Abstract Title Page

Title: Accounting for one-group clustering in effect-size estimation

Authors and Affiliations: Martyna Citkowicz & Larry V. Hedges, Northwestern University

Abstract Body

Background / Context:

In some instances, intentionally or not, study designs are such that there is clustering in one group but not in the other. For example, a treatment may be administered to subgroups of individuals by doctors, therapists, or teachers. The subgroups create statistical clusters whose additional between-group variance needs to be taken into account when estimating effect sizes and their sampling variance. The treatments might then be compared to a waitlist or no treatment control group. In that case, the design is such that there is clustering in only one group (the treatment group).

These one-group cluster-randomized designs (sometimes called partially clustered designs) can be found in areas such as medicine, psychology, and education. However, clustering is often ignored in the analyses and the data are assessed as if the individuals within the treatment subgroups are independent of one another. In a review of six public health and behavioral health journals spanning the years 2002 to 2006, Pals, Murray, Alfano, Shadish, Hannan, and Baker (2008) found that 32 out of 34 articles ignored the group-level entirely, analyzing the data at the individual level. Ignoring this between-cluster variation underestimates the error term, inflates the magnitude of the effect size, and can lead to inflated Type I errors (or chances of concluding that an effect is present when it is actually not; see, e.g., Wampold & Serlin, 2000). Thus, assuming independence of observations may lead to wrong conclusions. Naturally, making these incorrect assumptions impacts meta-analyses, which are syntheses of statistical estimates from a collection of studies used to summarize a particular topic (Lipsey & Wilson, 2001). The overall mean effect will be inflated if the magnitudes of the effects included in the meta-analysis are inflated.

Purpose / Objective / Research Question / Focus of Study:

This paper describes methods for computing effect size estimates and their variances when there is clustering in only one group and the analysis has not taken that clustering into account. We provide the effect size and sampling variance computations, along with the adjustments for the test of significance. In addition, we provide examples of how much the statistical results can be affected (or adjusted) once one-group clustering is taken into account depending on various data characteristics.

Significance / Novelty of study:

In 2007 and 2011, Hedges derived adjusted statistics that allow for the calculation of effect sizes and their sampling variances when the summary data comes from a clustered two-level design or a clustered three-level design. However, these statistics account for clustering when it exists in *all* treatment groups, not when it exists in only one. Moreover, 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), but none of the researchers provide solutions for computing appropriate effect sizes. The purpose of this paper is to do exactly that.

The corrections we provide to the statistical estimates will lead to more accurate effect size estimation, which should then lead to more appropriate conclusions in both individual studies as well as in meta-analyses.

Statistical, Measurement, or Econometric Model:

Estimating effect sizes. Effect sizes typically calculated in the social sciences are standardized mean differences (SMDs), defined as the difference between treatment and control group means over the standard deviation. Thus, here we focus only on deriving the statistics that correct SMDs and their sampling variances for one-group clustering. Let Y_{ij}^{T} be the j^{th} individual in the i^{th} treatment cluster and Y_i^C be the i^{th} individual in the control group (which has no clusters).

If clustering is not taken into account, the SMD may be computed by

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

where $\bar{Y}_{\bullet\bullet}^{T}$ and $\bar{Y}_{\bullet}^{C}$ are the overall (grand) means in the treatment and control groups, respectively, and S_T is the square root of the total pooled within-treatment group variance S_T^2

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

When clustering in the treatment group is taken into account, we multiply d_{Naive} by a correction factor so that the adjusted SMD d_T is

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

where N^C is the sample size in the control group, n is the subgroup sample size within the m^T treatment group clusters so that $N^T = nm^T$ is the sample size in the treatment group, N is the total sample size ($N = N^T + N^C$), and ρ is the intraclass correlation defined by

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

with σ_B^2 and σ_W^2 representing the between- and within-group variances, respectively. Note that when clustering does not exist, $\rho = 0$ and the correction factor is one, reducing d_T to d_{Naive}. But anytime $\rho > 0$, the correction factor will imply that $|d_T| \leq |\, d_{Naive}\,|$, producing a smaller effect size once one-group clustering is taken into account. So, for example, if $N^C = 10$, $m^T = 2$, and $n = 5$, so that $N^C = N^T = 10$, the ratio d_T/d_{Naive} ranges from 0.98 for $\rho = 0.05$ to 0.91 for $\rho = 0.25$.

The sampling variance of the effect must also be corrected from

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

to

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

$$= \frac{N(1-\rho)+N^C n\rho}{N^T N^C} + \frac{d_T^2}{2h},$$

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

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

Again, imagine that $N^C = 10$, $m^T = 2$, and $n = 5$. The ratio v_T/v_{Naive} ranges from 1.07 for $\rho = 0.05$ to 1.36 for $\rho = 0.25$. Notice that the adjustment is larger in magnitude for the variance than for the effect size, with the effect decreasing by 9% and variance increasing by 36% when $\rho = 0.25$.

Adjusting the significance test. In addition to inflating the effect size and underestimating the variance, studies that ignore one-group clustering in their analyses may compute tests of significance for the treatment effect that produce *t*- or *F*-statistics that are larger than the actual *t* or *F*. That may lead to inappropriately rejecting the null hypothesis that the treatment and control group means are equal, as the *p*-values will be smaller than those computed if clustering were not ignored.

When clustering is ignored, the usual Student's *t* is computed as

$$t_{Naive} = \left(\frac{\bar{Y}_{\bullet\bullet}^T - \bar{Y}_{\bullet}^C}{S_T}\right)\sqrt{\frac{N^T N^C}{N^T + N^C}}$$

with $N^T + N^C - 2$ degrees of freedom. Taking one-group clustering into account, the *t*-value must be adjusted by multiplying t_{Naive} by the square root of factor f

$$f = \frac{N\left[(N-2)(1-\rho)+\left(N^T - n\right)\rho\right]}{(N-2)\left[N(1-\rho)+nN^C\rho\right]},$$

which leads to the following expression for t_A

$$t_A = \left(\frac{\bar{Y}_{\bullet\bullet}^T - \bar{Y}_{\bullet}^C}{S_T}\right)\sqrt{\frac{N^T N^C}{N^T + N^C}}\sqrt{\frac{N\left[(N-2)(1-\rho)+\left(N^T - n\right)\rho\right]}{(N-2)\left[N(1-\rho)+nN^C\rho\right]}}.$$

t_A has a *t*-distribution with h degrees of freedom when the null hypothesis is true, where h is defined above.

Usefulness / Applicability of Method:

How much the results are affected when clustering is not taken into account differs depending on the data characteristics. Table 1 presents sample data results of how the effect size statistics change as ρ, m^T, n, and N^C increase. Columns 5 to 7 contain the results for the effect size,

variance, and degrees of freedom when clustering is ignored (i.e., unadjusted), Columns 8 to 10 denote the results when clustering is taken into account, and Columns 11 to 13 show the absolute value of the percent adjustment.

Overall, Table 1 shows that an increase in the intraclass correlation leads to a decrease in the statistics presented, while increases in the number of clusters and sample size lead to increases. So as ρ increases, d_T and h decrease in absolute magnitude and v_T increases substantially. As m^T and n increase, v_T always decreases and h always increases; however, the magnitude of the adjustment, which leads to decreases in d_T and and h and increases in v_T, is much larger with changes in n. Look, for example, at the cases where $\rho = 0.20$ and the sample sizes are equal at $N^C = N^T = 200$ ($m^T = 20$ with $n = 10$ and $m^T = 4$ with $n = 50$). The absolute magnitude of the adjustment for d_T, v_T, and h, is 5.4%, 72.2%, and 18.5% in the previous case and 6.4%, 434.6%, and 49.4% in the latter case, respectively. That suggests that changes in n drive the adjustment more than changes in m^T.

Next, we examined how the t-test estimates are adjusted as ρ, m^T, n, and N^C change. Table 2 presents those results, which show that when $\rho > 0$, nominally significant t-values are often actually insignificant after an adjustment for clustering in one group is made. Notice that, once again, n drives the adjustment more than m^T, as including more individuals in fewer clusters produces larger adjustments. For example, in the case where $p = 0.10$, when $m^T = 8$ and $n = 5$, the results are significant at the 0.05 level ($t_A = 2.027$, $h = 76.256$, $p = 0.0462$), but when $m^T = 5$ and $n = 8$, the results are no longer significant ($t_A = 1.902$, $h = 75.350$, $p = 0.0610$). In the case where $p = 0.20$, neither set of results are significant. Thus, the results suggest that changes in ρ and n appear to produce larger changes in t_A and h than do changes in m^T; however, changing all three values can change the results (and conclusions) drastically.

Conclusions:

Studies in fields such as education, psychology, and medicine sometimes require a design in which there is clustering in the treatment group, but not in the control group (a partially clustered design). Too often, this one-group clustering is ignored in statistical analyses, leading to inflated effect sizes, underestimated variances, and increased Type I errors. We derived adjusted statistics to account for this clustering in estimating effect sizes and their sampling variances, as well as produced appropriate t-test adjustments. In examining sample data, we found that not accounting for one-group clustering affects the variance and degrees of freedom more than the effect size. With those same degrees of freedom used in the t-test, it is easy to change the conclusion of the test from significant to non-significant once one adjusts for this clustering. This is especially true with larger values for the intraclass correlation and subgroup sample sizes, as the number of subgroups in the treatment group affects the results at a smaller magnitude. Last, note that even small changes in the effect size may affect meta-analyses, or syntheses of studies. If a portion of the effect sizes are adjusted toward zero in order to account for treatment group clustering, the combined mean effect will also decrease in size. Moreover, with substantially larger variances about the effects, the effect sizes will be less precise and therefore receive less weight in the combined estimate of effect size in a meta-analysis. Depending on the magnitude of both, the meta-analytic results may change considerably.

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. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341-370. doi: 10.3102/1076998606298043

Hedges, L. V. (2011). Effect sizes in three-level cluster-randomized experiments. *Journal of Educational and Behavioral Statistics*, *36*(3), 346-380. doi: 10.3102/1076998610376617

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.

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

Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5*(4), 425-433. doi: 10. I037111082-989X.5.4.425

Appendix B. Tables and Figures

Table 1. Changes in effect size statistics as functions of ρ, m^T, n, and N^C

				Unadjusted			Adjusted for Clustering			Percent Adjustment		
P	m^T	n	N^C	d_{Naive}	v_{Naive}	N-2	d_T	v_T	h	d	v	h
0.00	4	10	40	1.000	0.056	78	1.000	0.056	78.0	0.0	0.0	0.0
0.05	4	10	40	1.000	0.056	78	0.984	0.066	77.2	1.6	17.5	1.0
0.10	4	10	40	1.000	0.056	78	0.969	0.076	74.9	3.1	35.2	4.0
0.20	4	10	40	1.000	0.056	78	0.936	0.097	65.4	6.4	71.4	16.1
0.30	4	10	40	1.000	0.056	78	0.903	0.118	52.0	9.7	108.9	33.4
0.40	4	10	40	1.000	0.056	78	0.868	0.140	38.2	13.2	147.9	51.0
0.20	2	10	20	1.000	0.113	38	0.923	0.193	33.3	7.7	70.4	12.4
0.20	4	10	40	1.000	0.056	78	0.936	0.097	65.4	6.4	71.4	16.1
0.20	6	10	60	1.000	0.038	118	0.941	0.065	97.7	5.9	71.7	17.2
0.20	9	10	90	1.000	0.025	178	0.943	0.043	146.3	5.7	72.0	17.8
0.20	15	10	150	1.000	0.015	298	0.945	0.026	243.4	5.5	72.1	18.3
0.20	20	10	200	1.000	0.011	398	0.946	0.019	324.4	5.4	72.2	18.5
0.20	4	5	20	1.000	0.113	38	0.938	0.143	34.7	6.2	26.1	8.6
0.20	4	10	40	1.000	0.056	78	0.936	0.097	65.4	6.4	71.4	16.1
0.20	4	15	60	1.000	0.038	118	0.936	0.081	91.4	6.4	116.8	22.5
0.20	4	25	100	1.000	0.023	198	0.936	0.069	133.2	6.4	207.6	32.7
0.20	4	50	200	1.000	0.011	398	0.936	0.060	201.4	6.4	434.6	49.4
0.20	4	100	400	1.000	0.006	798	0.936	0.056	270.0	6.4	888.5	66.2

Note: $N^C = N^T = nm^T$; $N = N^C + N^T$.

Table 2. T-test adjustments as functions of ρ, m^T, n, and N^C

				Unadjusted			Adjusted for Clustering			Percent Adjustment		
P	m^T	n	N^C	t_{Naive}	N-2	p_{Naive}	t_A	H	p_A	t	h	p
0.00	4	10	40	2.236	78	0.0282	2.236	78.0	0.0282	0.0	0.0	0.0
0.05	4	10	40	2.236	78	0.0282	2.010	77.2	0.0479	10.1	1.0	69.9
0.10	4	10	40	2.236	78	0.0282	1.831	74.9	0.0711	18.1	4.0	152.0
0.20	4	10	40	2.236	78	0.0282	1.561	65.4	0.1233	30.2	16.1	337.3
0.30	4	10	40	2.236	78	0.0282	1.361	52.0	0.1794	39.1	33.4	535.9
0.40	4	10	40	2.236	78	0.0282	1.204	38.2	0.2360	46.2	51.0	736.6
0.10	2	20	40	2.236	78	0.0282	1.561	73.6	0.1228	30.2	5.6	335.3
0.10	4	10	40	2.236	78	0.0282	1.831	74.9	0.0711	18.1	4.0	152.0
0.10	5	8	40	2.236	78	0.0282	1.902	75.4	0.0610	14.9	3.4	116.2
0.10	8	5	40	2.236	78	0.0282	2.027	76.3	0.0462	9.4	2.2	63.6
0.10	10	4	40	2.236	78	0.0282	2.074	76.6	0.0414	7.3	1.8	46.9
0.10	20	2	40	2.236	78	0.0282	2.178	77.4	0.0325	2.6	0.8	15.0
0.20	2	20	40	2.236	78	0.0282	1.233	61.0	0.2223	44.9	21.8	688.1
0.20	4	10	40	2.236	78	0.0282	1.561	65.4	0.1233	30.2	16.1	337.3
0.20	5	8	40	2.236	78	0.0282	1.660	67.2	0.1016	25.8	13.8	260.1
0.20	8	5	40	2.236	78	0.0282	1.850	70.7	0.0685	17.3	9.4	142.8
0.20	10	4	40	2.236	78	0.0282	1.928	72.1	0.0578	13.8	7.6	104.9
0.20	20	2	40	2.236	78	0.0282	2.118	75.2	0.0375	5.3	3.5	32.9

Note: $N^C = N^T = nm^T$; $N = N^C + N^T$. N is kept constant to keep the unadjusted estimates constant for ease of interpretation.