Correcting for differential recruitment with respondent-driven sampling data

Isabelle Beaudry,
Pontificia Universidad Católica de Chile

Abstract

Respondent-driven sampling (RDS) is a sampling mechanism that has proven effective to sample hard-to-reach human populations connected through social networks, such as certain populations at higher risk of HIV/AIDS infection. Under RDS, a small number of individuals known to the researcher are initially sampled and asked to recruit a fixed small number of their contacts who belong to the target population. Subsequent sampling waves are produced by peer recruitment until the desired sample size is achieved. However, the researcher's lack of control over the sampling process has posed a number of challenges to valid statistical inference. For instance, it is often assumed that participants recruit completely at random among their contacts. However, participants may systematically over recruit individuals with a particular characteristic or with whom they engage in a given type of relationship. Literature suggests that most RDS prevalence estimators are greatly sensitive to this assumption. In this work, we define three forms of differential recruitment, provide methods for estimating them and introduce novel inferential methods to address the induced biases.