Analysis and sample size calculation with clustering effects in individually randomized trials.

Xingyan “David” Wang
Candidate for Master of Biostatistics
Department of Biostatistics and Bioinformatics,
Duke University
xingyan.wang@duke.edu

April 14, 2017
Outline

• Definition of trials

• Clustering Effects in RCT
## Definition of trials

<table>
<thead>
<tr>
<th>Trial Type</th>
<th>Before randomization</th>
<th>After randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRGT</td>
<td></td>
<td>😊</td>
</tr>
<tr>
<td>GRT</td>
<td>😊</td>
<td>😊</td>
</tr>
</tbody>
</table>

- Individually Randomized Clinical Trials (RCTs)
- Individually Randomized Group Treatment Trials (IRGTs)
- Group-Randomized Trials (GRTs)
Clustering effect in RCT

Definition of clustering

1. Observations are grouped together based upon common attributes;
   multicenter trials: Patients are grouped together within centers
   trails where the intervention is a type of surgery or therapy and patients are grouped together by surgeon.

2. Some non-standard situations;
   Examples: baseline factors—age, sex
   1) Pre-randomization
   Patients are grouped into clusters and then randomized.
   2) Post-randomization
   Patients are randomized and then assigned to clusters.
   Notes: Compared to IRGT & GRT
   Pre-randomization – GRT
   Post-randomization – IRGT
• Non-ignorable Clustering

Theoretical Definition

\[ Var(\text{trt effect}) = V_0 + V_E \]

Where:
1. \( V_0 \): Variance of treatment effect when clustering is not present;
2. \( V_E \): Additional variance/factor based on clustering.

• Non-ignorable clustering means \( V_E \neq 0 \);

• \( V_E \) is a function of:
  1. the correlation between outcomes for patients in the same cluster
     → Generally speaking: ICC (Intraclass correlation coefficient);
  2. Correlation between treatment assignments for patients in the same cluster.

• Non-ignorable clustering → both 1 and 2 are not zero
Notes on $V_E$:

\[ Var(\text{trt effect}) = V_0 + V_E \]

1. $V_E$ is a function of:
   1. Correlation between outcomes for patients in the same cluster $\rightarrow$ ICC
   2. Correlation between treatment assignments for patients in the same cluster.

   Reason 1: Patients with similar characteristics may be more likely to present to the same cluster;
   
   Reason 2: Clusters themselves exert some influence on outcome.

   - Correlation between treatment assignments for patients in the same cluster.
     Reason: Patients in certain clusters are more likely to be in a certain treatment group
• **Statistical Methods**

**Scenario 1:**
Both treatment arm & Control arm are group treatment

\[ y_{ij} = \mu + I_T \delta + x_{ij} \beta + u_i + e_{ij} \]

Where:
1. \( y_{ij} \): Continuous outcome of the \( j \)th patients in the \( i \)th clusters;
2. \( I_T \): Indicator function of treatment group;
3. \( \delta \): Treatment effect;
4. \( x_{ij}, \beta \): Baseline covariates and their coefficients;
5. \( u_i \sim N(0, \sigma_u^2) \): Between cluster variation;
6. \( e_{ij} \sim N(0, \sigma_e^2) \): Patient level error term.
\[ y_{ij} = \mu + I_T \delta + x_{ij} \beta + u_i + e_{ij} \]

\[ y_{ij} = \mu + \delta + x_{ij} \beta + u_i + e_{ij} \]

\[ y_{ij} = \mu + x_{ij} \beta + u_i + e_{ij} \]

\[ \text{Treatment} \quad \text{Control} \]

- **ICC:** \[ \rho = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_e} \]
- Disadvantage: between-treatment heteroscedasticity
  1. Can happen both between and within clusters;
  2. Example:
     - Standardize one type of therapist \( \rightarrow \) reduce between cluster variance since therapists become mutually more consistent. – between cluster
     - Standardize one type of therapist \( \rightarrow \) reduce variation in outcome of patients treated by the same therapist. – within cluster
\[ y_{ij} = \mu + I_T \delta + x_{ij} \beta + u_i + e_{ij} \]

Where:
1. \( V_i \sim N(0, \sigma_v^2) \): Random coefficients at cluster level;
2. \( \xi_{ij} \sim N(0, \sigma_\xi^2) \): Random coefficients at individual level

\[ \rho = \frac{\sigma_u^2 + \sigma_v^2}{\sigma_u^2 + \sigma_v^2 + \sigma_e^2 + \sigma_\xi^2} \]

Duke Medicine
Scenario 2:
Treatment arm is group treatment, while control arm is individual treatment \(\rightarrow\) random coefficient model

\[
y_{ij} = \mu + I_T \delta + x_{ij} \beta + I_T u_i + e_{ij}
\]

Treatment: Group Therapy
\[
y_{ij} = \mu + \delta + x_{ij} \beta + u_i + e_{ij}
\]

Control: Individual Therapy
\[
y_{ij} = \mu + x_{ij} \beta + e_{ij}
\]

\[
\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}
\]

\(\rho\) = standard deviation of treatment effect
• Sample Size Calculation

- Assumptions:
  1) All clusters in each treatment arm are of equal size; \( m_1, m_2 \)
  2) Both treatment groups are of equal size with equal variance in each arm and intracluster correlation coefficients (ICC); \( \sigma, \rho_1, \rho_2 \)

- Estimation of standard error of the treatment effect

\[
\sigma_{pooled} = \sigma \sqrt{\frac{(1 + (m_1 - 1)\rho_1)}{N_T R} + \frac{(1 + (m_2 - 1)\rho_2)}{N_T / R}}
\]
A large sample formula for computing the power

\[ z_\beta = \frac{\mu_1 - \mu_2}{\sigma_{pooled}} - z_{\alpha/2} \]

\[ \sigma_{pooled} = \sigma \sqrt{\frac{(1 + (m_1 - 1)\rho_1) + (1 + (m_2 - 1)\rho_2)}{N_TR + N_T/R}} \]

We can maximize power by finding a propitiate allocation ratio: \( R \)

\[ R = \sqrt{\frac{1 + (m_1 - 1)\rho_1}{1 + (m_2 - 1)\rho_2}} \]
Sample size

\[ z_\beta = \frac{\mu_1 - \mu_2}{\sigma_{\text{pooled}}} - z_{\alpha/2} \]

\[ \sigma_{\text{pooled}} = \sigma \sqrt{\frac{(1 + (m_1 - 1)\rho_1)}{N_T R} + \frac{(1 + (m_2 - 1)\rho_2)}{N_T/R}} \]

\[ R = \frac{\sqrt{1 + (m_1 - 1)\rho_1}}{\sqrt{1 + (m_2 - 1)\rho_2}} \]

Then we can solve the sample size \( N_T \):

\[ N_T = \sigma^2 \frac{\left( z_\beta + z_\alpha \right)^2}{(\mu_1 - \mu_2)^2} \left( \frac{1 + (m_1 - 1)\rho_1}{R} + \frac{1 + (m_2 - 1)\rho_2}{1/R} \right) \]
\[ N_T = \sigma^2 \left( \frac{z_\beta + z_\alpha}{2} \right)^2 \frac{(1 + (m_1 - 1)\rho_1)}{R} + \frac{(1 + (m_2 - 1)\rho_2)}{1/R} \]

Compared with standard two-arm sample size calculation:

\[ N = \frac{2\sigma^2 [z_{1-\alpha/2} + z_{1-\beta}]^2}{(\mu_1 - \mu_2)^2} \]

\[ \frac{(1 + (m_1 - 1)\rho_1)}{R} + \frac{(1 + (m_2 - 1)\rho_2)}{1/R} : \text{Weight Value for sample size calculation} \]
If control arm is individual treatment, then?

- The unit of cluster in control arm is individual patient, which means $m_2 = 1$;
- Allocation Ratio:
  
  $$R = \sqrt{\frac{1 + (m_1 - 1)\rho_1}{1 + (m_2 - 1)\rho_2}} = \sqrt{1 + (m_1 - 1)\rho_1}$$

- Sample size calculation:
  
  $$N_T = \sigma^2 \frac{(\frac{z_\beta + z_\alpha}{2})^2}{(\mu_1 - \mu_2)^2} \left( \frac{1 + (m_1 - 1)\rho_1}{R} + \frac{1}{1/R} \right)$$
Short summary

- Clustering effects in RCT
  1. Definition of clustering
  2. Non-ignorable clustering
  3. Statistical Methods (3 models, 2 scenarios)
  4. Sample size calculation
- What’s Next
  Equal sample size → Unequal sample sizes
  Ben’s presentation:
  Sample Size Calculations for Group Randomized Trials with Unequal Sample Sizes through Monte Carlo Simulations
Q&A
References

• David Murray’s online course: Pragmatic and Group-Randomized Trials in Public Health and Medicine
Thank You!