ABSTRACT

This talk will provide an overview of probability models and inferential methods for the analysis of data collected using Respondent Driven Sampling (RDS). RDS is an innovative sampling technique for studying hidden and hard-to-reach populations for which no sampling frame can be obtained. RDS has been widely used to sample populations at high risk of HIV infection and has also been used to survey undocumented workers and migrants. RDS solves the problem of sampling from hidden populations by replacing independent random sampling from a sampling frame by a referral chain of dependent observations: starting with a small group of seed respondents chosen by the researcher, the study participants themselves recruit additional survey respondents by referring their friends into the study. As an alternative to frame-based sampling, the chain-referral approach employed by RDS can be extremely successful as a means of recruiting respondents.

Current estimation relies on sampling weights estimated by treating the sampling process as a random walk on a graph, where the graph is the social network of relations among members of the target population.

These estimates are based on strong assumptions allowing the sample to be treated as a probability sample. In particular, the current estimator assumes a with-replacement sample or small sample fraction, while in practice samples are without-replacement, and often include a large fraction of the population. A large sample fraction, combined with different mean nodal degrees for infected and uninfected population members, induces substantial bias in the estimates. We introduce a new estimator which accounts for the without-replacement nature of the sampling process, and removes this bias. We then briefly introduce a further extension which uses a parametric model for the underlying social network to reduce the bias induced by the initial convenience sample.

This is joint work with Krista J. Gile, Nuffield College, Oxford.