Respondent-Driven Sampling: two views

Amber Tomas, Department of Statistics
Jonathan Forney, Department of Politics

April 24, 2014
Hard-to-Reach == Hidden

pathprogram.samhsa.gov
What is the Problem?

- No useful sampling frame
- Rare or hard-to-reach population
- Population or trait may be stigmatized
What is the Problem?

- No useful sampling frame
- Rare or hard-to-reach population
- Population or trait may be stigmatized

RDS attempts to take a representative sample from a hidden population which may be concerned about confidentiality
Possible Sampling Methods

- Time-Location sampling
- Targeted sampling
- Convenience sampling
- Snowball sampling
- Respondent-driven sampling
Hard-to-Reach?

FIGURE 2. Distribution of visible, semi-visible and non-visible FSWs in HCMC and Hai Phong, Vietnam; adjusted proportions (number). Visible FSWs are assessed to be the most visible of the FSWs in Vietnam; semi-visible FSWs are assessed as not always being easily identifiable; and non-visible FSWs are assessed as the most hidden types of FSWs in Vietnam.

From Johnston et al., 2006
Examples of Populations Sampled with RDS

- Homeless
- Female sex workers
- Men who have sex with men
- LGBTs of color
- Injection drug users
- Undocumented immigrants
Examples of Populations Sampled with RDS

- Homeless
- Female sex workers
- Men who have sex with men
- LGBTs of color
- Injection drug users
- Undocumented immigrants

Populations are usually limited geographically to one area, city, or neighborhood.
Examples of Populations Sampled with RDS

- Jazz musicians
- Recent veterans in New York City
- Polish migrants in Oslo/Copenhagen/Dublin
- Migrant workers from Myanmar and Cambodia on the Thai-Cambodia border
- Alcohol use among Muslim undergraduates at college
- Militia members in Sierra Leone
"Respondent Driven Sampling"
Scopus Article Count by Year

This chart shows the total number of documents for this query by Year.
So What is Respondent-Driven Sampling?!
Respondent-Driven Sampling

Trait A

Trait B

\[ N_A = 6 \]

\[ N_B = 12 \]
Respondent-Driven Sampling

Trait A

Trait B

$N_A = 6$

$N_B = 12$
Respondent-Driven Sampling

- RDS is a form of “link-tracing” sampling
- It relies on the population being well-connected, or “networked”
Respondent-Driven Sampling

- RDS is a form of “link-tracing” sampling
- It relies on the population being well-connected, or “networked”
- Members of the population should know others in the population
- They should also know how to contact each other
Respondent-Driven Sampling - Method

• First, “seeds” are selected to participate

• Thereafter, sampling is driven by the respondents

• Respondents are given a number of uniquely identifiable “coupons” to distribute to others they know in the population

• Requires no list, no contact information, no identifiers
Respondent-Driven Sampling

Each coupon has a unique barcode, and details of how to participate
Respondent-Driven Sampling

Each coupon has a unique barcode, and details of how to participate

NCHHSTP/DHAP-SE/BCSB Behavioral Surveillance Team
Respondent-Driven Sampling

First, select seeds

- Trait A
- Trait B

$N_A = 6$
$N_B = 12$

$n_A = 0$
$n_B = 2$
Respondent-Driven Sampling

They recruit their “friends”…

$N_A = 6$
$N_B = 12$
$n_A = 1$
$n_B = 5$
Respondent-Driven Sampling

Then they recruit their “friends”…
Respondent-Driven Sampling

People with lots of connections are more likely to be selected

$N_A = 6$
$N_B = 12$

$n_A = 2$
$n_B = 7$
Challenges

• Respondents do not pass on coupons

• Coupon recipients do not participate
  • Lack of incentives
  • Confidentiality concerns
  • Access to interview sites

• “Homophily”: people tend to know others similar to themselves, so samples can become unbalanced

• Large design effects, often ~ 5
Challenges

Figure 7 Recruitment chain of FSWs by place of work

Chinaglia et al., 2008
Design

• **Start with an ethnographic study of the population**
  - Assess connectivity
  - Investigate interest in the study, likely recruiting behaviors, value of incentives, confidentiality concerns, understanding of technical terms, etc

• Carefully select interview sites

• Train interviewers to train respondents

• Carefully consider selection of seeds
To reduce bias:

- Select diverse seeds
- Use 2 or 3 coupons, to encourage long (rather than fat) recruitment chains
- Sampling must be without replacement
- Try to ascertain how many other people in the population a respondent knows
To encourage recruitment:

- Respondents may need incentives to participate, and to recruit others
  - In the US, usually ~ $25 to participate, $10 for each recruit
  - Health screening may also act as an incentive

- Carefully consider the study sites and interview experience
What’s all this I hear about unbiased estimation?
“It relies on veterans’ social networks and is in part a social network method that allows researchers to obtain unbiased estimates for the target population”

– Alexander Bennett, Ph.D.

“In order to ensure the sample is representative, the investigators will use [RDS]... to obtain a sample that is theoretically independent of its non-random starting points.”

RDS Estimates Are Not Unbiased

• These claims are based on a mathematical model
RDS Estimates Are Not Unbiased

• These claims are based on a mathematical model
RDS Estimates Are Not Unbiased

- These claims are based on a mathematical model
RDS Estimates Are Not Unbiased

- RDS does not produce a probability sample
RDS Estimates Are Not Unbiased

- RDS does not produce a probability sample
- Models are used to **estimate** the selection probabilities based on:
  - The local network characteristics of respondents
  - Observed patterns in recruitment
RDS Estimates Are Not Unbiased

- RDS does not produce a probability sample
- Models are used to estimate the selection probabilities based on:
  - The local network characteristics of respondents
  - Observed patterns in recruitment
- The true selection probabilities will always depend on unknown social processes
A simple model assumes:

- Population is completely networked
- Seeds are selected at random, with probability proportional to their number of “friends”
- Each respondent recruits exactly 1 friend
- Recruitment is uniformly at random from friends
- All recruits respond
- Reported network size is accurate
- Sampling is with replacement
RDS Estimates Are Not Unbiased

- Under these assumptions the probability of selection is proportional to the number of friends

- Possibly the biggest source of bias is “differential recruitment”
  - people recruit others based on hidden characteristics
Conclusions

• RDS facilitates recruitment of well-networked hidden and hard-to-reach populations

• Estimated selection probabilities can be used to calculate population estimates

• The social and cultural characteristics of the population should always be considered
  • Design
  • Data collection
  • Analysis
References


More References


More References
