

STATISTICS COLLOQUIUM

MONDAY, NOVEMBER 16, 2009

TALK: 4:00 PM — SCIENCE CENTER, ROOM 309

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

“Respondent-Driven Sampling: Degrees of Uncertainty with Uncertain Degrees”

**Joseph Blitzstein
Statistics Department
Harvard University**

ABSTRACT

Respondent-driven sampling (RDS) is a contact-tracing sampling design (introduced by Heckathorn) which is widely and increasingly being used to study hard-to-reach populations such as injection drug users and individuals at high risk of HIV infection. The standard estimators used with RDS make many strong assumptions, including that the degree d of an individual in the sample is known exactly and that this individual is equally likely to recruit any of his or her d neighbors in the social network.

Classical issues of model-based vs. design-based inference arise here, with new challenges from having also to account for the network structure. We explore the bias-variance tradeoff encountered in this setting, and show how the uncertainty in the measurement of degrees propagates into uncertainty in the RDS estimates. These tradeoffs point to the need both for better models for the random walk through the network, and for estimators that are less sensitive to the strong assumptions mentioned above.