Primer on Group Randomized Trials

Group randomized trials are experiments in which the intervention occurs at the level of the group (typically physicians or clinics) but observations are made on individuals within the groups (e.g., patients). Because group randomized trials are increasingly common in health services research, critical readers should understand their rationale, the implications of group size vs. number of groups, and the limitations of the approach.

Why Randomize by Group?

Group randomization is particularly useful when there is a high risk for contamination if group members are randomized as individuals. For example, an investigator studying the effects of a clinical practice guideline can't assume that a provider caring for patients in the intervention arm will not apply this knowledge to the patients assigned to the control arm. Such contamination biases the study toward a finding of no effect. Randomizing at the level of the physician avoids this source of contamination because physicians are either exposed or not exposed to the intervention. If there are concerns that intervention physicians will contaminate control physicians in the same clinic, randomization should occur at the clinic level.

Group Size vs. Number of Groups

To illustrate some of the issues raised by group randomization, consider a trial to test a cholesterol management guideline. Physicians would be randomly assigned to a control or an intervention arm while the outcome (say, the mean change in cholesterol after 6 months) would be measured on their patients. As shown in the Figure, however, there are many possible combinations of group size and number of groups.

In each case we have 200 patient observations (100 patients in each arm), but as group size increases there are fewer physicians. With smaller group size, there is less information on many physicians; with larger group size, there is more information on only a few physicians. Because the study is intended to measure the impact of the guideline on physicians, the design

with 40 physicians is more likely to detect a significant intervention effect than the one with only 8 physicians—despite the equivalent size of the patient sample. In other words, collecting a large amount of information on patients in one physician practice allows something precise to be said about that physician but adds little to the ability to answer the study question.

Although ideally there should be as many physicians as possible, practical considerations often limit enrollment. The number of physicians available and willing to participate is often limited. It can be very expensive to enroll and train a physician. It is often easier to recruit many patients and a few physicians than it is to recruit many physicians. Thus, there is a trade-off between increasing group size (often the most expedient way to increase sample size) and increasing the number of groups (generally the most effective way to increase power).

Sample Size in Group Randomized Trials

The ability to make statistical inferences is inversely related to variability in the outcome measure. In this example, the variability in cholesterol can come from two sources: differences among patients and differences among physicians (presumably in their ability to influence the patients' cholesterol either through behavior modification or pharmacologic treatment). The *proportion* of cholesterol variability attributable to physicians is called the intraclass correlation (the term is a misnomer because it has nothing to do with correlation), or rho. As rho increases, a greater share of the variability comes from physicians, so that increasing the number of physicians will become more important. If rho is small, then increasing the number of patients per physician may be sufficient to increase the power to detect an effect. Rho can only be zero if there is no systematic difference between groups. In other words, 1) physicians do not differ in their response to education and 2) the patients of one physician do not differ systematically from those of another. A typical rho in this setting is between 0.01 and 0.04.

Table 1 illustrates how changes in the intraclass correlation affect the sample size needed to produce equivalent levels of
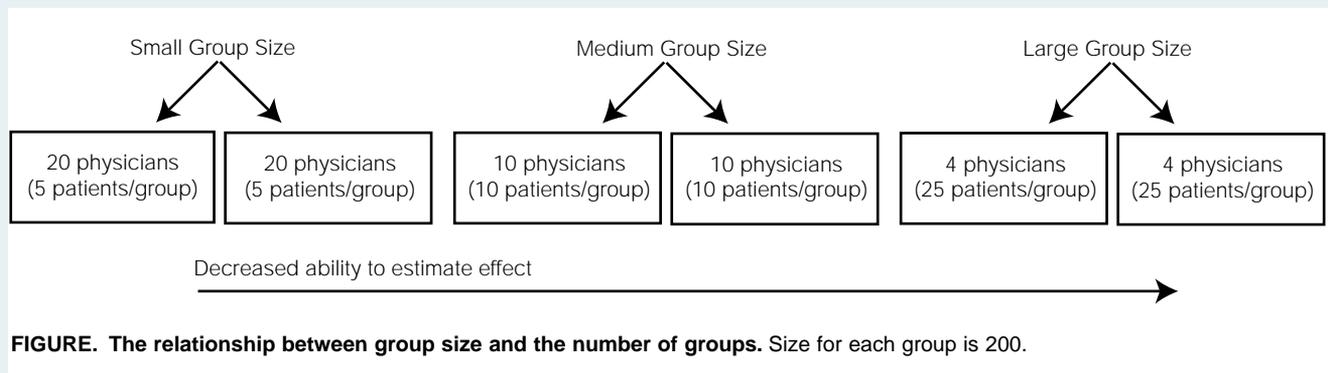

FIGURE. **The relationship between group size and the number of groups.** Size for each group is 200.

Relationship between Intraclass Correlation, Sample Size, and Number of Groups

NUMBER OF PHYSICIANS PER TREATMENT ARM	INTRACLASS CORRELATION (RHO)	SAMPLE SIZE (TOTAL)
4	0.0*	200
	0.01	396
	0.02	∞
	0.03	∞
10	0.0*	200
	0.01	248
	0.02	320
	0.03	486
20	0.0*	200
	0.01	220
	0.02	246
	0.03	278

No physician effect.

precision. As the intraclass correlation increases, the total number of patients needed also increases. In addition, Table 1 shows how the effect is modified by the number of physicians. When the intraclass correlation is 0.03, for example, a study with 10 physicians in each arm requires 486 patients to achieve the same precision as a study with 278 patients and 20 physicians in each arm. Notice that with 4 physicians in each arm, no number of patients would provide sufficient information to answer the study question. This illustrates a major limitation of group randomized trials: It may be impossible to collect enough data at the patient level to make up for a small number of groups. The important lesson here is that the *effective* sample size in a group randomized trial is not related only to the number of patients but depends on the number of groups and the intraclass correlation.

Comparability of Patients

One of the most important advantages of randomization is that, if the trial is large enough, it is fair to assume that the study groups will be comparable with respect to all variables (measured and unmeasured). This enhances our ability to make inferences about the effect of the intervention on the outcome. In contrast to randomized trials of individuals, group randomized trials involve only a limited number of groups—typically 15 or 20. Thus, there are rarely enough groups to ensure even distribution of variables that could confound the treatment effect and bias the outcomes comparison.

As a result, investigators need to collect information on important confounders and plan analyses that will control for these factors. These analyses require special techniques that directly incorporate the group structure (cluster analyses). It would be a mistake in our hypothetical example to simply compare the average cholesterol levels in the treatment and control group with, say, a standard *z*-test. For example, a study with rho = 0.03, 10 physicians per group, and 486 total patients would be equivalent to a study with rho = 0 and 200 total patients. A *z*-test would calculate a standard error based on 486 patients, when the effective sample size is only 200. Statistical analysis that ignores this fact can give falsely low *P* values and overly optimistic confidence intervals.

Policymakers and managers are increasingly interested in moving "hard science" to the vagaries of actual clinical practice. To help translate efficacy into effectiveness, interventions are being directed to physicians (or groups of physicians). Group randomization is the best approach to make valid inferences about their value.

This Primer was contributed by Michael L. Beach, MD, PhD, Dartmouth–Hitchcock Medical Center, Lebanon, New Hampshire.

A compendium of **ecp** *primers from past issues can be viewed and/or requested at http://www.acponline.org/journals/ecp/primers.htm.*