Cluster randomized trials in Epidemiology

Aspects of planning, field work, analysis, examples

Volker Arndt, M.D., M.P.H.
German Cancer Research Center
Ringvorlesung 08.11.2011
Learning objectives

• To distinguish experimental trials based on the unit of randomization (e.g. individual, family, community).

• To appreciate the consequences of cluster randomization on sample size estimation and data analysis.

• To identify potential sources of biases in cluster randomization trials.
Cluster Randomized Trials

... are clinical trials (experiments) in which **social units** or **clusters of individuals** rather than independent individuals are randomly allocated to intervention groups.
Examples of Cluster Randomized Trials

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Social unit / Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass education programs</td>
<td>Communities</td>
</tr>
<tr>
<td>Medical intervention</td>
<td>Clinics or hospitals</td>
</tr>
<tr>
<td>Smoking prevention programs</td>
<td>Schools</td>
</tr>
<tr>
<td>Dietary interventions</td>
<td>Families</td>
</tr>
</tbody>
</table>

**Note**: Repeated measures in longitudinal analysis or multiple measures on the same subject (teeth or paired organs) also represent „clusters“
Evolution of cluster randomized trials

Bland 2004/Donner 2008
Reasons for Cluster Randomization

- Administrative convenience
- To obtain cooperation of investigators
- Ethical considerations
- To enhance subject compliance
- To avoid treatment group contamination
- Intervention is naturally applied at the cluster level
Example 1:
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. (Lancet, 1986)
Example 2:
Study Purpose: To evaluate the effectiveness of treated nasal tissues versus standard tissues.

- 98 families were selected and randomized to one of the two intervention groups separately in each of three family size strata (2, 3, or 4 members per family).
- 24-week incidence of respiratory illness were compared in the two group

Farr et al. (Am. J. Epid., 1988)
Example 3:
Study Purpose: To evaluate the effectiveness of a breastfeeding promotion strategy.

- One member of each pair of 17 matched maternity hospitals in Belarus was randomly assigned to receive a breastfeeding promotion strategy, with the other member of the pair receiving a control condition based on usual practice (PROBIT trial).
- The rate of breastfeeding at 12 months was compared between the two groups.

Kramer et al. (JAMA, 2001)
Example 4:
Study Purpose: To evaluate the effectiveness of a structured group education program in people with newly diagnosed type 2 diabetes.

- 207 general practices were randomized to receive either a structured group education program or standard care offered to patients with newly diagnosed type 2 diabetes.
- A variety of response variables, including biomedical, lifestyle, and psychosocial measurements were collected over a one-year follow-up period.

Davies et al. (BMJ, 2008).
Example 5:
Study purpose: To promote smoking cessation using community resources.

- 11 paired communities were selected and one member of each pair was randomly assigned to the intervention group.
- 5-year smoking cessation rates were compared in the two groups.
- Communities were matched on demographic characteristics (e.g. size, population density) and geographical proximity.

Unit of Randomization vs. Unit of Analysis

- Theory of experimental design assumes that experimental unit which is randomized is also the unit of analysis.
- Inferences are frequently intended to apply at the individual level, while randomization is applied at the cluster level.
  - Individuals in the same cluster are assumed to be not independent (e.g. more similar)
  => Between-cluster variation
Potential Sources for Between-Cluster Variation

• Subjects frequently select the clusters to which they belong
  – e.g. Patient characteristics could be related to age and gender of physician

• 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

• Individuals within clusters frequently interact and, as a result, may respond similarly
  – e.g. Education strategies or therapies provided in a group setting could lead to sharing information or predispositions

• Tendency of infectious diseases to spread more rapidly within than among families or communities.
Impact of Between-cluster Variation on Design and Analysis

1. Application of standard sample size formulas will lead to underpowered studies

   Larger sample size

2. Application of standard statistical methods will tend to bias p-values downward risking a spurious claim of statistical significance.

   Sophisticated statistical methods

=> Reduction in effective sample size
Reduction in effective sample size

- Extent depends on degree of within-cluster correlation and on average cluster size.

- Within-cluster correlation (intracluster resemblance) typically measured by the magnitude of the intraclass correlation coefficient $\rho$
Intraclass correlation coefficient

- Measure of the degree of similarity among responses within a cluster
- May be interpreted as the standard Pearson correlation coefficient between any two responses in the same cluster.

\[ \rho \approx \frac{\sigma^2_B}{\sigma^2_B + \sigma^2_W} \]

- \( \sigma^2_B \) = variance component between clusters
- \( \sigma^2_W \) = variance component within clusters.
Measuring Clustering Effect

For sample size determination, "design effect" (variance inflation factor) is defined as:

\[
DE = 1 + (m - 1) \rho
\]

where \( \rho \) is the intraclass correlation coefficient and \( m \) is the cluster size.

It gives a measure of how much the **sample size in each group** has to be increased to achieve the same statistical power as would be obtained by individual level randomization.

When \( \rho = 0 \Rightarrow DE = 1 \Rightarrow \) the responses within clusters are independent.
Example

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

\[ \sigma^2_W = 2209, \sigma^2_B = 93 \]

\[ \rho = \sigma^2_B / (\sigma^2_B + \sigma^2_W) = 93 / (93 + 2209) = 0.04 \]

Assuming \( m = 70 \) subjects/worksite,

\[ DE = 1 + (m-1) \rho = 1 + (70 - 1) 0.04 = 3.76 \]
Are „small“ intracluster correlation coefficients negligible?

• Values of $\rho$ in practice tend to be small and positive:

  0.01 to 0.05 in primary care setting
  ~ 0.001 in trials randomizing intact communities (Campbell et al., 2000).

• However, dismissing small $\rho$ values as negligible can be seriously misleading, since the impact of clustering depends not only on the magnitude of $\rho$ but also on the sizes of the clusters enrolled in the trial.
Example

School-based trial where past experience suggests that $\rho$ is likely to be about 0.01.

The investigators decides to randomize schools of size $m=100$ to each of two intervention groups

$\Rightarrow \ DE = 1 + (m-1)\rho = \sim 2.0$

$\Rightarrow$ Variances of the resulting means and proportions will be underestimated by as much as 50%
General Issues in Sample Size Estimation

• Usual approach:
  – Calculate „standard sample size“ as for a non-clustered study
  – Multiply „standard sample size“ by VIF (DE)

• Varying cluster size:
  see Eldrigde et al. (Int J Epidemiol 2006)
Impact on Power of Increasing the Number of Clusters vs. Increasing Cluster Size:

Let \( d \) = mean difference between intervention groups and \( k \) the number of clusters then,

\[
\text{Var} (d) = \left( \frac{2\sigma^2}{km} \right) \left[ 1 + (m - 1) \rho \right]
\]

As the number of clusters \( k \to \infty \), \( \text{Var}(d) \to 0 \) but, as the cluster size size \( m \to \infty \), \( \text{Var} (d) \to \frac{2\sigma^2 \rho}{k} = \frac{2\sigma^2_B}{k} \)

Trial randomizing between 30 and 50 individuals will tend to have almost the same statistical power as trials randomizing the same number of much larger units. But clusters of larger size are often recruited for very practical reasons (to reduce contamination, to avoid logistic or ethical problems, etc.).

(Donner & Klar 2004)
Other Factors Influencing Power

1. Entire clusters, rather than just individuals may be lost to follow-up.
2. Interventions often applied on a group basis with little or no attention given to individual study participants.
3. Over-optimistic expectations regarding effect size.
Strategies for Improving Precision in Cluster Randomization Trials

1. Reduce between-cluster variability e.g., geographical restrictions.
2. Consider increasing the number of clusters randomized
3. Consider matching or stratifying in the design by baseline variable having prognostic importance.
4. Obtain baseline measurements on other potentially important prognostic variables.
5. Develop a detailed protocol for ensuring compliance and minimizing loss to follow-up.
Common Designs in Cluster Randomized Trials

- **Completely randomized**: intervention allocated at random to clusters.
  - Suitable when randomizing a fairly large number of clusters.

- **Matched pairs**: clusters are paired and the two clusters within each pair are allocated at random to the interventions.
  - Advantage: provides very tight and explicit balancing of potentially important prognostic factors at baseline.

- **Stratified**: clusters are grouped in homogenous strata and then they are allocated at random to interventions.
CRTs and Informed Consent

• Need and norms to obtain informed consent is discussed controversially
  – Great diversity in size and nature of units of randomization

• In theory, consent can be obtained at multiple levels:
  – Individual => rarely feasible
  – Gate-keeper (physician, mayor, school principal): safeguarding the interest of study participants.

• Data protection regulations?
Analysis: Incorporate clustering effects into standard statistical analysis

1. Analysis using summary measures on the cluster level, (such as mean, proportion)
   - Unit of randomization = unit of analysis
   - Loss of precision, no conclusion regarding individual effects
2. Use of robust standard errors,
   - e.g. Adjusted Chi-square statistics (Χ²/VIF)
3. Random effects (multilevel, hierarchical, mixed) models
   - Include a random cluster-level effect to allow the average response to vary between clusters
4. Generalized estimating equations (GEE)
   - GEE specify mean and variance of outcome rather than a full probability distribution
Bias

- Although inappropriate analysis of cluster trials appears to be very common (up to 40%), more concern is warranted regarding potential sources of bias pertinent to cluster randomized trials

- Inappropriate analysis $\Rightarrow$ misleading precision (note: the effect size estimate is still valid if randomization had been successful)

- Bias $\Rightarrow$ misleading effect size estimate
Selection bias

- Unbiased results rely on the prerequisite that study participants represent a random sample of all eligible participants
- Randomization aims to eliminate confounding and selection bias
- Selection bias is reintroduced within any trial in case of high loss to follow-up or failure to use intention to treat analysis.
Recruitment and Dilution Bias

• In CRTs, participant recruitment after randomization can introduce bias:

• Recruitment bias (Detect and treatment bias)
  – Foreknowledge of allocation: (Opportunistic) recruitment by unblinded personnel in medical practices
    • Motivation to recruit may depend on intervention arm
  – Motivation of subjects to participate may differ on a prior knowledge

• Dilution bias: Quality of intervention (adherence) depends on motivation (may apply to provider as well as to patient)
Strategies to reduce selection bias

• Recruitment by an independent person rather than by the physician who is providing the intervention

• Prior identification of participants e.g. in schools, workplace

• Motivation of subjects

• Intention to treat analysis
Hip protectors and prevention of fractures: Different results in individually and cluster randomized trials

Possible explanations
- Publication bias
- Different settings
- Intervention may work better using a clustered design
- Selection bias or poor randomization in CRTs

![Graph showing sample size vs. RR for individually and cluster randomised trials.](image)
Is fear of contamination in individually randomized trials overrated?

- „Fear of contamination“ most common stated reason for adopting cluster randomization
- 30% contamination (R) in a trial to detect a 20% difference in event rates requires a effective sample size to detect a difference of 14% => $1/(1-R)^2 = 204\%$ of original sample size.
- However, VIF associated with clustering (DE) may be much larger
- But, CRT may be preferrable to estimate the „uncontaminated“ effect size.
Conclusion

• Cluster randomized trials may pose a very promising alternative in certain settings
  – Conclusion may differ between cluster randomized and individually randomized trials
• Cluster effect has to be taken into account in the planning and analysis
• Precautions are necessary to avoid selection bias
References & Links


References & Links


