Reviewing the ins and outs of conducting cluster randomized control trials in primary care research and evaluation

School of Nursing
UNIVERSITY OF WISCONSIN–MADISON

Roger L. Brown, Ph.D.
Schools of Nursing, Medicine and Public Health, University of Wisconsin

School of Medicine and Public Health
UNIVERSITY OF WISCONSIN–MADISON

Abstract

Although most clinical trials/evaluations in the primary care setting involve the randomization of the patient, many times individual patient allocation may not be possible or not even desirable, and groups of individuals are randomized instead. This is known as cluster randomization. Based on this randomization approach a common design has evolved, known as the cluster randomized control trial (CRCT). Specifically, the cluster randomized designs introduce dependence (or clustering) between individual units sampled. The evaluation of certain types of interventions (such as those used in health promotion and educational interventions) a cluster randomized trial is virtually the only valid approach. Advantages of cluster randomized controlled trials over individually-randomized controlled trials include the ability to study interventions that cannot be directed toward selected individuals (e.g., a clinic quality improvement program) and the ability to control for "contamination" across study individuals. Disadvantages compared with individually-randomized controlled trials include greater complexity in design and analysis, and may make it particularly vulnerable to a range of threats that can introduce bias. The effects of cluster can be large, inflating Type I error rates, and may not be obvious to researchers. In this poster I discuss the issues that can lead to bias in cluster randomized trials and conclude with some suggestions for avoiding these problems.

What is a CRCT?

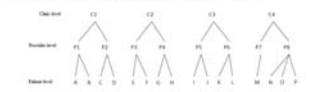
A Cluster Randomized Controlled Trial is a type of randomized controlled trial in which groups of subjects, as opposed to individual subjects are randomized (Bland, 2004).

Clustered Data Structures

Examples of clustering

– *Patients seen by a health care provider:*

– *Patients seen by provider in specific clinic:*



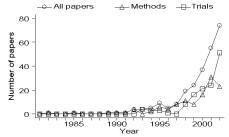
"Once you know hierarchies exist, you see them everywhere" (Kreft & de Leeuw, 1998).

Reasons for Adopting a Cluster Randomization

Need to minimize or remove contamination. Example: In a trial for the prevention of coronary heart disease, clinics were chosen as units of randomization to minimize the likelihood of subjects in different intervention groups sharing information concerning preventive advice on coronary risk factors.

Basic feasibility considerations. Example: Evaluate a program to assess the effectiveness of depression management. It was recognized that such a program would best function effectively if patients participated in a group session. Unit of randomization therapy group.

Only natural choice. Example: Intervention programs that use mass education. It is difficult to provide general recommendations concerning diet, smoking or exercise to some people and not to others in the same community.



Major issue with CRCTs

- **Analysis – aggregation bias (Ecological fallacy) , disaggregation bias**
- **Power Estimates – number of subjects per site and number of sites**
- **Randomization – Randomization type, Differential allocation ratios?**

Multilevel Analysis Model - Solution

$$Y_{ijk} = \beta_{0ijk} + \beta_1 X_{1ijk}$$

$$\beta_{0ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}$$

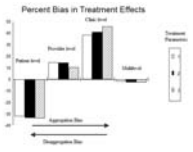
Intra-provider correlation = $\frac{\sigma_{u0}^2 + \sigma_{v0}^2}{\sigma_{u0}^2 + \sigma_{v0}^2 + \sigma_e^2}$

Intra-clinic correlation = $\frac{\sigma_{v0}^2}{\sigma_{u0}^2 + \sigma_{v0}^2 + \sigma_e^2}$

β_{0ijk} is the overall population intercept, which is allowed to vary across the levels,
X_1 is a covariate of interest (e.g., treatment versus control),
β_1 is the slope coefficient,
i level 1 (e.g., patient), j is level 2 (e.g., provider), and k is level 3 (e.g., clinic)
v_{0k} is the random effect at the clinic level, and allowed to vary as a departure from the grand mean,
u_{0jk} is the random effect at the provider level, a departure from the clinic effect,
e_{0ijk} is the random effect at the patient level, a departure from the provider effect within a clinic,
So:
Variance between clinics = $\mathrm{Var}(v_{0k}) = \sigma^2_{v0}$
Variance between providers within clinics = $\mathrm{Var}(u_{0jk}) = \sigma^2_{u0}$
Variance between patients within providers within clinics = $\mathrm{Var}(e_{0ijk}) = \sigma^2_e$
Variance between providers = $\sigma^2_{u0} - \sigma^2_{v0}$

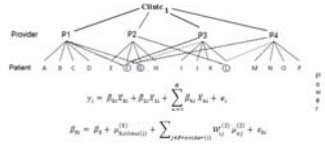
Analysis issues

Analyzed as single level



Brown, Baumann, and Cameron (1994)

Impure Clusters



$$y_i = \beta_{0i} X_{0i} + \beta_{1i} X_{1i} + \sum_{k=2}^{p} \beta_{ki} X_{ki} + e_i$$

$$\beta_{0i} = \beta_0 + \mu^{(3)}_{0,\mathrm{clinic}(i)} + \sum_{j \in \mathrm{Provider}(i)} W^{(3)}_{ij} \mu^{(2)}_{0j} + e_{0i}$$

$$y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \sum_{k=2}^{p} \beta_{ki} X_{ki} + \mu^{(3)}_{0,\mathrm{clinic}(i)} + \sum_{j \in \mathrm{Provider}(i)} W^{(3)}_{ij} \mu^{(2)}_{0j} + e_{0i}$$

Power Estimates issues

Power/Sample Size Estimates in CRCTs

power is the probability of rejecting the null hypothesis when it is false.

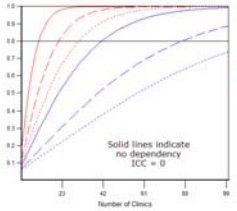
What happens with multilevel data?

We will here mainly consider 2-level models, so we have patients nested within clinics.

When deciding on a sampling scheme we have many choices:

- How many Clinics, N ?
- How many patients per clinic, n_j ?
- Should we collect the same size sample from each clinic?

Our decision will depend on which parameter we wish to estimate in the model.



Design and Randomization issues

Clustered Data Structures
Design Issues

Examples of pre-existing clusters
- – Patients clustered under physicians
- – Nurses nested in a hospital unit
- – Providers nested within clinics
- – Patients under providers nested within clinics
- – Repeated measurements clustered under patients

Examples of constructed cluster
- --- Groups put together for group treatment
- --- Other

The quest for comparable groups in trials

It has been known for centuries to properly evaluate something we need to compare groups that are similar and then expose one group to a treatment. In this way we can compare treatment effects. Without similar groups we cannot be sure any effects we see are treatment related.

Random
- Simple randomization
- Stratified randomization
- Paired randomization
- Differential allocation ratio randomization

Non-Random Methods
- Minimization

Conclusion

The purpose of this poster was to bring to the attention of researchers and trialists the benefits and limitations of using Cluster randomized control trials. The gains obtained in analytical power and precision from using multilevel models in CRCTs are not without a price. One disadvantage is an increase in the complexity of the models being analyzed. However, some may consider this an advantage, in that the researcher is forced to disentangle the complexity of the effects in the intervention model.