Randomization of Clusters Versus Randomization of Persons Within Clusters: Which Is Preferable?

Mirjam MOERBEEK

Many experiments aim at populations with persons nested within clusters. Randomization to treatment conditions can be done at the cluster level or at the person level within each cluster. The latter may result in control group contamination, and cluster randomization is therefore often preferred in practice. This article models the control group contamination, calculates the required sample sizes for both levels of randomization, and gives the degree of contamination for which cluster randomization is preferable above randomization of persons within clusters. Moreover, it provides examples of situations where one has to make a choice between both levels of randomization.

KEY WORDS: Experimental design; Intervention study; Intra-cluster correlation coefficient; Nested data; Relative efficiency.

1. INTRODUCTION

Experimenters in the biomedical and social sciences often face nested data structures with persons nested within clusters. Examples are school-based smoking prevention interventions with pupils nested within schools (Ausems, Mesters, Van Breukelen, and De Vries 2002), smoking cessation studies with employees nested within worksites (Hedeker, McMahon, Jason, and Salina 1994), and multicenter clinical trials to test the efficacy of new drugs with patients nested within centers (Hedeker, Gibbons, and Davis 1991; Bach et al. 1995).

A key issue that may arise when designing an experiment with nested data concerns the level of randomization. Although treatments are often aimed at the individual, randomization to treatment conditions may be possible at any level of the hierarchical data structure, that is, at the cluster level or at the person level within each cluster. In the first case, complete clusters are randomized to treatment conditions, so clusters are nested within treatment conditions and all persons within a cluster receive the same treatment. This design is often referred to as a cluster randomized trial. In the latter case, persons within clusters are randomized to treatment conditions, so clusters are crossed with treatment conditions and all treatment conditions are available within each cluster. This design may be referred to as a multicenter or a multisite trial (Raudenbush and Liu 2000).

In practice, cluster randomization is often preferred for ethical, political, administrative, or financial reasons (Gail, Mark, Carroll, and Green 1996). Moreover, for some types of trials there is no alternative to cluster randomization. An example is a community-based intervention where the intervention will necessarily affect all members of the community (Gail et al. 1996). Another reason to randomize complete clusters to treatment conditions is the need to avoid control group contamination, which occurs when information leaks from the experimental to the control group (Donner and Klar 1994). Control group contamination is an unwanted side-effect, because it ordinarily would result in an underestimation of the difference between the experimental and control group, as individuals in different treatment groups would now receive treatments that are more similar than intended (Plewis and Hurry 1998; Craven, Marsh, Debus, and Jayasinghe 2001). To compensate for control group contamination, a larger sample size is needed to detect a given sized treatment effect than would have been needed were control group contamination absent.

A design is preferred from a financial point of view when we can buy a smaller variance of the treatment effect estimator with available recourses using that design than we can for the alternative design. Randomization of persons within clusters has been shown to result in a more efficient design and therefore a higher power to detect treatment effects when control group contamination is absent, especially when the intra-cluster correlation coefficient is large and/or the costs to sample an additional cluster are large in relation to the costs to sample an additional person in an already sampled cluster (Moerbeek, van Breukelen, and Berger 2001). Little research has been done on the optimal level of randomization for experiments with control group contamination.

The goal of this article is to summarize the extent of control group contamination that can be tolerated and still result in randomization of persons within clusters being preferable to cluster randomization. The results may help investigators to weigh the advantages of higher efficiency that can be achieved by using a multisite trial, and the ethical, political, and administrative limitations of such a design. Investigators should clearly report the arguments for the chosen level of randomization in order to help other researchers in designing future trials. Also, the size of the intra-cluster correlation and the (anticipated) degree of control group contamination should be reported.

This article builds upon those by Slymen and Hovell (1997) and Torgerson (2001), who investigated the efficiency of a cluster randomized trial versus a trial in which persons are ran-

Mirjam Moerbeek is Assistant Professor, Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508 TC Utrecht, the Netherlands (E-mail: M.Moerbeek@fss.uu.nl). This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO).

  DOI: 10.1198/000313005X20727

domly sampled from the population under study and randomly allocated to treatment conditions. The comparison made in the present article is different because it compares a cluster randomized trial to a trial in which randomization is done at the person level within each cluster. Moreover, the present article accounts for possible partial contamination, which means there may be an effect representing some portion of the treatment versus control difference for individuals in the contaminated part of the control group.

The contents of the article are as follows. The next section presents a statistical model to analyze the data and summarizes the results for experiments without control group contamination. Section 3 models the control group contamination and derives the optimal level of randomization for different degrees of control group contamination. Section 4 illustrates the choice of level of randomization using some realistic examples of situations where one has to make a choice between a cluster randomized trial and a multisite trial. As will be shown, the choice should not be driven by the design efficiency only, but also by limitations from an ethical, political, or administrative point of view. The last section gives some conclusions and a discussion.

2. CONTROL GROUP CONTAMINATION ABSENT

Due to the nesting of persons within clusters, outcomes for cluster members may be influenced by the cluster membership (Goldstein 1995). Moreover, persons often choose the cluster to which they belong, leading to similarities between persons within a cluster. Clusters may be existing groups, such as schools, families, and neighborhoods, or may be established by the randomization of individuals to groups in which the treatment is administered, such as therapy groups or physician practices. An unbiased estimate of the treatment effect is obtained when therapists or physicians are assigned at random and the analysis treats the data as clustered by using an appropriate statistical model, such as the mixed-effects model. This model may be used to account for dependency of outcomes of persons within the same cluster because it accounts for random variation in the outcome at the person and cluster level.

When randomization to treatment conditions is done at the person level within each cluster, the interaction between cluster and treatment can be estimated since both treatment conditions are available within each cluster. Suppose outcome y_{ijk} of the ith person in the jth cluster in the kth experimental group $(k = 1, 2)$ is given by $y_{ijk} = \mu_k + u_j + (\mu u)_{jk} + e_{ijk}$. The fixed effect μ_2 is the expected outcome in the experimental group, and it should be replaced by μ_1 for a person in the control group. Let $u_j \sim N(0, \sigma^2_{u0})$ be the random cluster effect, $(\mu u)_{jk} \sim N(0, \sigma^2_{u1})$ be the random interaction between cluster and treatment, and $e_{ijk} \sim N(0, \sigma^2_e)$ be the random person effect. The total variance is equal to $\sigma^2_{u0} + \sigma^2_{u1} + \sigma^2_e$, and the proportion variation at the cluster level is given by the intra-cluster correlation coefficient $\rho = (\sigma^2_{u0}+\sigma^2_{u1})/(\sigma^2_{u0}+\sigma^2_{u1}+\sigma^2_e)$, which varies between 0 and 1. If $\rho = 1$, then all variability in the outcome is at the cluster level and persons within the same cluster respond identically. If $\rho = 0$, then all variability is at the person level and the outcomes of persons within the same cluster are no more correlated than outcomes of persons within different clusters. The intra-cluster correlation coefficient may be split into two parts that reflect the proportion of variance due to the random cluster and interaction effects: $\rho = \rho_0 + \rho_1$, where $\rho_0 = \sigma^2_{u0}/(\sigma^2_{u0} + \sigma^2_{u1} + \sigma^2_e)$ and $\rho_1 = \sigma^2_{u1}/(\sigma^2_{u0} + \sigma^2_{u1} + \sigma^2_e)$.

In order to derive simple formulas of practical use, we assume nonvarying cluster sizes and a balanced design. The number of persons per cluster is assumed to be fixed and denoted n_1, and the number of clusters is denoted n_2, so that the total number of persons is equal to $n_1 n_2$. For randomization of persons within clusters, $n_1/2$ persons per cluster have to be randomized to the control group and the others to the experimental group to achieve a balanced design. For simplicity, let n_1 be an even number. The treatment effect is equal to $\mu_2 - \mu_1$, with an unbiased estimate given by the difference of the mean outcomes across both treatment conditions. The variance of the treatment effect estimator is equal to

$$\frac{4(\sigma^2_{u0} + \sigma^2_{u1} + \sigma^2_e)}{n_1 n_2} (1 - \rho_0 + (n_1 - 1)\rho_1) . \qquad (1)$$

The first factor is equal to the variance of this estimator for a simple random sample ignoring the nesting of persons within clusters. The second factor may be larger or smaller than 1, depending on the values of ρ_0, ρ_1, and n_1. Given a sample size N for a simple random sample, a sample size $N_P = N(1-\rho_0+(n_1-1)\rho_1)$ is needed for a trial with randomization of persons within clusters to achieve the same power level.

For randomization at the cluster level, only one treatment condition is available within each cluster and so the interaction between cluster and treatment cannot be estimated. As a result, the variances σ^2_{u0} and σ^2_{u1} cannot be estimated separately. Instead, their sum is estimated which will be coded as $\sigma^2_u = \sigma^2_{u0} + \sigma^2_{u1}$ in this article. To achieve a balanced design the cluster size is fixed and equal to n_1, and there are $n_2/2$ clusters per treatment condition. For simplicity, let n_2 be an even number. The treatment effect is equal to $\mu_2 - \mu_1$, with an unbiased estimate given by the difference of the mean outcomes across both treatment conditions. The variance of the treatment effect estimator is

$$\frac{4\left(\sigma^2_u + \sigma^2_e\right)}{n_1 n_2} (1 + (n_1 - 1)\rho) . \qquad (2)$$

Given the underlying assumption that the outcomes will be positively correlated or uncorrelated within a cluster (i.e., $\rho \geq 0$), the second factor is always larger than or equal to 1 and is referred to as the variance inflation factor. If a total of N persons is needed in a simple random sample to achieve a certain power level for the test on treatment effect, then $N_C = N(1 + (n_1 - 1)\rho)$ persons are needed for a cluster randomized trial to reach the same power level (Donner, Birkett, and Buck 1981). So, cluster randomization leads to a larger sample size because the variability between clusters needs to be taken into account and only one treatment condition is available within each cluster.

The variance formulas (1) and (2) are a function of the cluster size n_1 and the number of clusters n_2. The optimal allocation of units depends on the costs c_1 to sample an additional person in an already sampled cluster, the costs c_2 to sample an additional cluster, and the available budget C to sample persons and clusters. In general, it is more expensive to sample a new cluster than to sample an additional person once a cluster has been sampled, and the remainder of the article restricts to the case where $c_2 > c_1$. The cost constraint is given by $c_1 n_1 n_2 + c_2 n_2 \leq C$, and

Table 1. Optimal Allocations of Units and Optimal Variance of the Treatment Effect Estimator

Level of randomization	Optimal n_1	Optimal n_2	Optimal variance treatment effect estimator
Person	$\sqrt{\frac{1-\rho}{\rho_1}\frac{c_2}{c_1}}$	$\frac{C}{\sqrt{\frac{1-\rho}{\rho_1}c_1c_2}+c_2}$	$4\sigma^2\frac{\left(\sqrt{\rho_1 c_2}+\sqrt{(1-\rho)c_1}\right)^2}{C}$
Cluster	$\sqrt{\frac{1-\rho}{\rho_0+\rho_1}\frac{c_2}{c_1}}$	$\frac{C}{\sqrt{\frac{1-\rho}{\rho_0+\rho_1}c_1c_2}+c_2}$	$4\sigma^2\frac{\left(\sqrt{(\rho_0+\rho_1)c_2}+\sqrt{(1-\rho)c_1}\right)^2}{C}$

NOTE: $\sigma^2 = \sigma^2_{u0} + \sigma^2_{u1} + \sigma^2_e$

for both levels of randomization the optimal allocation of units is given in Table 1. It follows that the optimal allocation demands a smaller cluster size in the case of cluster randomization than in the case of within-cluster randomization. Also, the optimal n_1 increases when the costs at the cluster level increase relative to the costs at the person level and/or when the variance within clusters increases relative to the variance between clusters. It should be noted that the optimal number of clusters is equal to zero when $\rho_1 = 0$ (randomization of persons within clusters), or when $\rho_0 + \rho_1 = 0$ (cluster randomization). In such cases, a design with a small number of cluster clusters is adopted, and then the results presented in the following hold approximately.

Substitution of the optimal sample sizes into the corresponding formulas (1) and (2) gives the optimal variances of the treatment effect estimator that are given in the last column of Table 1. Comparison of these variances shows that, for any budget C, randomization at the person level always results in a lower variance of the treatment effect estimator than a cluster randomized trial, and hence in a more efficient design. The efficiency of cluster randomization relative to randomization of persons within clusters is defined as the ratio of the variances of the treatment effect estimator, which is equal to

$$\frac{\left(\sqrt{\rho_1 c_2}+\sqrt{(1-\rho)c_1}\right)^2}{\left(\sqrt{(\rho_0+\rho_1)c_2}+\sqrt{(1-\rho)c_1}\right)^2}, \tag{3}$$

and varies between 0 and 1. Its inverse shows how often a cluster randomized trial should be replicated to do as well as a multisite trial. The left side of Figure 1 shows the relative efficiency for $0 \le \rho_0 \le 1$ and different values of the costs ratio c_2/c_1 when treatment by cluster interaction is small $(\rho_1 = .05)$. As can be seen, randomization at the person level is especially preferable when c_2/c_1 and/or ρ_0 are large. As an example consider an intervention with $c_2/c_1 = 25, C/c_1 = 5{,}000, \rho_0 = .1$, and $\rho_1 = .05$. The optimal sample sizes are $n_1 = 20.6$ and $n_2 = 109.6$ for randomization of persons within clusters, and $n_1 = 11.9$ and $n_2 = 135.5$ for cluster randomization. The relative efficiency is equal to .5, which means that the trial that randomizes complete clusters to treatment conditions needs to be replicated twice to do as well as a trial that randomizes persons within clusters. In-

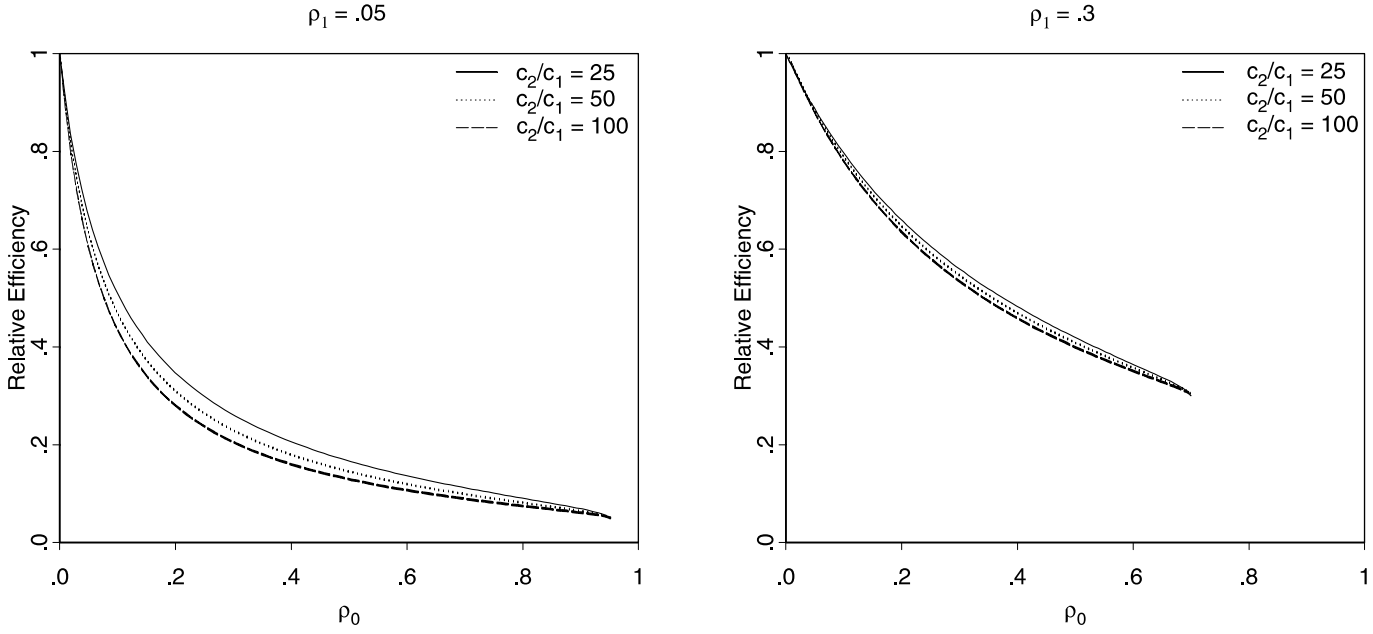

Figure 1. Efficiency of randomization of clusters relative to randomization of persons within clusters as a function of the proportion of variance due to the random cluster and interaction effects (ρ_0 and ρ_1), and the cost ratio c_2/c_1 in the absence of control group contamination.

creasing ρ_1 results in a larger relative efficiency, which depends on c_2/c_1 to a lesser degree. This is illustrated in the right side of Figure 1, for which $\rho_1 = .3$. Note that $\max(\rho_0) = 1 - \rho_1$.

3. CONTROL GROUP CONTAMINATION PRESENT

Control group contamination occurs when both treatments are available within each cluster and information leaks from the experimental to the control group. In the calculations that follow it is assumed that there is contamination in the control group but not in the experimental group, meaning there are no persons who are exposed to the control condition while being assigned to the experimental group. Furthermore, it is assumed that the degree of contamination does not vary across the clusters. We model control group contamination as follows. The expected outcome in the experimental group is simply equal to μ_2. Not necessarily all persons in the control group will receive information from the experimental group. Denote the proportion of the control group that is contaminated by $p(0 \leq p \leq 1)$. The expected outcome in the control group is then equal to

$$(1-p)\mu_1 + p\mu_1', \tag{4}$$

where μ_1 and μ_1' are the expected responses in the uncontaminated and contaminated parts of the control group, respectively. The contamination may be partial, meaning that the communication on the contents of the experimental condition between both treatment groups is incomplete, so that the persons in the contaminated part of the control group will only receive part of the treatment effect. Define μ_1' as the expected response in the uncontaminated part of the control group plus an additional fraction $f(0 \leq f \leq 1)$ of the treatment effect:

$$\mu_1' = \mu_1 + f(\mu_2 - \mu_1). \tag{5}$$

When $f = 1$, there is complete contamination, and when $f < 1$, there is partial contamination, meaning that a person in the contaminated part of the control group receives only part of the treatment effect. Thus, we refer to f as the completeness of the contamination. When $p = 0$ and/or $f = 0$, then there is no control group contamination and the efficiency of a cluster randomized trial relative to a multisite trial is calculated as shown in the previous section. Substitution of (5) into (4) and subtraction from μ_2 results in the treatment effect

$$(\mu_2 - \mu_1)(1 - pf), \tag{6}$$

which is smaller than the treatment effect without control group contamination.

To account for contamination, the variance of the treatment effect estimator for randomization of persons within clusters as calculated in the previous section needs to be divided by $(1-pf)^2$. The relative efficiency of cluster randomization versus randomization of persons within clusters is now equal to

$$\frac{\left(\sqrt{\rho_1 c_2} + \sqrt{(1-\rho)c_1}\right)^2}{\left(\sqrt{(\rho_0+\rho_1)c_2} + \sqrt{(1-\rho)c_1}\right)^2} \frac{1}{(1-pf)^2}, \tag{7}$$

which varies between 0 and ∞. Cluster randomization is more efficient than, as efficient as, or less efficient than randomization of persons within clusters when the relative efficiency is more than, equal to, or less than 1, respectively. Consequently, cluster randomization is preferable when

$$1 - pf < \frac{\sqrt{\rho_1 c_2} + \sqrt{(1-\rho)c_1}}{\sqrt{(\rho_0+\rho_1)c_2} + \sqrt{(1-\rho)c_1}}. \tag{8}$$

The curves in Figure 2 describe conditions for which cluster randomization is as efficient as randomization of persons within clusters. For each cost ratio c_2/c_1, the area above the corresponding curve reflects conditions for which cluster randomization is preferable. From Figure 2 it follows that a cluster randomized trial is more efficient when the cost ratio c_2/c_1 and the proportion of variance due to the random cluster effect ρ_0 are small and when the degree of contamination for a multisite trial (as indicated by p and f) is large. Comparison of the curves on the right side ($\rho_1 = .3$) to those on the left side ($\rho_1 = .05$) shows that the choice for cluster randomization becomes more tenable in terms of efficiency when treatment by cluster interaction becomes larger. Again, note that $\max(\rho_0) = 1 - \rho_1$.

4. CHOICE OF LEVEL OF RANDOMIZATION

Due to the control group contamination, the treatment effect for a multisite trial as estimated by the difference of the mean outcomes in both treatment conditions is an underestimate of the true treatment effect. The choice of the most efficient level of randomization depends on the degree of control group contamination. Moreover, the degree of contamination is needed to correct for the bias in the estimated treatment effect. Unfortunately, this degree is generally unknown. When control group contamination is expected to be absent or low, a multisite trial is preferable in terms of efficiency above a cluster randomized trial. Although the treatment effect may be slightly underestimated, it may still be statistically significantly different from no treatment effect. Moreover, an advantage of a multisite trial is that the treatment by cluster interaction can be estimated because both treatment conditions are available within each cluster. When, on the other hand, the degree of control group contamination is likely to be moderate to large, an unbiased estimate of the treatment effect can only be obtained when the degree is known. This is unlikely, and it is therefore better to choose a cluster randomized trial, although it may be less efficient than a multisite trial. Of course, the choice for the level of randomization should not be driven only by the efficiency of a design, but also by criteria from a practical nature as is illustrated by the examples that follow.

Consider a multicenter, randomized, double-blind, placebo-controlled clinical trial with patients nested within clinics. Patients are randomly assigned to either the experimental or control group and neither the patients nor the researchers know who belongs to which treatment condition. Such a study is feasible when the experimental treatment is a new drug that is administered to the patients using tablets or injections that differ from those given to the patients in the control group only by the amount of active substance. With such a study it is reasonable to randomize patients within clinics to treatment conditions because control group contamination is likely to be absent. Such a design is often used in the biomedical sciences. Another example is a new type of surgery for which the required equipment is expen-

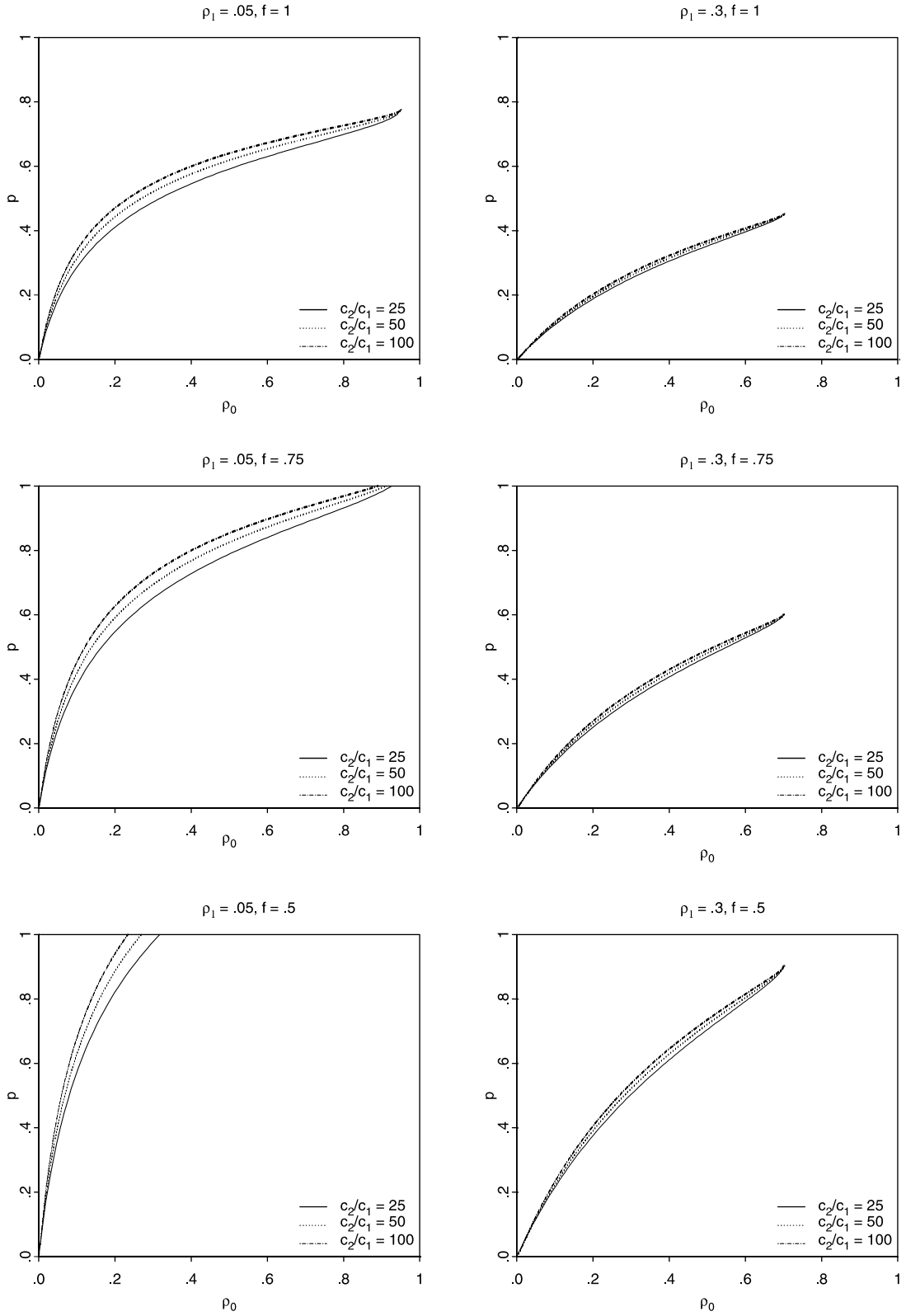

Figure 2. The curves indicate combinations of the cost ratio c_2/c_1, the proportion of variance due to the random cluster and interaction effects (ρ_0 and ρ_1), proportion p of the control group that is contaminated, and the completeness f of the contamination for which cluster randomization is as efficient as randomization of persons within clusters. For each c_2/c_1, cluster randomization is more efficient in the area above the corresponding curve.

sive. Cluster randomization is preferable from a financial point of view when a smaller variance of the treatment effect estimator is bought with available resources than for a multisite trial. In that case, the expensive equipment does not have to be purchased by each clinic that participates in the trial. Furthermore, there may be political or administrative reasons to randomize complete clusters, such as in a study on the impact of vitamin A supplementation on childhood mortality (Sommer et al. 1986). In this study it was not acceptable to treat some children in a given village and not others, and therefore complete villages were randomized to treatment conditions.

The use of blinding to avoid treatment group contamination is no option when the new treatment relies on interpersonal interactions, as with peer-pressure groups, risk-reduction sessions, or employment and training programs. In such cases the degree of control group contamination depends, among others, on the degree to which persons in the experimental group meet with those in the control group and exchange information on the intervention. In a multisite study on the effects of a behavioral intervention to reduce HIV risk behaviors (National Institute of Mental Health 1998) participants from 37 clinics were randomly assigned to an intervention group (HIV risk reduction sessions) or a control group. In three multisite trials (Greenberg, Meyer, and Wiseman 1994) the effects of employment and training programs on outcomes such as employment status, earnings, welfare status, and so forth were evaluated. Welfare recipients from different local administrative offices were assigned to the experimental or control group. In these two examples both treatment conditions were available within each site, and such a design is only tenable when the researchers are convinced and can motivate that control group contamination is negligible.

Control group contamination may occur when information on the contents of the intervention leaks from the individuals in the experimental group to those in the control group. In other examples, such as guideline trials, it may be due to the person delivering the intervention. An example is the education of family physicians about guidelines to reduce unhealthy life styles. When both a control and experimental group are available within each family practice, it would be difficult for the physician not to let patients in the control group benefit from the education. Therefore the appropriate unit of randomization is the physician and not the patient.

Cluster randomization is often the only viable option in situations where the clustering is such that persons within the same cluster meet regularly, such as families, schools, worksites, and churches. In such cases the degree of control group contamination is often too large to be negligible, especially when the intervention is considered to be peer pressure resulting from the program. Consequently the treatment effect may be severely underestimated when randomization is done at the person level within each cluster.

5. CONCLUSIONS AND DISCUSSION

In comparison to an experiment with a simple random sample, a cluster randomized trial typically needs a higher sample size to achieve the same level of precision on account of intra-cluster correlation. With a multisite trial, both treatment conditions are available within each cluster and the sample size is smaller than that of a cluster randomized trial, given the assumption that control group contamination is absent. Unfortunately, multisite trials may be subject to control group contamination, and the sample size would need to be increased to achieve the same efficiency as in an experiment without control group contamination. This article has modeled the control group contamination and explored the conditions under which a multisite trial is more efficient than a cluster randomized trial. It has shown that a multisite trial is preferable only when the degree of contamination is negligible or small. With more severe contamination, the treatment effect may be severely underestimated and cannot be corrected when the magnitude of contamination is unknown. If one is unaware of the level of control group contamination and resources permit exploration of this question, a pilot study may be conducted in which some clusters are located to a cluster randomized trial and some other clusters assigned to a multisite trial. The factor $(1 - pf)$ can then be estimated as the estimated treatment effect in the multisite trial divided by the estimated treatment effect in the cluster randomized trial. An estimate of the intra-cluster correlation coefficient ρ can also be obtained from such a pilot study, and Equation (7) may be used to form a decision on the preferred level of randomization for the experiment. Variability in the estimation of $(1 - pf)$ and ρ would imply uncertainty in such decisions; development of decision rules in such a context is left for future research.

This article has implicitly assumed that control group contamination does not occur when randomization is done at the cluster level. This assumption is not always true in practice. For instance, control group contamination in the case of a school-based smoking prevention intervention may occur when both control and experimental schools are located in the same neighborhood, or when children within a family attend different schools. Another example is a trial on obesity prevention and reduction, with primary care practices as the unit of randomization. Control group contamination may occur when some staff members work between several practices and distribute items such as dietary leaflets in control practices. In such cases, there is less of an imperative to implement a cluster randomized trial to avoid control group contamination.

[Received December 2003. Revised October 2004.]

REFERENCES

Ausems, M., Mesters, I., Van Breukelen, G., and De Vries, H. (2002), "Short-Term Effects of a Randomized Computer-Based Out-of-School Smoking Prevention Trial Aimed at Elementary School Children," *Preventive Medicine*, 34, 581–589.

Bach, D., Van den Berg-Segers, A., Hubner, A., Van Breukelen, G., Cesana, M., and Pletan, Y. (1995), "Refloxacin Once Daily Versus Ciprofloxacin Twice Daily in the Treatment of Patients with Acute Uncomplicated Pyelonephritis," *The Journal of Urology*, 154, 19–24.

Craven, R. G., Marsh, H. W., Debus, R. L., and Jayasinghe, U. (2001), "Diffusion Effects: Control Group Contamination Threats to the Validity of Teacher-Administered Interventions," *Journal of Educational Psychology*, 93, 639–645.

Donner, A., Birkett, N., and Buck, C. (1981), "Randomization by Cluster: Sample Size Requirements and Analysis," *American Journal of Epidemiology*, 114, 906–914.

Donner, A., and Klar, N. (1994), "Methods for Comparing Event Rates in Intervention Studies When the Unit of Allocation Is a Cluster," *American Journal of Epidemiology*, 140, 279–289.

Gail, M. H., Mark, S. D., Carroll, R. J., and Green, S. B. (1996), "On Design Considerations and Randomization-Based Inference for Community Intervention

Trials," *Statistics in Medicine*, 15, 1069–1092.

Goldstein, H. (1995), *Multilevel Statistical Models* (2nd ed.), London: Edward Arnold.

Greenberg, D., Meyer, R. H., and Wiseman, M. (1994), "Multisite Employment and Training Program Evaluations: A Tale of Three Studies," *Industrial and Labour Relations Review*, 47, 679–691.

Hedeker, D., Gibbons, R. D., and Davis, J. M. (1991), "Random Regression Models for Multicenter Clinical Trials Data," *Psychopharmacology Bulletin*, 27, 73–77.

Hedeker, D., McMahon, S. D., Jason, L. A., and Salina, D. (1994), "Analysis of Clustered Data in Community Psychology: With an Example from a Worksite Smoking Cessation Project," *American Journal of Community Psychology*, 22, 595–615.

Moerbeek, M., van Breukelen, G. J. P., and Berger, M. P. F. (2001), "Optimal Experimental Designs for Multilevel Models with Covariates," *Communications in Statistics, Theory and Methods*, 30, 2683–2697.

National Institute of Mental Health (1998), "The NIMH Multisite HIV Prevention Trial: Reducing HIV Sexual Risk Behavior," *Science*, 280, 1889–1894.

Plewis, I., and Hurry, J. (1998), "A Multilevel Perspective on the Design and Analysis of Intervention Studies," *Educational Research and Evaluation*, 4, 13–26.

Raudenbush, S. W., and Liu, X. (2000), "Statistical Power and Optimal Design for Multisite Randomized Trials," *Psychological Methods*, 5, 199–213.

Slymen, D. J., and Hovell, M. F. (1997), "Cluster Versus Individual Randomization in Adolescent Tobacco and Alcohol Studies: Illustrations for Designs Decisions," *International Journal of Epidemiology*, 26, 765–771.

Sommer, A., Tarwotjo, I., Djunaedi, E., West, K. P., Loeden, A. A., Tilden, R., and Mele, L. (1986), "Impact of Vitamin a Supplementation on Childhood Mortality. A Randomised Controlled Community Trial," *Lancet*, 1986, 1169–1173.

Torgerson, D. J. (2001), "Contamination in Trials: Is Cluster Randomisation the Answer?," *British Medical Journal*, 322, 355–357.