CHL 5225H
Advanced Statistical Methods for Clinical Trials: Cluster Trials

Prof. Kevin E. Thorpe

Dalla Lana School of Public Health
University of Toronto

What are Cluster Randomized Trials?

Cluster randomization trials are experiments in which intact social units or clusters of individuals rather than independent individuals are randomly allocated to interventions being studied.

Some Reasons for Cluster Randomized Trials

Examples

- Communities selected as the experimental unit in trials evaluating mass education programs.
- Primary care clinics as the experimental unit in trials evaluating lay health worker programmes.
- Wards selected as the experimental unit in trials evaluating inter-professional collaboration initiatives.

- 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.

Complications Arising from Clustering

- Although there are many situations where cluster trials are advised, in fact it may be the only feasible option to evaluate an intervention, complications are introduced.
  - Multiple levels of inference are possible.
  - Unit of randomization may differ from the unit of inference.
- These have implications for both trial design and analysis.

Multiple Levels of Inference

- In cluster trials, multiple levels of inference are possible while decisions regarding the appropriate unit of inference and the appropriate unit of analysis can be difficult in some trials.

Example

  - If schools are the unit of randomization, analyses could be carried out at the individual, classroom, or school level.
  - However, analyses conducted at the individual, or classroom level need to account for dependencies of responses for children from the same school.

- Note that the levels of inference as described in cluster trials are a different concept than the marginal (population) versus conditional (subject specific) model distinctions in the mixed-effects models literature.

Example

- A trial randomizing hospitals designed to assess the effect of obtaining a second clinical opinion on the decision to proceed with a Caesarian section operation.
- The target of the intervention may naturally be the hospital rate of Caesarian section, i.e., the trial may be intended to evaluate policy at the hospital level only.
- In this case, the hospital is the natural unit of inference and standard methods of sample size estimation and analysis would apply at the cluster level.

Unit of Randomization vs. Unit of Analysis

- Inferences are frequently intended to apply at the individual level but randomization is at the cluster or group level. Thus the unit of randomization is often different from the unit of analysis.
- Lack of independence among individuals in the same cluster (possibly), i.e., among-cluster variation, creates special methodologic challenges in both design and analysis.

Measuring the Effect of Clustering

- The effect of clustering is measured by the intra-class correlation coefficient (ICC) defined as

$$\rho = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_W^2}$$

  where σ_A^2 is the variance among the clusters and σ_W^2 is the variance within clusters.
- In a sense it measures how alike the clusters are.
- When $\rho = 0$, responses within the cluster are independent and clustering, effectively, has no effect.
- When $\rho = 1$, all members of each cluster respond the same, within cluster.

Reasons Clustering Can Have an Effect

- Subjects may select their own cluster. Subjects who select the same cluster are similar in at least one way (they selected the same cluster).
- Important covariates at the cluster level are likely to affect all individuals within the cluster in the same manner.
- Individuals within clusters may interact and, as a result, may respond similarly.

Implications of Clustering

- Reduction in effective sample size.
  - Extent depends on degree of within-cluster correlation and on average cluster size.
  - In particular, sample size for a cluster design needs to be increased by a factor of about $1 + (m - 1)\rho$, the "Variance Inflation Factor" or "Design Effect."
- Standard approaches for sample size estimation and analysis do not apply.
  - Application of standard sample size approaches leads to an underpowered study.
  - Application of standard statistical methods generally tend to bias p-values downwards (make them smaller) which could lead to spurious statistical significance.

Sample Size — Additional Considerations

- In addition to the usual necessities for sample size (significant difference, variance, control group event rate, etc.) estimates of average cluster size and the ICC are required.
- Problem: the ICC is even harder to come by than a variance estimate.

Power Considerations — Cluster Size versus Number of Clusters

- In general, there are greater power gains achieved by increasing the number of clusters compared to increasing the number of subjects per cluster.
- It is well worth the extra effort involved in finding more clusters compared to bigger clusters or a longer trial to achieve bigger clusters.

Power Considerations — Cluster Size versus Number of Clusters

Example

- An investigator wants to detect an improvement of response proportions from 0.1 to 0.15. They say they have 20 centres of size 20 per group available. The ICC is 0.01.
- A sample size calculation, assuming 80% power and $\alpha = 0.05$ gives 813 per group.
- With 20 centres of size 20, the power is 0.502.
- The investigator says lets recruit longer until 820 patients per group are recruited (41 per centre). Power is 0.737.
- If instead, the number of centres of size 20 is increased to 41, the power is 0.804.

Cluster Level Analysis

- If inference is to be cluster level, even if subject level data are collected, the data can be aggregated within the cluster (eg. cluster specific means or proportions).
- Then, standard statistical methods, such as t-tests or permutations tests, etc. can then be directly applied to the collapsed measures.
- The effect of clustering is effectively removed since subsequent significance tests and confidence intervals are based on variation between cluster summary values rather than on variation between individuals.
- Note that the sample size for this approach essentially becomes the number of clusters.

Multilevel Models

- When applicable, multilevel models offer several advantages over cluster-level analyses.
  - More naturally allows for modeling of random effects to obtain estimates of intra-cluster correlation used to design future studies.
  - Allows adjustment of the effect of intervention for (imbalance on) individual-level and cluster-level baseline predictors of outcome.
  - Simplifies examination of factors influencing the effect of intervention (i.e., interaction effects).

Types of Multilevel Models

- Continuous outcome.
  - This is usually a *linear mixed effect* model.
- Binary outcome.
  - Two types of models: *Generalized Estimating Equations (GEE)* and *Generalized Linear Mixed Models (GLMM)*.
  - The GEE is called a marginal model and estimates the average population level effects of the treatment and other factors. This is probably the model usually required. Typically, a compound symmetry correlation structure is used.
  - The GLMM is a conditional model and estimates the effect on "these" patients. This model could be considered if there are very few clusters.
  - For continuous data there is no difference between the marginal and conditional models with respect to the interpretation of the coefficients, but there is for binary data.

An Example

- The Clinical Problem
  - Elderly patients presenting to the hospital with minor fractures may not be appropriately screened or treated for oesteoporosis.
- The Question
  - Would the presence of an on-site oesteoporosis coordinator improve the screening and treatment for oesteoporosis?
- The Design
  - A cluster RCT where hospitals are randomized to either receive a coordinator, or not.

Some Results

- There were 37 hosptials with 269 patients. The *Intervention* group had 130 patients while the *Control* group had 139 patients.
- The outcome we will consider is whether or not patients received "appropriate" treatment. In *Intervention*, 45.4% of patients did while in *Control* 25.9% patients did.
- For the GEE analysis, $p = 0.00196$, whereas a standard χ^2 test gave $p = 0.000834$ (no CC) and $p = 0.00131$ (with CC).

GEE versus Logistic Regression

	β	se(β)	OR	p
GEE	0.849	0.274	2.337	0.00196
Logistic	0.866	0.262	2.378	0.000938

Reporting of Cluster Randomization Trials

- 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.

Reporting of Cluster Randomization Trials

- 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 Cluster Randomization Trials

- 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.
- Provide a clear and explicit description of the method used to account for clustering.
- Report empirical estimates of ρ.

Reference

Donner A. and Klar N., *Design and Analysis of Cluster Randomized Trials in Health Research*