Sample Size Calculation for Cluster-Randomized Trials

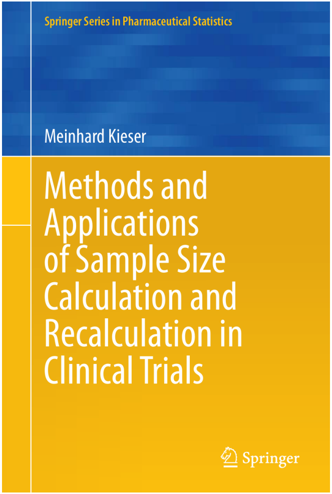
Chapter 13
of the book "Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials"

Alireza Ghorbani

Ludwig Maximilian University of Munich, Faculty of Medicine, IBE

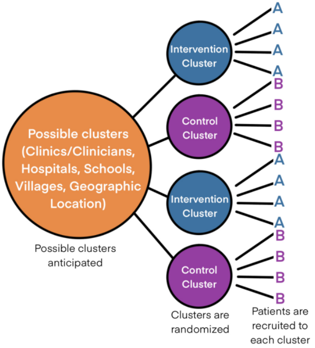
March 12, 2025

Overview


Springer Series in Pharmaceutical Statistics
Meinhard Kieser
Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials
Springer

Why Cluster Randomization?

- **Gold standard**: Individual randomization isn't always feasible
- **Common applications**:
  - Public health interventions (clinics, households)
  - e.g., water sanitation programs are often applied to entire communities.
- **Key challenge**: Individuals of the same cluster tend to be more similar



Key Concepts in Cluster Randomization

Model Structure

$$X_{ijk} = \mu_i + Y_{ik} + \epsilon_{ijk}$$

Where μ_i is the treatment effect, $Y_{ik} \sim N(0, \sigma_b^2)$ is the cluster effect and $\epsilon_{ijk} \sim N(0, \sigma_w^2)$ is the variation within each cluster.

Intra-Cluster Correlation (ICC)

$$\text{ICC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

- σ_b^2 = Between-cluster variance
- σ_w^2 = Within-cluster variance
- ICC quantifies how much outcomes depend on cluster membership

Design Effect (DE)

$$\text{DE} = 1 + (m - 1) \times \text{ICC}$$

- m = cluster size
- DE ≥ 1 always

Cluster-Level vs Individual-Level Analysis

Two Analytical Approaches for Sample Size Calculation:

Cluster-Level Analysis

- *(Aggregated Approach)*
- Uses cluster means
- Simple implementation
- No individual covariates

Individual-Level Analysis

- *(Non-aggregated Approach)*
- Uses all raw data
- Allows covariates
- Requires advanced methods:
  - **GEE models**
  - Mixed models
  - Multilevel modeling

Cluster-Level Analysis

Individual Randomization (non-clustered data)

$$n_{\text{indiv}} = \frac{(1+r)^2}{r} \times (z_{1-\alpha/2} + z_{1-\beta})^2 \times \frac{\sigma^2}{\Delta_A^2}$$

Cluster Randomization

$$n_{\text{cluster}} = \text{DE} \times n_{\text{indiv}}, \; K = \frac{n_{\text{cluster}}}{m}$$

Design Effect

$$\text{DE} = 1 + (m-1) \times \text{ICC}$$

Key Components:

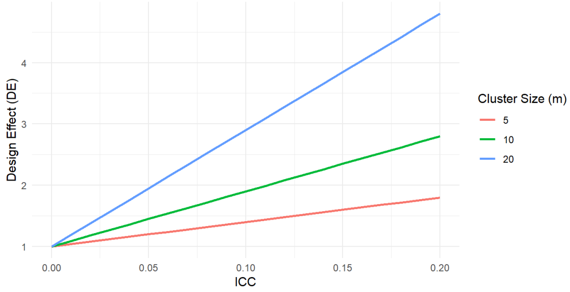
- $r = K_E/K_C$ (allocation ratio)
- σ^2 = total variance
- Δ_A = effect size
- m = cluster size
- K = Number of clusters

Variance of Clusters Means

$$Var(\bar{X}_{ik}) = \sigma_b^2 + \frac{\sigma_w^2}{m} \Rightarrow$$
$$\sigma^2 = \frac{m \times Var(\bar{X}_{ik})}{DE}$$

- $\bar{X}_{ik}$ = cluster mean

Design Effect & Intra-cluster Correlation Coefficient



Design Effect

$$DE = 1 + (m-1) \times ICC$$

When ICC $\to$ 0

- $\sigma_b^2 \approx 0$
- All variance is within clusters
- Design Effect DE $= 1$

Example

For ICC $= 0.02$ and $m = 10$: DE $= 1.18 \Rightarrow$ Requires 18% more subjects

ChroPac Trial Example

Table: Sample Size Requirements for Different Cluster Sizes

Cluster Size (m)	6	12	24
DE	1.05	1.11	1.23
Total Sample Size (n)	192	216	288
Clusters Needed (K)	32	18	12

Assumptions: ICC=0.01, α=0.05, Power=0.90, Δ_A=10, σ=20

Key Observations:

- Cluster size ↑ = Clusters needed ↓ but total subjects ↑
- DE grows with cluster size (1.05→1.23)
- Small ICC (0.01) still impacts requirements

Individual-Level Analysis with Generalized Estimating Equations

Key Insight (Shih 1997)

$$n_{\text{indiv}} = \frac{(1+r)^2}{r} \times (z_{1-\alpha/2} + z_{1-\beta})^2 \times \frac{\sigma^2}{\Delta_A^2}$$

$$n_{\text{GEE}} = \text{DE} \times n_{\text{indiv}}$$

- Same formula as cluster-level analysis
- $DE = 1 + (m-1) \times \text{ICC}$
- Requires same sample size increase

Practical Implication

GEE provides covariate flexibility while needing the same sample size adjustment as cluster-level methods

References:

- Hardin & Hilbe (2012)
- Shih (1997)

Handling Unequal Cluster Sizes

When to Adjust? [1]

- Original formulas assume equal cluster sizes (m)
- Adjustment needed if: $\frac{\mathrm{SD}(m_k)}{\bar{m}} > 0.23$ ($\bar{m}$ = mean cluster size)

- DE calculation method depends on available cluster size information:
  - Known sizes: Sample size for each cluster is known
  - Unknown sizes: Mean and standard deviation of cluster sample sizes are known

[1] Eldridge et al. (2006)

Handling Unequal Cluster Sizes

Known Cluster Sizes

$$\text{DE} = \frac{K \times \bar{m}}{\sum_{i=E,C} \sum_{k_i=1}^{K_i} \frac{m_{k_i}}{1 + (m_{k_i} - 1) \times \text{ICC}}}$$

Example:

- $m_k = \{5, 7, 6\}$, $\bar{m} = 6$
- $\text{ICC} = 0.02$
- $\text{DE} \approx 1.12$

Unknown Cluster Sizes

$$\text{DE} = 1 + \left[(\text{CV}^2 + 1) \times \bar{m} - 1\right] \times \text{ICC}$$

Where:

- $CV = \sigma_m / \mu_m$
- σ_m = SD of cluster sizes
- μ_m = mean cluster size

Other Outcome Types

General Rule

$n_{\text{cluster}} = \text{DE} \times n_{\text{indiv}}$, where $\text{DE} = 1 + (m-1) \times \text{ICC}$ applies to diverse outcome types when cluster sizes are constant.

Key Insight

The same DE multiplier applies across all outcome types, but analysis methods differ!
(Refer to original formulas for individual-level calculations first)

Key Takeaways

Core Principles

- **Design Effect Rule**: DE ≥ 1 (Always!)
    - ICC $> 0 \Rightarrow$ Larger sample
    - Cluster size (m) matters

Even ICC $= 0.01$ might requires larger sample (ChroPac example)

Critical Checks

- **Analysis choice**:
    - Cluster-level (simple)
    - Individual-level (flexible)
- **Cluster variability**: $\frac{\mathrm{SD}(m_k)}{\bar{m}} > 0.23$

Thank You for Your Attention

Questions?