Institute for Policy Research
Northwestern University
Working Paper Series

WP-06-13

Effect Sizes in Cluster-Randomized Designs

Larry V. Hedges
Faculty Fellow, Institute for Policy Research
Board of Trustees Professor of Statistics and Social Policy
Northwestern University

DRAFT
Please do not quote or distribute without permission.

2040 Sheridan Rd. ◆ Evanston, IL 60208-4100 ◆ Tel: 847-491-3395 Fax: 847-491-9916
www.northwestern.edu/ipr ◆ ipr@northwestern.edu

Abstract

Multisite research designs involving cluster randomization are becoming increasingly important in educational and behavioral research. Researchers would like to compute effect-size indices based on the standardized mean difference to compare the results of cluster randomized studies (and corresponding quasi-experiments) with other studies and to combine information across studies in meta-analyses. This working paper addresses the problem of defining effect sizes in multilevel designs—and computing estimates of those effect sizes and their standard errors—from information that is likely to be reported in journal articles. Three effect sizes are defined corresponding to different standardizations. Estimators of each effect size index are also presented along with their sampling distributions (including standard errors).

Effect Sizes in Cluster Randomized Designs

Multi-site studies are frequently used to evaluate the effects of educational treatments (for example, interventions, products or technologies). One common design assigns entire *sites* (often schools) to the same treatment group, with different sites assigned to different treatments. This design is often called a *cluster randomized* design because sites such as schools correspond to statistical clusters. Several analysis strategies for cluster randomized trials are possible. The simplest is to treat the clusters as units of analysis, that is, to compute mean scores on the outcome (and all other variables that may be involved in the analysis) and carry out the statistical analysis as if the site (cluster) means were the data. A more sophisticated alternative is to use a hierarchical linear modeling scheme with clusters as one level in the model (see, e.g., Raudenbush and Bryk, 2002). Many authors have commented on the problems of analyses of cluster-randomized trials (e.g., Raudenbush and Bryk, 2002; Donner and Klar, 2000; Klar and Donner, 2001; Murray, Varnell, and Blitstein, 2004).

Problems of representation of the results of cluster randomized trials (and the corresponding quasi-experiments) in the form of effect sizes and combining them across studies in meta-analyses have received less attention. The problem of meta-analysis of cluster randomized trials was considered by Rooney and Murray (1996), who called attention to the problem of effect size estimation in cluster randomized trials and suggested that conventional estimates were not appropriate and their standard errors were incorrect. Donner and Klar (2002) suggested that corrections for the effects of clustering should be introduced in meta-analyses of cluster randomized experiments. Laopaiboon (2003) reviewed the methods used in 25 published meta-analyses involving cluster randomized experiments, and found that only 3 used methods to account for clustering in their analysis. All of these three were meta-analyses of health care studies using binary outcomes. Of the six meta-analyses involving education, none used methods that addressed the impact of clustering.

This work was stimulated by problems faced by the US Department of Education's What Works Clearinghouse, whose mission is to evaluate, compare, and synthesize evidence of effectiveness of educational programs, products, practices, and policies. The What Works Clearinghouse reviewers found that the majority of the high quality studies they were examining involved assignment of treatment by clusters, which needed to be taken into account in computing an estimate of effect size and its uncertainty. This paper has two purposes. One is to examine the problem of defining effect sizes for cluster randomized trials. The second is to examine how to estimate these effect sizes and obtain standard errors for them from statistics that are typically given in reports of research (that is, without a reanalysis of the raw data)..

Model and Notation

Let Y_{ij}^T ($i = 1, \ldots, m^T$; $j = 1, \ldots, n_i^T$) and Y_{ij}^C ($i = 1, \ldots, m^C$; $j = 1, \ldots, n_i^C$) be the j^{th} observation in the i^{th} cluster in the treatment and control groups respectively, so that there are m^T clusters in the treatment group and m^C clusters in the control group, and a total of $M = m^T + m^C$ clusters with n observations each. Thus the sample size is

$$N^T = \sum_{i=1}^{m^T} n_i^T$$

in the treatment group,

$$N^C = \sum_{i=1}^{m^C} n_i^C$$

in the control group, and the total sample size is $N = N^T + N^C$.

Let $\overline{Y}_{i\bullet}^T$ (i = 1, …, m^T) and $\overline{Y}_{i\bullet}^C$ (i = 1, …, m^C) be the means of the i^{th} cluster in the treatment and control groups, respectively, and let $\overline{Y}_{\bullet\bullet}^T$ and $\overline{Y}_{\bullet\bullet}^C$ be the overall (grand) means in the treatment and control groups, respectively. Define the (pooled) within-cluster sample variance S_W^2 via

$$S_W^2 = \frac{\sum_{i=1}^{m^T}\sum_{j=1}^{n_i^T}(Y_{ij}^T - \overline{Y}_{i\bullet}^T)^2 + \sum_{i=1}^{m^C}\sum_{j=1}^{n_i^C}(Y_{ij}^C - \overline{Y}_{i\bullet}^C)^2}{N - M}$$

and the total pooled within-treatment group variance S_T^2 via

$$S_T^2 = \frac{\sum_{i=1}^{m^T}\sum_{j=1}^{n_i^T}(Y_{ij}^T - \overline{Y}_{\bullet\bullet}^T)^2 + \sum_{i=1}^{m^C}\sum_{j=1}^{n_i^C}(Y_{ij}^C - \overline{Y}_{\bullet\bullet}^C)^2}{N - 2}.$$

Let S_B be the pooled within treatment-groups standard deviation of the cluster means given by

$$S_B^2 = \frac{\sum_{i=1}^{m^T}(\overline{Y}_{i\bullet}^T - \overline{Y}_{*\bullet}^T)^2 + \sum_{i=1}^{m^C}(\overline{Y}_{i\bullet}^C - \overline{Y}_{*\bullet}^C)^2}{m^T + m^C - 2},$$

where $\overline{Y}_{*\bullet}^T$ is the (unweighted) mean of the m^T cluster means in the treatment group, and $\overline{Y}_{*\bullet}^C$ is the (unweighted) mean of the m^C cluster means in the control group. That is,

$$\overline{Y}_{*\bullet}^T = \frac{1}{m^T}\sum_{i=1}^{m^T}\overline{Y}_{i\bullet}^T$$

and

$$\overline{Y}_{*\bullet}^{C} = \frac{1}{m^{C}} \sum_{i=1}^{m^{C}} \overline{Y}_{i\bullet}^{C} .$$

Note that when cluster sample sizes are unequal, $\overline{Y}_{*\bullet}^{T}$ need not equal $\overline{Y}_{\bullet\bullet}^{T}$, the grand mean of the treatment group and $\overline{Y}_{*\bullet}^{C}$ need not equal $\overline{Y}_{\bullet\bullet}^{C}$, the grand mean of the control group. However, when cluster sample sizes are all equal $\overline{Y}_{*\bullet}^{T} = \overline{Y}_{\bullet\bullet}^{T}$ and $\overline{Y}_{*\bullet}^{C} = \overline{Y}_{\bullet\bullet}^{C}$.

Suppose that observations within the treatment and control group clusters are normally distributed about cluster means μ_i^T and μ_i^C with a common within-cluster variance σ_W^2. That is

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

and

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

Suppose further that the clusters are random effects (for example they are considered a sample from a population of clusters) so that the cluster means themselves have a normal sampling distribution with means $\mu_\bullet^T$ and $\mu_\bullet^C$ and common variance σ_B^2. That is

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

and

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

Note that in this formulation, σ_B^2 represents true variation of the population means of clusters over and above the variation in sample means that would be expected from variation in the sampling of observations into clusters.

These assumptions correspond to the usual assumptions that would be made in the analysis of a multi-site trial by a hierarchical linear models analysis, an analysis of variance (with treatment as a fixed effect and cluster as a nested random effect), or a *t*-test using the cluster means in treatment and control group as the unit of analysis.

In principle there are three different within-treatment group standard deviations, σ_B, σ_W, and σ_T, the latter defined by

$$\sigma_T^2 = \sigma_B^2 + \sigma_W^2 .$$

In most educational data when clusters are schools, σ_B^2 is considerably smaller than σ_W^2. Obviously, if the between cluster variance σ_B^2 is small, then σ_T^2 will be very similar to σ_W^2.

A parameter that summarizes the relationship between the three variances is called the intraclass correlation ρ, which is defined by

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2} . \tag{1}$$

The intraclass correlation ρ can be used to obtain one of these variances from any of the others, since $\sigma_W^2 = (1 - \rho)\sigma_B^2/\rho$, $\sigma_W^2 = (1 - \rho)\sigma_T^2$, and $\sigma_B^2 = \rho\sigma_T^2$.

Effect Sizes

The effect sizes typically used in educational and psychological research are standardized mean differences, defined as the ratio of a difference between treatment and control group means to a standard deviation. In single site designs or designs where there is no statistical clustering, the notion of standardized mean difference is often unambiguous: There is only one possibility. In multi-site designs such as cluster randomized trials, there are several possible standardized mean differences. In this section we clarify the possibilities.

The three possibilities for the standard deviation lead to different possible definitions for the population effect size in this clustered design. The choice of one of these effect sizes should be determined on the basis of the inference of interest to the researcher. If the effect size measures are to be used in meta-analysis, an important inference goal may be to estimate parameters that are comparable with those that can be estimated in other studies. In such cases, the standard deviation may be chosen to be the same kind of standard deviation used in the effect sizes of other studies to which this study will be compared. We focus on three effect sizes that seem likely to be the most useful (meaning the most widely used).

If $\sigma_W \neq 0$ (and hence $\rho \neq 1$), one effect size parameter has the form

$$\delta_W = \frac{\mu_\bullet^T - \mu_\bullet^C}{\sigma_W}. \tag{2}$$

This effect size might be of interest, for example, in a meta-analysis where the other studies to which the current study is compared are typically single site studies. In such studies δ_W may (implicitly) be the effect size estimated and hence δ_W an might be the effect size most comparable with that in other studies.

A second effect size parameter is of the form

$$\delta_T = \frac{\mu_\bullet^T - \mu_\bullet^C}{\sigma_T}. \tag{3}$$

This effect size might be of interest in a meta-analysis where the other studies are multi-site studies or studies that sample from a broader population but do not include clusters in the sampling design (this would typically imply that they used an individual, rather than a cluster, assignment strategy). In such cases, δ_T might be the most comparable with the effect sizes in other studies.

If $\sigma_B \neq 0$ (and hence $\rho \neq 0$), a third possible effect size parameter would be

$$\delta_B = \frac{\mu_{\bullet}^{T} - \mu_{\bullet}^{C}}{\sigma_B}. \tag{4}$$

This effect size is less likely to be of general interest, but it might be of interest in cases where the treatment effect is defined at the level of clusters and the individual observations are of interest because the average defines an aggregate property. The effect size δ_B may also be of interest in a meta-analysis where the studies being compared are typically multi-site studies that have been analyzed by using cluster means as the unit of analysis.

Note, however, that although δ_W and δ_T will often be similar in magnitude, δ_B will typically be much larger (in the same population) than either δ_W or δ_T, because σ_B is typically considerably smaller than σ_W and σ_T. Note also that if all of the effect sizes are defined (that is, if $0 < \rho < 1$), and ρ is known, any one of these effect sizes may be obtained from any of the others. In particular, both δ_W and δ_T can be obtained from δ_B and ρ since

$$\delta_W = \delta_B \sqrt{\frac{\rho}{1-\rho}} = \frac{\delta_T}{\sqrt{1-\rho}} \tag{5}$$

and

$$\delta_T = \delta_B \sqrt{\rho} = \delta_W \sqrt{1-\rho}\,. \tag{6}$$

Estimates of Effect Sizes: Equal Cluster Sample Sizes

While it is easy to define different effect sizes in multi-level (e.g., clustered) designs, the nested variance structure makes estimation somewhat less intuitive than in single level designs. In this section we present estimates of the effect sizes and their approximate sampling distributions when the cluster sample sizes are equal to n, that is when $n_i^T = n$, $i = 1, \ldots, m^T$ and $n_i^C = n$, $i = 1, \ldots, m^C$. In this case $N^T = nm^T$, $N^C = nm^C$, and $N = N^T + N^C = n(m^T + m^C) = nM$. Derivations and the details of the small sample distribution of the effect size estimators are given in the Appendix.

We present results explicitly for the case of equal cluster sample sizes for two reasons. The first reason is that most designs attempt to achieve equal cluster sample sizes but specific (realized) cluster sample sizes are rarely reported, so that the equal sample size formulas will be of most practical use. The second reason is that the results become considerably more complicated when cluster sample sizes are unequal—sufficiently complicated that it is difficult to obtain much insight from examining the formulas in the unequal cluster sample size case.

Estimation of δ_W

We start with estimation of δ_W, which is the most straightforward. If $\rho \neq 1$, the estimate

$$d_W = \frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_W} \tag{7}$$

is a consistent estimator of δ_W. The estimator d_W is approximately normally distributed about δ_W with variance

$$V\{d_W\} = \left(\frac{N^T + N^C}{N^T N^C}\right)\left(\frac{1+(n-1)\rho}{1-\rho}\right) + \frac{\delta_W^2}{2(N-M)}. \tag{8}$$

An estimate v_W of the variance of d_W can be computed by substituting the consistent estimate d_W for δ_W in equation (8) above. Note that the presence of the factor $(1 - \rho)$ in the denominator of the first term is possible since δ_W is defined only if $\rho \neq 1$.

Note that if $\rho = 0$ and there is no clustering, equation (8) reduces to the variance of a mean difference divided by a standard deviation with $(N - M)$ degrees of freedom (see, Hedges, 1981). The leading term of the variance in equation (8) arises from uncertainty in the mean difference. Note that it is $[1 + (n - 1)\rho]/(1 - \rho)]$ as large as would be expected if there were no clustering in the sample (that is if $\rho = 0$). Thus $[1 + (n - 1)\rho]/(1 - \rho)]$ is a kind of variance inflation factor for the variance of the effect size estimate d_W.

Estimation of δ_B

Estimation of the other δ_B and δ_T is less intuitive than that of δ_W. For example, one might expect that

$$\frac{\bar{Y}^T_{\bullet\bullet} - \bar{Y}^C_{\bullet\bullet}}{S_B},$$

would be that natural estimator of δ_B, but this is not the case. The reason is that S_B is not a pure estimate of σ_B, since it is inflated by the within-cluster variability. In particular, the expected value of S_B^2 is

$$\sigma_B^2 + \frac{\sigma_W^2}{n}.$$

If an estimate S_W^2 of the (average) within-cluster variance is reported, then it is possible to obtain an estimate $\hat{\sigma}_B^2$ of σ_B^2 by subtraction, namely

$$\hat{\sigma}_B^2 = S_B^2 - \frac{S_W^2}{n} \quad ,$$

whenever this quantity is nonnegative and zero otherwise. Whenever $\hat{\sigma}_B^2$ is nonzero, one estimate of δ_B is therefore

$$d_{B1} = \frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{\hat{\sigma}_B}. \tag{9}$$

Whenever δ_B is defined (that is, when $\rho \neq 0$) d_{B1} is normally distributed in large samples with variance

$$V\{d_{B1}\} = \left(\frac{m^T + m^C}{m^T m^C}\right)\left(\frac{1+(n-1)\rho}{n\rho}\right) + \left[\frac{[1+(n-1)\rho]^2}{2(M-2)n^2\rho^2} + \frac{(1-\rho)^2}{2n^2(N-M)\rho^2}\right]\delta_B^2. \tag{10}$$

An estimate v_{B1} of the variance of d_{B1} can be computed by substituting the consistent estimate d_{B1} for δ_B in equation (10) above. The estimate d_{B1} has the virtue that it can be computed without an external estimate of ρ. Note that the presence of ρ in the denominators of the variance terms is possible since δ_B is only defined if $\rho \neq 0$.

Alternatively, (again assuming that $\rho \neq 0$ so that δ_B is defined), the intraclass correlation may be used to obtain an estimate of δ_B using S_B. A direct argument shows that

$$d_{B2} = \frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_B}\sqrt{\frac{1+(n-1)\rho}{n\rho}} \tag{11}$$

is a consistent estimate of δ_B. This estimate is normally distributed in large samples with variance

$$V\{d_{B2}\} = \left(\frac{m^T + m^C}{m^T m^C}\right)\left(\frac{1+(n-1)\rho}{n\rho}\right) + \frac{[1+(n-1)\rho]\delta_B^2}{2n\rho(M-2)}. \tag{12}$$

An estimate v_{B2} of the variance of d_{B2} can be computed by substituting the consistent estimate d_{B2} for δ_B in equation (12) above. Note that the presence of ρ in the denominators of the variance terms is possible since δ_B is only defined if $\rho \neq 0$.

The variance in equation (12) is $[1 + (n - 1)\rho]/n\rho$ as large as the variance of the standardized mean difference computed from an analysis using cluster means as the unit of analysis, that is, applying the usual formula for the variance of the standardized difference between the means of the cluster means in the treatment and control group. Thus $[1 + (n - 1)\rho]/n\rho$ is a kind of variance inflation factor for the variance of effect size estimates like d_B compared to this alternative effect size estimate.

Estimation of δ_T

An estimate of δ_T can also be obtained in either of two ways. If the pooled within-treatment groups variance of the cluster means S_B^2 and the pooled within-cluster variance S_W^2 are both known, then an estimate of (σ_T) can be constructed as

$$\hat{\sigma}_T = \sqrt{S_B^2 + \left(\frac{n-1}{n}\right) S_W^2} \; .$$

The estimator

$$d_{T1} = \frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{\hat{\sigma}_T} \tag{13}$$

is a consistent estimator of δ_T. This estimate is normally distributed about δ_T in large samples with variance

$$V\{d_{T1}\} = \left(\frac{N^T + N^C}{N^T N^C}\right)\left(1 + (n-1)\rho\right) + \left[\frac{[1+(n-1)\rho]^2}{2n^2(M-2)} + \frac{(n-1)^2(1-\rho)^2}{2n^2(N-M)}\right]\delta_T^2 . \tag{14}$$

An estimate v_{T1} of the variance of d_{T1} can be computed by substituting the consistent estimate d_{T1} for δ_T in equation (14) above. The estimate d_{T1} has the virtue that it can be computed without an external estimate of ρ.

Alternatively, the intraclass correlation and S_T may be used to obtain an estimate of δ_T. A direct argument shows that a consistent estimator of δ_T is

$$d_{T2} = \left(\frac{\bar{Y}_{\bullet\bullet}^{T} - \bar{Y}_{\bullet\bullet}^{C}}{S_T} \right) \sqrt{1 - \frac{2(n-1)\rho}{N-2}} \; . \tag{15}$$

It is normally distributed in large samples with variance

$$\mathrm{V}\{d_{T2}\} = \left(\frac{N^T + N^C}{N^T N^C} \right) \left(1 + (n-1)\rho\right) + \delta_T^2 \left(\frac{(N-2)(1-\rho)^2 + n(N-2n)\rho^2 + 2(N-2n)\rho(1-\rho)}{2(N-2)\left[(N-2) - 2(n-1)\rho\right]} \right). \tag{16}$$

An estimate v_{T2} of the variance of d_{T2} can be computed by substituting the consistent estimate d_{T2} for δ_T in equation (16) above. Note that if $\rho = 0$ and there is no clustering, d_{T2} reduces to the conventional standardized mean difference and equation (16) reduces to the usual expression for the variance of the standardized mean difference (see Hedges, 1981).

The leading term of the variance in equations (14) and (16) arise from uncertainty in the mean difference. Note that this leading term is $[1 + (n - 1)\rho]$ as large as would be expected if there were no clustering in the sample (that is if $\rho = 0$). The expression $[1 + (n - 1)\rho]$ is the variance inflation factor mentioned by Donner (1981) and the design effect mentioned by Kish (1965) for the variance of means in clustered samples and also corresponds to a variance inflation factor for the effect size estimates like d_{T1} and d_{T2}.

Confidence Intervals for δ_W, δ_B, and δ_T

The results in this paper can also be used to compute confidence intervals for effect sizes. If δ is any one of the effect sizes mentioned, d is a corresponding estimate, and v_d is the estimated variance of d, then a 100(1 – α) percent confidence interval for δ based on d and v_d is given by

$$d - c_{\alpha/2}v_d \leq \delta \leq d + c_{\alpha/2}v_d, \tag{17}$$

where $c_{\alpha/2}$ is the 100(1 – α/2) percent point of the standard normal distribution (e.g., 1.96 for $\alpha/2 = 0.05/2 = 0.025$).

Estimates of Effect Size: Unequal Cluster Sample Sizes

When cluster sample sizes are unequal, expressions for the effect size estimators and their variances are considerably more complex. In this section we give estimates of the effect sizes and their sampling distributions when cluster sample sizes are not equal. These expressions may be of use when cluster sample sizes are unequal and are reported explicitly. They also give some insight about what single "compromise" sample size might give most accurate results (for example in computing the variances of estimates) when substituted into the equal sample size formulas.

Estimation of δ_W

When $\rho \neq 1$, the estimator d_W of δ_W is the same as in the case of equal cluster sample sizes, but the variance of the estimator is given by

$$\boldsymbol{V}\{\boldsymbol{d}_W\} = \left(\frac{\boldsymbol{N}^T + \boldsymbol{N}^C}{\boldsymbol{N}^T\boldsymbol{N}^C}\right)\left(\frac{1+(\tilde{\boldsymbol{n}}-1)\rho}{1-\rho}\right) + \frac{\delta_W^2}{2(\boldsymbol{N}-\boldsymbol{M})}, \tag{18}$$

where

$$\tilde{\boldsymbol{n}} = \frac{\boldsymbol{N}^C \sum_{i=1}^{m^T}\left(\boldsymbol{n}_i^T\right)^2}{\boldsymbol{N}^T\boldsymbol{N}} + \frac{\boldsymbol{N}^T \sum_{i=1}^{m^C}\left(\boldsymbol{n}_i^C\right)^2}{\boldsymbol{N}^C\boldsymbol{N}}.$$

When all of the n_i^T and n_i^C are equal to n, $\tilde{n} = n$ and (18) reduces to (8).

Estimation of δ_T

The form of the estimator d_{T2} is somewhat different when cluster sample sizes are unequal. In this case the estimator becomes

$$d_{T2} = \left(\frac{\bar{Y}_{\bullet\bullet}^T - \bar{Y}_{\bullet\bullet}^C}{S_T}\right)\sqrt{1-\rho\left(\frac{(N - n_U^T m^T - n_U^C m^C) + n_U^T + n_U^C - 2}{N-2}\right)}, \tag{19}$$

where

$$n_U^T = \frac{\left(N^T\right)^2 - \sum_{i=1}^{m^T}\left(n_i^T\right)^2}{N^T(m^T - 1)},$$

and

$$n_U^C = \frac{\left(N^C\right)^2 - \sum_{i=1}^{m^C}\left(n_i^C\right)^2}{N^C(m^C - 1)}.$$

When all of the n_i^T and n_i^C are equal to n, $n_U^T = n_U^C = n$ and (19) reduces to (15).

The variance of d_{T2} is somewhat more complex. It is given by

$$V\{d_{T2}\} = \left(\frac{N^T + N^C}{N^T N^C}\right)\left(1 + (\tilde{n} - 1)\rho\right) + \frac{\left[(N-2)(1-\rho)^2 + A\rho^2 + 2B\rho\left(1-\rho\right)\right]\delta^2}{2\left(N-2\right)[(N-2) - \rho(N-2-B)]}, \tag{20}$$

where the auxiliary constants A and B are defined by $A = A^T + A^C$,

$$A^T = \frac{\left(N^T\right)^2 \sum_{i=1}^{m^T}\left(n_i^T\right)^2 + \left(\sum_{i=1}^{m^T}\left(n_i^T\right)^2\right)^2 - 2N^T \sum_{i=1}^{m^T}\left(n_i^T\right)^3}{\left(N^T\right)^2},$$

$$A^C = \frac{\left(N^C\right)^2 \sum_{i=1}^{m^C}\left(n_i^C\right)^2 + \left(\sum_{i=1}^{m^C}\left(n_i^C\right)^2\right)^2 - 2N^C \sum_{i=1}^{m^C}\left(n_i^C\right)^3}{\left(N^C\right)^2},$$

and

$$\boldsymbol{B} = \boldsymbol{n}_U^{\boldsymbol{T}}(\boldsymbol{m}^{\boldsymbol{T}} - 1) + \boldsymbol{n}_U^{\boldsymbol{C}}(\boldsymbol{m}^{\boldsymbol{C}} - 1).$$

When n_i^T and n_i^C are equal to n, $A = n(N - 2n)$, $n_U^T = n_U^C = n$ and $B = (N - 2n)$ so that (20) reduces to (16). These expressions suggest that if cluster sample sizes are unequal, substituting the average of n_U^T and n_U^C into (15) and (16) would give results that are quite close to the exact values.

Estimation of δ_B

There is more than one way to generalize the estimator d_{B2} to the case of unequal cluster sample sizes. One possibility is to use the means of the cluster means in the treatment and control group in the numerator and standard deviation of the cluster means in the denominator. This corresponds to using the cluster means as the unit of analysis. Another possibility for the numerator is to use the grand means in the treatment and control groups. Similarly, there are multiple possibilities for the denominator, such as some function of the mean square between groups. When cluster sample sizes are identical, then all of these approaches are equivalent in the sense that the effect size estimates are identical. When the cluster sample sizes are not identical, the resulting estimators are not the same. Because the use of cluster means as the unit of analysis is a common approach, we believe that the means and standard deviations of cluster means are most likely to be available and hence we give the sampling distribution of the effect size estimate based on the standard deviation of the cluster means.

When $\rho \neq 0$, an estimator of δ_B which is a generalization of (11) is given by

$$d_B = \left(\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_B}\right)\sqrt{\frac{1 + (\overline{n}_B - 1)\rho}{\overline{n}_B\rho}}, \tag{21}$$

where

$$\bar{n}_B = \left(\frac{(m^T - 1)\bar{n}_I^T + (m^C - 1)\bar{n}_I^C}{M - 2} \right)^{-1},$$

$$\bar{n}_I^T = \frac{1}{m^T} \sum_{i=1}^{m^T} (1/n_i^T),$$

and

$$\bar{n}_I^C = \frac{1}{m^C} \sum_{i=1}^{m^C} (1/n_i^C).$$

The variance of d_B is approximately

$$V\{d_B\} = \left(\frac{m^T + m^C}{m^T m^C} \right) \left(\frac{1 + (\tilde{n}_B - 1)\rho}{\tilde{n}_B \rho} \right) + \frac{\bar{n}_B C \delta_B^2}{2(M-2)^2 \rho [1 + (\bar{n}_B - 1)\rho]}, \qquad (22)$$

where

$$\tilde{n}_B = \left(\frac{m^C \bar{n}_I^T + m^T \bar{n}_I^C}{M} \right)^{-1},$$

$$C = (M - 2)\rho^2 + 2[(m^T - 1)\bar{n}_I^T + (m^C - 1)\bar{n}_I^C]\rho(1 - \rho)$$
$$+ [(m^T - 2)\bar{n}_I^{T2} + (m^C - 2)\bar{n}_I^{C2} + (\bar{n}_I^T)^2 + (\bar{n}_I^C)^2](1 - \rho)^2,$$

$$\bar{n}_I^{T2} = \frac{1}{m^T} \sum_{i=1}^{m^T} (1/n_i^T)^2,$$

and

$$\bar{n}_I^{C2} = \frac{1}{m^C} \sum_{i=1}^{m^C} (1/n_i^C)^2.$$

Note that when the n_i^T and n_i^C are all equal to n, $\bar{n}_B = n$, $\tilde{n}_B = n$,

$\bar{n}_I^T = \bar{n}_I^C = 1/n$, $\bar{n}_I^{T2} = \bar{n}_I^{C2} = 1/n^2$, and

$$C = \frac{(M-2)[1+(n-1)\rho]^2}{n^2}$$

so that (21) reduces to (11) and (22) reduces to (12). These expressions suggest that, when cluster sample sizes are unequal, substituting $\overline{n}_B$ for n in (11) and (12) would give results that are quite close to the exact values.

Applications in Meta-analysis

The statistical results in this paper should be useful in deciding what effect sizes are desirable in a cluster randomized experiment. They should also be useful for finding ways to compute effect size estimates and their variances from data that may be reported. We illustrate applications in some examples in the sections that follow.

Intraclass correlations are needed for the methods described in this paper are often not reported. However, because plausible values of ρ are essential for power and sample size computations in planning cluster randomized experiments, there have been systematic efforts to obtain information about reasonable values of ρ in realistic situations. Some information about reasonable values of ρ comes from cluster randomized trials that have been conducted. For example, Murray and Blitstein (2003) reported a summary of intraclass correlations obtained from 17 articles reporting cluster randomized trials in psychology and public health and Murray, Varnell, and Blitstein (2004) give references to 14 very recent studies that provide data on intraclass correlations for health related outcomes. Other information on reasonable values of ρ comes from sample surveys that use clustered sampling designs. For example Guilliford, Ukoumunne, and Chinn (1999) and Verma and Lee (1996) presented values of intraclass correlations based on surveys of health outcomes. Hedberg, Santana, and Hedges (2004) presented a compendium of several hundred intraclass correlations for academic achievement computed from national probability samples at various grade levels. This later compendium provides national values for intraclass correlations as well as values for regions of the country and subsets of regions differing in level of urbanicity.

Computing Effect Sizes When Individuals are the Unit of Analysis

The results given in this paper can be used to produce effect size estimates and their variances from studies that incorrectly analyze cluster randomized trials as if individuals were randomized. The required means, standard deviations, and sample sizes can usually be extracted from what may be reported.

Suppose it is decided that the effect size δ_T is appropriate because most other studies both assign and sample individually from a clustered population. Suppose that the data are analyzed by ignoring clustering, then the test statistic is likely to be either

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

or

$$F = \left(\frac{N^T N^C}{N^T + N^C} \right) \left(\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_T} \right)^2 .$$

Either can be solved for

$$\left(\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_T} \right),$$

which can then be inserted into equation (15) along with ρ to obtain d_{T2}. This estimate (d_{T2}) of δ_T can then be inserted into equation (16) to obtain v_{T2}, an estimate of the variance of d_{T2}.

Alternatively, suppose it is decided that the effect size δ_W is appropriate because most other studies involve only a single site. We may begin by computing d_{T2} and v_{T2} as before. Because we want an estimate of δ_W, not δ_T, we use the fact given in equation (5) that

$$\delta_W = \frac{\delta_T}{\sqrt{1-\rho}}$$

and therefore

$$\frac{d_{T2}}{\sqrt{1-\rho}} \tag{23}$$

is an estimate of δ_W with a variance of

$$\frac{v_{T2}}{1-\rho}. \tag{24}$$

Example. An evaluation of the connected mathematics curriculum reported by Ridgway, et al. (2002) compared the achievement of $m^T = 18$ classrooms of 6th grade students who used connected mathematics with that of $m^C = 9$ classrooms in a comparison group that did not use connected mathematics. In this quasi-experimental design the clusters were classrooms. The cluster sizes were not identical but the average cluster size in the treatment groups was $N^T/m^T = 338/18 = 18.8$ and N^C/m^C $162/18 = 18$ in the control group. The exact sizes of all the clusters were not reported, but here we treat the cluster sizes as if they were equal and choose $n = 18$ as a slightly conservative sample size. The mean difference between treatment and control groups is $\bar{Y}^T_{\bullet\bullet}$ - $Y^C_{\bullet\bullet} = 1.9$, the pooled within-groups standard deviation $S_T = 12.37$. This evaluation involved sites in all regions of the country and it was intended to be nationally representative. Ridgeway et al. did not give an estimate of the intraclass correlation based on their sample. Hedberg, Santana, and Hedges (2004) provide an estimate of the grade 6 intraclass correlation in mathematics achievement for the nation as a whole (based on a national probability sample) of 0.264 with a standard error of 0.019. For this example we assume that the intraclass correlation is identical to that estimate, namely $\rho = 0.264$.

Suppose that the analysis ignored clustering and compared the mean of all of the students in the treatment with the mean of all of the students in the control group. This leads to a value of the standardized mean difference of

$$\frac{\overline{Y}_{\bullet\bullet}^{T} - \overline{Y}_{\bullet\bullet}^{C}}{S_T} = 0.1536,$$

which is not an estimate of any of the three effect sizes considered here. If an estimate of the effect size δ_T is desired, and we had imputed an intraclass correlation of $\rho = 0.264$, then we use equation (15) to obtain

$$d_{T2} = (0.1536)(0.9907) = 0.1522.$$

The effect size estimate is very close to the original standardized mean difference because the amount of clustering in this case is rather small. However even this small amount of clustering has a substantial effect on the variance of the effect size estimate. The variance of the standardized mean difference ignoring clustering is

$$\frac{324+162}{324*162} + \frac{0.1531^2}{2(324+162-2)} = 0.009259.$$

However, computing the variance of d_{T2} using equation (16) with $\rho = 0.264$, we obtain a variance estimate of 0.050865, which is 549 percent of the variance ignoring clustering. A 95 percent confidence interval for δ_T is given by

$$-0.2899 = 0.1522 - 1.96\sqrt{0.050865} \le \delta_T \le 0.1522 + 1.96\sqrt{0.050865} = 0.5942.$$

If clustering had been ignored, the confidence interval for the population effect size would have been -0.0350 to 0.3422.

If we wanted to estimate δ_W, then an estimate of δ_W given by expression (23) is

$$\frac{0.1522}{\sqrt{1-0.264}} = 0.1774\,,$$

with variance given by expression (24) as

0.050865/(1 - 0.264) = 0.06911,

and a 95 percent confidence interval for δ_W based on (17) would be

$$-0.3379 = 0.1774 - 1.96\sqrt{0.06911} \le \delta_W \le 0.1774 + 1.96\sqrt{0.06911} = 0.6926\,.$$

Computing Effect Sizes When Clusters are the Unit of Analysis

The results given in this paper can also be used to obtain different effects size estimates when the data have been analyzed with the cluster mean as the unit of analysis. In such cases, the researcher might report a *t*-test or an analysis of variance carried out on cluster means, but we wish to estimate δ_T. In this case the test statistic reported will either be

$$t = \sqrt{\frac{m^T m^C}{m^T + m^C}}\left(\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_B}\right)$$

or

$$F = \left(\frac{m^T m^C}{m^T + m^C}\right)\left(\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_B}\right)^2.$$

Either can be solved for

$$\left(\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_B}\right),$$

which can then be inserted into equation (11) along with ρ to obtain d_{B2}. This estimate of d_{B2} can then be inserted into equation (12) to obtain v_{B2}, an estimate of the variance of d_{B2}. Because we want an estimate of δ_T, not δ_B, we use the fact given in equation (6) that

$$\delta_T = \delta_B \sqrt{\rho}$$

and therefore

$$d_{B2}\sqrt{\rho} \tag{25}$$

is an estimate of δ_T with a variance of

$$\rho v_{B2}\,. \tag{26}$$

Alternatively, suppose it is decided that the effect size δ_W is the desired effect size. We may begin by computing d_{B2} and v_{B2} as before. Because we want an estimate of δ_W, not δ_B, we use the fact given in equation (5) that

$$\delta_W = \delta_B \sqrt{\frac{\rho}{1-\rho}}$$

and therefore

$$d_{B2}\sqrt{\frac{\rho}{1-\rho}} \tag{27}$$

is an estimate of δ_W with a variance of

$$\frac{\rho v_{B2}}{1-\rho}\,. \tag{28}$$

Example. An evaluation of UCSMP Geometry reported by Senk (2002) compared the results of $m^T = 8$ classrooms using UCSMP Geometry curriculum with $m^C = 8$ comparison classrooms that did not use the UCSMP curriculum. In this quasi-experimental design, clusters (classrooms) were the unit of analysis. The cluster sizes were not identical but the average cluster size in the treatment group was N^T/m^T = 139/8 = 17.4 and N^C/m^C 115/8 = 14.4 in the comparison group. The exact sizes of all the clusters were reported, but here we treat the cluster sizes as if they were equal and choose n = 15 as a slightly conservative compromise sample size. The mean difference between treatment and control groups is $\overline{Y}_{\bullet\bullet}^T$ - $Y_{\bullet\bullet}^C$ = -0.84, the pooled within-groups standard deviation S_B = 2.034. This evaluation involved sites in several regions of the country and it was intended to be nationally representative. Senk did not give an estimate of the intraclass correlation based on their sample. Hedberg, Santana, and Hedges (2004) provide an estimate of the intraclass correlation in mathematics achievement in grade 10 for the nation as a whole (based on a national probability sample) of 0.234 with a standard error of 0.010. For this example we assume that the intraclass correlation is identical to that estimate, namely ρ = 0.264.

These values lead to a value of the standardized mean difference of

$$\frac{\overline{Y}_{\bullet\bullet}^T - \overline{Y}_{\bullet\bullet}^C}{S_B} = -0.4130\,,$$

which is not an estimate of any of the three effect sizes considered here. If an estimate of the effect size δ_B is desired, and we had imputed an intraclass correlation of ρ = 0.234, then we use equation (11) to obtain

$$d_{B2} = (-0.4130)(1.2649) = -0.4558,$$

which is 26% larger than the unadjusted standardized mean difference. The variance of the standardized mean difference ignoring clustering is

$$\frac{8+8}{8\times 8} + \frac{(-0.4130)^2}{2(8+8-2)} = 0.2577\,.$$

However, computing the variance of d_{B2} using equation (12) with ρ = 0.234, we obtain a variance estimate of 0.3239, which is about 60% larger than the variance computed ignoring clustering. A 95 percent confidence interval for δ_B based on (17) is

$$-1.5712 = -0.4558 - 1.96\sqrt{0.3239} \le \delta_W \le -0.4558 + 1.96\sqrt{0.3239} = 0.6596 .$$

If we wanted to estimate δ_T, using expression (25) with ρ = 0.234 we obtain

$$-0.4558\sqrt{0.234} = -0.2205$$

as an estimate of δ_T with a variance given by expression (26) as

0.3239(0.234) = 0.0758.

If we wanted to estimate δ_W, then using expression (27) with ρ = 0.234 we obtain

$$-0.4558\sqrt{\frac{0.234}{1-0.234}} = -0.2519 ,$$

as an estimate of δ_W with a variance given by expression (28) as

(0.3239)[0.234/(1 - 0.234)] = 0.0989.

The report of this study (Senk, 2002) gives the sample sizes for each cluster, which range from 5 to 25 and are therefore are not nearly all equal. Because the individual cluster sample sizes are all given, it is possible to compute d_B and its variance using the formulas for unequal sample sizes. Using the data in Table 1, we compute $\overline{n}_B$ = 12.997, and using (21) we compute

d_B = (-0.4139)(1.1189) = -0.4621.

We also compute $\tilde{n}_B$ = 12.997, $\overline{n}_I^{T2}$ = 0.008342, $\overline{n}_I^{C2}$ = 0.007089, and A = 0.4185, so that the estimate of v_B using (22) is 0.2820. Comparing the values of d_B (-0.4558 versus -0.4621) and estimates of the variance (0.3239 versus 0.2820) assuming equal cluster sample sizes with those using the exact cluster sample sizes, we see that even with these large discrepancies among sample sizes, the values of d_B and the variance estimates assuming equal cluster sample sizes are within 15 percent of the actual values. If the value n = 13 had been used for the (common) cluster sample size (approximately the value of $\bar{n}_B$ or $\tilde{n}_B$) the results using the equal sample size formulas would have been quite close to the exact values.

Conclusion

This paper has provided definitions of three different effect sizes that can be estimated in studies using cluster randomization. Alternative methods of estimation are provided for each effect size, and the sampling variances are also given. The sampling distribution of each estimator is shown to be a constant times a noncentral *t*-distribution and simple normal approximations are given in each case. Because these approximations have been extensively studied in the context of simpler effect size estimates and power analysis, there is reason to believe that they are reasonably accurate unless sample sizes are quite small (which is unlikely in cluster randomized designs). Simulation studies (not reported here) evaluating the accuracy of these approximations confirm expectations.

The analytic work shows that clustering can have a substantial effect on the variance of effect sizes estimates in cluster randomized designs. The example provided illustrates that small amounts of clustering can have a large effect on the variance of effect sizes, even if the effect on the expected value of the estimates is modest. The results given in this paper can be used to estimate the effect sizes (and their variances) in cluster randomized trials that have been improperly analyzed by ignoring clustering, provided an intraclass correlation is known or can be imputed. The effect size estimates can then be used in meta-analyses along with any other effect size estimates of the same conceptual parameter, using the variances of the estimates to compute weights in the usual way.

The results given in this paper require that a value of the intraclass correlation parameter ρ be known or imputed for sensitivity analysis. In some cases external data about ρ may be available (e.g., from previous studies or compendia such as that of Hedberg, Santana, and Hedges, 2004). It is important to use external values of ρ with considerable caution, because the value of ρ has substantial influence on the results of analyses. In particular, it would be difficult to justify the use of the methods described in this paper using estimates of ρ obtained from small samples (small numbers of clusters) because those estimates are likely to be subject to considerable sampling error. Similarly, it would be difficult to justify the use of external estimates of ρ, even from large sample sizes if those estimates were not based on a similar sampling strategy, with similar populations, and similar outcome measures. However, making no correction for the effects of clustering at all corresponds to assuming that $\rho = 0$. The assumption that $\rho = 0$ is often very far from the case and thus it may introduce more serious biases in the computation of variances than using values of ρ that are slightly in error.

Appendix: Derivation of Sampling Distributions of Effect Size Estimates

The sampling distribution of the effect size estimates proposed in this paper all follow from the same theorem, given below.

Theorem: Suppose that $Y \sim \mathrm{N}(\mu, a\sigma^2/\tilde{N})$ and that S^2 is a quadratic form in normal variates that is independent of Y, so that the $\mathrm{E}\{S^2\} = b\sigma^2$, and $\mathrm{V}\{S^2\} = 2c\sigma^4$, where a, b, c, and $\tilde{N}$ are known constants. Then

$$T = \sqrt{\frac{\tilde{N}b}{a}}\left(\frac{Y}{S}\right)$$

has approximately the noncentral t-distribution with b^2/c degrees of freedom and noncentrality parameter

$$\theta = \sqrt{\frac{\tilde{N}b}{a}}\left(\frac{\mu}{\sigma}\right) = \sqrt{\frac{\tilde{N}b}{a}}\delta\,,$$

where $\delta = \mu/\sigma$. Consequently

$$D = \frac{Y\sqrt{b}}{S} = T\sqrt{\frac{a}{\tilde{N}}}$$

is a consistent estimate of the effect size δ with approximate variance

$$\frac{a}{\tilde{N}} + \frac{c\delta^2}{2b}\,. \tag{29}$$

An approximately unbiased estimate of δ is given by $D\mathrm{J}(b^2/c)$, where the function $\mathrm{J}(x)$ is given by

$$\mathrm{J}(x) = 1 - \frac{3}{4x-1}\,.$$

Proof: First obtain the approximate sampling distribution of S^2. Box (1954) gives the approximate sampling distribution of quadratic forms in normal variables (such as S^2, which is a linear combination of chi-squares) in terms of the first two cumulants of the quadratic form. Theorem 3.1 in Box (1954) implies that S^2 is distributed as approximately a constant g times chi-square with h degrees of freedom, where g and h are given by $g = V\{S^2\}/2E\{S^2\} = c\sigma^2/b$ and $h = 2(E\{S^2\})^2/V\{S^2\} = b^2/c$, where $E\{X\}$ and $V\{X\}$ are the expected value and the variance of X. Therefore we have that $S^2/gh = S^2/b\sigma^2$ is distributed as a chi-square with h degrees of freedom divided by h. This approximation is generally excellent and is the basis, for example, of the standard tests used in repeated measures analysis of variance (e.g., Geisser and Greenhouse, 1958).

By the definition of the noncentral t-distribution (see, e.g., Johnson and Kotz, 1970), it follows that

$$\frac{Y\sqrt{\tilde{N}/a\sigma^2}}{\sqrt{S^2/b\sigma^2}} = \sqrt{\frac{\tilde{N}b}{a}}\left(\frac{Y}{S}\right)$$

has (approximately) the noncentral t-distribution with b^2/c degrees of freedom and noncentrality parameter

$$\theta = \sqrt{\frac{\tilde{N}b}{a}}\left(\frac{\mu}{\sigma}\right).$$

Using properties of the noncentral t-distribution, via arguments that parallel those in Hedges (1981), it follows that D is a consistent estimator of δ, that $DJ(b^2/c)$ is an unbiased estimator of δ, and the variance of D is approximately given by (29). □

The theorem can be applied to obtain the sampling distribution of each of the effect size estimators given in this paper, using some elementary facts. In each case we apply the theorem with $Y = \bar{Y}^T_{\bullet\bullet} - \bar{Y}^C_{\bullet\bullet}$, $\mu = \mu^T_{\bullet} - \mu^C_{\bullet}$, and $\tilde{N} = N^T N^C/(N^T + N^C)$, but with different definitions of S and σ. Therefore in each case, we use the fact that the mean of $\bar{Y}_{\bullet\bullet}^{T} - \bar{Y}_{\bullet\bullet}^{C}$ is given by

$$\mathrm{E}\left\{\bar{Y}_{\bullet\bullet}^{T} - \bar{Y}_{\bullet\bullet}^{C}\right\} = \mu_{\bullet}^{T} - \mu_{\bullet}^{C}.$$

However the variance of $\bar{Y}_{\bullet\bullet}^{T} - \bar{Y}_{\bullet\bullet}^{C}$ and the mean and various choices of S require different derivations in the balanced (equal cluster sample size) and unbalanced (unequal cluster sample size) cases.

Equal Cluster Sample Sizes

In the case of equal cluster sample sizes, a direct argument gives the variance of the mean difference as

$$\mathrm{V}\left\{\bar{Y}_{\bullet\bullet}^{T} - \bar{Y}_{\bullet\bullet}^{C}\right\} = \left(\frac{N^{T} N^{C}}{N^{T} + N^{C}}\right)^{-1}\left(\sigma_{W}^{2} + n\sigma_{B}^{2}\right).$$

We also use the moments of S_B^2, S_W^2 , and S_T^2, which are derived from their relation to sums of squares (see, e.g., Snedecor, 1956). Specifically, because $n(M-2)S_B^2/(n\sigma_B^2 + \sigma_W^2)$ has a chi-squared distribution with $(M-2)$ degrees of freedom, the mean of S_B^2 is

$$\mathrm{E}\left\{S_{B}^{2}\right\} = \sigma_{B}^{2} + \frac{\sigma_{W}^{2}}{n} \tag{30}$$

and the variance of S_B^2 is

$$\mathrm{V}\left\{S_{B}^{2}\right\} = \frac{2(n\sigma_{B}^{2} + \sigma_{W}^{2})^{2}}{n^{2}(M-2)}. \tag{31}$$

Similarly, because $(N - M)S_W^2/\sigma_W^2$ has the chi-square distribution with $(N - M)$ degrees of freedom, the mean of S_W^2 is σ_W^2 and the variance of S_W^2 is

$$\mathrm{V}\left\{S_W^2\right\} = \frac{2\sigma_W^4}{N-M}. \tag{32}$$

Because

$$S_T^2 = \frac{n(M-2)S_B^2 + (N-M)S_W^2}{N-2},$$

the expected value of S_T^2 follows from the expected values of S_B^2 and S_W^2, namely

$$\mathrm{E}\left\{S_T^2\right\} = \sigma_W^2 + \left(\frac{N-2n}{N-2}\right)\sigma_B^2, \tag{33}$$

and the variance of S_T^2 follows from the variances of S_B^2 and S_W^2, namely

$$\mathrm{V}\left\{S_T^2\right\} = \frac{2(N-2)\sigma_W^4 + 2n(N-2n)\sigma_B^4 + 4(N-2n)\sigma_B^2\sigma_W^2}{(N-2)^2}. \tag{34}$$

The distribution of d_W. In this case we apply the theorem with $\sigma^2 = \sigma_W^2$ and $S^2 = S_W^2$. Here

$$a = \frac{\sigma_W^2 + n\sigma_B^2}{\sigma_W^2} = \frac{1+(n-1)\rho}{1-\rho}.$$

Because the expected value of S_W^2 is σ_W^2, it follows that $b = 1$. Because the variance of S_W^2 is $2\sigma_W^4/(N-M)$, it follows that $c = 1/(N-M)$. Substituting the expressions for a, b, and c into (29), noting that $\sigma_B^2/\sigma_W^2 = \rho/(1-\rho)$, and simplifying, gives the result in expression (8). Since S^2 involves only a single chi-square, it follows that the t-statistic corresponding to d_W has exactly the noncentral t-distribution with $(N-M)$ degrees of freedom.

The distribution of d_{B1}. In this case we apply the theorem with $\sigma^2 = \sigma_B^2$ and $S^2 = \hat{\sigma}_T^2 = S_B^2 - S_W^2/n$. Here,

$$a = \frac{n\sigma_B^2 + \sigma_W^2}{\sigma_B^2} = \frac{1+(n-1)\rho}{\rho}.$$

The expected value of S^2 is just σ_B^2, so $b = 1$. The variance of S^2 is

$$V\left\{S^2\right\} = V\left\{S_B^2\right\} + \left(\frac{1}{n}\right)^2 V\left\{S_W^2\right\} = \frac{2(n\sigma_B^2 + \sigma_W^2)^2}{n^2(M-2)} + \frac{2\sigma_W^4}{n^2(N-M)}$$

so that

$$c = \frac{(n\sigma_B^2 + \sigma_W^2)^2}{n^2(M-2)\sigma_B^4} + \frac{\sigma_W^4}{n^2(N-M)\sigma_B^4}.$$

Substituting the expressions for a, b, and c into (29), noting that $\sigma_W^2/\sigma_B^2 = (1-\rho)/\rho$, and simplifying, gives the result in expression (10).

The distribution of d_{B2}. In this case we apply the theorem with $\sigma^2 = \sigma_B^2$ and $S^2 = S_B^2$. Here, as in d_{B1},

$$a = \frac{\sigma_B^2 + \frac{\sigma_W^2}{n}}{\sigma_B^2} = \frac{1+(n-1)\rho}{\rho}.$$

Using the expected value of S_B^2 given in (30), gives

$$b = \frac{1+(n-1)\rho}{n\rho}.$$

Using the variance of S_B^2 given in (31) yields

$$c = \frac{(n\sigma_B^2 + \sigma_W^2)^2}{n^2(M-2)\sigma_B^2}.$$

Substituting the expressions for a, b, and c into (29), noting that $\sigma_W^2/\sigma_B^2 = (1-\rho)/\rho$, and simplifying, gives the result in expression (12). Since S^2 involves only a single chi-square, it follows that the t-statistic corresponding to d_{B2} has exactly the noncentral t-distribution with $(M-2)$ degrees of freedom.

The distribution of d_{T1}. In this case we apply the theorem with $\sigma^2 = \sigma_T^2 = \sigma_W^2 + \sigma_B^2$ and $S^2 = \hat{\sigma}_T^2$. Here

$$a = \frac{\sigma_W^2 + n\sigma_B^2}{\sigma_W^2 + \sigma_B^T} = 1 + (n-1)\rho .$$

The expected value of S^2 is just σ_T^2, so $b = 1$. The variance of S^2 is

$$V\left\{S^2\right\} = V\left\{S_B^2\right\} + \left(\frac{n-1}{n}\right)^2 V\left\{S_W^2\right\} = \frac{2(n\sigma_B^2 + \sigma_W^2)^2}{n^2(M-2)} + \left(\frac{n-1}{n}\right)^2 \left(\frac{2\sigma_W^4}{N-M}\right),$$

so that

$$c = \frac{2(n\sigma_B^2 + \sigma_W^2)^2}{n^2(M-2)(\sigma_B^2 + \sigma_W^2)^2} + \left(\frac{n-1}{n}\right)^2 \left(\frac{2\sigma_W^4}{(N-M)(\sigma_B^2 + \sigma_W^2)^2}\right).$$

Substituting the expressions for a, b, and c into (29), noting that $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_W^2)$ and $1 - \rho = \sigma_W^2/(\sigma_B^2 + \sigma_W^2)$, and simplifying, gives the result in expression (14).

The distribution of d_{T2}. In this case we apply the theorem with $\sigma^2 = \sigma_T^2 = \sigma_W^2 + \sigma_B^2$ and $S^2 = S_T^2$. Here, as in d_{T1},

$$a = \frac{\sigma_W^2 + n\sigma_B^2}{\sigma_W^2 + \sigma_B^T} = 1 + (n-1)\rho .$$

Using the expected value of S_T^2 given in (33), we compute

$$b = \frac{\sigma_W^2 + \left(\frac{N-2n}{N-2}\right)\sigma_B^2}{\sigma_W^2 + \sigma_B^2} = 1 - \frac{2(n-1)\rho}{N-2}.$$

Using the variance of S_T^2 given in (34), we compute

$$c = \frac{(N-2)\sigma_W^4 + n(N-2n)\sigma_B^4 + 2(N-2n)\sigma_B^2\sigma_W^2}{(N-2)^2(\sigma_W^2 + \sigma_B^2)^2}.$$

Substituting the expressions for *a*, *b*, and *c* into (29), dividing the numerator and denominator by σ_W^2, noting that noting that $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_W^2)$ and $1 - \rho = \sigma_W^2/(\sigma_B^2 + \sigma_W^2)$, and simplifying, gives the result in expression (16).

Unequal Cluster Sample Sizes

When cluster sample sizes are unequal, expressions for the effect size estimators and their variances are more complex. We first derive the variance of the mean differences. A direct argument leads to

$$\mathrm{V}\left\{\overline{Y}_{\bullet\bullet}^{T} - \overline{Y}_{\bullet\bullet}^{C}\right\} = \left(\frac{N^T N^C}{N^T + N^C}\right)^{-1}\left(\sigma_W^2 + \tilde{n}\sigma_B^2\right) \tag{35}$$

and

$$\mathrm{V}\left\{\overline{Y}_{*\bullet}^{T} - \overline{Y}_{*\bullet}^{C}\right\} = \left(\frac{m^T m^C}{m^T + m^C}\right)^{-1}\left(\sigma_B^2 + \overline{n}_B\sigma_B^2\right), \tag{36}$$

where $\tilde{n}$ and $\overline{n}_B$ are defined in the text. The expected value and variance of S_T^2 can be calculated from the analysis of variance across clusters within the treatment groups. When cluster sample sizes are unequal, the between and within cluster sums of squares are still independent, and the within cluster sum of squares has a chi-square distribution, but if $\rho \neq 1$ the between cluster sum of squares does not have a chi-square distribution. However because the between cluster sum of squares is quadratic form, the methods used in this paper apply and the distribution of effect size estimates can be obtained. To obtain the expected value of S_T^2, use the fact that

$$S_T^2 = \frac{SSB^T + SSW^T + SSB^C + SSW^C}{N-2},$$

where SSB^T and SSW^T and SSB^C and SSW^C are the sums of squares between and within clusters in the treatment and control groups, respectively. Using the expected values of the *SSB*'s and *SSW*'s given, for example, in equations 77 and 78 on page 70 of Searle, Casella, and McCulloch (1992), we obtain

$$E\left\{S_T^2\right\} = \sigma_W^2 + \frac{B\sigma_B^2}{N-2}, \tag{37}$$

where B is the auxiliary constant defined in (20). Because the between clusters variance component estimates in the treatment and control groups are

$$\left(\hat{\sigma}_B^T\right)^2 = \frac{MSB^T - MSW^T}{n_U^T}$$

and

$$\left(\hat{\sigma}_B^C\right)^2 = \frac{MSB^C - MSW^C}{n_U^C},$$

it follows that S_T^2 can be written as a function of between and within cluster variance components

$$S_T^2 = \frac{SSW^T + SSW^C + A^T\left(\hat{\sigma}_B^T\right)^2 + A^C\left(\hat{\sigma}_B^C\right)^2}{N-2}.$$

Therefore the variance of S_T^2 is given by

$$\begin{aligned}(N-2)^2 V\left\{S_T^2\right\} = {} & V\left\{SSW^T\right\} + V\left\{SSW^C\right\} + \left(A^T\right)^2 V\left\{\left(\hat{\sigma}_B^T\right)^2\right\} + \left(A^C\right)^2 V\left\{\left(\hat{\sigma}_B^C\right)^2\right\} \\ & + 2Cov\left(SSW^T, \left(\hat{\sigma}_B^T\right)^2\right) + 2Cov\left(SSW^C, \left(\hat{\sigma}_B^C\right)^2\right)\end{aligned}.$$

Using expressions 95 and 102 for the variances of the sums of squares and the variance component estimates and expression 96 for the covariance term from pages 74 and 75 of Searle, Casella, and McCulloch (1992), and simplifying yields

$$V\left\{S_T^2\right\} = \frac{2\sigma_W^4}{N-2} + \frac{2B\sigma_B^2\sigma_W^2}{(N-2)^2} + \frac{2A\sigma_B^4}{(n-2)^2}, \tag{38}$$

where A and B are the auxiliary constants in (20).

The expected value and variance of S_B^2 can be derived directly. Writing S_B^2 = $(\mathbf{y_T}'\mathbf{A_T}\mathbf{y_T} + \mathbf{y_C}'\mathbf{A_C}\mathbf{y_C})/(M - 2)$ where $\mathbf{y_T}$ and $\mathbf{y_C}$ are vectors of treatment and control group cluster means, respectively, and $\mathbf{A_T}$ and $\mathbf{A_C}$ are m^T by m^T and m^C by m^C matrices defined by $\mathbf{A_T} = \mathbf{I} - \mathbf{11}'/m^T$ **and** $\mathbf{A_C} = \mathbf{I} - \mathbf{11}'/m^C$ respectively where $\mathbf{I}$ is an identity matrix and $\mathbf{1}$ is a column vector of 1's of appropriate dimensions. A direct, but tedious application of a theorem on the mean and variance of variance of quadratic forms in normal variables (see, e.g , Searle, 1971, p. 57) gives the mean of S_B^2 as $\{\text{trace}(\mathbf{A_T V_T}) + \text{trace}(\mathbf{A_C V_C})\}/(M-2)$ and variance of S_B^2 as $2\{\text{trace}(\mathbf{A_T V_T A_T V_T}) + \text{trace}(\mathbf{A_C V_C A_C V_C})\}/(M-2)^2$, where $\mathbf{V_T}$ and $\mathbf{V_C}$ are the covariance matrices of $\mathbf{y_T}$ and $\mathbf{y_C}$. Here $\mathbf{V_T}$ is an m^T by m^T diagonal matrix whose i^{th} diagonal element is $\sigma_B^2 + \sigma_W^2/n_i^T$, and $\mathbf{V_C}$ is an m^C by m^C diagonal matrix whose i^{th} diagonal element is $\sigma_B^2 + \sigma_W^2/n_i^C$. Using this theorem we obtain

$$E\left\{S_B^2\right\} = \sigma_B^2 + \frac{\sigma_W^2}{\bar{n}_B} \tag{39}$$

and

$$V\left\{S_B^2\right\} = \frac{2(M-2)\sigma_B^4 + 2C_1\sigma_W^4 + 4C_2\sigma_B^2\sigma_W^2}{(M-2)^2} \tag{40}$$

where $C_1 = (m^T - 2)\bar{n}_I^{T2} + (m^C - 2)\bar{n}_I^{C2} + (\bar{n}_I^T)^2 + (\bar{n}_I^C)^2$ and $C_2 = (m^T - 1)\bar{n}_I^T + (m^C - 1)\bar{n}_I^C$.

The distribution of d_W. To obtain the distribution of d_W, apply the theorem with $\sigma^2 = \sigma_W^2$ and $S^2 = S_W^2$. Here

$$a = \frac{\sigma_W^2 + \tilde{n}\sigma_B^2}{\sigma_W^2} = \frac{1 + (\tilde{n} - 1)\rho}{1 - \rho},$$

$b = E\{S_W^2\}/\sigma_W^2 = 1$, and $c = V\{S_W^2\}/2\sigma_W^2 = 1/(N - M)$. Substituting the expressions for a, b, and c into (29) and simplifying gives (18). Note that since S^2 involves only a single chi-square, it follows that the t-statistic corresponding to d_W has exactly the noncentral t-distribution with $(N - M)$ degrees of freedom.

The distribution of d_T. In this case we apply the theorem with $\sigma^2 = \sigma_T^2 = \sigma_B^2 + \sigma_W^2$ and $S^2 = S_T^2$. Here

$$a = \frac{\sigma_W^2 + \tilde{n}\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = 1 + (\tilde{n} - 1)\rho .$$

Using the expected value of S_T^2 given in (37), compute

$$b = \frac{\sigma_W^2 + \dfrac{B\sigma_B^2}{N-2}}{\sigma_W^2 + \sigma_B^2} = 1 - \rho\left(\frac{N - 2 - B}{N - 2}\right).$$

Using the variance of S_T^2 given in (38) compute

$$c = \frac{(N-2)\sigma_W^4 + A\sigma_B^4 + 2B\sigma_B^2\sigma_W^2}{(N-2)^2(\sigma_W^2 + \sigma_B^2)^2} = \frac{(N-2)(1-\rho)^2 + A\rho^2 + 2B\rho(1-\rho)}{(N-2)^2}.$$

Substituting the expressions for a, b, and c into (29) and simplifying, gives the result in expression (20).

The distribution of d_B. To obtain the distribution of d_B, apply the theorem with $\sigma^2 = \sigma_B^2$ and $S^2 = S_B^2$. Here

$$a = \frac{\sigma_W^2 + \overline{n}_B \sigma_B^2}{\sigma_B^2} = \frac{1 + (\overline{n}_B - 1)\rho}{\rho}.$$

Using the expected value of $S_B{}^2$ given in (39) gives

$$b = \frac{\sigma_B^2 + \dfrac{\sigma_W^2}{\overline{n}_B}}{\sigma_B^2} = \frac{1 + (\overline{n}_B - 1)\rho}{\overline{n}_B \rho}.$$

Using the variance of $S_B{}^2$ given in (40), compute $c = \mathrm{V}\{S_B{}^2\}/2\sigma_B{}^4$ as

$$c = \frac{(M-2)\sigma_B^4 + C_1\sigma_W^4 + 2C_2\sigma_B^2\sigma_W^2}{(M-2)^2\sigma_B^4} = \frac{(M-2)\rho^2 + C_1(1-\rho)^2 + 2C_2\rho(1-\rho)}{(M-2)^2},$$

where C_1 and C_2 are given above. Substituting a, b, and c into (29) and simplifying yields (22).

References

Box, G. E. P. (1954). Some theorems on quadratic forms applied to the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, *25*, 290-302.

Donner, N., Birkett, N., & Buck, C. (1981). Randomization by cluster. *American Journal of Epidemiology*, 114, 906-914.

Donner, A. & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.

Donner, A. & Klar, N. (2002). Issues in the meta-analysis of meta-analysis of cluster randomized trials. *Statistics in Medicine*, *21*, 1971-2980.

Geisser, S. & Greenhouse, S. W. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics*, *29*, 885-891.

Guilliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies. Data from the Health Survey for England 1994. *American Journal of Epidemiology*, *149*, 876-883.

Hedberg, E. C. Santana, R., & Hedges, L. V. (2004). *The variance structure of academic achievement in America*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128.

Kish, L. (1965). *Survey sampling*. New York: John Wiley.

Klar, N. & Donner, A. (2001). Current and future challenges in the design and analysis of cluster randomization trials. *Statistics in Medicine*, *20*, 3729-3740.

Laopaiboon, M. (2003). Meta-analyses involving cluster randomization trials: A review of published literature in health care. *Statistical Methods in Medical Research*, *12*, 515-530.

Murray, D. M. & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials, *Evaluation Review*, *27*, 79-103.

Murray, D. M., Varnell, S. P., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health*, *94*, 423-432.

Ridgeway, J. E., Zawgewski, J. S., Hoover, M. N., & Lambdin, D. V. (2002). Student attainment in connected mathematics curriculum. Pages 193-224 in S. L. Senk & D. R. Thompson (Eds.) *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Erlbaum.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.

Rooney, B. L. & Murray, D. M. (1996). A meta-analysis of smoking prevention programs after adjustment for errors in the unit of analysis. *Health Education Quarterly*, *23*, 48-64.

Searle, S. R. (1971). *Linear models*. New York: John Wiley.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: John Wiley.

Senk, S. L. (2002). The effects of the UCSMP secondary school curriculum on students' achievement. Pages 425-456 in S. L. Senk & D. R. Thompson (Eds.) *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Erlbaum.

Snedecor, G. W. (1956). *Statistical methods applied to experiments in agriculture and biology*. Ames, IA: Iowa State University Press.

Verma, V. & Lee, T. (1996). An analysis of sampling errors for demographic and health surveys. *International Statistical Review*, *64*, 265-294.

Table 1
Data from the Evaluation of UCSMP Geometry Second Edition: HSST Geometry test

UCSMP			Comparison		
n	**Mean**	**SD**	**n**	**Mean**	**SD**
9	34.4	10.1	7	39.3	13.7
5	29.0	8.9	9	36.7	14.1
22	50.3	12.7	13	43.5	12.7
20	48.3	9.3	17	42.2	12.3
20	46.8	15.0	19	48.4	14.8
17	47.5	10.7	15	49.7	10.2
25	40.4	10.0	14	38.6	15.0
21	33.8	10.8	21	38.8	17.1

Note: These data are from Senk, 2002.