentric-overload resistance exercise in men and women. *European journal of applied physiology*, 114(5):1075–1084, 2014.

[20] Julius Fink, Naoki Kikuchi, Shou Yoshida, Kentaro Terada, and Koichi Nakazato. Impact of high versus low fixed loads and non-linear training loads on muscle hypertrophy, strength and force development. *Springerplus*, 5(1):698, 2016.

[21] Emerson Franchini, Braulio M Branco, Marcus F Agostinho, Michel Calmet, and Robin Candau. Influence of linear and undulating strength periodization on physical fitness, physiological, and performance responses to simulated judo matches. *The Journal of Strength & Conditioning Research*, 29(2):358–367, 2015.

[22] Paulo Gentil, James Steele, Maria C Pereira, Rafael PM Castanheira, Antonio Paoli, and Martim Bottaro. Comparison of upper body strength gains between men and women after 10 weeks of resistance training. *PeerJ*, 4:e1627, 2016.

[23] Gederson K Gomes, Cristiane M Franco, Paulo Ricardo P Nunes, and Fábio L Orsatti. High-frequency resistance training is not more effective than low-frequency resistance training in increasing muscle mass and strength in well-trained men. *The Journal of Strength & Conditioning Research*, 33:S130–S139, 2019.

[24] Amelia Guadalupe-Grau, Jorge Perez-Gomez, Hugo Olmedillas, Javier Chavarren, Cecilia Dorado, Alfredo Santana, Jose Antonio Serrano-Sanchez, and Jose AL Calbet. Strength training combined with plyometric jumps in adults: sex differences in fat-bone axis adaptations. *Journal of applied physiology*, 106(4):1100–1111, 2009.

[25] Simon K Harries, David R Lubans, and Robin Callister. Comparison of resistance training progression models on maximal strength in sub-elite adolescent rugby union players. *Journal of Science and Medicine in Sport*, 19(2):163–169, 2016.

[26] Glenn R Harris, MICHAEL H STONE, HAROLD S O'BRYANT, CHRISTOPHER M PROULX, and ROBERT L JOHNSON. Short-term performance effects of high power, high force, or combined weight-training methods. *The Journal of Strength & Conditioning Research*, 14(1):14–20, 2000.

[27] Hagen Hartmann, Andreas Bob, Klaus Wirth, and Dietmar Schmidtbleicher. Effects of different periodization models on rate of force development and power ability of the upper extremity. *The Journal of Strength & Conditioning Research*, 23(7):1921–1932, 2009.

[28] Andrew B Herrick and William J Stone. The effects of periodization versus progressive resistance exercise on upper and lower body strength in women. *The Journal of Strength & Conditioning Research*, 10(2):72–76, 1996.

[29] Jay R Hoffman, Nicholas A Ratamess, Marc Klatt, Avery D Faigenbaum, Ryan E Ross, Nicholas M Tranchina, Robert C McCurley, Jie Kang, and William J Kraemer. Comparison between different off-season resistance training programs in division iii american college football players. *The Journal of Strength & Conditioning Research*, 23(1):11–19, 2009.

[30] Jay R Hoffman, Michael Wendell, Joshua Cooper, and Jie Kang. Comparison between linear and nonlinear in-season training programs in freshman football players. *Journal of strength and conditioning research*, 17(3):561–565, 2003.

[31] David Hostler, Matthew T Crill, Fredrick C Hagerman, and Robert S Staron. The effectiveness of 0.5-lb increments in progressive resistance exercise. *Journal of strength and conditioning research*, 15(1):86–91, 2001.

[32] MONICA J HUBAL, HEATHER Gordish-Dressman, PAUL D THOMPSON, THOMAS B PRICE, ERIC P HOFFMAN, THEODORE J ANGELOPOULOS, PAUL M GORDON, NIALL M MOYNA, LINDA S PESCATELLO, PAUL S VISICH, et al. Variability in muscle size and strength gain after unilateral resistance training. *Medicine & Science in Sports & Exercise*, 37(6):964–972, 2005.

[33] Gary R Hunter. Changes in body composition, body build and performance associated with different weight training frequencies in males and females. *Strength & Conditioning Journal*, 7(1):26–28, 1985.

[34] DE Hurlbut, ME Lott, AS Ryan, RE Ferrell, SM Roth, FM Ivey, GF Martel, JT Lemmer, JL Fleg, and BF Hurley. Does age, sex, or ace genotype affect glucose and insulin responses to strength training? *Journal of Applied Physiology*, 92(2):643–650, 2002.

[35] AC Jozsi, WW Campbell, L Joseph, SL Davey, and WJ Evans. Changes in power with resistance training in older and younger men and women. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 54(11):M591–M596, 1999.

[36] Robert T Kell. The influence of periodized resistance training on strength changes in men and women. *The Journal of Strength & Conditioning Research*, 25(3):735–744, 2011.

[37] Lian-Yee Kok, Peter W Hamer, and David J Bishop. Enhancing muscular qualities in untrained women: linear versus undulating periodization. *Medicine & Science in Sports & Exercise*, 41(9):1797–1807, 2009.

[38] David J Kosek, Jeong-su Kim, John K Petrella, James M Cross, and Marcas M Bamman. Efficacy of 3 days/wk resistance training on myofiber hypertrophy and myogenic mechanisms in young vs. older adults. *Journal of applied physiology*, 101(2):531–544, 2006.

[39] William J Kraemer, Keijo Häkkinen, N Travis Triplett-McBride, Andrew C Fry, L Perry Koziris, Nicholas A Ratamess, JEFFREY E BAUER, Jeff S Volek, Tim McConnell, Robert U Newton, et al. Physiological changes with periodized resistance training in women tennis players. *Medicine & Science in Sports & Exercise*, 35(1):157–168, 2003.

[40] James B Kramer, Michael H Stone, Harold S O'Bryant, Michael S Conley, Robert L Johnson, David C Nieman, Darren R Honeycutt, and Thomas P Hoke. Effects of single vs. multiple sets of weight training: impact of volume, intensity, and variation. *Journal of strength and Conditioning Research*, 11:143–147, 1997.

[41] Kathleen A Martin Ginis, Jeff J Eng, Kelly P Arbour, Joseph W Hartman, and Stuart M Phillips. Mind over muscle?: Sex differences in the relationship between body image change and subjective and objective physical changes following a 12-week strength-training program. *Body image*, 2(4):363–372, 2005.

[42] JOHN R McLESTER, E Bishop, ME Guilliams, et al. Comparison of 1 day and 3 days per week of equal-volume resistance training in experienced subjects. *The Journal of Strength & Conditioning Research*, 14(3):273–281, 2000.

[43] Fabrício Miranda, Roberto Simão, Matthew Rhea, Derek Bunker, Jonato Prestes, Richard Diego Leite, Humberto Miranda, Belmiro Freitas de Salles, and Jefferson Novaes. Effects of linear vs. daily undulatory periodized resistance training on maximal and submaximal strength gains. *The Journal of strength & conditioning research*, 25(7):1824–1830, 2011.

[44] Artur G Monteiro, Marcelo S Aoki, Alexandre L Evangelista, Daniel A Alveno, Gizele A Monteiro, Ivan da Cruz Piçarro, and Carlos Ugrinowitsch. Nonlinear periodization maximizes strength gains in split resistance training routines. *The Journal of Strength & Conditioning Research*, 23(4):1321–1326, 2009.

[45] Harold S O'Bryant, Ronald Byrd, and Michael H Stone. Cycle ergometer performance and maximum leg and hip strength adaptations to two different methods of weight-training. *The Journal of Strength & Conditioning Research*, 2(2):27–30, 1988.

[46] Eisuke Ochi, Masataka Maruo, Yosuke Tsuchiya, Naokata Ishii, Koji Miura, and Kazushige Sasaki. Higher training frequency is important for gaining muscular strength under volume-matched training. *Frontiers in physiology*, 9:744, 2018.

[47] Mark D Peterson, Daniel J Dodd, Brent A Alvar, Matthew R Rhea, and Mike Favre. Undulation training for development of hierarchical fitness and improved firefighter job performance. *The Journal of Strength & Conditioning Research*, 22(5):1683–1695, 2008.

[48] Jonato Prestes, Anelena B Frollini, Cristiane de Lima, Felipe F Donatto, Denis Foschini, Rita de Cássia Marqueti, Aylton Figueira Jr, and Steven J Fleck. Comparison between linear and daily undulating periodized resistance training to increase strength. *The Journal of Strength & Conditioning Research*, 23(9):2437–2442, 2009.

[49] Matthew R Rhea, Stephen D Ball, Wayne T Phillips, and Lee N Burkett. A comparison of linear and daily undulating periodized programs with equated volume and intensity for strength. *The Journal of strength & conditioning research*, 16(2):250–255, 2002.

[50] MATTHEW R RHEA, WAYNE T PHILLIPS, LEE N BURKETT, WILLIAM J STONE, STEPHEN D BALL, BRENT A ALVAR, and AARON B THOMAS. A comparison of linear and daily undulating periodized programs with equated volume and intensity for local muscular endurance. *The Journal of Strength & Conditioning Research*, 17(1):82–87, 2003.

[51] Alex S Ribeiro, Ademar Avelar, Brad J Schoenfeld, Michele CC Trindade, Raphael M Ritti-Dias, Leandro R Altimari, and Edilson S Cyrino. Effect of 16 weeks of resistance training on fatigue resistance in men and women. *Journal of human kinetics*, 42(1):165–174, 2014.

[52] OM Rutherford and DA Jones. The role of learning and coordination in strength training. *European journal of applied physiology and occupational physiology*, 55(1):100–105, 1986.

[53] Emanuel Pericles Salvador, Raphael Mendes Ritti Dias, André Luiz Demantova Gurjão, Ademar Avelar, Luiz Gustavo Pinto, and Edilson Serpeloni Cyrino. Effect of eight weeks of strength training on fatigue resistance in men and women. *Isokinetics and Exercise Science*, 17(2):101–106, 2009.

[54] Matthew K Schiotz, Jeffrey A Potteiger, Phillip G Huntsinger, and Lt Col Donald C Denmark. The short-term effects of periodized and constant-intensity training on body composition, strength, and performance. *The Journal of Strength & Conditioning Research*, 12(3):173–178, 1998.

[55] Dan Schmidt, Kaz Anderson, Marissa Graff, and Victoria Strutz. The effect of high-intensity circuit training on physical fitness. *The Journal of sports medicine and physical fitness*, 56(5):534–540, 2016.

[56] Brad J Schoenfeld, Bret Contreras, D Ogborn, A Galpin, J Krieger, and GT Sonmez. Effects of varied versus constant loading zones on muscular adaptations in trained men. *International journal of sports medicine*, 37(06):442–447, 2016.

[57] Brad J Schoenfeld, Nicholas A Ratamess, Mark D Peterson, Bret Contreras, and Gul Tiryaki-Sonmez. Influence of resistance training frequency on muscular adaptations in well-trained men. *The Journal of Strength & Conditioning Research*, 29(7):1821–1829, 2015.

[58] Roberto Simão, Juliano Spineti, Belmiro Freitas de Salles, Thiago Matta, Liliam Fernandes, Steven J Fleck, Matthew R Rhea, and Helen E Strom-Olsen. Comparison between nonlinear and linear periodized resistance training: hypertrophic and strength effects. *The Journal of strength & conditioning research*, 26(5):1389–1395, 2012.

[59] Eduardo O Souza, Carlos Ugrinowitsch, Valmor Tricoli, Hamilton Roschel, Ryan P Lowery, André Y Aihara, Alberto RS Leão, and Jacob M Wilson. Early adaptations to six weeks of non-periodized and periodized strength training regimens in recreational males. *Journal of sports science & medicine*, 13(3):604, 2014.

[60] Roberto Simao, Tiago Figueiredo, Juliano Spineti, Liliam Fernandes, Humberto Miranda, and Belmiro Freitas de Salles. The effects of exercise order and periodized resistance training on maximum strength and muscle thickness. *International SportMed Journal*, 15(4):374–390, 2014.

[61] MICHAEL H STONE, JEFF A POTTEIGER, KYLE C PIERCE, CHRIS M PROULX, HAROLD S O'BRYANT, ROBERT L JOHNSON, and MEG E STONE. Comparison of the effects of three different weight-training programs on the one repetition maximum squat. *The Journal of Strength & Conditioning Research*, 14(3):332–337, 2000.

[62] Amr Hassan Tammam and Enas Mohamed Hashem. Comparison between daily and weekly undulating periodized resistance training to increase muscular strength for volleyball players. *Journal of Applied Sports Science*, 5(3), 2015.

[63] Michael H Thomas and Steve P Burns. Increasing lean mass and strength: A comparison of high frequency strength training to lower frequency strength training. *International journal of exercise science*, 9(2):159, 2016.

[64] Lawrence W Weiss, Frank C Clark, and David G Howard. Effects of heavy-resistance triceps surae muscle training on strength and muscularity of men and women. *Physical therapy*, 68(2):208–213, 1988.

[65] DL Williamson, PM Gallagher, CC Carroll, U Raue, and SW Trappe. Reduction in hybrid single muscle fiber proportions with resistance training in humans. *Journal of Applied Physiology*, 91(5):1955–1961, 2001.

[66] JACK WILMORE. Alterations in strength, body composition and anthropometric measurements consequent to a 10-week weight training program. *Medicine and Science in Sports*, 6(2):133–138, 1974.

[67] Fu Yue, Bettina Karsten, Eneko Larumbe-Zabala, Marcos Seijo, and Fernando Naclerio. Comparison of 2 weekly-equalized volume resistance-training routines using different frequencies on body composition and performance in trained males. *Applied Physiology, Nutrition, and Metabolism*, 43(5):475–481, 2018.

[68] Rafael S Zaroni, Felipe A Brigatto, Brad J Schoenfeld, Tiago V Braz, Júlio C Benvenutti, Moisés D Germano, Paulo H Marchetti, Marcelo S Aoki, and Charles R Lopes. High resistance-training frequency enhances muscle thickness in resistance-trained men. *The Journal of Strength & Conditioning Research*, 33:S140–S151, 2019.

[69] Matheus Barbalho, Victor Silveira Coswig, James Steele, James P Fisher, Antonio Paoli, and Paulo Gentil. Evidence for an upper threshold for resistance training volume in trained women. *Med Sci Sports Exerc*, 51(3):515–522, 2019.

[70] Matheus Barbalho, Victor S Coswig, James Steele, James P Fisher, Jurgen Giessing, and Paulo Gentil. Evidence of a ceiling effect for training volume in muscle hypertrophy and strength in trained men—less is more? *International journal of sports physiology and performance*, 15(2):268–277, 2020.

[71] Matheus Barbalho, Victor Coswig, Daniel Souza, Julio Cerca Serrão, Mário Hebling Campos, and Paulo Gentil. Back squat vs. hip thrust resistance-training programs in well-trained women. *International Journal of Sports Medicine*, 2020.

[72] Matheus Barbalho, Paulo Gentil, Rodolfo Raiol, James Fisher, James Steele, and Victor Coswig. Influence of adding single-joint exercise to a multijoint resistance training program in untrained young women. *J Strength Cond Res*, 10, 2018.

[73] Matheus Barbalho, Daniel Souza, Victor Coswig, James Steele, James Fisher, Odilon Abrahin, Antonio Paoli, and Paulo Gentil. The effects of resistance exercise selection on muscle size and strength in trained women. *International Journal of Sports Medicine*, 2020.

[74] Matheus Barbalho, Victor Silveira Coswig, Rodolfo Raiol, James Steele, James Fisher, Antonio Paoli, and Paulo Gentil. Effects of adding single joint exercises to a resistance training programme in trained women. *Sports*, 6(4):160, 2018.

[75] Matheus Barbalho, Victor Silveira Coswig, Rodolfo Raiol, James Steele, James P Fisher, Antonio Paoli, Antonino Bianco, and Paulo Gentil. Does the addition of single joint exercises to a resistance training program improve changes in performance and anthropometric measures in untrained men? *European journal of translational myology*, 28(4), 2018.

[76] Matheus Barbalho, Victor Coswig, Rodolfo Raiol, James Fisher, James Steele, Antonino Bianco, and Paulo Gentil. Single joint exercises do not provide benefits in performance and anthropometric changes in recreational bodybuilders. *European journal of sport science*, 20(1):72–79, 2020.

[77] Charles Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15(1):72–101.

[78] Matheus de Siqueira Mendes Barbalho, Paulo Gentil, Mikel Izquierdo, James Fisher, James Steele, and Rodolfo de Azevedo Raiol. There are no no-responders to low or high resistance training volumes among older women. *Experimental gerontology*, 99:18–26, 2017.

[79] Damon L Swift, Joshua E McGee, Conrad P Earnest, Erica Carlisle, Madison Nygard, and Neil M Johannsen. The effects of exercise and physical activity on weight loss and maintenance. *Progress in cardiovascular diseases*, 61(2):206–213, 2018.

[80] J. S. Garrow and C. D. Summerbell. Meta-analysis: effect of exercise, with or without dieting, on the body composition of overweight subjects. *European Journal of Clinical Nutrition*, 49(1):1–10, Jan 1995.

[81] John A. Batsis, Lydia E. Gill, Rebecca K. Masutani, Anna M. Adachi-Mejia, Heather B. Blunt, Pamela J. Bagley, Francisco Lopez-Jimenez, and Stephen J. Bartels. Weight Loss Interventions in Older Adults with Obesity: A Systematic Review of Randomized Controlled Trials Since 2005. *Journal of the American Geriatrics Society*, 65(2):257–268, Feb 2017.

[82] Gary R. Hunter, David R. Bryan, Carla J. Wetzstein, Paul A. Zuckerman, and Marcas M. Bamman. Resistance training and intra-abdominal adipose tissue in older men and women. *Medicine and Science in Sports and Exercise*, 34(6):1023–1028, Jun 2002.

[83] K. H. Schmitz, M. D. Jensen, K. C. Kugler, R. W. Jeffery, and A. S. Leon. Strength training for obesity prevention in midlife women. *International Journal of Obesity and Related Metabolic Disorders : Journal of the International Association for the Study of Obesity*, 27(3):326–333, Mar 2003.

[84] M. S. Treuth, G. R. Hunter, T. Kekes-Szabo, R. L. Weinsier, M. I. Goran, and L. Berland. Reduction in intra-abdominal adipose tissue after strength training in older women. *Journal of Applied Physiology*, Apr 1995.

[85] Leon Mabire, Ramakrishnan Mani, Lizhou Liu, Hilda Mulligan, and David Baxter. The Influence of Age, Sex and Body Mass Index on the Effectiveness of Brisk Walking for Obesity Management in Adults: A Systematic Review and Meta-Analysis. *Journal of Physical Activity & Health*, 14(5):389–407, May 2017.

[86] E. M. Inelmen, E. D. Toffanello, G. Enzi, G. Gasparini, F. Miotto, G. Sergi, and L. Busetto. Predictors of drop-out in overweight and obese outpatients. *International Journal of Obesity (2005)*, 29(1):122–128, Jan 2005.

[87] E. Burgess, P. Hassmén, M. Welvaert, and K. L. Pumpa. Behavioural treatment strategies improve adherence to lifestyle intervention programmes in adults with obesity: a systematic review and meta-analysis. *Clinical Obesity*, 7(2):105–114, Apr 2017.

[88] Brenda M. Davy, Richard A. Winett, Jyoti Savla, Elaina L. Marinik, Mary Elizabeth Baugh, Kyle D. Flack, Tanya M. Halliday, Sarah A. Kelleher, Sheila G. Winett, David M. Williams, and Soheir Boshra. Resist diabetes: A randomized clinical trial for resistance training maintenance in adults with prediabetes. *PloS One*, 12(2):e0172610., Feb 2017.

[89] Tanya M. Halliday, Brenda M. Davy, Adrienne G. Clark, Mary Elizabeth Baugh, Valisa E. Hedrick, Elaina L. Marinik, Kyle D. Flack, J. Savla, Sheila Winett, and Richard A. Winett. Dietary intake modification in response to a participation in a resistance training program for sedentary older adults with prediabetes: Findings from the Resist Diabetes study. *Eating Behaviors*, 15(3):379–382, Aug 2014.

[90] Tanya M. Halliday, Jyoti Savla, Elaina L. Marinik, Valisa E. Hedrick, Richard A. Winett, and Brenda M. Davy. Resistance training is associated with spontaneous changes in aerobic physical activity but not overall diet quality in adults with prediabetes. *Physiology & Behavior*, 177:49–56, Aug 2017.

[91] Cleiton Silva Correa, Bruno Costa Teixeira, Roberto Carlos Rebolledo Cobos, Rodrigo Cauduro Oliveira Macedo, Renata Lopes Kruger, Randall Bruce Kreismann Carteri, Régis Radaelli, Julia Silveira Gross, Ronei Silveira Pinto, and Álvaro Reischak-Oliveira. High-volume resistance training reduces postprandial lipaemia in postmenopausal women. *Journal of Sports Sciences*, 33(18):1890–1901, 2015.

[92] Jennifer W. Bea, Ellen C. Cussler, Scott B. Going, Robert M. Blew, Lauve L. Metcalfe, and Timothy G. Lohman. Resistance training predicts 6-yr body composition change in postmenopausal women. *Medicine and Science in Sports and Exercise*, 42(7):1286–1295, Jul 2010.

[93] CHRISTOPHER B SCOTT and CONRAD P EARNEST. Resistance exercise energy expenditure is greater with fatigue as compared to non-fatigue. *Journal of Exercise Physiology Online*, 14(1), 2011.

[94] George J. Abboud, Beau K. Greer, Sara C. Campbell, and Lynn B. Panton. Effects of load-volume on EPOC after acute bouts of resistance training in resistance-trained men. *Journal of Strength and Conditioning Research*, 27(7):1936–1941, Jul 2013.

[95] Wanderson Divino Nilo Dos Santos, Carlos A. Vieira, Martim Bottaro, Vitória A. Nunes, Rodrigo Ramirez-Campillo, James Steele, James P. Fisher, and Paulo Gentil. Resistance Training Performed to Failure or Not to Failure Results in Similar Total Volume, but With Different Fatigue and Discomfort Levels. *Journal of Strength and Conditioning Research*, Jan 2019.

[96] Body Weight Planner | NIDDK, Jul 2020. [Online; accessed 16. Jul. 2020].

[97] E. J. Dhurandhar, K. A. Kaiser, J. A. Dawson, A. S. Alcorn, K. D. Keating, and D. B. Allison. Predicting adult weight change in the real world: a systematic review and meta-analysis accounting for compensatory changes in energy intake or expenditure. *International Journal of Obesity*, 39(8):1181–1187, Oct 2014.

[98] James Steele, Rebecca Conway, Sarah Weber, Jordan Niblock, James Fisher, Paulo Gentil, Christopher Scott, and Hayao Ozaki. Energy expenditure, physiological, and perceptual responses to a brief, simple bodyweight resistance training protocol in previously sedentary adults. *ResearchGate*, Oct 2018.

[99] Diego T. Brunelli, Enrico A. R. Finardi, Ivan L. P. Bonfante, Arthur F. Gáspari, Amanda V. Sardeli, Thiago M. F. Souza, Mara P. T. Chacon-Mikahil, and Claudia R. Cavaglieri. Acute low- compared to high-load resistance training to failure results in greater energy expenditure during exercise in healthy young men. *PLOS ONE*, 14(11):e0224801, Nov 2019.

[100] Jason Robert Lytle, Danielle M. Kravits, Steven E. Martin, John S. Green, Stephen F. Crouse, and Brad S. Lambert. Predicting Energy Expenditure of an Acute Resistance Exercise Bout in Men and Women. *Medicine and Science in Sports and Exercise*, 51(7):1532–1537, Jul 2019.

[101] S. P. Brown, J. M. Clemons, Q. He, and S. Liu. Prediction of the oxygen cost of the deadlift exercise. *Journal of Sports Sciences*, 12(4):371–375, Aug 1994.

[102] James Steele, Kristin Raubold, Wolfgang Kemmler, James Fisher, Paulo Gentil, and Jürgen Giessing. The Effects of 6 Months of Progressive High Effort Resistance Training Methods upon Strength, Body Composition, Function, and Wellbeing of Elderly Adults. *BioMed Research International*, 2017(2541090.):;, 2017.

[103] Zvi Drezner and Dawit Zerom. A simple and effective discretization of a continuous random variable. *Communications in Statistics-Simulation and Computation*, 45(10):3798–3810, 2016.