A Murmuration of Survey Sample Size Planning Parameters - Observed Values of Vaccination Coverage Intracluster Correlation Coefficients from Many Recent Studies

Dale A. Rhoda[1], Caitlin B. Clary[1], Becca Robinson[1], Mary Kay Trimner[1], Shiraz Johnson[1,2]

[1] Biostat Global Consulting, Worthington, Ohio, USA
[2] Intern from Grinnell College Mathematics and Statistics Department, Grinnell, Iowa, USA

Since the 1980s the World Health Organization has promoted household surveys with a cluster sample design as one method to estimate vaccination coverage in low- and middle-income countries. When calculating the survey sample size, planners begin with numeric targets for the outcome proportion and its precision, and they calculate an effective sample size that would succeed with a simple random sample. Then they account for complex sampling using a guesstimated design effect that depends on the unknown spatial heterogeneity of the outcome as parameterized with an intracluster correlation coefficient (ICC). In this talk we use scatterplots to graphically summarize hundreds of observed values of ICC (and other planning parameters) from surveys conducted in recent years. We stratify the results by survey type and sampling design, and we showcase the generic user-written Stata program that we shared so collaborators around the world may extract parameters from their datasets using consistent calculations. Summaries from our (still growing) collective dataset will help planners balance their tolerance for risk of fielding too small a survey to meet the stakeholders' precision goals versus one that is larger and more costly (and perhaps of lower quality) than necessary. Of note: the program that extracts parameters includes some modifications to Stata's estimation commands and would be generically useful for cluster surveys related to other topics.