NIH Collaboratory Rethinking Clinical Trials™
Health Care Systems Research Collaboratory

The Intraclass Correlation Coefficient (ICC)

A working document from the NIH Collaboratory Biostatistics/Study Design Core. This work was supported by a cooperative agreement (U54 AT007748) from the NIH Common Fund for the NIH Health Care Systems Research Collaboratory. The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Background

The intraclass correlation coefficient (ICC), related to the design effect (DEFF) [1] as:

$$\mathrm{DEFF} = 1 + (\mathrm{n} - 1)\mathrm{ICC}$$

is a key parameter in the design and analysis of group- or cluster-randomized trials (GRTs or CRTs). The ICC, together with the degrees of freedom (df) based on the number of groups or clusters, is commonly used to calculate how much the sample size of a CRT should be inflated compared with a simple individual-randomized trial. Because multiple expressions and estimators of ICC exist, it is important to understand that the selection of a particular ICC estimate during the planning stage of the study should be tailored to the study design and planned analysis.

General definition of the ICC

In applications to CRTs, Eldridge et al. [2] provide a general definition of the ICC as a common correlation coefficient between responses of any two subjects from the same cluster. The actual expression for the ICC depends on the type of the outcome and the model describing the data. In the case of a continuous outcome, the value of ICC is constrained between $\frac{-1}{n_{max}-1}$ and 1, where n_{max} is the maximum cluster size. As n_{max} becomes arbitrarily large, the actual lower bound for the ICC approaches zero. When hierarchical models are used to describe the data structure, the ICC can be expressed as the ratio of the outcome variance between clusters to the total subject variance, which is essentially equal to the sum of the variance between cluster means and the average variance between subjects within a cluster.

Previous research has shown that it is important to allow a negative variance estimate for the variance between clusters. As a result, the lower bound for the ICC may be slightly negative. Negative estimates are common when the true ICC is close to zero, and if they are constrained to be non-negative, the type 1 error rate can be suppressed, with adverse effects on statistical power [3,4] when data are analyzed using methods commonly applied in CRTs. Recent work suggests that ICC estimates may be constrained to be positive if the analysis employs the

Prepared by: Elizabeth DeLong and Yuliya Lokhnygina
Reviewed by: NIH Collaboratory Biostatistics/Study Design Core
Version: 1.0, last updated June 26, 2014

Kenward-Roger method for df in conjunction with those same analysis methods [5]; additional work is needed to evaluate the generalizability of that finding.

Estimating the ICC for binary data

The ANOVA estimator used for continuous variables may also be used for binary data [4,6] to estimate the ICC without transforming the data. Another expression [7] quantifies the extent of the probability of agreement in response from two subjects from the same cluster over the agreement between two subjects from different clusters, divided by the maximum value of this difference. Either expression is the same as the well-known Kappa index.

Many hierarchical regression programs analyze binary data using a log link and binomial error distribution, based on the generalized linear mixed model [8]. In this case, the ICC cannot be calculated as the ratio of the between-clusters variance to the total subject variance because those variance components are not on the linear scale. However, with proper transformation, an ICC estimate is obtained that agrees closely with the ANOVA estimator [4].

Methods for obtaining estimates of the ICC

Multiple methods of estimating ICC have been described in the literature [9,10]. Properties of different estimators depend on study characteristics such as the balance of the design, the number and size of clusters, and the presence of covariates. In the case of binary data, the estimators depend upon the underlying data model and can produce quite different results [11]. Therefore, when planning a new study, investigators should be aware of the modeling approach used when selecting a value for the ICC from prior studies.

Numerous studies have published ICC estimates, as is now encouraged by the CONSORT Statement. In selecting an estimate from the published literature, or choosing an estimate from pilot data, investigators should seek to select an estimate that reflects the key properties of the trial they are planning. Therefore, the estimate should be based on the proposed dependent variable, measured using the proposed methods, and taken from a similar population aggregated in similar clusters or groups. Where multiple estimates are available, investigators can pool them using meta-analytic approaches to obtain a single estimate with greater precision [12].

References

1. Kish L. *Survey Sampling.* New York, NY: John Wiley & Sons; 1965.

2. Eldridge SM, Ukoumunne OC, Carlin JB. The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. Int Statist Rev 2009;77(3):378–94. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2009.00092.x/abstract. Accessed May 14, 2014.

3. Swallow WH, Monahan JF. Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. Technometrics 1984;26(1):47–57. Available at: http://www.jstor.org/stable/1268415. Accessed May 14, 2014.

4. Murray DM. *Design and Analysis of Group-Randomized Trials.* New York, NY: Oxford University Press; 1998.

5. Andridge RR, Shoben AB, Muller KE, Murray DM. Analytic methods for individually randomized group treatment trials and group-randomized trials when subjects belong to multiple groups. Stat Med 2014;33(13):2178 –90. PMID: 24399701. doi: 10.1002/sim.6083.

6. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research.* London: Arnold; 2000.

7. Mak TK. Analyzing intraclass correlation for dichotomous-variables. Appl Stat J Roy Stat Soc Ser C 1988;37(3):344–52. doi: 10.2307/2347309. Available at: http://www.jstor.org/stable/2347309. Accessed May 14, 2014.

8. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall; 1989.

9. Donner A, Wells G. A comparison of confidence-interval methods for the intraclass correlation-coefficient. Biometrics 1986; 42(2):401–12. doi: 10.2307/2531060. Available at: http://www.jstor.org/stable/2531060. Accessed May 14, 2014.

10. Ridout MS, Demetrio CGB, Firth D. Estimating intraclass correlation for binary data. Biometrics 1999;55(1):137–48. Available at: http://www.jstor.org/stable/2533904. Accessed May 14, 2014.

11. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemp Clin Trials 2012;33:869–80. PMID: 22627076. doi: 10.1016/j.cct.2012.05.004. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3426610/. Accessed May 14, 2014.

12. Blitstein JL, Hannan PJ, Murray DM, Shadish WR. Increasing the degrees of freedom in existing group randomized trials: the DF* approach. Eval Rev 2005;29(3):241–67. PMID: 15860765. doi: 10.1177/0193841X04273257.