**Comparing Different Sampling Schemes to Estimate the Population Mean:**

**A Simulation Study**

**Indrani Sarkar1, Pratibha Shrestha2, Cornelius Rosenbaum1, Evrim Oral1\***

1LSUHSC School of Public Health, Biostatistics Program, New Orleans

2LSUHSC School of Public Health, Epidemiology Program, New Orleans

The three most common methods of probability sampling are Simple Random Sampling (SRS), Stratified Random Sampling (StRS), and Cluster Sampling (CS). In SRS, every member of the population is chosen randomly and has an equal chance of being selected. Although SRS is the least complicated probability sampling method, to be able to utilize it, an exhaustive and definitive sampling frame is needed, which can be challenging to obtain. StRS is used when the sampling units associated with the population can be separated into two or more homogeneous subpopulations (strata) where response variation within stratum is less than response variation between stratum.​ StRS ensures a better representation of population compared to SRS; however, it cannot be used when researchers cannot confidently classify every member of the population into a subgroup. In CS, the population is divided into subpopulations (clusters) which are externally homogeneous but internally heterogenous. CS is a very convenient and effective probability sampling scheme for geographically dispersed populations where a sampling frame is not available. Although it is the least expensive and most feasible sampling scheme among the three, it provides less precise estimates. To our knowledge, there has been no study that compared the empirical biases and variances from SRS, StRS, and CS to date. This study compares the bias and variance of the sample mean from SRS, StRS, and CS via an extensive Monte-Carlo simulation study. We show that StRS performs better than both SRS and CS in terms of bias and relative efficiency; thus, when researchers have a sampling frame, they should utilize StRS method with appropriately selected strata. We also show that CS should be avoided unless there is no reliable/available sampling frame.