Sampling and Nested Data in Practice-Based Research

Stephen Zyzanski, PhD
Department of Family Medicine
Case Western Reserve University
School of Medicine
Randomization by cluster accompanied by analysis appropriate to randomization by individual is an exercise in self-deception and should be discouraged.

Cornfield, 1978
Group-Randomized trials are comparative studies in which:

1) the units of assignment are identifiable groups, e.g., physicians, schools, cities, etc.

2) the units of observation are members of those groups, e.g., patients, students, residents, etc.
Unique Characteristics of Group-Randomized Trials

1) **Unit of assignment** is the group rather than the individual (physicians not patients)

2) Different units of assignment are allocated to each study condition

3) **Units of observation** are members of the groups that serve as the units of assignment

4) Group randomized trials typically involve a limited number of assignment units in each study condition.
Impact of these Unique Characteristics

The largest impact is on:

1) sample size consideration and on

2) the required approach to the statistical analysis of the data.
What are Nested (Multilevel) Data?

Observations about sub-units (patients) nested within larger units (physicians/practices).

Nested or multilevel data refer to data which contain a multilevel (hierarchical or nested) structure.

Multilevel structure indicates that data to be analyzed were obtained from various levels, and these levels are nested within each other.
Examples of Nested Data

Patients nested within physicians (2-level)
Data are collected from both
patient (micro) and
physician (macro) levels
Patients nested within physicians within practices (3-level)
Time points within patients (repeated measures)
Subjects within studies (meta-analysis)
Table 5.1 Selected sources for estimates of intracluster correlation coefficients and design effects (Koepsell 1998)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcomes</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. (1997)</td>
<td>Cholesterol levels among children</td>
<td>Family</td>
</tr>
<tr>
<td>Vachon et al. (1998)</td>
<td>Dietary intakes among postmenopausal women</td>
<td>Family</td>
</tr>
<tr>
<td>Katz et al. (1993a, b), Katz and Zeger (1994)</td>
<td>Childhood illness</td>
<td>Family, village</td>
</tr>
<tr>
<td>Siddiqui et al. (1996)</td>
<td>Tobacco use among adolescents</td>
<td>Classroom, school</td>
</tr>
<tr>
<td>Murray and Hannan (1990), Murray et al. (1994)</td>
<td>Tobacco and drug use among adolescents</td>
<td>School</td>
</tr>
<tr>
<td>Slymen and Hovell (1997)</td>
<td>Tobacco and alcohol use among adolescents</td>
<td>Orthodontist offices</td>
</tr>
<tr>
<td>Kelder et al. (1993)</td>
<td>Physical health and tobacco use</td>
<td>Worksites</td>
</tr>
<tr>
<td>Gulliford et al. (1999)</td>
<td>Lifestyle risk factors and health outcomes</td>
<td>Neighbourhood, household</td>
</tr>
<tr>
<td>Verma and Le (1996)</td>
<td>Fertility rates, family and child health</td>
<td>Neighbourhood</td>
</tr>
<tr>
<td>Feldman and McKinlay (1994)</td>
<td>Height, weight, body mass index, blood pressure, cholesterol levels</td>
<td>Community</td>
</tr>
<tr>
<td>Hannan et al. (1994)</td>
<td>Behavioural risk factors, knowledge and attitudes concerning heart disease</td>
<td>Community</td>
</tr>
<tr>
<td>Murray and Short (1995)</td>
<td>Alcohol use among adolescents</td>
<td>Community</td>
</tr>
<tr>
<td>Murray and Short (1997)</td>
<td>Tobacco use among adolescents</td>
<td>Community</td>
</tr>
<tr>
<td>Mickey et al. (1991), Mickey and Goodwin (1993)</td>
<td>Mortality from cardiovascular disease and cancer</td>
<td>County</td>
</tr>
</tbody>
</table>
Multi-Level Data Structure

A multilevel data structure allows for an understanding of how person-level variables (e.g., patient gender) as well as group-level variables (e.g., physician characteristics) can be used to explain an outcome (measured at the lowest level, e.g., patient)

Information from both levels are used in the analysis.
Why are Nested Data a Problem?

Research has shown that people within a particular group (physician) tend to be more similar to each other in terms of the outcome variable than they do to people in a different group (physician).

This correlation violates the assumption of independence necessary for a traditional linear models approach.
Nested Data

The extent of clustering is reflected in a statistic called the intra-class correlation (ICC).

Even mild violations can lead to severe problems with inflation of Type I error.
Intraclass Correlation

The ICC may be interpreted as the usual pair-wise correlation coefficient between any two members of the same cluster.

Also, it is the proportion of total variance in data that is attributable to the between group variation (e.g., physician).
**Intraclass Correlation**

When ICC=0, this corresponds to statistical independence among members of a group.

When ICC=1.0, this corresponds to total dependence, i.e., the information supplied by the cluster is that supplied by a single member, i.e., the effective group size is one.
Intraclass Correlation

Hox, 2002 recommends using:

.05, .10 and .15 as small, medium and large values for the ICC in general cases and

.1, .2 and .3 in those cases where on a priori grounds higher ICC’s appear reasonable, e.g., small group and family research.
Design Effect (Variance Inflation Factor)

The design effect or VIF is a number that indicates how much the sample size is to be adjusted due to the clustering of observations within groups.

For a 2-level sample, the design effect is:

\[(1 + (m-1)ICC); m=\text{group sample size}\]
Intraclass Correlation

When ICC > 0, such clustering leads to a reduction in the effective sample size for a study. Application of standard sample size formulas will lead to underpowered studies.

The degree of reduction is measured by the “design effect” or “Variance Inflation Factor” (VIF).
Design Effect (VIF)

1) \( (1 + (m-1)\text{ICC}) \), \( m=\text{obs/group} \)

2) When ICC>0 the VIF increases both as the ICC increases and as the number of observation units in each assignment unit increases.

3) Small values of ICC combined with large cluster size can yield sizeable design effects.
Effective Sample Size

A convenient formula to compute the effective sample size is:

\[ \text{Neff} = \frac{N}{1 + (m-1) \text{ICC}} \] or \[ \frac{N}{\text{VIF}} \]

where \( N \) is the total sample size; \( \text{Neff} \) is the effective sample size.

The effective cluster size is given by the formula:

\[ \frac{m}{1 + (m-1) \text{ICC}} \] or \[ \frac{m}{\text{VIF}} \]
Example of Effective Sample Size

Given a sample of 10 physicians, each with 20 patients, equals a total sample of 200. Given an ICC of 0.10

Effective sample size is:
\[ \frac{200}{1 + (20-1)(0.10)} = 69 \]
which is much less than the apparent total sample size of 200.
Example of “Effective Sample Size”

Cohen tables show that one needs $N=73$ patients to detect a correlation (say between age and satisfaction) of 0.3 (medium effect) with power of 80% and $\alpha=0.05$, w/simple random sampling.

Now let’s assume the data come from a 3 physician practice where: ICC is 0.10 and the average physician patient sample is $73/3=24.3$
Effective Sample Size Example Con’t.

To compute the “effective sample size” i.e., one adjusted for clustering we use:

\[ \text{Neff} = \frac{N}{VIF} \]
\[ = \frac{73}{1 + (24.3-1)(0.1)} \]
\[ = \frac{73}{1 + 3.33} \]
\[ = \frac{73}{3.33} \]
\[ = 22 \]

\[ \text{N} = (N)(VIF) = (73)(1 + (23.3)(0.1)) \]
\[ = 73 + 168 = 241 \text{ or about 10 MD’s} \]

Considerably more than the \( N=73 \) for simple random sampling
Intraclass correlation coefficients for cardiovascular measures from the Cholesterol Education and Research Trial (CEART)

Parker DR, Eaton CB, Evangelou E

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\rho$</th>
<th>Design Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0652791</td>
<td>4.394514</td>
</tr>
<tr>
<td>Gender</td>
<td>0.0106866</td>
<td>1.555702</td>
</tr>
<tr>
<td>Marital Status</td>
<td>0.0374449</td>
<td>2.947137</td>
</tr>
<tr>
<td>Smoking Status</td>
<td>0.0649154</td>
<td>4.375603</td>
</tr>
<tr>
<td>Weight</td>
<td>0.0058421</td>
<td>1.303788</td>
</tr>
<tr>
<td>BMI</td>
<td>0.0222298</td>
<td>2.155947</td>
</tr>
<tr>
<td>SBP</td>
<td>0.0019579</td>
<td>1.101813</td>
</tr>
<tr>
<td>DBP</td>
<td>0.0532843</td>
<td>3.770786</td>
</tr>
<tr>
<td>Total Cholesterol</td>
<td>0.0001121</td>
<td>1.005829</td>
</tr>
<tr>
<td>HDL Cholesterol</td>
<td>0.0519004</td>
<td>3.69882</td>
</tr>
<tr>
<td>LDL Cholesterol</td>
<td>0.0025644</td>
<td>1.133351</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.0570509</td>
<td>3.966647</td>
</tr>
<tr>
<td>TC/HDL ratio</td>
<td>0.042257</td>
<td>3.197377</td>
</tr>
</tbody>
</table>
A Priori Power Analysis

The proper use of the sample size formulas requires some prior assessment of the ICC.

The power of a cluster randomization trial depends more on the number of units randomized than on their size.
A Priori Power Analysis

The effect of clustering on sample size depends on the joint influence of both cluster size and ICC.

The same or similar power values may be obtained with different numbers of groups and group sizes.
A Priori Power Analysis

Kreft 1996 suggests a 30/30 rule of thumb. Researchers should strive for a sample of at least 30 groups with at least 30 individuals per group. Best for fixed parameters.

If the interest is in cross-level interactions, the numbers of groups should be larger which leads to the 50/20 rule-about 50 groups with 20 persons per group.
Raudenbush’s Website

Includes a free power analysis program for multilevel models. This is the program used in the following examples. The website is:

http://www-personal.umich.edu/~rauden/

Click on “Optimal Design Software.”
A Priori Power Analysis

The sample estimation software requires:
1) Cluster size (n)
2) Number of clusters (J)
4) Intra-class correlation (p)
5) Effect size (δ)
6) Power (e.g., 80%)
7) Level of significance (.05)
Effect of number of subjects per cluster on power (10 clusters)

- $\alpha = 0.050$
- $J = 10$
- $\varepsilon = 0.20, \rho = 0.10$
- $\varepsilon = 0.50, \rho = 0.10$

Power

Number of subjects per cluster

$\delta = 0.50$

$\delta = 0.20$
Effect of number of subjects per cluster on power (20 clusters)

\[ \alpha = 0.050 \]

\[ J = 20 \]

\[ \varepsilon = 0.20, \rho = 0.10 \]

\[ \varepsilon = 0.50, \rho = 0.10 \]

Number of subjects per cluster
Effect of number of subjects per cluster on power (40 clusters)

$J = \alpha \sigma^2 \xi = 0.050$

$J = 40$

- $\xi = 0.20, \rho = 0.10$
- $\xi = 0.50, \rho = 0.0$

Number of subjects per cluster

Power
The effect of number of clusters on power (n=10)

\[ \delta \leq 0.50 \quad \alpha = 0.050 \]
\[ n = 10 \]

\[ \delta = 0.20, \rho = 0.10 \]
\[ \delta = 0.50, \rho = 0.10 \]
The effect of number of clusters on power (n=20)

- $\alpha = 0.050$
- $n = 20$
- $\delta = 0.20, \rho = 0.10$
- $\delta = 0.50, \rho = 0.10$
The effect of number of clusters on power (n=50)

\( \alpha = 0.050 \)

\( \alpha = 0.20, \rho = 0.10 \)

\( \alpha = 0.50, \rho = 0. \)
Take Home Messages

1. Nested data are correlated data
2. Correlated data violate statistical assumptions
3. The magnitude of the correlation is represented by the ICC
4. Small ICC correlations can have a big impact
5. Small (.05); medium (.10); large (.15)
6. Study sample size requires adjustment for the degree of nesting using the design effect/VIF
7. Software requires additional estimates of cluster size, number of clusters, and the ICC