Cluster versus Individual Randomization in Adolescent Tobacco and Alcohol Studies: Illustrations for Design Decisions

DONALD J SLYMEN AND MELBOURNE F HOVELL

Slymen D J (Center for Behavioral Epidemiology, Graduate School of Public Health, San Diego State University, San Diego, CA 92182–4162, USA) and Hovell M F. Cluster versus individual randomization in adolescent tobacco and alcohol studies: Illustrations for design decisions. International Journal of Epidemiology 1997; 26: 765–771.

Background. The decision to randomize by clusters of subjects such as a classroom or clinic versus individual randomization where some contamination may occur is examined within the framework of sample size issues. Estimates for background rates and intraclass correlations are also provided for adolescent tobacco and alcohol outcomes derived from a recent study using cluster randomization.

Methods. A ratio of adjusted sample sizes is derived which is a function of the intraclass correlation and cluster size for cluster randomization and total amount of contamination for individual randomization. Using estimated incidence rates and intraclass correlations, we provide a comparison of sample sizes for two plausible study outcomes.

Results. Small clusters such as a family or small classroom tend to have stronger within cluster dependence and cluster randomization would be clearly favoured over individual randomization. For moderately sized clusters, if contamination levels are likely to be high then cluster randomization would be a better choice. However in some situations where lower levels of contamination are expected, individual randomization may be preferred. With larger clusters, individual randomization should be considered when contamination rates are expected to be low.

Conclusions. Investigators must carefully consider the choice of cluster randomization versus individual randomization in the context of likely contamination. In this paper we provided a basis for making this decision as well as examples to illustrate these decisions, and parameter estimates that will be especially useful for investigators in adolescent tobacco and alcohol studies.

Keywords: adolescent tobacco and alcohol studies, clinical trials, cluster randomization, contamination, intraclass correlation, sample size

Large-scale clinical trials face a number of design decisions that may be critical to the internal as well as the external validity of results. Among these decisions is random assignment of individuals or cluster groups to experimental conditions. This decision may be based on practical considerations of cost/feasibility in contrast with risks of experimental contamination. Ordinarily, assigning individuals to conditions, where it can be assumed that all individuals are independent, is the preferred design and the design that may be the most efficient. In a cost sensitive world, the more efficient design is the more likely to be feasible and actually conducted.

However, with non-therapeutic trials it may not be practical or feasible to use this approach.

Assumptions of independence may not be reasonable and it may not be possible to guarantee independence if individuals are assigned at random to experimental conditions. With lifestyle or behavioural interventions, although the outcomes are usually measured at the individual level, individual randomization may lead to high rates of contamination. Contamination occurs when members of the control group are exposed to the experimental intervention and/or members of the experimental group are exposed to the control either by direct exposure to an attention control or a diminution of the experimental intervention with behaviour similar to that of control subjects. The more likely form of contamination is where the control subjects are exposed to the experimental conditions—either directly or indirectly. This may occur when an intervention is delivered within a school or a health clinic where experimental and control subjects are in close proximity and information...
may be shared. Youth encouraged to exercise while at recess (or in physical education classes) may serve as both models and as sources of encouragement for other youth in the same class to exercise. To obviate this problem an increasingly popular design is to randomize the entire school or clinic (i.e. the cluster) although outcomes are still measured at the subject level. This design has received considerable attention in recent literature, however, the decision whether to use cluster or individual randomization has not been given due consideration.

As pointed out in a recent article investigators should carefully consider whether cluster randomization is preferable to individual randomization given the loss of power. However, the degree of contamination must also be considered in this decision. This paper will provide guidance within the framework of sample size issues.

We also provide some empirical results on how the selection of outcome measures can affect sample size with cluster randomization. Although a few papers have published estimates of the degree of clustering as measured by the intraclass correlation for selected outcomes, this information is difficult to find in the literature and is very useful for planning future studies. We focus on tobacco and alcohol measures obtained from a recent study employing cluster randomization.

CLUSTER VERSUS INDIVIDUAL RANDOMIZATION

RANDOMIZATION WITH CONTAMINATION

Consider a two-group study design to compare an experimental with a control group using a continuous outcome. It is well known that the usual sample size formula to compare two population means assuming individual randomization (N) must be adjusted using an inflation factor to accommodate cluster randomization. Let m represent the cluster size and ρ the intraclass correlation. The parameter $\rho$ measures the degree of dependence within the cluster. If $N'$ represents the adjusted sample size per group, then $N' = N[1 + (m - 1) \rho]$. The inflation factor (IF), $1 + (m - 1) \rho$, is affected by the degree of dependence within cluster as well as the size of the cluster. In large clusters such as a health clinic or school it is not unusual to observe low correlations of 0.01 or less. However, the cluster size may be in the hundreds. Thus, the adjustment may be substantial. For example, if $\rho = 0.01$ and $m = 200$, the sample size must be nearly tripled to overcome the clustering even though the intraclass correlation is relatively small. In clusters such as a family, $\rho$ is typically higher but $m$ much smaller. We might anticipate $\rho$ at 0.25 for some outcomes with an average family size of, say four. The IF of 1.75 suggests the sample size must be inflated by 75% to overcome randomizing by families.

Cluster randomization is compared to individual randomization in the presence of contamination. In some studies individual randomization may not be an option; all of the control subjects would be exposed to the experimental condition and contamination would be complete. This might be true, for instance, in community-based media interventions where almost all individuals within a given community will come in contact with media and it would not be possible to direct media events to individuals assigned to conditions. However, in other circumstances the rate of contamination may be small enough where individual randomization could be considered. This might be true for clinical medicine studies where the degree of similarity among patients may be limited and where the intervention can be implemented for one individual within a given cluster. Giving some patients vitamins, while not doing so for others may be feasible within a single medical office, with individuals assigned at random to conditions. Some contamination might result, for example, from general advertisements for vitamins seen by study participants in the community, but the overall impact would be expected to be small considering the intervention is targeted at the individual patient.

A simple and common approach to account for contamination or dropouts is to assume that a certain proportion of experimental subjects will ‘behave’ like control subjects and therefore dilute the mean outcome for the experimental group. This could be the case, as in the above example, where even though assigned to a group that should receive vitamins, an occasional patient is seen and the clinician fails to prescribe vitamins. Let $\mu_E$ and $\mu_C$ represent the anticipated mean rates for the experimental and control groups respectively. Let $p_1$ be the rate of contamination for dropouts. By dropouts we are assuming individuals who are no longer exposed to the intervention but who, nevertheless, have their outcomes measured and will be included in an intention-to-treat analysis. Dropouts should be distinguished from subjects lost to follow-up whose outcomes are not measured. The effective treatment mean becomes

$$\mu_E' = (1 - p_1) \mu_E + p_1 \mu_C.$$

Likewise it is possible that some of the control subjects may follow a programme regimen similar to that of the experimental subjects. This is not unusual in lifestyle intervention studies involving exercise or diet. Again, using the example above, this might occur if the clinician mistakenly prescribed vitamins to a patient in the control condition, or it might take place if an experimental patient discussed the vitamin prescription...
and recommended to another patient that he/she attain the same vitamins at the local pharmacy. Let $p_2$ represent the contamination rate due to ‘dropins’. The effective control mean becomes $m_C' = p_2 m_E + (1 - p_2) m_C$. The effective treatment difference can be written as $m_E' - m_C' = (m_E - m_C)(1 - p_1 - p_2)$. Thus, the anticipated mean difference must be adjusted lower by a factor of $(1 - p_1 - p_2)$. This adjustment may be somewhat conservative in that it does not account for possible ‘partial’ contamination in which a subject receives a portion of the intended intervention effect rather than none of it. To account for contamination, the usual sample size must be adjusted higher by dividing by $(1 - p_1 - p_2)^2$. If $N^*$ represents the adjusted sample size, then $N^* = N/(1 - p_1 - p_2)^2$. A similar expression yielding an approximate adjusted sample size estimate may be obtained when comparing two proportions if the contamination rates are low.\(^{10}\)

By taking the ratio of $N$ to $N^*$ we have a basis for comparing the two approaches on sample size:

$$\frac{N'}{N^*} = [1 + (m - 1)p](1 - p_1 - p_2)^2 = IF[1 - (p_1 + p_2)]^2.$$  

Table 1 displays this ratio for selected values of $m$, $p$, and $p_1 + p_2$ (total amount of contamination). Values were selected based on realistic combinations. Smaller cluster sizes are combined with larger intraclass correlations. If the ratio is less than one, sample size requirements favour cluster randomization (i.e. required sample size is smaller). A ratio greater than one favours individual randomization.

When cluster sizes are small, cluster randomization tends to be favoured. In studies where family is the cluster, contamination is likely to be quite high; clearly cluster randomization should be implemented. For cluster sizes of 10–30 such as the classroom setting, if contamination levels are moderate to large, cluster randomization would be a better choice. For instance, if a teacher tests a nutrition intervention aimed at reducing fat intake for grade school students, he/she might provide formal nutrition education and differentially reinforce youth in an experimental condition when they eat low fat foods, fruits and vegetables. Presenting the nutrition education may not be possible for selected students, thereby exposing all control students to the same procedures. Similarly, the youth who receive differential reinforcement for low fat consumption may model and formally encourage other students to do the same. Thus, in small clusters, such as classrooms, both direct and indirect forms of contamination may be unavoidable.

With larger clusters, if possible, individual randomization should be favoured unless contamination is

<table>
<thead>
<tr>
<th>Cluster size</th>
<th>Intraclass correlation</th>
<th>Inflation factor</th>
<th>Total amount of contamination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>1.05</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1.10</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>1.20</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>1.40</td>
<td>1.13</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>1.09</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>1.45</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1.90</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>2.80</td>
<td>2.27</td>
</tr>
<tr>
<td>30</td>
<td>0.01</td>
<td>1.29</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>1.98</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>3.16</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>6.80</td>
<td>5.51</td>
</tr>
<tr>
<td>100</td>
<td>0.01</td>
<td>1.99</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>5.95</td>
<td>4.82</td>
</tr>
<tr>
<td>500</td>
<td>0.005</td>
<td>3.50</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>5.99</td>
<td>4.85</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>25.95</td>
<td>21.02</td>
</tr>
</tbody>
</table>

A ratio less than one indicates the sample size requirement for cluster randomization is smaller than that for individual randomization given the set of parameters indicated. A ratio greater than one indicates a smaller sample size for individual randomization.

\(^{10}\) A ratio less than one indicates the sample size requirement for cluster randomization is smaller than that for individual randomization given the set of parameters indicated. A ratio greater than one indicates a smaller sample size for individual randomization.
expected to be so severe that this is not a practical alternative, or when contamination is not the primary consideration. However, even with a contamination rate of 30% a case could be made for individual randomization. This might be true where an intervention is aimed at increasing unpopular behaviour and/or behaviour that might be practiced individually. For instance, a teacher might provide students assigned individually to an experimental condition with extra credit toward their semester grade for exercising at home. Other students assigned to the control would not receive extra credit. Even though some students may discuss or even model and encourage peers in the control group to exercise, the difficulty level of many forms of exercise is so great that most individuals will not do so reliably without definitive reinforcement. The level of contamination might be so low as to justify random assignment of individuals within the class to conditions. Studies where this situation may be the case, might best employ random assignment of individuals to conditions. This can be estimated a priori based on reasonable estimates of contamination in the context of cluster size.

Suppose we set $N'^{-1}/N^* = 1$ which indicates combinations favouring neither approach; estimates that fall above or below 1 can be used as a guide for design considerations. Then solving for the inflation factor, we have $IF = 1/[1 - (p_1 + p_2)]$. Figure 1 displays a plot of this relationship. For combinations of IF and total contamination above the curve, individual randomization is favoured, whereas, combinations below the curve favours cluster randomization. This figure emphasizes the decision-making aspects and relies on the inflation factor to quantify the impact of cluster randomization.

**INTRACLASS CORRELATIONS FOR SMOKING AND ALCOHOL OUTCOMES**

One of the difficulties of planning a study with cluster randomization is finding good estimates of the intraclass correlation. In the future, such estimates should be routinely reported. Based on our recently completed trial on smoking prevention in adolescents, we report the estimated correlations for a number of outcomes typically used in tobacco and alcohol studies.

The study was a randomized controlled trial to test a clinician-delivered tobacco-use prevention programme. A total of 154 orthodontic offices was randomly assigned in equal numbers to experimental or control conditions. The intervention consisted of exposure to anti-tobacco materials and delivery of prescriptions to the patients containing anti-tobacco messages. Cluster randomization was necessary since part of the intervention consisted of creating a tobacco-free environment for the office including displaying tobacco prevention materials. Moreover, to randomize by individual would have required all offices to keep track of both experimental and control subjects which would have put a greater burden on the orthodontists and their staff and probably severely curtailed recruitment of participating offices. Subjects 11–18 years old were surveyed at baseline and 2 years later. The survey covered demographics, health-related behaviours and attitudes including questions about tobacco and alcohol use. Of the 16,915 subjects interviewed at baseline, 92.5% were interviewed at the 2-year follow-up. Although the study focused on tobacco use, we also report outcomes for alcohol use since several questions on the survey addressed this issue. Prevalence outcomes were determined from the baseline survey.

The intraclass correlation was calculated using the formula:

$$[MS_{Between} - MS_{Within}] / [MS_{Between} + (m - 1)MS_{Within}]$$

where $m$ is the average cluster size and $MS_{Between}$ and $MS_{Within}$ represent the mean squares between and within office respectively from a one-factor analysis of variance. This estimate is appropriate for dichotomous as well as continuous outcomes. The standard error was estimated based on the variance formula in Murray, Rooney et al.

Table 2 displays the prevalence and incidence rates, intraclass correlations and their standard errors for
10 tobacco and alcohol outcomes. The range of correlations from 0.0026 to 0.016 is typical for clusters of this type where the degree of dependence is largely driven by demographic similarities within office and limited interaction among patients. Although the correlations are low, the average number of subjects per office is large enough that inflation factors cannot be ignored. They range from 1.28 for the prevalence of using more than 100 tobacco products to 2.66 for prevalence of ever using alcohol. Not surprisingly, the prevalence and incidence rates vary substantially.

### SAMPLE SIZE AND CHOICE OF OUTCOMES

Although much attention has focused on the intraclass correlation and its effect on sample size requirements, the background prevalence or incidence rate is a crucial piece of information for determining the effect size used in the calculation. In this section we examine sample size using the estimates in Table 2. A more complete description of sample size issues for various designs will not be discussed here since other authors have addressed these problems.\(^2\),\(^11\) As an illustration we assess sample size based on two outcomes: past 30-day tobacco incidence and using more than 100 tobacco products in the past 2 years. These measures differ with respect to background incidence rates (12.6% versus 7.6%) and intraclass correlations (0.00893 versus 0.00562).

Suppose a two-group randomized trial is proposed to compare incidence rates on tobacco use between experimental and control groups using cluster randomization. We use a formula to compare two binomial proportions\(^11\) assuming equal sample sizes and apply the inflation factor.

For both outcomes set the significance level at 0.05, the power at 0.90 and the cluster size (m) at 100. Suppose we would like to detect a 30% reduction in the control rate. Using the data in Table 2, for past 30-days use, the control rate is 12.6% and the experimental rate is 12.6(1 – .3) = 8.7%. The required number of subjects per group is now 2644. Therefore, 27 offices per group would be needed. If we select as our outcome using more than 100 tobacco products, then the control rate is 7.6% and the experimental rate is 5.32%. With \(\rho\) at 0.00562, the required number of subjects per group is now 3796 and the required number of offices is 38. The impact of the background incidence rate is substantial and overrides the larger intraclass correlation for 30-days use. Thus, the background rate may be of greater concern than the size of the intraclass correlation.
CONCLUDING REMARKS
Cluster randomization has become an important design feature in many primary prevention trials. However, investigators must carefully evaluate the potentially sizeable impact this approach has on sample size and, therefore, on the cost and time constraints of the proposed project. The alternative for some studies is to use individual randomization even though some contamination is unavoidable. The dilemma faced by investigators could be viewed in terms of choosing between some loss of precision represented by the inflation factor with cluster randomization versus bias where contamination with individual randomization attenuates the true intervention effect. In this paper we have tried to provide some guidance in choosing between these two approaches. To make informed choices, investigators must rely on published information on intraclass correlations and contamination rates. In the second part of this paper we provide estimates on adolescent tobacco and alcohol use for background rates and intraclass correlations. Unfortunately, good information on contamination rates is difficult to find. Detailed information concerning the exposure of both experimental and control subjects to the intervention is necessary to estimate contamination rates. For lifestyle intervention studies such detailed record-keeping is difficult but is of great value in assessing the success of intervention delivery as well as providing estimates of contamination for future studies.

We recognize that for many studies, there is no choice but to use cluster randomization. This is often the case in certain institutional settings, such as schools, where the institution may preclude random assignment of individuals. Most schools, for instance, will not allow students to be reassigned to different classes in order to receive control or experimental procedures. In other instances part or all of the intervention to be tested may be applied to a group. This was true for Hovell et al., for example, where anti-tobacco materials were placed in the offices, and cluster randomization was mandatory. Even if these materials were eliminated from the intervention, individual randomization would have required that each orthodontist keep track of experimental and control subjects within the practice substantially increasing the amount of work required, and with questionable compliance by doctors and staff, risks contamination. The degree to which the compliance with design procedures is questionable may dictate randomization by clusters instead of individuals.

Another form of contamination, which is difficult to quantify, occurs when the same clinician administers the intervention to both treatment and control subjects. When this risk is high the most effective means of prevention is some form of masking (experimental blinding) of clinicians regarding the subject assignment and interventions delivered. However, if masking is not possible, cluster designs may be required to avoid this form of contamination. The anticipated within cluster correlation would be based on the degree to which individuals treated by the same clinician are similar. The similarities may include demographic characteristics. However, the provider or clinician effect will likely account for the largest portion of the clustering. In other instances where the source or sources of contamination are difficult to identify or quantify, cluster randomization may be preferred.

The intraclass correlation estimates for tobacco and alcohol outcomes provided in this paper should aid investigators in appropriate choices for study outcome measures in a study with cluster randomization. Clearly, the background rate for prevalence or incidence has a tremendous impact on sample size requirements and should play a role in choice of outcome measure. Future studies should be encouraged to report intraclass correlations for common cluster units to aid design decisions for future clinical trials.

ACKNOWLEDGEMENT
Research was supported by funds provided by the Cigarette and Tobacco Surtax Fund of the State of California through the Tobacco-Related Disease Research Program of the University of California, Grant #2RT0106, awarded to Dr Hovell.

REFERENCES


(Revised version received January 1997)