Relationship between Intra-Cluster Correlation and Sample Size Requirement for Cluster Randomized Trial

Hrishikesh Chakraborty
Statistics and Epidemiology, RTI International

Sample size calculation for cluster randomized trials (CRT) primarily depends on the primary outcome variable's distribution, effect size, average cluster size and intra-cluster correlation (ICC) estimates. Furthermore, the ICC estimate depends on the outcome variable's distribution, cluster size, and number of clusters. Researchers often design CRTs based on ICC estimates from previous trials or from a simulation. Furthermore, since the ICC estimate depends on the outcome variable's distribution, the ICC estimate will often change at the end of a trial. This change in the ICC estimate effects the sample size requirement and power for the trial. We developed a simulation based technique to estimate the ICC and confidence intervals for ICC for binomial variables and also determine the change in ICC at the end of the trial for different scenarios before the trial starts. This is a very helpful technique to ensure power at the end of the trial in CRTs. We apply this simulation technique to calculate sample size for assessing the effect of neonatal resuscitation on 7 day neonatal mortality for a multi-country CRT where communities within a geographical area are clusters. We conclude that if the ICC estimate at the end of the trial is less than the ICC value used for sample size calculation then the trial will be over powered. On the other hand, if the ICC estimate at the end of the trial is greater than the ICC value used for sample size calculation then the inferences will be based on an under powered trial. Therefore, we need to adjust for this predicted ICC change during the design phase when calculating the required sample size.