

Sampling and Nested Data in Practice- Based Research

Stephen Zyzanski, PhD

Department of Family Medicine

Case Western Reserve University

School of Medicine

September, 2008

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)

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 pairwise 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

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 (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 =group sample size

Design Effect (VIF)

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

$N_{\text{eff}} = N / (1 + (m-1) \text{ICC})$ or N / VIF where N is the total sample size; N_{eff} is the effective sample size.

The effective cluster size is given by the formula:
 $m / (1 + (m-1) \text{ICC})$ or m / 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:

$200 / (1 + (20-1)(.10)) = 69$ which is much less than the apparent total sample size of 200

Intraclass correlation coefficients for cardiovascular measures from the Cholesterol Education and Research Trial (CEART)

Parker DR, Eaton CB, Evangelou E

<u>Variable</u>	<u>ρ</u>	<u>Design Effect</u>
Age	.0652791	4.394514
Gender	.0106866	1.555702
Marital Status	.0374449	2.947137
Smoking Status	.0649154	4.375603
Weight	.0058421	1.303788
BMI	.0222298	2.155947
SBP	.0019579	1.101813
DBP	.0532843	3.770786
Total Cholesterol	.0001121	1.005829
HDL Cholesterol	.0519004	3.69882
LDL Cholesterol	.0025644	1.133351
Glucose	.0570509	3.966647
TC/HDL ratio	0.042257	3.197377

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.

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/>

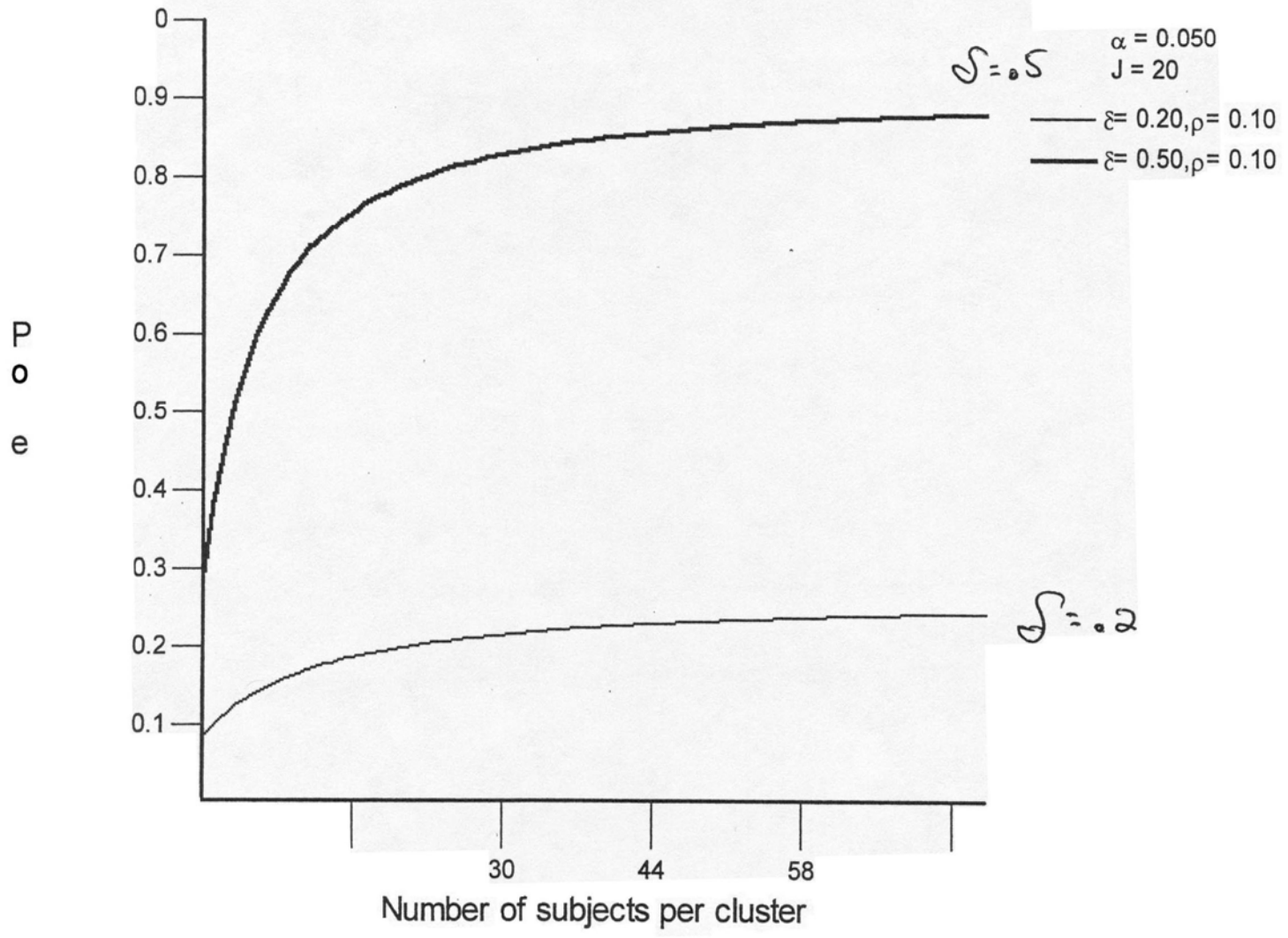
Click on "Optimal Design Software."

A Priori Power Analysis

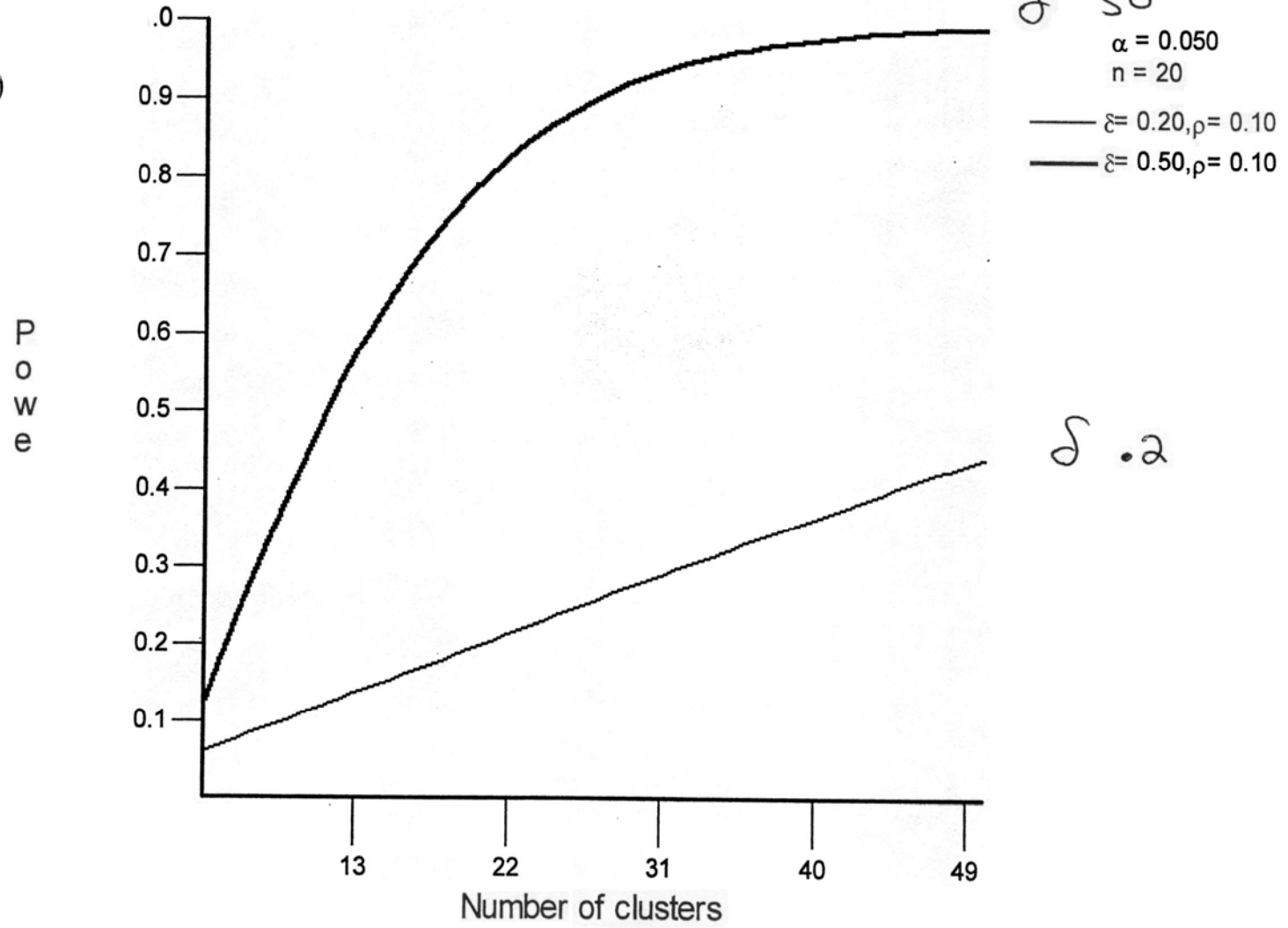
The sample estimation software requires:

- 1) Cluster size (n)
- 2) Number of clusters (J)
- 4) Intra-class correlation (ρ)
- 5) Effect size (δ)
- 6) Power (e.g., 80%)
- 7) Level of significance (.05)

Effect of number of subjects per cluster on power (20 clusters)



The effect of number of clusters on power (n=20)



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

Intraclass Correlation

Psychosocial variables (attitudes/beliefs) tend to have higher ICC's than behavioral measures which in turn have have higher ICC's than physiological risk factors.

Thus, the precision of the intervention effect could be expected to be better for physiological measures than psychosocial measures.

