Current Issues in the Design and Analysis of Cluster Randomization Trials

Allan Donner, Ph.D.

Department of Epidemiology and Biostatistics
The University of Western Ontario
London, Canada

Robarts Clinical Trials,
Robarts Research Institute
London, Canada
COURSE BASED IN PART ON:


ORDER FROM:

Oxford University Press U.S.
www.oup.com/us
Hoddereducation.co.uk
Amazon.co.uk
What Are Cluster Randomization Trials

Cluster randomization trials are experiments in which intact social units or clusters of individuals rather than independent individuals are randomly allocated to intervention groups.
Examples:

- Medical practices selected as the randomization unit in trials evaluating the efficacy of disease screening programs

- Communities selected as the randomization unit in trials evaluating the effectiveness of new vaccines in developing countries

- Hospitals selected as the randomization unit in trials evaluating educational guidelines directed at physicians and/or administrators
Reasons for Adopting Cluster Randomization

- Administrative convenience
- To obtain cooperation of investigators
- Ethical considerations
- To enhance subject compliance
- To avoid treatment group contamination
- Intervention naturally applied at the cluster level
CLUSTER RANDOMIZATION TRIALS
PUBLISHED 1981-2003

Bland (BMJ, 2004)
A key property of cluster randomization trials is that inferences are frequently intended to apply at the individual level while randomization is at the cluster or group level. Thus the unit of randomization may be different from the unit of analysis.

In this case, the lack of independence among individuals in the same cluster, i.e., intracluster correlation, creates special methodologic challenges in both design and analysis.
Implications of Intracluster Correlation

Reduction in effective sample size (ESS):

- Extent depends on size of the intracluster correlation coefficient and on average cluster size.

If clusters of size are randomized to each of two treatment groups,

then
Application of standard sample size approaches leads to an underpowered study (Type II error)

Application of standard statistical methods generally tends to bias p-values downwards, i.e., could lead to spurious statistical significance (Type I error)
Possible reasons for Intracluster Correlation

1. Subjects frequently select the clusters to which they belong
   e.g., patient characteristics could be related to age or sex differences among physicians

2. Important covariates at the cluster level affect all individuals within the cluster in the same manner
   e.g., differences in temperature between nurseries may be related to infection rates
3. Individuals within clusters frequently interact and, as a result, may respond similarly e.g., education strategies or therapies provided in a group setting.

4. Tendency of infectious diseases to spread more rapidly within than among families or hospital wards.
Design Issues:

Choice of Unit of Inference

The choice of unit of inference has profound implications for both the design and analysis of a cluster randomized trial.

Key question:
Are the primary inferences of interest targeted at i) the individual level or ii) at the cluster level?
Example 1

The intervention consists of a blood pressure screening program intended to lower the risk of cardiovascular mortality

Bass et al (1986)

Example 2

The intervention consists of educational guidelines for the management of hyperlipidaemia intended to increase the proportion of eligible patients prescribed lipid-lowering drugs

What is the unit of inference in these two trials?

How does this guide the approach to the trial design and analysis?
Avoiding Recruitment Bias

Consider a trial randomizing general practices in which physicians are asked to identify as well as to treat selected patients (e.g. Kinmouth et al [1998]).

Can cluster randomization introduce bias through the way patients are differentially recruited across treatment groups?
Source of bias:

- If the physician’s practice has already been randomized, recruitment for patient participation cannot be done blindly with respect to intervention group.

- If physicians in the experimental group are more diligent in seeking out patients than in the control group, or tend to identify patients who are less ill, bias may result.
Unbiased estimates of the effect of intervention can be assured only if analyses are based on data from all cluster members or, alternatively, from a random sub-sample of cluster members.

**Possible Solutions:**

- Identify eligible patients in each practice prior to randomization.

- If eligible patients are identified after randomization, recruitment should be done by an individual independent of the trial.

Torgerson (2001)
Farrin et al (2005)
Requirements for Obtaining Informed Consent

Two distinct levels of informed consent must be distinguished in cluster randomization trials:

(i) informed consent for randomization (usually provided by a ‘decision-maker’)

(ii) informed consent for participants given that randomization has occurred.
By analogy to current ethical requirements for clinical trials, it would be unethical not to obtain informed consent from every cluster member prior to random assignment.

Is such a strict analogy required for cluster randomization trials?
Some Relevant Issues

• To what extent, if any, should distinctions in study design (e.g., unit of randomization assignment, level of intervention) inform ethical considerations for the design of a randomized trial?

• How should the risk/benefit ratio experienced by individual study subjects factor into such a discussion?

• How should decision makers who can agree to random assignment of clusters be identified?
• What should be the responsibilities and requirements of such decision makers (e.g., completion of an informed consent document)?

“Ethical advice indicated that, since we were only providing information to clinicians, there was no reason to seek patient consent”

An Acceptable Approach?

- Health services researchers in the U.S. Department of Veterans Affairs may be seen, by some, as taking a somewhat different approach (Henderson et al. [1998]):

  “In studies in which an administrative intervention does not directly interpose between the physician and the patient, the patient’s treatment remains under the direction of the physician and is not removed by the process of randomization. Moreover eliciting consent from the many thousands of patients involved would create severe budgetary and logistic obstacles. The central human rights committee at the Hines Coordinating Center and those at the participating sites have deemed this type of quality improvement research exempt from direct patients consent”.
Benefit to Control Group Subjects

- It has been argued that patients randomly assigned to a placebo control group in therapeutic trials may still see some benefit if only as a consequence of concern shown by health care providers to follow the standard of care specified by protocol.

- Will similar benefits be enjoyed by participants in health promotion trials comparing an experimental intervention to a usual care control group?

- Concern for fair distribution of benefits and burdens of research among study participants suggest use of minimal control interventions (e.g., an educational brochure) or wait-list control groups.

Glanz et al. (1996)
The Effect of Level of Intervention

- Edwards et al. (1999) distinguished cluster randomization trials by the level of intervention:
  - for “individual-cluster” trials intervention is provided directly to individual study subjects (e.g., vitamin A supplementation);
  - for “cluster-cluster” trials intervention is provided at the cluster level (e.g., effect of local medical opinion leaders on patient treatment).
• The need for consent by individual study subjects is deemed of particular concern for “individual cluster” trials.

• However these distinctions may only be relevant when randomizing larger clusters (e.g., work sites, classrooms, communities).
Cluster Randomization Trials and the Zelen Double Consent Design

• The requirements of informed consent are often perceived to be a barrier to patient accrual. Randomized consent designs were introduced by Zelen (1990) to lower this barrier, thus hastening trial completion.

• In the double consent design patients are randomly assigned to an intervention group prior to asking for consent. Patients are then asked if they consent to receive the intervention to which they have been assigned. If not, they may be offered the alternative therapy under consideration.
• Analyses are conducted based on the intervention to which patients were initially assigned, i.e., analysis by intention to treat.

• Randomized consent designs have proved quite controversial. Reluctance to use randomized consent arises, in part, because of concern for the ethical implications of randomizing subject prior to obtaining their consent.

• But this strategy is typical of many cluster randomized trials!
ISSUES IN SAMPLE SIZE ESTIMATION
Sample Size Requirements for Completely Randomized Designs

Comparison of Means

Suppose \( k \) clusters of size \( m \) are to be assigned to each of two intervention groups.

We wish to determine the size of trial needed to test
The number of subjects required per intervention group is given by

where

Then the number of required clusters per group is given by

For variable sized clusters, replace \( \alpha \) by where \( CV = \) coefficient of variation of the cluster sizes (Manatunga et al [2001]).
Example 1:

Hsieh (1988) reported on the results of a pilot study for a planned 5-year trial examining cardiovascular risk factors, obtaining cholesterol levels from 754 individuals in 4 worksites.

Estimated variance components were
Assuming 70 subjects/worksite,
To obtain 80% power at \( (2\text{-sided}) \) for detecting a mean difference of 20 mg/dl between intervention groups, the number of required worksites per group is given by

\[
\frac{50.4}{20} \leq k
\]

To adjust for the use of normal distribution critical values, and possible loss to follow-up, might enroll 7 clusters per group.
IMPACT ON POWER OF INCREASING THE NUMBER OF CLUSTERS VS. INCREASING CLUSTER SIZE:

Let

As

As
Comparison of Proportions:

No. of subjects required per intervention group to test is given by

No. of clusters required is given by
Example 2:

Murray et al. (1992) reported on a study evaluating the effect of school-based interventions in reducing adolescent tobacco use. An estimate of \( \rho \) as obtained from 24 schools is calculated as 0.01.

Suppose we wish to detect a reduction in the proportion of students using tobacco from 0.06 to 0.04.
If the number of students per schools is approximately 100, then the total number of students required per intervention arm at \( (2\text{-sided}) \), is given by

\[
\sqrt{2^{0.06} \cdot 0.01 + 100} = 0.05 \beta
\]

or 38 schools per intervention group.
Strategies for Increasing Precision

- Restrict inclusion criteria
  e.g. recruit practices of same size, with physicians having similar years of experience in a controlled geographic area.

- Conduct a baseline survey of the prevalence of the trial outcome
  e.g. In smoking cessation trials, perform baseline survey of smoking habits in different communities
  - serves as powerful risk factor for outcome
  - provides estimate of
  - allows trial personnel to gain experience
  - can help to identify potential stratification factors
Increasing cluster size provides diminishing returns in statistical power if e.g. If increasing beyond 100 provides little statistical gain.

Consider sample size re-estimation as part of an interim analysis.

Lake et al (2001)

Recognize likely size of $\rho$ in primary care research (0.01 to 0.05 for outcome variables; 0.05 to 0.15 for process variables).

Campbell et al (2001)
Is fear of contamination overrated as a reason for randomizing clusters?

Suppose under individual randomization a proportion of $R$ of the control group patients experience the same success rate $P_1$ as seen among experimental patients.
Then the difference that can be detected is reduced to

and the required sample size must be inflated by the factor $IF = 1/ (1-R)^2$

e.g. If $R = 0.30$, then $IF = 2.04$

But under cluster randomization, the inflation factor might be much more!

Torgerson (2001)
Choice of Design

Most Frequently Adopted Cluster Randomization Designs

- Completely randomized

- Matched pair (in which one of two clusters in a stratum are randomly assigned to each intervention).

- Stratified (involving the assignment of two or more clusters to at least some combinations of intervention and stratum).
Examples:

COMPLETELY RANDOMIZED DESIGN

Study Purpose:

To evaluate the effectiveness of vitamin A supplements on childhood mortality.

450 villages in Indonesia were randomly assigned to either participate in a vitamin A supplementation scheme, or serve as a control. One year mortality rates were compared in the two groups.

Sommer et al. (1986)
MATCHED PAIR DESIGN

Study Purpose:

To evaluate the effectiveness of preventive advice in reducing incidence of coronary heart disease at six years.

Unit of Randomization: Factory

Number of Matched Pairs: 40

Matching Variables: Geographic area, factory size

World Health Organization (1986)
MATCHED PAIR DESIGN

Study Purpose:

The COMMIT community intervention trial was designed to promote smoking cessation using a variety of community resources. The primary outcome measure was the 5-year smoking cessation rate.

*Unit of Randomization:* Community

*Number of Matched Pairs:* 11
Matching Variables: Community size, population density, demographic profile, community structure, geographical proximity.

STRATIFIED DESIGN

Study Purpose:

To evaluate the effectiveness of treated nasal tissues versus placebo tissues in reducing incidence of respiratory illness at 24 weeks.

*Unit of Randomization:* Family
*Number of Strata:* 3
*Number of Clusters per Stratum:* 30
*Stratification Variable:* Family size (2, 3, or 4 members)

Farr et al. (1988)
Examples:

Study Purpose:

The purpose of the WHO antenatal trial was to compare the impact of two programmes of antenatal care on the health of mothers and newborns.

Unit of Randomization:  Antenatal care clinic.
**Stratification Variables:** Primary stratification was by country: Thailand, Cuba, Argentina, Saudi Arabia.

**Number of Clusters per Stratum:** Ranged from 12 to 17.

Villar et al. (2001)
Benefits of Matching

1. If an effective matching factor can be found, the matched design gains power relative to the completely randomized design.

2. Failure to match may lead to a poor randomization, particularly in smaller samples, thus undermining the trial’s credibility.

In general, matching tends to be a more important design issue in trials randomizing intact clusters than in trials randomizing individuals, since the number of clusters is often relatively small.
Evaluation of the Effectiveness of Matching

There has been considerable research comparing the power of completely randomized and matched pair designs, (e.g., Martin et al. [1993], Klar and Donner [1997]). Some findings from this research are as follows:

If the total study size is 20 or fewer clusters (10 treated and 10 control) then matching should be used only if the investigators are confident that the correlation due to matching is at least 0.20.

“It is unlikely that effective matching would be possible for small studies. Matching may be overused as a design tool”.
<table>
<thead>
<tr>
<th>Community Pair</th>
<th>Proportion of Heavy Smokers Who Quit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
</tr>
<tr>
<td>1</td>
<td>0.139</td>
</tr>
<tr>
<td>2</td>
<td>0.163</td>
</tr>
<tr>
<td>3</td>
<td>0.164</td>
</tr>
<tr>
<td>4</td>
<td>0.204</td>
</tr>
<tr>
<td>5</td>
<td>0.183</td>
</tr>
<tr>
<td>6</td>
<td>0.164</td>
</tr>
<tr>
<td>7</td>
<td>0.262</td>
</tr>
<tr>
<td>8</td>
<td>0.193</td>
</tr>
<tr>
<td>9</td>
<td>0.215</td>
</tr>
<tr>
<td>10</td>
<td>0.136</td>
</tr>
<tr>
<td>11</td>
<td>0.155</td>
</tr>
<tr>
<td>Source</td>
<td>Unit of Randomization</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Stanton &amp; Clemens (1987)</td>
<td>Cluster of Families</td>
</tr>
<tr>
<td>Kidane &amp; Morrow (2002)</td>
<td>Cluster of villages</td>
</tr>
<tr>
<td>Thompson et al. (1997)</td>
<td>Physician Practice</td>
</tr>
<tr>
<td>Ray et al. (1997)</td>
<td>Nursing Home</td>
</tr>
<tr>
<td>Peterson et al (2002)</td>
<td>School district</td>
</tr>
<tr>
<td>Haggerty et al. (1994)</td>
<td>Community</td>
</tr>
<tr>
<td>Grosskurth et al. (1995)</td>
<td>Community</td>
</tr>
<tr>
<td>The COMMIT Research Group (1995)</td>
<td>Community</td>
</tr>
</tbody>
</table>
The overall conclusion from this research is that matching should be used cautiously in cluster randomization trials, particularly if the total number of matched pairs is small.

Other disadvantages arise because the effect of intervention is totally confounded with the natural variation between the clusters within each matched pair. Thus, the intracluster correlation cannot be directly estimated.

Klar and Donner (1997)
Some consequences of the inability to estimate from the matched pairs design are as follows:

- Information on necessary for the planning of future trials in the same application area will not be routinely available.

- Regression analyses whose aim is to explore the effect of individual-level covariates on outcome cannot be conducted using data from the matched pairs design.

- A test of heterogeneity in the effect of intervention across the matched pairs cannot be constructed.
These consequences do not occur for data arising from either the completely randomized design or the stratified design. This is because these designs allow some replication of clusters within a given combination of intervention and stratum.

For the matched pair design, there is no such replication. As a result, the inherent variation in response between clusters in a matched pair is totally confounded with the effect of intervention.

Thus it is impossible to obtain a valid estimate of for this design except under the null hypothesis of no intervention effect or without making special assumptions.
Features of the Stratified Design

- Provides some control in the design on factors known to be related to outcome. Imbalances on remaining variables can be adjusted for.

- Permits an estimate of $\rho$ to be computed without the need to make special assumptions, thus facilitating secondary regression analyses using individual-level covariates.
Decreasing the number of strata while increasing the number of clusters per stratum tends to reduce the closeness of the matching but increases the degrees of freedom available for the estimation of error.

When the number of pairs is fairly large (>20), it may be difficult to create matches that represent distinct rather than similar levels of risk across all strata. In this case, stratification with two or more clusters assigned to each treatment within strata may be preferable.
Conclusions

1. It is known that in small samples it is unlikely that effective matching is possible. This is largely because the required size of the correlation due to matching is difficult to achieve in practice.

2. In large samples, the pair-matched design may still have some practical and theoretical drawbacks as compared to the stratified design. This will be particularly true if prior information on potentially important matching variables is limited or it is difficult to obtain close matches on these variables.
Analysis Issues:

What methods should be used to compare the proportion of successes over all individuals in an intervention and control group when clusters are the unit of randomization?

Example:

Data obtained from a study evaluating the effect of school-based interventions in reducing adolescent tobacco use.
12 schools were randomly assigned to each of four groups, including three intervention groups and a control group (existing curriculum).

For each school, the number of students engaged in different types of smoking behavior was recorded at various follow-up times.

We compare here the effect of the SFG (Smoke Free Generation) intervention to the existing curriculum (EC) with respect to reducing the proportion of children who report using smokeless tobacco after four years of follow-up.

Donner and Klar (1994)
## DATA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/55</td>
<td>23/225</td>
<td>16/125</td>
<td>12/207</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXPERIMENTAL (SFG)</th>
<th>0/42</th>
<th>1/84</th>
<th>9/149</th>
<th>11/136</th>
<th>4/58</th>
<th>1/55</th>
<th>10/219</th>
<th>4/160</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2/63</td>
<td>5/85</td>
<td>1/96</td>
<td>10/194</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### OVERALL RATES OF TOBACCO USE

Control: \( \frac{91}{1479} = 0.062 \)

Experimental: \( \frac{58}{1341} = 0.043 \)
Procedures Reviewed

- Standard chi-square test
- Two sample t-test on school-specific event rates
- Wilcoxon rank-sum test
- Direct adjustment of standard chi-square test
- Ratio estimator chi-square test
- Likelihood ratio test
- Generalized estimating equations approach
RATES OF TOBACCO USE

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Experimental Group</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMOKERS</td>
<td>91 (.062)</td>
<td>58 (.043)</td>
<td>149 (.053)</td>
</tr>
<tr>
<td>NON-SMOKERS</td>
<td>1388</td>
<td>1283</td>
<td>2671</td>
</tr>
<tr>
<td>TOTALS</td>
<td>1479</td>
<td>1341</td>
<td>2820</td>
</tr>
</tbody>
</table>

Are the overall event rates the same in the two groups?
The value of Pearson’s chi-square statistic, with one degree of freedom, is given by

\[ \chi^2 = \sum (O - E)^2 / E \]

But this test is invalid, since it ignores the within-cluster correlation.
## TWO SAMPLE t-TEST

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Mean Event Rate</td>
<td>.060</td>
<td>.039</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.035</td>
<td>.026</td>
</tr>
</tbody>
</table>

Pooled variance =
COMMENT:

Assumptions required (normality, homogeneity of variance) not strictly satisfied, although robustness of t-test assures results will be accurate to a reasonable approximation.
WILCOXON RANK-SUM TEST

Nonparametric test on the school-specific rates based on ranks; makes no distributional assumptions.

Procedure requires that the two samples be pooled and the success rates ranked by size.

If the two samples come from the same population, sum of the ranks in each sample should be the same.
RESULT:

Exact p-value (StatXact: 0.11)

COMMENTS:

1. Valid, but lacks power.

2. Alternative nonparametric approach would be to apply a two-sample permutation test, also available in StatXact.
ADJUSTED CHI-SQUARE TEST

Compute clustering correction factors for each group given by:

\[ \frac{1}{\hat{\rho}_{11}} \sum \sum - \frac{1}{\hat{\rho}_{21}} \sum \sum = \frac{1}{\hat{\rho}_{11}} - \frac{1}{\hat{\rho}_{21}} \]

where \( \hat{\rho}_{11} \) and \( \hat{\rho}_{21} \) are the cluster sizes within groups 1 and 2, respectively, and \( \hat{\rho}_{11} \) is a pooled estimate of intracluster correlation in the two groups.
Then the adjusted one degree of freedom chi-square statistic is given by:
COMMENTS:

1. Reduces to standard Pearson chi-square test when

2. Does not assume any two observations in a given cluster have the same correlation; only that the average correlation is the same from cluster to cluster.

3. Assumes that $\hat{\rho}$ and $\rho$ are not significantly different, i.e., estimates the same population design effect.
Let overall proportion of successes in sample

estimated binomial variance

estimated variance regarding as a ratio
estimated “design effect” in sample

Define

Then the one degree of freedom ratio estimator chi-square (Rao and Scott [1992]) is given by:
RESULT:

COMMENTS:

1. Does not involve estimation of the intracluster correlation coefficient.

2. Requires a large number of clusters per group

3. Best suited to observational comparisons.
Likelihood Ratio Test Based on Parametric Modelling

Several parametric extensions of standard logistic regression have been developed.

For inferences concerning the effect of intervention these models impose distributional assumptions on the cluster-specific event rates.

As an example, consider the logistic-normal model given by

\[ \log \frac{\rho_{ij}}{\rho} = \beta_{ij} + \epsilon \]

where \( \rho_{ij} \) and \( \rho \) are the event rates in the intervention and control groups, respectively.
where

The odds ratio for the effect of intervention is given by

Parameters may be estimated using the method of maximum likelihood, leading to a likelihood ratio test of $H_0 : \beta_1 = 0$. 
The resulting chi-square statistic, with one degree of freedom, for testing is given by

**COMMENTS:**

1. The logistic-normal likelihood ratio test statistic can be obtained using the statistical packages SAS or STATA.

2. The logistic-normal model can be extended to model dependence on individual-level covariates.
3. A limitation of the logistic-normal model is that, as a “cluster-specific” model, the resulting estimates of intervention effect may not be simple to interpret (see Neuhaus, 1992).

4. A general weakness of the likelihood ratio test is that it requires a large number of clusters, and may not be robust to departures from the underlying model.
Generalized Estimating Equations Approach

- The generalized estimating equations (GEE) procedure, developed by Liang and Zeger (1986), as a “population-averaged” approach, can be used to construct an extension of standard logistic regression which adjusts for the effect of clustering, and does not require parametric assumptions.

- Consider the logistic-regression model given by
The odds ratio for the effect of intervention is given by

Two distinct strategies are available to calculate

— Model based, requiring correct specification of a correlation matrix which describes the pattern of correlation between responses of cluster members.

— Robust variance estimation, using a working correlation matrix and implemented using between-cluster information.
RESULT:

The resulting robust one degree of freedom chi-square statistic, constructed using a working exchangeable correlation matrix is given by
COMMENTS:

1. The GEE extensions of logistic regression were fit using the statistical package SAS (also available in STATA, SUDAAN).

2. Robust statistical inferences are valid so long as there are large number of clusters (e.g. ≥40) even if the working correlation is not correctly specified. Small sample adjustments have been proposed by Pan and Wall (2002) and Bell and McCaffrey (2002).
3. Can be extended to model dependence on individual level covariates.

4. Does not involve estimation of the intracluster correlation coefficient.

5. Odds ratios obtained using a logistic-normal model will be equivalent to those obtained using GEE under $H_0$ or when the probability of a positive response is small.
## SUMMARY OF RESULTS

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Test Statistic</th>
<th>P</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard chi-square test for comparing proportions</td>
<td></td>
<td>0.03</td>
<td>Biased in the presence of clustering</td>
</tr>
<tr>
<td>Two-sample t-test</td>
<td></td>
<td>0.11</td>
<td>Required assumptions not strictly satisfied</td>
</tr>
<tr>
<td>Wilcoxon Rank-sum Test</td>
<td></td>
<td>0.16</td>
<td>Valid, but lacks power</td>
</tr>
<tr>
<td>Direct adjustment of standard chi-square test</td>
<td></td>
<td>0.18</td>
<td>Reduces to standard chi-square test at ; most suited to randomized studies</td>
</tr>
</tbody>
</table>
### SUMMARY OF RESULTS - *continued*

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Test Statistic</th>
<th>P</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio estimate chi-square test</td>
<td></td>
<td>0.15</td>
<td>Most suited to non-randomized studies; requires large numbers of clusters</td>
</tr>
<tr>
<td>Likelihood ratio test based on logistic normal model</td>
<td></td>
<td>0.16</td>
<td>“Cluster-specific” approach; requires large number of clusters</td>
</tr>
<tr>
<td>Generalized estimating equations</td>
<td></td>
<td>0.11</td>
<td>“Population-averaged” approach; requires large numbers of clusters</td>
</tr>
</tbody>
</table>
Issues Arising in Covariate Adjustment

- Random assignment ensures, on average, that the distribution of baseline covariates will be balanced across intervention groups. However in any one trial some substantively important residual imbalance may still occur by chance.
• Such imbalance may also arise in trials randomizing individuals. However for a given total number of individuals the probability of an important imbalance will be higher in a trial randomizing clusters, owing to the smaller effective sample size.

• Note that adjustment for individual-level covariates such as age and gender might improve precision in the trial even though these variables are well-balanced across intervention groups.
# Unadjusted and Adjusted Inferences for the Effect of Intervention on Smokeless Tobacco Use

<table>
<thead>
<tr>
<th></th>
<th>UNADJUSTED</th>
<th>ADJUSTED FOR AGE AND SEX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>95% Confidence Interval</td>
</tr>
<tr>
<td>Effect of intervention (SFG vs. EC)</td>
<td>0.67</td>
<td>(0.41, 1.09)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.24</td>
<td>(0.60, 1.09)</td>
</tr>
<tr>
<td>Gender (Male vs. Female)</td>
<td>31.86</td>
<td>(14.51, 79.00)</td>
</tr>
</tbody>
</table>
Association at the Individual and Cluster Levels

A complicating feature is that there may be different associations at the individual and cluster levels. For example, a child’s decision to use smokeless tobacco might not only be influenced by his or her gender but also by the overall proportion of boys in the school.

The possibility of distinct individual-level and cluster-level gender associations can be explored by including both a child’s gender and the proportion of boys per school in the same model, (e.g., Neuhaus and Kalbfleisch [1998]).
Using data from the school-based trial reported by Murray et al. (1992), the observed associations with smokeless tobacco use are qualitatively similar whether or not one allows for different associations at the individual or cluster level.
Choice of Regression Model

• Two modeling approaches widely used for regression analysis of correlated outcome binary data in cluster randomization trials are that of generalized estimating equations (GEE) and logistic normal regression.

• The GEE approach may be characterized as “population-averaged” in that it measures the expected (marginal) change in a response as the value of the covariate increases by one unit. With one observation per cluster, it reduces to standard multiple logistic regression.
• The logistic-normal model may be characterized as “cluster-specific” or conditional in that it measures the expected change in response within a cluster as the value of a covariate increases by one unit.

• Both approaches estimate the same population parameter when the outcome is normally distributed. However this equivalence disappears in the case of a binary outcome variable.

Under what circumstances should models which provide marginal or conditional measures of covariate effects be preferred?

Omar and Thompson (2000)
Neuhaus (1992)
Stopping Rules for Interim Analyses

- Efficacy monitoring in accordance with a predetermined plan, unlike the case for most large-scale clinical trials, is not a common feature of most cluster randomization trials.

- This is at least in part because:
  (i) Standard group sequential stopping plans (e.g., O’Brien – Fleming) assume individual randomization.
  (ii) Ethical need to terminate a trial prematurely due to unexpected early benefits may be perceived to be slight.
• However issues of both safety and efficacy would seem to imply that interim analyses are ethically imperative in some trials (e.g., where mortality is the endpoint).

• Recent research has shown that standard group sequential stopping plans can be safely applied to the monitoring of cluster randomization trials.

Zou, Donner and Klar (2005)
Issues in the Meta-Analysis of Cluster Randomization Trials

Meta-analysis combining cluster randomization trials are being increasingly reported. These analyses raise methodologic issues beyond those raised by meta-analyses which include individually randomized trials:

- Increased Trial Heterogeneity.
- Difficulties in Estimating Design Effects from Individual Trials.
- Special Issues Involving Publication Bias.
- Choice of Statistical Method.
- Assessment of Trial Quality.

Donner and Klar (2002)
Donner, Piaggio and Villar (2001)
Reporting of Cluster Randomization Trials

Reporting of Study Design

- State clearly the justification for employing cluster randomization.
- Provide clear rationale for the selection of any matching or stratification factors.
- Clarify the choice of inferential unit.
- Describe in detail the content of the interventions.
- Describe the clusters that meet the eligibility criteria for the trial, but declined to participate.

- State the methods used to obtain informed consent.

- Describe clearly how the trial sample size was determined.

- Describe the method of randomization in context of the selected design.

- Where relevant, discuss the steps taken to minimize the risk of contamination and loss of follow-up.
Reporting of Study Results

- Include a table showing baseline characteristics by intervention group, separately for individual-level and cluster-level characteristics.

- Avoid the use of significance-tests in comparing baseline characteristics.

- Report the observed distribution of cluster sizes (median, range)

- Report on loss to follow-up at both the individual level and the cluster level
Use caution in reporting standard deviations of individual baseline characteristics.

Provide a clear and explicit description of the method used to account for between-cluster variation and for the influence of baseline prognostic factors.

Report empirical estimates of

EXTENSION OF CONSORT STATEMENT TO CLUSTER RANDOMIZATION TRIALS (CAMPBELL ET AL, BMJ, 2004)

- Includes a checklist of items that should be included in a trial report

- Extends CONSORT statement for reporting of individually randomized trials (Begg et al, JAMA, 1996)

- Key Points
  - Rationale for adopting cluster design
  - Incorporation of clustering into sample size estimation and analysis
  - Chart showing flow of clusters through the trial, from assignment to analysis
REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Torgerson, D.J. (2001), Contamination in trials: is cluster randomisation the answer?, *British Medical Journal*, 322, 355-357.
REFERENCES


