NIH Collaboratory *Rethinking Clinical Trials®*

Health Care Systems Research Collaboratory

Intraclass Correlation Coefficient Cheat Sheet

PURPOSE

This document provides an introductory description of the intraclass correlation coefficient (ICC), a descriptive statistic that is important for the design and analysis of cluster-randomized trials. In a cluster randomized trial, instead of being randomized by individual participant, the unit of randomization is a cluster, such as a group of participants being seen at a hospital, clinic, or primary-care practice, although the outcomes may still be measured at an individual level.

DEFINITION

The intraclass correlation coefficient (ICC) is a descriptive statistic that describes the extent to which outcomes 1) *within* each cluster are likely to be *similar* or 2) *between* different clusters are likely to be *different* from each other, relative to outcomes from other clusters. The ICC is an important tool for cluster-randomized pragmatic trials because this value helps determine the sample size needed to detect a treatment effect. Although it ranges from 0 to 1 theoretically, the ICC for most pragmatic cluster-randomized trials is typically <0.2; commonly around 0.01 to 0.05.

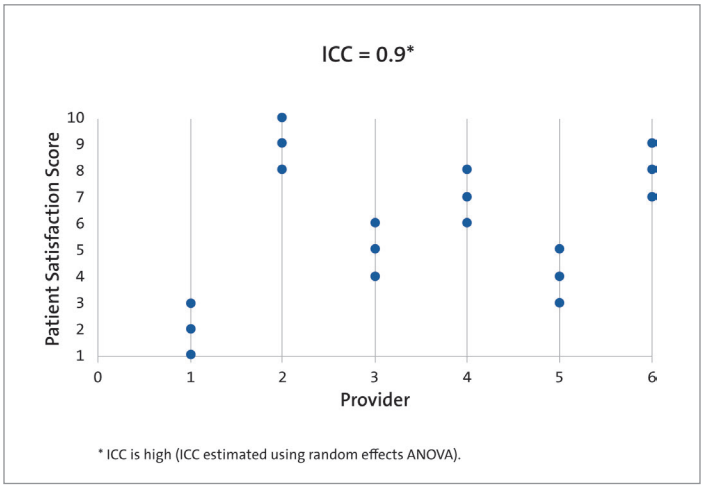
EXAMPLES

In cluster-randomized trials where groups of individuals are randomized to treatment arms, when outcomes within clusters are highly correlated and when the magnitude of outcomes across clusters is quite different, then participants within the cluster are likely to have similar outcomes and the ICC will be large. When this is the case, the data from one member of the cluster provides almost as much information as if all of the members are included. Hence, the effective sample size is closer to the number of clusters as opposed to the entire sample size of study participants.

To demonstrate why this is relevant, let's consider two examples:

1. In a dietary intake study, the data from several members of the same family would likely be very similar and would differ from that of other families. Hence there may be little gain from sampling more than one member. On the other hand, if a cluster is an entire city and subjects within the city are randomly sampled, one might expect relatively little similarity from subject to subject relative to the rest of the sample. In this case, each individual subject would likely contribute "independent" information.
2. Suppose we have 6 providers, each with 3 eligible participants for a pragmatic cluster-randomized trial. In this hypothetical case, the outcome is patient satisfaction rated on a scale from 1 to 10 with an outcome distribution as shown in Figure 1. One might expect that patients seen by a specific provider will have more similar levels of satisfaction to each other than to patients from other providers and that some providers will have consistently high patient satisfaction (e.g. provider 2) whereas others will have consistently low patient satisfaction (e.g. provider 1). This is an example of how outcomes within each cluster are likely to be similar. Thus, the ICC is high, and adding individuals to the cluster does not provide much additional information.

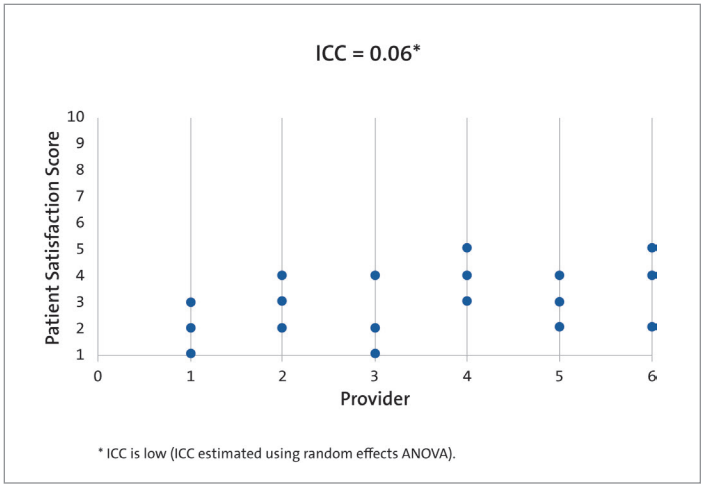
Figure 1. Patient Satisfaction



Conversely, if all providers have fairly low average patient satisfaction and the within-provider variability is similar to that in Figure 1, then the ICC will be smaller than in Figure 1. Figure 2 provides an example of this type of situation. Note that no patient provides a satisfaction score above 5, the overall variability of the data is lower than in the previous figure, and there is much lower between-provider variability in these data. Here, the ICC is lower because the outcomes across different clusters are not likely to be different from each other.

Finally, when the outcomes within a cluster are no more similar to each other than they are to members of any other cluster, there is no structure provided by the clusters and therefore the the overall group of participants looks like a random sample of individuals rather than a sample from different clusters. In this case, the effective sample size is close to the total number of study participants and the ICC is close to 0.

Figure 2. Patient Satisfaction



POWER CALCULATIONS

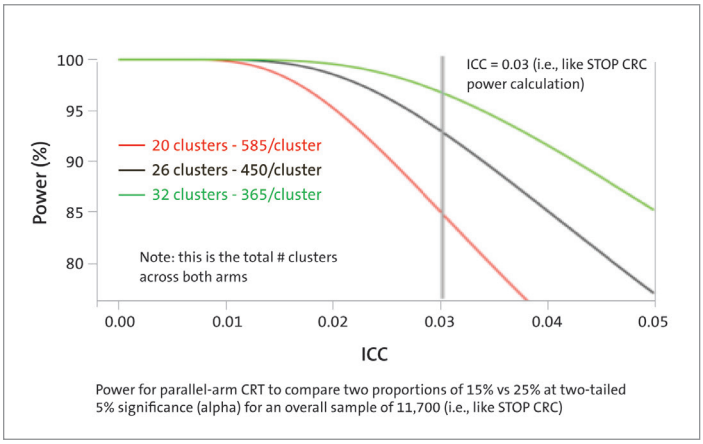
Typically, investigators of individually randomized trials think of statistical power as being related to the number of patients. With cluster-randomized trials, the number of clusters and the level of the ICC drive the needed sample size to obtain reasonable statistical power.

In the figure below, each trial has the same total number of people (11,700) but different numbers of clusters and, therefore, different numbers of participants in each cluster. The ICC is the X axis and power is the Y axis.

Specifically, the Y axis is the power to detect a difference in proportions between treatment and control arms of 10% (for a comparison of 15% vs. 25%) under different assumptions. As seen in the figure, with increasing ICC, the modifier of power is the number of clusters. For example, at ICC=0.03, a cluster-randomized trial with 11,700 participants divided equally between a total of 32 clusters (i.e. 365 members per cluster) has more than 95% power, whereas for that same ICC and same total sample size (11,700) now divided equally between a total of 20 clusters (i.e. 585 members per cluster), power is greatly reduced to 85%.

Finally, it is important to partner with a statistician to consider the implications of the magnitude of the ICC. Data for estimating the ICC can be derived from previous CRTs with similar outcomes in a similar context. It is important to note that if pilot study data is available, it may not provide a reliable estimate of the ICC due to the small sample size. For example, if the ICC estimate to be used to design the CRT is too small, then the CRT may be underpowered as a consequence. Similarly, it is important to be cautious when dealing with binary and count outcomes as there are multiple definitions of the ICC for these outcomes. For such outcomes, it is important to know which one is used in the selected sample size formula.

Figure 3. Clustering: impact on power



For more on the ICC, see the Intraclass Correlation section in the Living Textbook or this working document from the Biostatistics and Study Design Core. If you have questions, feedback or suggestions regarding this tool, please contact us at nih-collaboratory@dm.duke.edu.