Supplementary Material

Summary of Results of Plotkin et al 2022

Aaron R. Caldwell

Table of contents

Scripts and Accessibility

This reproducible analysis document runs source scripts. These can be found on the independent, version-controlled, data repository for this project.

Introduction

This document is a short summary of the re-analysis of the 1-RM back squat outcome that was reported in the manuscript of Plotkin et al. (2022). For this re-analysis, we replicate the model presented in the manuscript wherein the change score was the outcome, the independent variable was the training group (LOAD vs REPS), and two covariates (sex and pre-intervention 1-RM) were included in the model.

We present the average treatment effect using the estimated marginal means using the `emmeans` R package (Lenth 2024). Following this, we present various tests of heteroskedasticity and the bounds for the heterogeneity of treatment effects (Gadbury, Iyer, and Allison 2001).

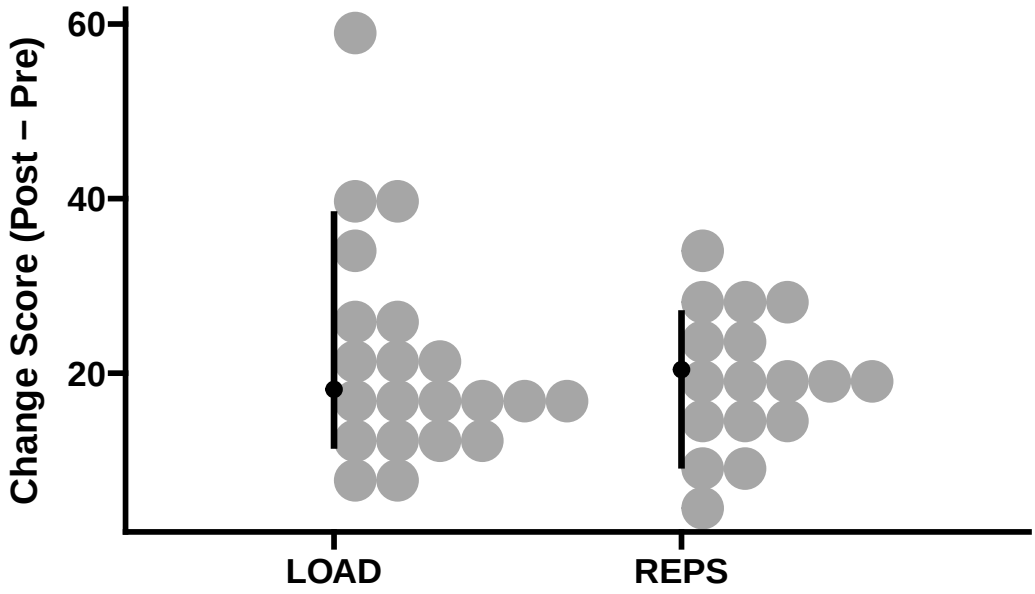
Summary Statistics

This match the publication, barring one SD which appears to be a typo in the original publication[1]

Characteristic	LOAD N = 21[1]	REPS N = 17[1]
Pre	76.9 (29.9)	86.8 (27.1)
Post	98.7 (30.8)	106.2 (29.6)
Change	21.8 (12.2)	19.3 (7.7)

[1]Mean (SD)

Plot of Change Scores



[1]The delta in the LOAD group is incorrectly presented as 21.2 rather than 12.2 in Table 1 of Plotkin et al. (2022)

Model Summary

Note: pre-scores not centered or scaled.

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	23.737	5.480	4.331	0.0001 ***
x1rm_pre	-0.050	0.081	-0.617	0.5413
sexM	3.364	4.589	0.733	0.4685
groupREPS	-2.030	3.563	-0.570	0.5726

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 10.67 on 34 degrees of freedom

Multiple R-squared: 0.0307, Adjusted R-squared: -0.05482

F-statistic: 0.359 on 34 and 3 DF, p-value: 0.7830

Average Treatment Effect

The average treatment effect can be provided by a contrast using the estimated marginal means (`emmeans` package). The differences differ slightly from the publication because we utilize a parametric model while the original authors calculated confidence intervals using bootstrap methods.

contrast	estimate	SE	df	lower.CL	upper.CL
LOAD - REPS	2.030124	3.562845	34	-3.994377	8.054625

Tests for Heteroskedasticity

The results of the different tests are mostly in agreement. However, note that the s_{IR} (last test) is negative which, under the logic of s_{IR} proponents, would end any further investigation of heterogeneity of treatment effects. Additionally, none of the tests are "significant", but such tests would be quite underpowered given the study's sample size.

Breusch-Pagan Test (Model Residuals)

```
# A tibble: 1 x 5
  statistic p.value parameter method                          alternative
      <dbl>   <dbl>     <dbl> <chr>                           <chr>
1      3.39  0.0657         1 Breusch-Pagan (non-studentised) greater
```

Variance Ratio Test

```
    F test to compare two variances

data:  subset(df, group == "REPS")$x1rm_change and subset(df, group == "LOAD")$x1rm_change
F = 0.39076, num df = 16, denom df = 20, p-value = 0.0613
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1534472 1.0475450
sample estimates:
ratio of variances
         0.3907594
```

Log Variability (SD) Test

```
    log transformed variability ratio

data:  SD1 = 7.65405617568727, SD2 = 12.2443837698202
z = -1.9546, p-value = 0.05063
alternative hypothesis: true log ratio of SD ratio is not equal to 0
sample estimates:
    log VR
-0.4635816
attr(,"ni")
[1] 38
attr(,"measure")
[1] "VR"
```

Standard Deviation of Individual Response

```
    Difference in Variances

data:  SD1 = 7.65405617568727, SD2 = 12.2443837698202
z = -1.7655, p-value = 0.07749
alternative hypothesis: true Standard Deviation of the Individual Response is not equal to 0
sample estimates:
Standard Deviation of the Individual Response
                                    -9.557215
```

Bounds on HTE

We can then use the work of Gadbury, Iyer, and Allison (2001) to get the bounds on standard deviation of the treatment effect. A figure and table of the results are below. The average treatment from the model was utilized to calculate the P_{harmed}. Also, we opted to use the method of creating bounds that does *not* utilize covariate information, but the code to do such an analysis is included on our independent repository.

Method to calculate s_D and its confidence intervals:

Let:

- n_x = size of treatment group
- n_y = size of control group
- $\hat{\sigma}_x$ = sample standard deviation of treatment group
- $\hat{\sigma}_y$ = sample standard deviation of control group
- ρ_{xy} = correlation between treatment and control measurements

The standard deviation of individual responses ($\hat{\sigma}_D$) is estimated as follows:

$\hat{\sigma}_D^2 = \hat{\sigma}_x^2 + \hat{\sigma}_y^2 - 2\hat{\sigma}_x\hat{\sigma}_y\rho_{xy}$

The variance of this estimate is given by: $Var(\hat{\sigma}_d^2) = 2\left[\frac{\hat{\sigma}_x^2}{n_x}(\hat{\sigma}_x - \rho_{xy}\hat{\sigma}_y)^2 + \frac{\hat{\sigma}_y^2}{n_y}(\hat{\sigma}_y - \rho_{xy}\hat{\sigma}_x)^2\right]$

For a confidence level of $1-\alpha$, the confidence interval is constructed as:

$\hat{\sigma}_D^2 \pm z_{1-\alpha/2}\sqrt{Var(\hat{\sigma}_d^2)}$

where $z_{1-\alpha/2}$ is the critical value from the standard normal distribution.

The final estimate of the standard deviation of individual responses and its confidence interval are obtained by taking the square root: $\hat{\sigma}_D = \sqrt{|\hat{\sigma}_D^2|} \cdot sign(\hat{\sigma}_D^2)$

For the confidence interval bounds:

$CI_{lower,upper} = \sqrt{|\hat{\sigma}d^2 \pm z1-\alpha/2\sqrt{Var(\hat{\sigma}_d^2)}|} \cdot sign(\hat{\sigma}d^2 \pm z1-\alpha/2\sqrt{Var(\hat{\sigma}_d^2)})$

Method to calculate P_{harmed} and its confidence intervals:

Let's define the additional variables:

- $\hat{\mu}_D$ = estimated mean treatment effect
- $\hat{\sigma}_D$ = estimated standard deviation of individual responses
- $\phi(\cdot)$ = standard normal probability density function
- $\Phi(\cdot)$ = standard normal cumulative distribution function

The probability of harm (unfavorable effect) is estimated as:

$\hat{P}_{harmed} = \Phi\left(\frac{\hat{\mu}_D}{\hat{\sigma}_D}\right)$

The variance of this estimate requires several components:

First, the variance of the mean treatment effect:

$Var(\hat{\mu}_D) = \frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}$

Then, using the delta method, the variance of $\hat{P}_{\_}$ is:

$Var(\hat{P}_{harmed}) = \frac{\phi^2(\hat{\mu}_D/\hat{\sigma}_D)}{\hat{\sigma}_D^2}\left[Var(\hat{\mu}_D) + \frac{\hat{\mu}_D^2 \cdot Var(\hat{\sigma}_D^2)}{4\hat{\sigma}_D^4}\right]$

where $\phi(\hat{\mu}_D/\hat{\sigma}_D)$ is the standard normal density evaluated at $\hat{\mu}_D/\hat{\sigma}_D$

For a confidence level of $1-\alpha$, the confidence interval is constructed as:

$\hat{P}_{harmed} \pm z_{1-\alpha/2}\sqrt{Var(\hat{P}_{harmed})}$

where $z_{1-\alpha/2}$ is the critical value from the standard normal distribution. This formulation provides an estimate of the probability that an individual will experience an unfavorable response to treatment, along with uncertainty bounds around that estimate.

Note: The confidence interval bounds should be constrained to [0,1] since we are estimating a probability. However, as we can see in the visualizations, with small sample sizes this approach can produce bounds outside the interval.

Visualization of s_D and P_{harmed}

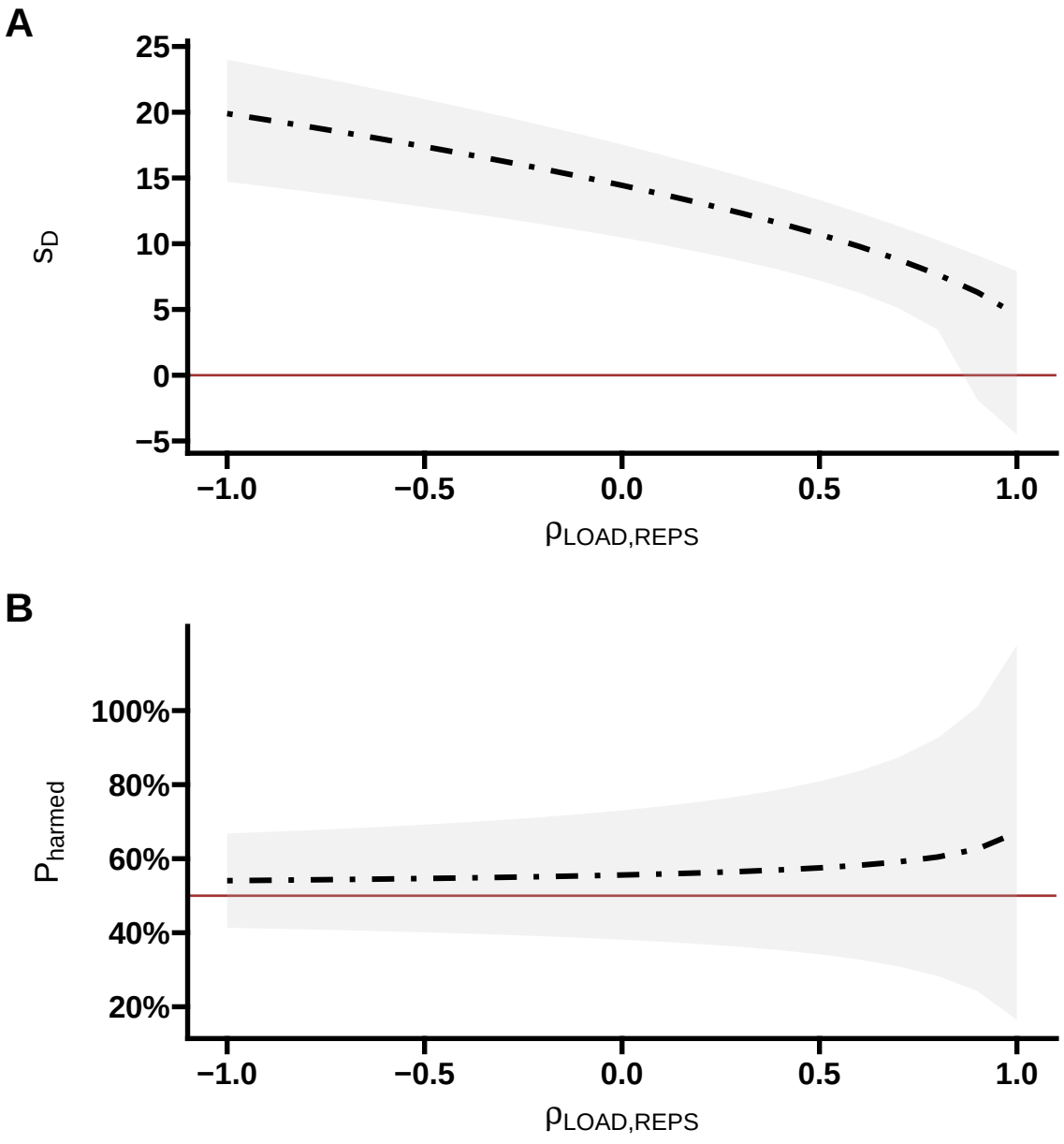


Table of Bounds

rho	sigma_d	sigma_d_lower	sigma_d_upper	p_minus	p_minus_lower	p_minus_upper
-1.00	19.90	14.71	23.99	0.54	0.41	0.67
-0.90	19.42	14.35	23.42	0.54	0.41	0.67
-0.80	18.93	13.98	22.84	0.54	0.41	0.68
-0.70	18.43	13.60	22.24	0.54	0.41	0.68
-0.60	17.92	13.20	21.62	0.55	0.40	0.69
-0.50	17.38	12.79	20.99	0.55	0.40	0.69
-0.40	16.84	12.37	20.35	0.55	0.40	0.70
-0.30	16.27	11.93	19.68	0.55	0.39	0.70
-0.20	15.68	11.47	18.99	0.55	0.39	0.71
-0.10	15.07	10.98	18.27	0.55	0.39	0.72
0.00	14.44	10.47	17.53	0.56	0.38	0.73
0.10	13.78	9.93	16.76	0.56	0.38	0.74
0.20	13.08	9.34	15.96	0.56	0.37	0.75
0.30	12.34	8.71	15.12	0.57	0.36	0.77
0.40	11.56	8.00	14.25	0.57	0.35	0.79
0.50	10.71	7.21	13.33	0.58	0.34	0.81
0.60	9.80	6.27	12.36	0.58	0.33	0.84
0.70	8.79	5.11	11.34	0.59	0.31	0.87
0.80	7.65	3.45	10.26	0.60	0.28	0.93
0.90	6.31	-1.86	9.12	0.63	0.24	1.01
1.00	4.59	-4.51	7.90	0.67	0.16	1.18

References

Gadbury, Gary L., Hari K. Iyer, and David B. Allison. 2001. "Evaluating Subject-Treatment Interaction When Comparing Two Treatments." *Journal of Biopharmaceutical Statistics* 11 (4): 313–33. https://doi.org/10.1081/bip-120008851.

Lenth, Russell V. 2024. "Emmeans: Estimated Marginal Means, Aka Least-Squares Means." https://CRAN.R-project.org/package=emmeans.

Plotkin, Daniel, Max Coleman, Derrick Van Every, Jaime Maldonado, Douglas Oberlin, Michael Israetel, Jared Feather, Andrew Alto, Andrew D. Vigotsky, and Brad J. Schoenfeld. 2022. "Progressive Overload Without Progressing Load? The Effects of Load or Repetition Progression on Muscular Adaptations." *PeerJ* 10 (September): e14142. https://doi.org/10.7717/peerj.14142.