

Merli, M. Giovanna

Proposed Project dates 04/01/2011 – 03/31/2015

R01

Using Multiple Data Sources to Improve Respondent Driven Sampling Estimation

ABSTRACT:

This study addresses efforts to obtain valid estimates of the prevalence of sexually transmitted disease (STD) infection and risky and preventive health behaviors in a hidden population, female sex workers in China. We take advantage of multiple observations schemas to improve the utility of Respondent Driven Sampling (RDS). RDS is an increasingly popular sampling method used to recruit samples of hidden populations with the aim to provide a probability-based inferential structure for representations of populations such as injection drug users, sex workers, men who have sex with men and population groups whose status characteristics are not likely to be revealed by omnibus survey research because they are rare and socially stigmatized and/or illegal. RDS capitalizes on the social network structure of the hidden population to identify and interview participants. Its validity rests on stringent theoretical assumptions about the referral practices of participants to new participants and the structure of the underlying network that are not observed. Despite significant investments by CDC and similar organizations in RDS, we have few empirical evaluations of its effectiveness at keeping its representation promise. Here we propose to improve RDS for representation of female sex workers in China by moving considerations regarding real-world referral processes from the theoretical to the empirical realms. We accomplish this with a combination of analyses of data we have recently collected through two RDS studies and a venue-based sampling approach in Shanghai and Liuzhou (Guangxi Province). We use this overlapping data collection to observe the social network information embedded in the RDS recruitment process and to realistically simulate RDS settings in order to develop improved RDS estimates adaptive to the observed network referral process. We distill guidelines for researchers using RDS methods on needed steps to improve RDS estimation for representation of other hidden populations.