DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 8 | Issue 2 Article 27

11-1-2009

Estimation of the Standardized Mean Difference for Repeated Measures Designs

Lindsey J. Wolff Smith
The University of Texas at Austin, ljwolff@hotmail.com

S. Natasha Beretvas
The University of Texas at Austin, tasha.beretvas@mail.utexas.edu

Journal of Modern Applied Statistical Methods
November 2009, Vol. 8, No. 2, 600-609

1538 – 9472/09/$95.00

Estimation of the Standardized Mean Difference for Repeated Measures Designs

Lindsey J. Wolff Smith S. Natasha Beretvas
The University of Texas at Austin

This simulation study modified the repeated measures mean difference effect size, $d_{RM}^{=}$, for scenarios with unequal pre- and post-test score variances. Relative parameter and *SE* bias were calculated for $d_{RM}^{\neq}$ versus $d_{RM}^{=}$. Results consistently favored $d_{RM}^{\neq}$ over $d_{RM}^{=}$ with worse positive parameter and negative *SE* bias identified for $d_{RM}^{=}$ for increasingly heterogeneous variance conditions.

Key words: meta-analysis, repeated measures, effect size

Introduction

Meta-analysis (Glass, 1976) entails pooling of results from related studies in an effort to synthesize the research results. Studies typically use various experimental designs and thus various effect size measures. In quantitative meta-analysis, a primary goal is to combine effect sizes to produce an overall effect size.

An effect size (ES) index is used to quantify the strength of the relationship between two variables. Each study's finding can be represented as an ES. The use of the ES is important as it allows for the comparison of multiple studies' results. ES indices do, however, differ depending on the type of study performed (e.g., repeated measures, independent groups, etc.). Although multiple effect sizes can be handled using meta-analysis, the effect size of interest in this study is the standardized mean difference for repeated measures designs, δ_{RM}.

Lindsey J. Wolff Smith is a doctoral student in Quantitative Methods in the Department of Educational Psychology. Her interests are in the evaluation of statistical models using simulation research. Email her at: ljwolff@hotmail.com. S. Natasha Beretvas is an Associate Professor and chair of Quantitative Methods in the Department of Educational Psychology. Her interests are in multilevel and meta-analytic modeling techniques. Email her at: tasha.beretvas@mail.utexas.edu.

The formula for the δ_{RM} and its associated variance have been derived by Becker (1988) and Morris and DeShon (2002). The δ_{RM} is necessary for summarizing results from a repeated measures (RM) design in which the same subjects are measured before and after a treatment is administered. Many primary studies employ the RM design. This design allows the researcher to assess change in an outcome that occurs within a subject as a result of what happens between a pre- and post-test. Little research has been done to assess the relative parameter and standard error bias of δ_{RM} estimates.

In the RM design, one group of subjects is measured before and after a treatment is administered. The RM design's ES measure is defined as follows:

$$\delta_{RM} = \frac{\mu_{post} - \mu_{pre}}{\sigma_D} = \frac{\mu_D}{\sigma_D} \tag{1}$$

where μ_{pre} and μ_{post} are the population means of the pre- and post-test scores, respectively, μ_D is the population mean difference in the pre- and post-test scores, and σ_D is the standard deviation of change scores (Gibbons, Hedeker, & Davis, 1993). The associated sample estimate is calculated as follows:

$$d_{RM} = \frac{\bar{X}_{post} - \bar{X}_{pre}}{s_D}, \tag{2}$$

where $\overline{X}_{pre}$ and $\overline{X}_{post}$ are the sample means of the pre- and post-test scores, respectively, and s_D is the sample standard deviation of difference scores.

The sampling variance formula for δ_{RM} is:

$$\sigma^2_{\delta_{RM}} = \left(\frac{1}{n}\right)\left(\frac{n-1}{n-3}\right)\left(1+n\delta^2_{RM}\right)-\frac{\delta^2_{RM}}{\left[c\left(n-1\right)\right]^2} \tag{3}$$

where n is the number of paired observations in the RM design study (Morris & DeShon, 2002) with a corresponding formula used for sample estimates:

$$s^2_{d_{RM}} = \left(\frac{1}{n}\right)\left(\frac{n-1}{n-3}\right)\left(1+nd^2_{RM}\right)-\frac{d^2_{RM}}{\left[c\left(n-1\right)\right]^2}. \tag{4}$$

Equations 3 and 4 also contain the bias correction factor, $c(n-1)$, that is approximated by

$$c\left(n-1\right)=1-\frac{3}{4\left(n-1\right)-1} \tag{5}$$

(Hedges, 1982).

Calculation of σ_D is necessary to obtain δ_{RM} (see Equation 1). Morris and DeShon (2002) presented the following relationship between the standard deviation of difference scores, σ_D, and the standard deviation of the original scores, σ:

$$\sigma^{=}_{D} = \sigma\sqrt{2\left(1-\rho\right)} \tag{6}$$

where ρ is the correlation between the pre- and post-test scores. The corresponding sample estimate is:

$$s^{=}_{D} = s\sqrt{2(1-r)} \tag{7}$$

with r representing the sample correlation. Both formulas (Equations 6 and 7) are founded on the assumption that the population standard deviations for the pre- and post-test scores are equal (i.e., $\sigma_{pre}=\sigma_{post}=\sigma$). Thus, the notation of including a superscript with = was adopted to distinguish the relevant formula when $\sigma_{pre}=\sigma_{post}$ is assumed from scenarios in which $\sigma_{pre} \neq \sigma_{post}$ is assumed.

If $\sigma_{pre} \neq \sigma_{post}$, another formula for σ_D must be employed that does not assume equal variances, namely:

$$\sigma^{\neq}_{D} = \sqrt{\sigma^2_{pre}+\sigma^2_{post}-2\sigma_{pre,post}} \tag{8}$$

where σ^2_{pre} and σ^2_{post} are the population variances of the pre- and post-groups, respectively, and $\sigma_{pre,post}$ is the covariance between the pre- and post-test scores such that:

$$\sigma_{pre,post}=\rho\sigma_{pre}\sigma_{post}. \tag{9}$$

Therefore, the equation for $\sigma^{\neq}_{D}$ (see Equation 8) becomes:

$$\sigma^{\neq}_{D}=\sqrt{\sigma^2_{pre}+\sigma^2_{post}-2\rho\sigma_{pre}\sigma_{post}}. \tag{10}$$

The corresponding sample estimate is then:

$$s^{\neq}_{D} = \sqrt{s^2_{pre}+s^2_{post}-2rs_{pre}s_{post}}. \tag{11}$$

Note that when $\sigma_{pre}=\sigma_{post}$, Equations 10 and 11 reduce to the corresponding (population and sample) homogeneous variances formula for σ_D (and s_D) (see Equations 6 and 7, respectively).

This leads to the two primary foci of this study. First, empirical research has not been conducted to assess how well the formulas for δ_{RM} and for $\sigma^2_{\delta_{RM}}$ work in terms of parameter and standard error (*SE*) bias when pre- and post-test scores are and are not homogeneous. Second, applied meta-analysts assume the homogeneity of the pre- and post-test scores and use the $s^{=}_{D}$ formula (Equation 7) as opposed to $s^{\neq}_{D}$ (Equation 11) when calculating the estimate of δ_{RM} (Equation 2). Thus, this study also investigated the effect of using the conventional formula for $s^{=}_{D}$ (Equation 7) when the homogeneity of variance assumption is violated and the modified formula for s_D (i.e., $s_D^{\neq}$ in Equation 11) should be used.

In the current simulation study, four design factors were manipulated, including: the true value of δ_{RM}, the correlation between pre- and post-test scores, sample size, and values for the pre- and post-test score standard deviations to assess the effect of these factors on parameter and *SE* estimates of δ_{RM}. Results were compared when the pre- and post-test scores were assumed to have equal variances ($\sigma_{pre}^2 = \sigma_{post}^2$), thus $s_D^{=}$ was used to calculate d_{RM} (i.e., providing $d_{RM}^{=}$) with the results based on the assumption that $\sigma_{pre}^2 \neq \sigma_{post}^2$ for which $s_D^{\neq}$ was calculated and used to obtain the associated d_{RM} (i.e., $d_{RM}^{\neq}$).

Methodology

A Monte Carlo simulation study was conducted to assess the relative parameter and *SE* bias of the two estimates of δ_{RM}. The two estimates, $d_{RM}^{=}$ and $d_{RM}^{\neq}$, are distinguished by the formula used to calculate the sample standard deviation of the difference (Equation 7 versus Equation 11). Four design factors were manipulated in this study and are described in detail below. R software version 2.8.1 was used to generate the data and to estimate and summarize all relevant parameters.

δ_{RM}

True values of δ_{RM} were manipulated to assess their effect on parameter and *SE* estimation. These values included: no effect, and small, moderate, and large effects (δ_{RM} = 0, 0.2, 0.5, and 0.8, respectively).

Correlation Between Pre- and Post-Test Scores

The following values of the true correlation, ρ, between pre- and post-test scores were manipulated to evaluate the effect of no, a small, moderate, and large correlation (ρ = 0, 0.2, 0.5, and 0.8, respectively).

Sample Size

Sample size was investigated at three levels including a small, moderate, and moderately large sample size (n = 10, 20, and 60, respectively). Note that the sample sizes used were the same for each of the pre- and post-test groups.

Ratio of the Pre- and Post-Test Scores' Standard Deviations

Five different values of the ratio of the pre- and post-test scores' standard deviations were investigated. The following patterns were evaluated: $\sigma_{pre} = \sigma_{post}$, $\sigma_{pre} < \sigma_{post}$, and $\sigma_{pre} > \sigma_{post}$. For the two unequal standard deviations' conditions, the degree of the difference was also manipulated, with the following four unequal combinations of values for $\sigma_{pre}:\sigma_{post}$ investigated: 0.8:1.2, 0.5:1.5, 1.2:0.8, and 1.5:0.5. For the $\sigma_{pre} = \sigma_{post}$ conditions, both pre- and post-test true standard deviations were generated to be one (i.e., $\sigma_{pre} = \sigma_{post} = 1$).

Repeated Measures Effect Size

To manipulate the true value of δ_{RM}, the value of μ_{pre} was set to zero across conditions and the value of μ_{post} was derived to result in the following values for δ_{RM}: 0, 0.2, 0.5, and 0.8. Specifically, μ_{post} is a function of δ_{RM}, μ_{pre}, and σ_D (see Equation 1) and thus can be derived because

$$\mu_{post} = (\delta_{RM})(\sigma_D) + \mu_{pre} \qquad (12)$$

and the values of δ_{RM} and σ_D are determined by the relevant conditions with μ_{pre} always set to zero.

Estimates of δ_{RM}

For each generated dataset, Equation 2 was used to calculate the sample standardized mean difference effect size for RM designs. Two values for s_D ($s_D^{=}$ and $s_D^{\neq}$) were used with the former based on the assumption of equal pre- and post-test score variances (Equation 7) and the latter based on the assumption that $\sigma_{pre}^2 \neq \sigma_{post}^2$ (Equation 11). The resulting estimates were termed $d_{RM}^{=}$ and $d_{RM}^{\neq}$, respectively.

Data Generation

For each set of conditions, a set of random, bivariate normally distributed scores (correlated in the population with the condition's value for ρ) were generated to provide the pre- and post-test scores for that condition's replication. Two values of d_{RM} ($d_{RM}^{=}$ and $d_{RM}^{\neq}$) were calculated using each dataset as described above. Ten thousand replication datasets were generated for each combination of conditions.

Bias Assessment

Relative parameter and *SE* estimation bias of each d_{RM} ($d_{RM}^{\neq}$ and $d_{RM}^{=}$) was summarized and assessed using Hoogland and Boomsma's (1998) formulas and criteria. More specifically, relative parameter bias was calculated using the following formula:

$$B\left(\hat{\theta}_j\right)=\frac{\left(\bar{\hat{\theta}}_j-\theta_j\right)}{\theta_j} \tag{13}$$

where θ_j represents the j^{th} parameter's true value and $\bar{\hat{\theta}}_j$ is the mean estimate of parameter j averaged across the 10,000 replications per condition. Hoogland and Boomsma recommended considering a parameter's estimate as substantially biased if its relative parameter bias exceeds 0.05 in magnitude. This cutoff means that estimates that differ from their parameter's true value by more than five percent should be considered substantially biased.

Hoogland and Boomsma's (1998) commonly used formulation of relative standard error bias is as follows:

$$B\left(s_{\hat{\theta}_j}\right)=\frac{\left(\bar{\hat{s}}_{\hat{\theta}_j}-\sigma_{\hat{\theta}_j}\right)}{\sigma_{\hat{\theta}_j}} \tag{14}$$

where $\bar{\hat{s}}_{\hat{\theta}_j}$ is the mean of the *SE* estimates associated with parameter estimates of θ_j and $\sigma_{\hat{\theta}_j}$ is the empirically true standard error of the distribution of $\hat{\theta}_j$ s calculated by computing the standard deviation of each conditions' 10,000 $\hat{\theta}_j$ s. Hoogland and Boomsma recommended using a cutoff of magnitude 0.10 indicating substantial relative *SE* bias. Note that, for conditions in which the true parameter, δ_{RM}, was zero, simple parameter estimation bias was calculated.

Results

Results are presented in three sections, one for each of the three sample size conditions. Note that relative parameter bias is not calculable if the true parameter value is zero (see Hoogland & Boomsma, 1998), thus, simple bias rather than relative bias is calculated for conditions in which the true δ_{RM} is zero.

Sample Size = 10: Relative Parameter Bias

Substantial positive relative parameter bias was identified for all non-zero values of δ_{RM} and ρ. No substantial bias was found in the $\rho = 0$ conditions. In all cases, the positive bias identified was greater when $d_{RM}^{=}$ was used rather than $d_{RM}^{\neq}$ (see Table 1). No criterion exists to indicate whether simple bias is substantial or not, however, the simple bias values seem small for the $\delta_{RM} = 0$ conditions. When $d_{RM}^{=}$ was used, the more the ratio of $\sigma_{pre} : \sigma_{post}$ values diverged from 1:1, the worse the bias. Similarly, the stronger the ρ, the worse the bias for the $d_{RM}^{=}$ estimate.

The $d_{RM}^{\neq}$ estimator was unaffected by the $\sigma_{pre}:\sigma_{post}$ and ρ values. However, substantial bias was detected for both $d_{RM}^{\neq}$ and $d_{RM}^{=}$ even when $\sigma_{pre}:\sigma_{post}$ was 1:1. Patterns of bias identified for a given $\sigma_{pre}:\sigma_{post}$ ratio closely mimicked patterns identified for the inverse ratio. Thus, across conditions, results found for the 1.5:0.5 ratio matched those for the 0.5:1.5 ratio. Similarly, results for the 0.8:1.2 ratio conditions matched those for the 1.2:0.8 ratio. This result held across all conditions including the three sample sizes and thus will not be mentioned further. Parameter estimation performance of both the $d_{RM}^{\neq}$ and $d_{RM}^{=}$ estimators was unaffected by the true δ_{RM} value (see Table 1). The positive parameter estimation bias of the $d_{RM}^{\neq}$ estimator was pretty consistently close to 10% across the n = 10 conditions.

Table 1: Summary of Relative Parameter Estimation Bias by Generating Condition for n = 10 Conditions

	$\sigma_{pre}:\sigma_{post}$ Ratio Value									
	1:1		0.8:1.2		0.5:1.5		1.2:0.8		1.5:0.5	
Condition	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$
ρ Value										
0	0.002	0.003	-0.003	-0.003	0.002	0.004	0.002	0.002	0.002	0.004
0.2	***0.090***	***0.106***	***0.097***	***0.130***	***0.096***	***0.199***	***0.093***	***0.125***	***0.087***	***0.190***
0.5	***0.092***	***0.126***	***0.099***	***0.182***	***0.087***	***0.355***	***0.086***	***0.168***	***0.092***	***0.358***
0.8	***0.105***	***0.165***	***0.097***	***0.325***	***0.100***	***0.896***	***0.088***	***0.311***	***0.089***	***0.866***
δ_{RM} Value										
0[a]	0.002	0.003	-0.003	-0.003	0.002	0.004	0.002	0.002	0.002	0.004
0.2	***0.103***	***0.133***	***0.104***	***0.194***	***0.091***	***0.397***	***0.090***	***0.178***	***0.072***	***0.360***
0.5	***0.089***	***0.118***	***0.101***	***0.190***	***0.094***	***0.392***	***0.088***	***0.175***	***0.092***	***0.390***
0.8	***0.092***	***0.121***	***0.093***	***0.181***	***0.093***	***0.394***	***0.088***	***0.176***	***0.096***	***0.399***
Overall[b]	***0.095***	***0.124***	***0.099***	***0.188***	***0.093***	***0.394***	***0.089***	***0.177***	***0.087***	***0.383***

Notes: Substantial relative parameter bias values are highlighted in the table; [a]Mean simple bias is presented for δ_{RM} = 0 conditions; [b] Overall = mean relative parameter bias across all δ_{RM} conditions excluding δ_{RM} = 0 conditions.

Table 2: Summary of Relative Standard Error Estimation Bias by Generating Condition for n = 10 Conditions

	$\sigma_{pre}:\sigma_{post}$ Ratio Value									
	1:1		0.8:1.2		0.5:1.5		1.2:0.8		1.5:0.5	
Condition	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$
ρ Value										
0	0.048	0.032	0.046	0.023	0.051	-0.012	0.051	0.027	0.049	-0.011
0.2	0.048	0.062	0.046	0.039	0.051	0.048	0.051	0.053	0.049	0.047
0.5	0.051	0.012	0.041	-0.034	0.046	***-0.147***	0.046	-0.028	0.039	***-0.152***
0.8	0.043	-0.013	0.048	***-0.112***	0.056	***-0.308***	0.047	***-0.110***	0.042	***-0.317***
δ_{RM} Value										
0	0.046	0.016	0.044	-0.031	0.043	***-0.143***	0.042	-0.032	0.038	***-0.145***
0.2	0.041	0.011	0.036	-0.039	0.049	***-0.135***	0.042	-0.030	0.042	***-0.142***
0.5	0.057	0.022	0.047	-0.025	0.049	***-0.134***	0.051	-0.023	0.046	***-0.129***
0.8	0.060	0.020	0.046	-0.027	0.060	***-0.111***	0.061	-0.014	0.052	***-0.120***
Overall[a]	0.051	0.017	0.043	-0.031	0.050	***-0.131***	0.049	-0.025	0.044	***-0.134***

Notes: Substantial relative *SE* bias values are highlighted in the table; [a]Overall = mean relative *SE* bias across all δ_{RM} conditions excluding δ_{RM} = 0 conditions.

Sample Size = 10: Relative *SE* Bias

No relative *SE* bias was found for $d_{RM}^{\neq}$ for the n = 10 conditions (see Table 2). For $d_{RM}^{=}$, however, substantial negative bias was identified in certain conditions. Substantial negative bias (i.e., $\left|B\left(s_{\hat{\theta}_j}\right)\right| > 0.10$, see Equation 14) was found at the most extreme $\sigma_{pre}:\sigma_{post}$ values (i.e., when $\sigma_{pre}:\sigma_{post}$ = 1.5:0.5 and $\sigma_{pre}:\sigma_{post}$ = 0.5:1.5). This bias occurred for conditions in which ρ = 0.5 or larger and the magnitude of the bias seemed to be slightly larger for smaller δ_{RM} (see Table 2). Substantial negative parameter estimation bias was also detected for $d_{RM}^{=}$ for $\sigma_{pre}:\sigma_{post}$ = 0.8:1.2 and for $\sigma_{pre}:\sigma_{post}$ = 1.2:0.8 for the largest ρ condition (i.e., when ρ = 0.8).

Sample Size = 20: Relative Parameter Bias

No substantial parameter bias was identified when $d_{RM}^{\neq}$ was used to estimate δ_{RM} across the n = 20 conditions (see Table 3). Substantial positive relative parameter bias was found when $d_{RM}^{=}$ was used to estimate δ_{RM}, however, the degree of parameter bias was lower for the n = 20 conditions (see Table 3) than was observed for the n = 10 conditions (in Table 1).

No substantial relative parameter bias was found in the ρ = 0 conditions for $d_{RM}^{=}$. With the slightly larger sample size, no substantial bias was detected when the $\sigma_{pre}:\sigma_{post}$ ratio was 1:1. Otherwise, the pattern of the bias found matched that noted for the n = 10 conditions. The more the value of the $\sigma_{pre}:\sigma_{post}$ ratio diverged from 1:1 (and for larger ρ values), the more the degree of substantial parameter bias increased. Values of δ_{RM} did not seem to affect the degree of bias (see Table 3).

Sample Size = 20: Relative *SE* Bias

The relative *SE* bias results for the n = 20 conditions (see Table 4) very closely matched those described for the n = 10 conditions (see Table 2). No substantial relative *SE* bias was found when using $d_{RM}^{\neq}$ to estimate δ_{RM}. For $d_{RM}^{=}$, however, in the most extreme $\sigma_{pre}:\sigma_{post}$

Table 3: Summary of Relative Parameter Estimation Bias by Generating Condition for n = 20 Conditions

Condition	$\sigma_{pre}:\sigma_{post}$ Ratio Value									
	1:1		0.8:1.2		0.5:1.5		1.2:0.8		1.5:0.5	
	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$
ρ Value										
0	-0.001	-0.001	-0.001	-0.001	0.001	0.002	0.002	0.002	<0.001	<0.001
0.2	0.048	***0.053***	0.038	***0.056***	0.049	***0.120***	0.049	***0.067***	0.038	***0.107***
0.5	0.034	0.046	0.043	***0.097***	0.036	***0.253***	0.043	***0.097***	0.042	***0.258***
0.8	0.038	***0.060***	0.042	***0.219***	0.043	***0.724***	0.039	***0.216***	0.041	***0.722***
δ_{RM} Value										
0[a]	-0.001	-0.001	-0.001	-0.001	0.001	0.002	0.002	0.002	<0.001	<0.001
0.2	0.037	0.047	0.031	***0.094***	0.040	***0.285***	0.040	***0.103***	0.037	***0.283***
0.5	0.043	***0.053***	0.045	***0.110***	0.043	***0.290***	0.048	***0.112***	0.044	***0.288***
0.8	0.045	***0.055***	0.040	***0.103***	0.041	***0.286***	0.042	***0.105***	0.042	***0.288***
Overall[b]	0.041	***0.052***	0.039	***0.102***	0.041	***0.287***	0.043	***0.106***	0.041	***0.287***

Notes: Substantial relative parameter bias values are highlighted in the table; [a]Mean simple bias is presented for δ_{RM} = 0 conditions; [b]Overall = mean relative parameter bias across all δ_{RM} conditions excluding δ_{RM} = 0 conditions.

Table 4: Summary of Relative Standard Error Estimation Bias by Generating Condition for n = 20 Conditions

Condition	$\sigma_{pre}:\sigma_{post}$ Ratio Value									
	1:1		0.8:1.2		0.5:1.5		1.2:0.8		1.5:0.5	
	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$
ρ Value										
0	0.020	0.016	0.017	0.007	0.027	-0.006	0.030	0.019	0.010	-0.024
0.2	0.020	0.018	0.017	0.023	0.027	0.010	0.030	0.016	0.010	0.018
0.5	0.017	0.003	0.017	-0.034	0.018	***-0.151***	0.019	-0.031	0.016	***-0.150***
0.8	0.023	0.001	0.011	***-0.118***	0.013	***-0.330***	0.018	***-0.115***	0.018	***-0.329***
δ_{RM} Value										
0	0.018	0.008	0.020	-0.034	0.017	***-0.144***	0.021	-0.034	0.010	***-0.152***
0.2	0.020	0.010	0.010	-0.044	0.016	***-0.144***	0.014	-0.039	0.012	***-0.145***
0.5	0.015	0.003	0.013	-0.041	0.011	***-0.142***	0.021	-0.033	0.019	***-0.135***
0.8	0.025	0.011	0.025	-0.025	0.024	***-0.120***	0.028	-0.026	0.021	***-0.127***
Overall[a]	0.019	0.008	0.017	-0.036	0.017	***-0.138***	0.021	-0.033	0.016	***-0.140***

Notes: Substantial relative *SE* bias values are highlighted in the table; [a]Overall = average relative *SE* estimation bias across δ_{RM} and ρ conditions.

ratio value conditions, substantial negative bias was again found for the stronger ρ conditions (i.e., when ρ = 0.5 and 0.8). The negative relative *SE* bias was slightly worse for smaller δ_{RM} values (see Table 4). Last, substantial negative *SE* bias was also identified for the $\sigma_{pre}:\sigma_{post}$ = 0.8:1.2 and $\sigma_{pre}:\sigma_{post}$ = 1.2:0.8 conditions in the ρ = 0.8 conditions. Again, slightly worse substantial negative bias was noted for lower true δ_{RM} values.

Sample Size = 60: Relative Parameter Bias

With the larger sample size (n = 60) conditions, the degree of bias decreased further (see Table 5). As with the n = 20 conditions, no substantial bias was detected when $d_{RM}^{\neq}$ was used to estimate δ_{RM}. Substantial positive relative parameter bias was only found in certain conditions when using $d_{RM}^{=}$ to estimate δ_{RM}. Specifically, substantial positive bias was found in the most extreme $\sigma_{pre}:\sigma_{post}$ ratio value conditions (i.e., when $\sigma_{pre}:\sigma_{post}$ = 1.5:0.5 and $\sigma_{pre}:\sigma_{post}$ = 0.5:1.5) and for the ρ = 0.8 conditions when $\sigma_{pre}:\sigma_{post}$ = 1.2:0.8 and $\sigma_{pre}:\sigma_{post}$ = 0.8:1.2.

The positive bias for ρ = 0.5 paired with the $\sigma_{pre}:\sigma_{post}$ = 1.2:0.8 and $\sigma_{pre}:\sigma_{post}$ = 0.8:1.2 conditions only just exceeded Hoogland and Boomsma's substantial relative parameter bias criterion. The magnitude of the bias increased for larger ρ values and was unaffected by δ_{RM} values.

Sample Size = 60: Relative *SE* Bias

For the n = 60 conditions, no substantial relative *SE* bias was found with $d_{RM}^{\neq}$ (see Table 6). The same pattern and degree of substantial negative relative *SE* bias as was found for the n = 10 and n = 20 conditions was noted when using $d_{RM}^{=}$ to estimate δ_{RM}. Consistent bias was found for the most extreme $\sigma_{pre}:\sigma_{post}$ values when ρ = 0.5 and 0.8 and in the $\sigma_{pre}:\sigma_{post}$ = 0.8:1.2 and $\sigma_{pre}:\sigma_{post}$ = 1.2:0.8 conditions when ρ = 0.8. The bias was worse within $\sigma_{pre}:\sigma_{post}$ values for higher ρ conditions. There seemed to be a very slight effect of δ_{RM} value on the bias with lower δ_{RM} values associated with slightly larger degrees of negative bias (see Table 6).

Table 5: Summary of Relative Parameter Estimation Bias by Generating Condition for n = 60 Conditions

Condition	$\sigma_{pre}:\sigma_{post}$ Ratio Value									
	1:1		0.8:1.2		0.5:1.5		1.2:0.8		1.5:0.5	
	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$
ρ Value										
0	<0.001	<0.001	-0.001	-0.001	<0.001	<0.001	-0.001	-0.001	-0.001	-0.001
0.2	0.015	0.017	0.018	0.030	0.007	***0.061***	0.010	0.022	0.010	***0.065***
0.5	0.012	0.016	0.010	***0.053***	0.011	***0.204***	0.013	***0.055***	0.007	***0.200***
0.8	0.015	0.021	0.012	***0.165***	0.010	***0.643***	0.009	***0.162***	0.014	***0.647***
δ_{RM} Value										
0[a]	<0.001	<0.001	-0.001	-0.001	<0.001	<0.001	-0.001	-0.001	-0.001	-0.001
0.2	0.014	0.017	0.010	***0.062***	0.007	***0.229***	0.009	***0.061***	0.013	***0.235***
0.5	0.016	0.019	0.014	***0.066***	0.009	***0.229***	0.012	***0.065***	0.012	***0.232***
0.8	0.013	0.016	0.014	***0.067***	0.013	***0.235***	0.010	***0.062***	0.012	***0.233***
Overall[b]	0.014	0.017	0.013	***0.065***	0.010	***0.231***	0.011	***0.063***	0.012	***0.233***

Notes: Substantial relative parameter bias values are highlighted in the table; [a]Mean simple bias is presented for δ_{RM} = 0 conditions; [b]Overall = mean relative parameter bias across all δ_{RM} conditions except for δ_{RM} = 0 conditions.

Table 6: Summary of Relative Standard Error Estimation Bias by Generating Condition for n = 60 Conditions

Condition	$\sigma_{pre}:\sigma_{post}$ Ratio Value									
	1:1		0.8:1.2		0.5:1.5		1.2:0.8		1.5:0.5	
	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$	$d_{RM}^{\neq}$	$d_{RM}^{=}$
ρ Value										
0	0.001	0.001	0.009	0.004	0.004	-0.016	0.015	0.009	-0.003	-0.023
0.2	0.001	0.009	0.009	0.003	0.004	0.004	0.015	0.005	-0.003	0.010
0.5	0.001	-0.003	0.010	-0.030	0.007	***-0.148***	0.001	-0.039	0.009	***-0.146***
0.8	0.014	0.007	0.006	***-0.114***	0.005	***-0.338***	0.013	***-0.109***	0.005	***-0.336***
δ_{RM} Value										
0	0.004	0.001	0.011	-0.036	0.004	***-0.148***	0.008	-0.038	0.008	***-0.142***
0.2	0.006	0.003	0.005	-0.041	0.005	***-0.144***	0.005	-0.042	0.003	***-0.147***
0.5	0.006	0.003	0.010	-0.035	0.012	***-0.132***	0.009	-0.035	0.004	***-0.139***
0.8	0.009	0.005	0.002	-0.040	<0.001	***-0.134***	0.010	-0.033	0.006	***-0.129***
Overall[a]	0.006	0.003	0.007	-0.038	0.005	***-0.140***	0.008	-0.037	0.005	***-0.139***

Notes: Substantial relative *SE* bias values are highlighted in the table; [a]Overall = average relative *SE* estimation bias across δ_{RM} and ρ conditions.

Conclusion

The purpose of this study was to compare estimation of the repeated measures design standardized mean difference effect size, δ_{RM}, using the conventional $d^{=}_{RM}$ estimator with the newly derived $d^{\neq}_{RM}$ modification under a variety of conditions including unequal pre- and post-test score variances. The $d^{\neq}_{RM}$ estimator was designed to correct the standard deviation of the difference scores used in the calculation of δ_{RM} (see Equation 1). The correction recognizes potential differences in the population variances of the pre- and post-test scores. Most statistical tests of differences are based on the assumption that pre- and post-test score variances are equal. However, it is reasonable to assume that this assumption is commonly violated. This study assessed the robustness of the $d^{=}_{RM}$ and $d^{\neq}_{RM}$ estimates of δ_{RM} in scenarios with unequal variances.

Overall, the results convincingly supported use of the suggested modification, $d^{\neq}_{RM}$, as an improved estimator of δ_{RM}. Neither substantial parameter nor *SE* bias was noted for this estimate for sample sizes of 20 or 60 across the spectrum of δ_{RM} and ρ values investigated. In comparison, use of the conventional $d^{=}_{RM}$ estimator, however, cannot be recommended. Substantial positive parameter estimation bias was noted when using the $d^{=}_{RM}$ estimator even in the equal variance conditions (i.e., when $\sigma_{pre}=\sigma_{post}$) for n = 10 and n = 20. Substantial bias was also found across the unequal variance conditions. Negative standard error bias was noted when using the $d^{=}_{RM}$ estimator regardless of sample size. Given the consistency of the degree of *SE* bias across sample sizes of 10, 20, and 60 for the $d^{=}_{RM}$ estimator, it is anticipated that this pattern would be maintained for samples larger than 60.

Substantial parameter bias was identified for the $d^{\neq}_{RM}$ estimator in all of the smallest sample size (n = 10) conditions. (Note that no substantial standard error bias was noted across conditions for the $d^{\neq}_{RM}$ estimator.) The degree of parameter estimation bias in the $d^{\neq}_{RM}$ estimator remained around ten percent across δ_{RM} and ρ values. In other words, the bias was unaffected by the degree of correlation between pre- and post-test scores and by the magnitude of the effect size.

Across conditions, the degree of positive relative parameter bias noted for the $d^{=}_{RM}$ estimator was consistently greater than that noted for the $d^{\neq}_{RM}$ estimator. In addition, the bias detected for the $d^{=}_{RM}$ estimator was affected by the magnitude of ρ. The larger the correlation between pre- and post-test scores, the worse the bias was in the $d^{=}_{RM}$ estimate. The overall degree of positive bias found in the $d^{=}_{RM}$ estimator was greater for smaller sample sizes. But even with samples as large as n = 60, substantial bias was still noted in certain conditions.

The source of the bias noted for the $d^{\neq}_{RM}$ estimator for samples of n = 10 (and not when n was 20 or 60), likely originates in the negative relationship between sample size and degree of bias in the estimation of ρ. Specifically, the conventional estimator, r, (the one used herein) is a biased under-estimate of ρ. Olkin and Pratt (1958) derived an unbiased estimate of ρ, $\hat{\rho}$, that is closely approximated by:

$$\hat{\rho} = r + \frac{r(1-r^2)}{2(n-4)}. \tag{15}$$

Clearly, the degree of bias exhibited when using r to estimate ρ is represented by $r(1\text{-}r^2)/[2(n\text{-}4)]$ which becomes more substantial with smaller n. Small-sample bias in the estimation of ρ will negatively impact estimation of both $\sigma^{\neq}_{D}$ (see Equation 8) and $\sigma^{=}_{D}$ (see Equation 6), ultimately increasing bias in the estimation of δ_{RM} (see Equation 1) for both estimators. Bias in r's estimation of ρ rapidly decreases for larger n which seems to explain the corresponding rapid decrement in the bias of $d^{\neq}_{RM}$'s estimation of δ_{RM}. However, while bias in r's estimation of ρ contributes to the bias noted in $d^{=}_{RM}$'s estimation of δ_{RM}, it cannot fully explain it given $d^{=}_{RM}$'s bias decreases less rapidly than that of $d^{\neq}_{RM}$ for larger n.

Given the consistency in the degree of bias noted for $d_{RM}^{\neq}$ across conditions when n = 10, applied researchers and meta-analysts using $d_{RM}^{\neq}$ as an estimate of δ_{RM} should recognize that, if it is necessary to calculate the repeated measures design standardized mean difference for a sample as small as 10, then it will be over-inflated by about ten percent. Thus, optimally $d_{RM}^{\neq}$ should only be used with sample sizes larger than 10.

Future research should extend this assessment of how well $d_{RM}^{\neq}$ works with smaller sample sizes and should investigate other potential factors that might influence its performance. In addition, future research should extend formulation of the standardized mean difference effect size for repeated measures designs with heterogeneous variances for use with independent groups, repeated measures designs (i.e., for designs with pre- and post-test measures for the treatment and control groups).

A current policy movement encouraging evidence-based practice is leading to an increased use of meta-analysis across the spectrum of medical, educational, and general social science research. Effect sizes summarizing results from studies that have been conducted using repeated measures research designs must also be synthesized to contribute to the evidence base for programs and interventions. While it is commonly assumed that interventions lead to changes in means, not in variances, this is not always the case. This study introduced and validated a correction to the estimate of δ_{RM} that can be used to handle potentially unequal pre- and post-test variances. The new estimator, $d_{RM}^{\neq}$, was found to work better than the conventional one ($d_{RM}^{=}$) across conditions including equal variance conditions. Given the consistently superior performance of $d_{RM}^{\neq}$ over that of the $d_{RM}^{=}$ estimate, applied researchers are encouraged to begin using the $d_{RM}^{=}$ estimator as a less biased estimate of δ_{RM}.

References

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257-278.

Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, *18*, 271-279.

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, *5*, 3-8.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*, 490-499.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*(*3*), 329-367.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics*, *29*, 201-211.