Effect sizes for paired data should use the change score variability rather than the pre-test variability:

A Paper by Dr. Scott Dankel and Dr. Jeremy P. Loenneke

Alex Gould

2024-10-03

Full Citation

Dankel, SJ and Loenneke, JP. Effect sizes for paired data should use the change score variability rather than the pre-test variability. J Strength Cond Res 35(6): 1773–1778, 2021—}

What is effect size

- Variable that provides an overall measure for magnitude of change (Dankel and Loenneke (2021))
- Differs from a T-statistic because sample size is not included
- Quantifies how much the mean of the post-treatment differs from the mean of the baseline score in terms of a certain standard deviation.
    - Specifically, they are looking at this comparison from the lens of meta-analyses for exercise science and sports medicine

Authors' Aims

- To convince the audience through analysis that baseline and post-test standard deviations (study sample measures of variability) don't tell the full story on the overall variability of the intervention.
- To convince the readers that the heterogeneity of the study sample can play a part in unintentionally influencing effect size measurements.

How would one describe the two types of aforementioned variability?

Variability of the Study Sample

- Any measure of difference between subjects in a given treatment group
- Represented by the Baseline and Post-treatment Standard Deviation.
- Dankel and Loenneke (2021) and his team claim that the use of this type of variability in paired-sample studies is useless as it has nothing to do with the treatment itself

Variability of the Intervention

- Any measure of difference between baseline and post-treatment measure
- Represented in this case by the Standard Deviation of Change Scores (I will elaborate on this later)
- Dankel and his team prefer this method of assessing variability

Dr. Scott Dankel

- Professor at Rowan University, a public research university in New Jersey
- Attended the University of Mississippi to pursue a Masters and PhD in Exercise Science
- Research Interests include acute and chronic adaptations to blood flow restricted exercise (Lab 2024)

Dr. Jeremy Paul Loenneke

- Professor at The University of Mississippi
- Attended Southeast Missouri State for his Bachelors and Masters in Nutrition and Exercise Science
- Eventually got his PhD in Exercise Physiology at the University of Oklahoma
- Research Discipline is in Skeletal Muscle Plasticity (Ole Miss 2024)
- Regarding the disciplines of the authors, this paper was published in The Journal of Strength and Conditioning Research. This is a good example of the use of statistics as an interdisciplinary tool

Introduction

Specific Effect Size Measures

The author's claim that the common effect size measures listed below are used exhaustively in meta-analyses in the exercise science discipline.

- Cohen's d (Cite)
- Hedge's g (Cite)
- Glass delta (Cite)
- Each use some combination of baseline standard deviation and post-treatment standard deviation.
- Measures of variability of the study sample

Paired Data vs. Independent Data

Independent Data

- Data collected through an Independent design
  - Each subject is only measured once
  - Subjects are allocated into a baseline group and a post-treatment group
  - Study sample variability is more important
  - The pooled standard error is the way to assess this variability

Paired Data

- Data that is collected through a Paired Sample design
  - Same subject is assessed at both time points.
  - Since its based on the same subject, this data is not independent
  - In this type of Design, study sample variability is irrelevant
  - Variability of assessed by standard error of the change scores

Since most meta-analysis data is paired, Dankel's analysis focuses on primarily paired-sample designs. Therefore, the authors believe that intervention variability is the best measure for this specific analysis.

Methods

Preliminary measures

$$M_{change} = \text{Difference between means of Posttreatment group and baseline group in an indepenc}$$

$$SD_{bsl} = \text{Standard Deviation of the baseline group in an independent design}$$

$$SD_{post} = \text{Standard Deviation of the posttreatment group in an independent design}$$

$$n_{bsl} = \text{The sample size of the baseline group}$$

$$n_{post} = \text{The sample size of the posttreatment group}$$

$$SD_{\text{pooled}} = \sqrt{\frac{(n_{\text{bsl}} - 1)SD^2_{\text{bsl}} + (n_{\text{post}} - 1)SD^2_{\text{post}}}{n_{\text{bsl}} + n_{\text{post}} - 2}}$$

Calculations of Common Effect Size measures

Cohen's $d = \frac{M_{\text{change}}}{SD_{\text{pooled}}}$

Glass's $\delta = \frac{M_{\text{change}}}{SD_{\text{bsl}}}$

Hedge's $g = C * \frac{M_{\text{change}}}{SD_{\text{pooled}}}$

Where C is a factor depending on n multiplied to account for small sample sizes

We can calculate an Independent t test statistic:

$$T_{\text{indep}} = \frac{M_{\text{change}}}{\sqrt{SD_{\text{pooled}}\left(\frac{1}{n_{\text{bsl}}} + \frac{1}{n_{\text{post}}}\right)}}$$

Where d represents Cohen's d. From this it is apparent how using the incorrect effect size can be detrimental to a given analysis when using a Paired Design.

Analysis and Procedure

Figure 1

- Describes a dataset with two interventions which have equivalent pre and post scores
- One intervention has a pre-post correlation of 0.99 and the other one has a pre-post correlation of -0.99.
- Seek to prove that pre and post scores dont tell the full story on intervention variability

Figure 2

- Describes a dataset with two interventions such that one intervention has a higher pre and post standard deviation than intervention 2
- They are both correlated the same way
- Seek to prove that heterogeneity of groups can have a profound effect on Common effect size measures but not on the measure that they define as the "appropriate" effect size

Figure 1

Graph A shows that both interventions have the same pre and post means and standard deviations. Graph B shows the graph between pre and post, outlining opposite correlations, and Graph C is a representation of the standard error for each of the intervention groups, showing high variability for intervention 2 and low variability for intervention 1. Supports Aim 1.

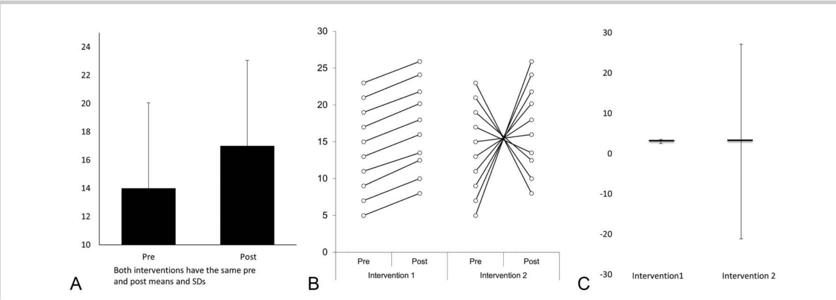
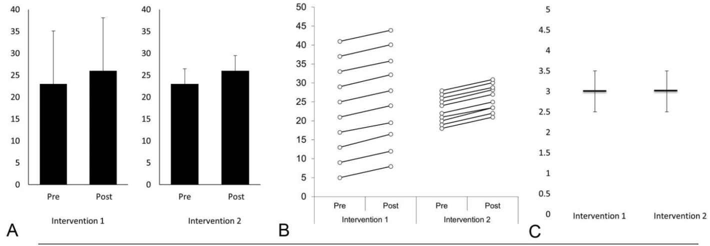


Figure 2

Graph A shows that each intervention has the same pre and post mean but different pre-post standard deviations. Graph B shows the graph between pre and post, outlining the same correlations but different pre/post variabilities, and Graph 3 illustrates the variability further by showing that the intervention variability is the same. Supports Aim 2.



Results (Dankel et. al)

	Pre	Post	r	Change Score	Cohen's d	Glass delta	Hedges g	T statistic	P-value	Appropriate effect size
Intervention 1	14 (SD: 6.05)	17.0 (SD: 6.03)	0.99	3.0 (SD: 0.2582)	0.496	0.497	0.475	36.74	<0.001	11.618
Intervention 2	14 (SD: 6.05)	17.0 (SD: 6.03)	-0.99	3.0 (SD: 12.084)	0.496	0.497	0.475	0.785	0.453	0.248

	Pre	Post	r	Change Score	Cohen's d	Glass delta	Hedges g	T statistic	P-value	Appropriate effect size
Intervention 1	23 (SD: 12.11)	26 (SD: 12.08)	0.99	3.0 (SD: 0.2582)	0.248	0.248	0.237	36.74	<0.001	11.618
Intervention 2	23 (SD: 3.49)	26 (SD: 3.48)	0.99	3.0 (SD: 0.2582)	0.86	0.862	0.824	36.74	<0.001	11.618

- Respective to the figures above
- Top attempted to prove aim 1
- Bottom attempted to prove aim 2
- In both, 11.618 is mentioned as the "Appropriate effect size" even though effect sizes rarely range over 4.
- Standard Deviation of the change scores is suspiciously low as well

My Results

Recalculate Cohen's dz and show if it changes the significance at all. mention: $SD_{\text{change}}` = \sqrt{SD_{\text{pre}}^2 + SD_{\text{post}}^2 - 2rSD_{pre}SD_{post}}$ this is the formula used in the article

Discussion

General

Impossible to quantify variability of actual intervention when using pre-post scores
Figure 1: Intervention 1 produces positive effect where intervention two does not
Common effect size measures are reliant on the heterogeneity of the study population
Figure 2 shows identical results in 2 diff groups with the only difference being that intervention 1 used a more heterogeneous population

Other Issues/Notes

Notes: normalizing using the preSD only is dependent on the subjects recruited rather than effectiveness
Also used in triangulation when trying to obtain sample size calculations

(“Any Questions Image,” n.d.)

References

"Any Questions Image." n.d. Webpage. https://mavink.com/explore/Images-for-Any-Questions.

Dankel, Scott J., and Jeremy P. Loenneke. 2021. "EFFECT SIZES FOR PAIRED DATA SHOULD USE THE CHANGE SCORE VARIABILITY RATHER THAN THE PRE- TEST VARIABILITY." *Journal of Strength and Conditioning Research* 6 (35): 1773–78.

Lab, Scott J. Dankel. 2024. "About Dr. Dankel." https://research.rowan.edu/research-areas/hes/dankel/about.html.

Ole Miss, Jeremy Paul Loenneke Lab at. 2024. "Jeremy Paul Loenneke Introduction." https://hesrm.olemiss.edu/people/jeremy-paul-loenneke/.