Abstract Title Page

Title: Consequences of not accounting for one-group clustering in meta-analysis

Authors and Affiliations:
Martyna Citkowicz, Northwestern University
Joshua R. Polanin, Vanderbilt University

Abstract Body

Background / Context:
Meta-analyses are syntheses of effect-size estimates obtained from a collection of studies to summarize a particular field or topic (Hedges, 1992; Lipsey & Wilson, 2001). These reviews are used to integrate knowledge that can inform both scientific inquiry and public policy (e.g., The Campbell Collaboration, 2010; The Cochrane Collaboration, 2010). Therefore, it is important to ensure that the estimates of the effect sizes are both robust and free of bias.

One such scenario under which bias can occur is from a failure to account for clustering. Clustering occurs when, for example, schools are assigned to conditions and students are grouped together within each school. Recent research has shown that a failure to account for between-group clustering has the potential to bias effect sizes (Hedges, 2009). As a result, corrections have been proposed for such clustered designs.

The typical design is such that clustering exists for each condition. Occasionally, however, study designs are such that clustering exists in one condition but not the other. For example, students may be assigned to a set of tutors or patients may receive counseling from a designated set of therapists. In these scenarios, the individuals in the treatment group are nested within subgroups of either tutors or therapists; however, control group individuals lack clustered effects as they do not receive any treatment.

These studies may be referred to as one-group cluster-randomized designs or partially clustered designs. The subgroups in the treatment group create statistical clusters whose additional between-group variance needs to be taken into account during analysis. Clustering is often ignored, however, and the data are assessed as if the individuals within the treatment subgroups are independent of one another (Pals, Murray, Alfano, Shadish, Hannan, & Baker, 2008).

In order to account for this clustering, researchers have developed multi-level random- and mixed-effects models that take one-group clustering into account in order to test for individual study significance (e.g., Bauer, Sterba, & Hallfors, 2008; Hoover, 2002; Lee & Thompson, 2005; Roberts & Roberts, 2005). Only recently were methods developed to calculate appropriately adjusted *effect sizes* from such designs. Hedges and Citkowicz (under review) derived adjusted statistics that allow for the calculation of effect sizes and their sampling variances when the summary data is from a one-group clustered, two-level design study.

Despite these advances, we do not yet have an understanding of the degree to which *not* accounting for clustering will bias the overall meta-analytic mean effect and its variance. Assessing this bias will help to encourage future meta-analysts to account for this bias as well as caution policymakers and practitioners when considering meta-analytic results that fail to adjust for clustering.

Purpose / Objective / Research Question / Focus of Study:
The purpose of this study is to investigate the consequences of not accounting for one-group clustering in meta-analyses. We conducted a series of simulations in order to find out how distorted meta-analytic results can be when analysts fail to correct effect sizes for one-group

clustering. The results of these analyses, in addition, demonstrate the extent to which we may be confident in previously conducted meta-analyses that did not have the tools to adjust the effects and variance in such designs.

Significance / Novelty of study:

Before Hedges and Citkowicz's (under review) derivations, statistical adjustments to the effect sizes and variances of one-group clustered designs were not available. Thus, previously conducted meta-analyses of studies that included such cluster-randomized designs will have inflated mean effects. To our knowledge, no one has investigated to what extent these unadjusted, synthesized results may be inflated. The purpose of this project is to compare the adjusted and unadjusted (or naïve) meta-analytic results.

Statistical, Measurement, or Econometric Model:

Model. When clustering exists in the treatment group, one may denote the j^{th} observation in the i^{th} cluster by Y_{ij}^T (i = 1, …, m; j = 1, …, n), so that there are m clusters of size n in the treatment group. With no clustering in the control group, the i^{th} observation may be represented by Y_i^C (i = 1, …, N^C). The mean of the i^{th} cluster in the treatment group is $\bar{Y}_{i\bullet}^T$ (i = 1, …, m), and the overall (grand) means in the treatment and control groups may be represented by $\bar{Y}_{\bullet\bullet}^T$ and $\bar{Y}_{\bullet}^C$, respectively. The sample sizes are $N^T = nm$ and N^C in the treatment and control groups, respectively, while the total sample size is $N = N^T + N^C = nm + N^C$. Using those values, the total pooled within-treatment group variance S_T^2 may be estimated via

$$S_T^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij}^T - \bar{Y}_{\bullet\bullet}^T\right)^2 + \sum_{i=1}^{N^C}\left(Y_i^C - \bar{Y}_{\bullet}^C\right)^2}{N-2}.$$

In order to estimate the model and generate observed effects (discussed below), we assume that the observations are normally distributed within both the treatment group clusters and the individual observations of the control group. That is

$$Y_{ij}^T \sim N(\mu_i^T, \sigma_W^2),\ i = 1, \ldots, m;\ j = 1, \ldots, n,$$

$$Y_i^C \sim N(\mu^C, \sigma_W^2),\ i = 1, \ldots, N^C.$$

Also assume that the clusters in the treatment group have random between-cluster effects so that the cluster means themselves are normally distributed:

$$\mu_i^T \sim N(\mu_{\bullet}^T, \sigma_B^2),\ i = 1, \ldots, m.$$

Thus, the variance in the control group is simply σ_W^2, while the variance in the treatment group is $\sigma_T^2 = \sigma_W^2 + \sigma_B^2$. The parameter used to summarize the relationship between the variances is called the intraclass correlation coefficient ρ and is defined by

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2}.$$

One-group correction. Hedges and Citkowicz (under review) derived statistical adjustments to the standardized mean difference effects (SMDs are the effect sizes typically calculated in the social sciences) of one-group clustered designs. We define those below.

If clustering is not taken into account, a naïve estimate of SMD may be computed by

$$d_{Naive} = \frac{\bar{Y}^{T}_{\bullet\bullet} - \bar{Y}^{C}_{\bullet}}{S_T}.$$

In order to account for clustering in the treatment group, we multiply d_{Naive} by a correction factor so that the adjusted SMD d_T is

$$d_T = \left(\frac{\bar{Y}^{T}_{\bullet\bullet} - \bar{Y}^{C}_{\bullet}}{S_T}\right)\sqrt{1 - \frac{(N^C + n - 2)\rho}{N - 2}},$$

where ρ is the intraclass correlation defined above.

The naïve sampling variance of the effect is

$$v_{Naive} = \frac{N}{N^T N^C} + \frac{d^2_{Naive}}{2\left(N^T + N^C - 2\right)},$$

but may also be corrected to

$$v_T = \frac{N}{N^T N^C}\left[1 + \left(\frac{nN^C}{N} - 1\right)\rho\right] + \frac{\delta^2_T}{2h},$$

where h is the effective degrees of freedom of S_T^2 given by

$$h = \frac{\left[(N-2)(1-\rho) + \left(N^T - n\right)\rho\right]^2}{(N-2)(1-\rho)^2 + \left(N^T - n\right)n\rho^2 + 2\left(N^T - n\right)(1-\rho)\rho}.$$

Usefulness / Applicability of Method:

We used the model defined above to generate pilot data that results in clustered effects in the treatment group. We varied the following data characteristics to investigate how their differences impact the meta-analytic results:

Level 1:

- Treatment subgroup sample size: n = 20 or 100;
- Intraclass correlation: ρ = 0.05 or 0.20.

Level 2:

- Heterogeneity: τ^2 = 0 or 0.05;
- Number of effects included in the meta-analysis: k = 10 or 60;
- Proportion of effects that include one-group clustering: clustering = 0.10 or 0.90.

The characteristics we did not vary include the true mean effect (μ = 0.20) and the number of clusters (or subgroups) in the treatment group (m = 30). The level-1 parameter combinations were derived using observed statistics in education (Hedges & Hedberg, 2007). Heterogeneity statistics were assumed from a recent review of the Cochrane database (Turner, Davey, Clarke, Thompson, & Higgins, 2012), and the number of studies included in the meta-analysis are the 25% and 75% quartiles of studies included in recent social science meta-analyses (Polanin, 2013). The proportion of one-group clustering represents a hypothesized range of scenarios.

The combination of the five factors resulted in 32 cells for which we ran 5,000 replications each.[1] We then used both the naïve and adjusted formulas to estimate effects and variances. We combined those effects using both meta-analytic fixed- and random-effects models.[2] We then compared the results of the models that include only the naïve effect-size data (d_{naive}) to those of the true effect-size data (d_{true}, i.e., naïve when no clustering exists in the data and adjusted when clustering exists).

Findings / Results:
We first examined the results of the overall mean effect estimated using the random effects model. Table 1 shows that the coverage rates of 95% confidence intervals for the mean effect are similar (and close to the nominal level of 0.95) for the naïve and true effects data in almost all the cells. Moreover, bias (Table 2) and root mean squared errors (RMSE, see Table 3) are close to zero for both data sets, indicating little deviance of the estimated naïve and true mean effects from the true mean effect ($\mu = 0.20$). The reason for the similarity in mean effect results for the naïve and true data is that the between-study variance component (τ^2) is taking in any extra variability that clustering produced in the naïve data. Thus, $\hat{\tau}^2$is slightly overestimated in the naïve data meta-analyses.

Next we assessed the results for the fixed-effects model. Table 4 shows that the coverage rates of 95% confidence intervals for the mean effect are close to 0.95 for the true data, but decrease substantially for the naïve data as the proportion of clustering, ρ, and n increase. This occurs because the standard errors for the naïve mean effects are underestimated, producing confidence intervals that are too narrow. Bias (Table 5) and RMSE (Table 6) are slightly larger for the naïve mean effects, particularly when the proportion of clustering and ρ are larger. However, with values still relatively close to zero, the naïve mean effects do not deviate from the true mean effect very much (and the estimated true mean effects do not deviate at all).

Last, we investigated heterogeneity for the fixed-effects model where between-study variability (τ^2) was *not* generated into the meta-analytic data. The various heterogeneity diagnostics, found in Table 7, indicate that increases in the proportion of clustering, ρ, and n produce significant results for the naïve data, suggesting that there is a substantially amount of heterogeneity when, in fact, no between-study variability exists. Thus, because the naïve data does not account for one-group clustering, the heterogeneity tests inappropriately detect excess variability.

Conclusions:
The goal of our study was to investigate the consequences of not accounting for one-group clustering in meta-analysis. We found that when one conducts a meta-analysis using a random-effects model, the overall mean effect may be fairly accurate, but the variance component will be overestimated as it takes in the extra variability due to clustering. Running fixed-effects models may produce mean effects that also resemble true mean effects, but the confidence intervals will be much too narrow when most of the studies contain one-group clustering and have larger n's and ρ's, leading one to believe the estimates are more precise than they actually are (and potentially leading to inappropriate significance conclusions). Moreover, the between-study heterogeneity tests will be wrong most of the time.

---

[1] We plan on expanding this pilot study by including more gradations in values for each of the parameters we vary.
[2] Note that only 16 cells were run for the fixed-effects models as they do not include a variance component.

Appendices

Appendix A. References

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43*(2), 210-236. doi: 10.1080/00273170802034810

Hedges, L. V. (1992). Meta-analysis. *Journal of Education Statistics, 17* (4), 279–296.

Hedges, L.V. (2009). Effect sizes in nested designs. In Cooper, H., Hedges, L.V., & Valentine, J.C. (Eds.), *The handbook of research synthesis and meta-analysis*, (pp. 337-357), New York, NY: Sage Publication.

Hedges, L. V., & Citkowicz, M. (under revision). *Estimating effect size when there is clustering in one treatment group.*

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60-87.

Hoover, D. R. (2002). Clinical trials of behavioural interventions with heterogeneous teaching subgroup effects. *Statistics in Medicine, 21*, 1351-1364. doi: 10.1002/sim.1139

Lee, K. J., & Thompson, S. G. (2005). The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials, 2*, 163-173. doi: 10.1191/1740774505 cn082oa

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications, Inc.

Pals, S. L., Murray, D. M., Alfano, C. M., Shadish, W. R., Hannan, P. J., & Baker, W. L. (2008). Individually randomized group treatment trials: A critical appraisal of frequently used design and analytical approaches. *American Journal of Public Health, 98*(8), 1418-1424.

Polanin, J.R. (2013). *Addressing the issue of meta-analysis multiplicity in education and psychology*. (Doctoral Dissertation). Loyola University Chicago, Chicago, IL.

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials, 2*, 152-162. doi: 10.1191/1740774505cn076oa

The Campbell Collaboration. (2010). *About us.* Retrieved from http://www.campbellcollaboration .org /about_us/index.php

The Cochrane Collaboration. (2010). *About us.* Retrieved from http://www.cochrane.org/about-us

Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., & Higgins, J. P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International journal of epidemiology*, *41*(3), 818-827.

Appendix B. Table and Figures

Table 1. Coverage rates of 95% confidence intervals for the random-effects mean

Clustering	*k*	τ^2	ρ	*n*	d_{true}	d_{naive}
0.10	10	0	0.05	20	0.957	0.956
				100	0.955	0.945
			0.20	20	0.957	0.948
				100	0.960	0.947
		0.05	0.05	20	0.907	0.906
				100	0.904	0.903
			0.20	20	0.904	0.903
				100	0.905	0.905
	60	0	0.05	20	0.955	0.951
				100	0.957	0.949
			0.20	20	0.956	0.940
				100	0.959	0.945
		0.05	0.05	20	0.943	0.943
				100	0.941	0.941
			0.20	20	0.944	0.945
				100	0.943	0.943
0.90	10	0	0.05	20	0.958	0.931
				100	0.958	0.898
			0.20	20	0.960	0.898
				100	0.954	0.885
		0.05	0.05	20	0.905	0.905
				100	0.906	0.907
			0.20	20	0.909	0.905
				100	0.906	0.901
	60	0	0.05	20	0.954	0.938
				100	0.957	0.925
			0.20	20	0.954	0.861
				100	0.961	0.840
		0.05	0.05	20	0.944	0.945
				100	0.947	0.947
			0.20	20	0.945	0.932
				100	0.944	0.931

Note: $N^C = N^T = nm$, where $m = 30$ is constant.

Table 2. Bias results for the random-effects mean

Clustering	*k*	τ^2	ρ	*n*	d_{true}	d_{naive}
0.10	10	0	0.05	20	0.0002	0.0005
				100	0.0000	0.0003
			0.20	20	-0.0005	0.0006
				100	-0.0001	0.0011
		0.05	0.05	20	-0.0008	-0.0005
				100	0.0010	0.0012
			0.20	20	-0.0015	-0.0003
				100	-0.0002	0.0010
	60	0	0.05	20	-0.0002	0.0000
				100	0.0000	0.0002
			0.20	20	-0.0002	0.0009
				100	0.0000	0.0012
		0.05	0.05	20	-0.0003	0.0000
				100	-0.0001	0.0001
			0.20	20	0.0007	0.0018
				100	0.0006	0.0017
0.90	10	0	0.05	20	0.0001	0.0025
				100	-0.0001	0.0023
			0.20	20	0.0004	0.0106
				100	0.0001	0.0104
		0.05	0.05	20	0.0008	0.0032
				100	0.0005	0.0028
			0.20	20	-0.0002	0.0099
				100	0.0011	0.0113
	60	0	0.05	20	0.0002	0.0025
				100	0.0000	0.0024
			0.20	20	0.0002	0.0103
				100	0.0001	0.0103
		0.05	0.05	20	0.0001	0.0024
				100	-0.0001	0.0023
			0.20	20	-0.0003	0.0098
				100	0.0003	0.0104

Note: $N^C = N^T = nm$, where $m = 30$ is constant.

Table 3. RMSE results for the random-effects mean

Clustering	*k*	τ^2	ρ	*n*	d_{true}	d_{naive}
0.10	10	0	0.05	20	0.019	0.019
				100	0.009	0.009
			0.20	20	0.019	0.020
				100	0.009	0.012
		0.05	0.05	20	0.073	0.073
				100	0.071	0.071
			0.20	20	0.073	0.073
				100	0.072	0.072
	60	0	0.05	20	0.008	0.008
				100	0.003	0.004
			0.20	20	0.008	0.008
				100	0.003	0.005
		0.05	0.05	20	0.030	0.030
				100	0.029	0.029
			0.20	20	0.030	0.030
				100	0.029	0.029
0.90	10	0	0.05	20	0.022	0.022
				100	0.014	0.015
			0.20	20	0.028	0.033
				100	0.021	0.029
		0.05	0.05	20	0.073	0.074
				100	0.071	0.072
			0.20	20	0.073	0.077
				100	0.072	0.077
	60	0	0.05	20	0.009	0.009
				100	0.006	0.006
			0.20	20	0.012	0.016
				100	0.008	0.015
		0.05	0.05	20	0.030	0.030
				100	0.029	0.029
			0.20	20	0.030	0.033
				100	0.029	0.032

Note: $N^C = N^T = nm$, where $m = 30$ is constant.

Table 4. Coverage rates of 95% confidence intervals for the fixed-effects mean

Clustering	*k*	ρ	*n*	d_{true}	d_{naive}
0.10	10	0.05	20	0.953	0.948
			100	0.949	0.921
		0.20	20	0.952	0.926
			100	0.948	0.825
	60	0.05	20	0.952	0.946
			100	0.956	0.923
		0.20	20	0.953	0.926
			100	0.955	0.821
0.90	10	0.05	20	0.951	0.898
			100	0.952	0.726
		0.20	20	0.949	0.717
			100	0.951	0.429
	60	0.05	20	0.951	0.883
			100	0.952	0.697
		0.20	20	0.950	0.619
			100	0.951	0.326

Note: $N^C = N^T = nm$, where $m = 30$ is constant.

Table 5. Bias results for the fixed-effects mean

Clustering	*k*	*ρ*	*n*	d_{true}	d_{naive}
0.10	10	0.05	20	0.0004	0.0006
			100	-0.0002	0.0001
		0.20	20	0.0004	0.0014
			100	-0.0002	0.0009
	60	0.05	20	-0.0001	0.0002
			100	-0.0001	0.0002
		0.20	20	-0.0001	0.0011
			100	-0.0001	0.0010
0.90	10	0.05	20	0.0002	0.0025
			100	-0.0001	0.0022
		0.20	20	0.0002	0.0101
			100	-0.0001	0.0096
	60	0.05	20	-0.0001	0.0022
			100	-0.0001	0.0023
		0.20	20	-0.0002	0.0098
			100	-0.0001	0.0098

Note: $N^C = N^T = nm$, where $m = 30$ is constant.

Table 6. RMSE results for the fixed-effects mean

Clustering	*k*	ρ	*n*	d_{true}	d_{naive}
0.10	10	0.05	20	0.019	0.019
			100	0.009	0.009
		0.20	20	0.019	0.020
			100	0.009	0.012
	60	0.05	20	0.008	0.008
			100	0.003	0.004
		0.20	20	0.008	0.008
			100	0.003	0.005
0.90	10	0.05	20	0.022	0.022
			100	0.013	0.015
		0.20	20	0.028	0.033
			100	0.019	0.028
	60	0.05	20	0.009	0.009
			100	0.005	0.006
		0.20	20	0.012	0.016
			100	0.008	0.015

Note: $N^C = N^T = nm$, where $m = 30$ is constant.

Table 7. Heterogeneity diagnostics for the fixed-effects model

				Q-test p-value		I^2		*H*	
Clustering	*k*	ρ	*n*	d_{true}	d_{naive}	d_{true}	d_{naive}	d_{true}	d_{naive}
0.10	10	0.05	20	0.504	0.478	11.291	12.879	0.970	0.992
			100	0.493	0.387	11.890	19.985	0.978	1.087
		0.20	20	0.504	0.408	11.257	18.127	0.970	1.063
			100	0.494	0.253	11.848	36.057	0.978	1.361
	60	0.05	20	0.501	0.431	5.978	8.177	0.996	1.019
			100	0.502	0.220	5.870	18.468	0.995	1.112
		0.20	20	0.501	0.250	5.975	16.574	0.995	1.095
			100	0.501	0.026	5.916	47.402	0.996	1.433
0.90	10	0.05	20	0.496	0.295	11.561	26.070	0.978	1.172
			100	0.494	0.049	11.578	61.560	0.976	1.763
		0.20	20	0.495	0.061	11.386	57.573	0.977	1.664
			100	0.496	0.001	11.555	87.971	0.975	3.211
	60	0.05	20	0.493	0.081	6.059	28.606	0.998	1.196
			100	0.505	0.000	5.752	68.246	0.994	1.799
		0.20	20	0.488	0.000	6.237	64.533	0.999	1.702
			100	0.506	0.000	5.757	90.436	0.994	3.280

Note: $N^C = N^T = nm$, where $m = 30$ is constant.