Respondent-driven sampling (RDS) is a link-tracing survey method for sampling members of a hidden or hard-to-reach population such as drug users, sex workers, or homeless people via their social network. Starting with a set of "seed" subjects, participants use a small number of coupons tagged with a unique code to recruit their social contacts by giving them a coupon. Subjects report their network degree, but not the identities of their contacts. RDS is controversial and researchers disagree about whether it can be used to estimate population-level characteristics of hidden risk groups. In this presentation, I outline four results that permit principled network-based epidemiology from RDS. First, I show that a simple continuous-time model of RDS recruitment implies a well-defined probability distribution on the recruitment-induced subgraph of respondents; the resulting distribution is an exponential random graph model (ERGM). I develop a computationally efficient method for estimating the hidden graph. Second, I show that two sources of dependence in the RDS sample - network homophily and preferential recruitment - are cofounded. However, it is still possible to make valid inferences via nonparametric graph-theoretic identification regions that permit hypothesis testing. Third, I derive conservative standard errors via graph-theoretic bounds for statistical functionals of the induced subgraph and traits of sampled subjects, including estimators of the population mean. Fourth, I describe a simple technique - based on capture-recapture and the network scale-up method - for estimating the size of a hidden population from an RDS sample. I apply these techniques to RDS studies of drug users in Eastern Europe, Russia, and Lebanon.

“Learning about hidden networks by tracing links: statistical approaches for network-based social epidemiology and public health”

Forrest W. Crawford PhD is Assistant Professor, Department of Biostatistics, Yale School of Public Health and Department of Ecology & Evolutionary Biology, Yale University. He is affiliated with the Center for Interdisciplinary Research on AIDS, the Institute for Network Science, the Computational Biology and Bioinformatics program, and the Operations doctoral program at the Yale School of Management. His research interests include networks, graphs, stochastic processes, and optimization for applications in epidemiology, public health, and social science.