Class 13: Respondent-driven sampling

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University

1/3 Background and data collection

Introduction

There are an estimated 38 million people [31.6 million–44.5 million] living with HIV in 2019. In most countries, the disease is concentrated in three high risk groups:

- ▶ drug users
- ▶ commercial sex workers
- ▶ men who have sex with men

Better information about these group can be used to understand and control the spread of HIV/AIDS: "know your epidemic"

Hidden populations

These "hidden" populations are hard to sample because:

► No sampling frame

► Small proportion of the general population

► In some cases, desire to remain anonymous

Previous approaches to the study of hidden populations

Institutional sampling: Samples drug injectors in treatment.

Previous approaches to the study of hidden populations

Institutional sampling: Samples drug injectors in treatment.

Relatively inexpensive, but difficult to generalize from institutionalized population to the non-institutionalized population.

Previous approaches to the study of hidden populations

Targeted sampling (Watters and Biernacki, 1989): Through a combination of methods, accesses populations members outside of institutional settings.

Previous approaches to the study of hidden populations

Targeted sampling (Watters and Biernacki, 1989): Through a combination of methods, accesses populations members outside of institutional settings.

Despite wider coverage, it is not possible clear how to the hidden population because probability of selection is not known.

Previous approaches to the study of hidden populations

Time-location sampling: Construct sampling frame of time/locations,

Previous approaches to the study of hidden populations

Time-location sampling: Construct sampling frame of time/locations, but
- ▶ Data collection is expensive and time consuming

Previous approaches to the study of hidden populations

Time-location sampling: Construct sampling frame of time/locations, but

- ▶ Data collection is expensive and time consuming
- ▶ Requires adjustments for oversampling people who frequently attend elements on sampling frame

Previous approaches to the study of hidden populations

Time-location sampling: Construct sampling frame of time/locations, but
- ▶ Data collection is expensive and time consuming
- ▶ Requires adjustments for oversampling people who frequently attend elements on sampling frame
- ▶ Not all time/location combinations are accessible to researchers

In some cases, it may not be possible to generalize to the hidden population. For example, drug injectors who attend venues accessible to researchers may be different from those who don't.

Another approach: snowball sampling

Instead of thinking of people as atomized units on a sampling frame, think of people as embedded in networks. Friends recruit friends and the sample progresses through the social network. But, conventional wisdom was that it was not possible to make unbiased estimates from snowball samples because they:

- ▶ oversample popular people
- ▶ non-independence of observations (people are similar to their friends)
- ▶ depended on the choice of seeds

Respondent-driven sampling

It turns out that unbiased estimation is possible under certain conditions (Salganik and Heckathorn, 2004), and the process is cheaper and faster than existing methods. Respondent-driven sampling is now being used by many researchers:

Respondent-driven sampling

It turns out that unbiased estimation is possible under certain conditions (Salganik and Heckathorn, 2004), and the process is cheaper and faster than existing methods. Respondent-driven sampling is now being used by many researchers:

- More than 100 studies studies in 29 countries involving more then 30,000 respondents (Malekinejad, et al, 2007)

Respondent-driven sampling

It turns out that unbiased estimation is possible under certain conditions (Salganik and Heckathorn, 2004), and the process is cheaper and faster than existing methods. Respondent-driven sampling is now being used by many researchers:

▶ More than 100 studies studies in 29 countries involving more then 30,000 respondents (Malekinejad, et al, 2007)

▶ Drug injector portion of the U.S. Center for Disease Control (CDC) National HIV Behavioral Surveillance System, repeated cross-sectional study in 25 largest U.S. cities (samples of 500 in each city)
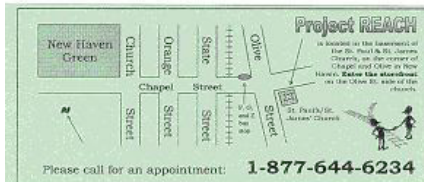
Respondent-driven sampling describes both:

- a method of data collection
- a method of estimation

Sampling

Sample progresses using dual-incentive system (respondents are paid to participant and to recruit others). Participants come to store-front location with coupon (Heckathorn 1997).

Sampling

Sample progresses using dual-incentive system (respondents are paid to participant and to recruit others). Participants come to store-front location with coupon (Heckathorn 1997).

Two additional steps:
- non-duplication

Sampling

Sample progresses using dual-incentive system (respondents are paid to participant and to recruit others). Participants come to store-front location with coupon (Heckathorn 1997).
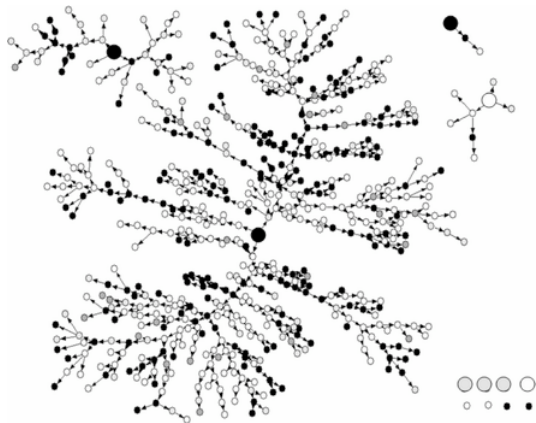
Two additional steps:
- non-duplication
- population verification

Sampling

Recruitment network from a study of drug users in New York City (Abdul-Quader et al., 2006)

- ▶ 8 seeds → 618 drug users
- ▶ 13 weeks

Reach of the sampling method

RDS as a sampling methods seems to work very well in a range of contexts

Reach of the sampling method

RDS as a sampling methods seems to work very well in a range of contexts

- ▶ It has been able to reach populations that the researchers had trouble accessing: men who have sex with men in Kampala, Uganda (Kajubi et al. 2007)

Reach of the sampling method

RDS as a sampling methods seems to work very well in a range of contexts

- ▶ It has been able to reach populations that the researchers had trouble accessing: men who have sex with men in Kampala, Uganda (Kajubi et al. 2007)
- ▶ Seems to work better for drug injectors than sex workers

Reach of the sampling method

RDS as a sampling methods seems to work very well in a range of contexts

- ▶ It has been able to reach populations that the researchers had trouble accessing: men who have sex with men in Kampala, Uganda (Kajubi et al. 2007)
- ▶ Seems to work better for drug injectors than sex workers
- ▶ Even if there are problems with statistical inference, still provides access to the population (McFarland)

But, this sample is not a simple random sample. In the next video we will learn how to make estimates from this kind of data.