Two Applications of Respondent Driven Sampling: Ethnic Minorities and Illicit Substance Users

Sunghee Lee, Ai Rene Ong, Michael Elliott
University of Michigan

Workshop on Improving Health Research for Small Populations
National Academy of Sciences, Engineering and Medicine
January 18, 2018
Introduction
Respondent Driven Sampling – 1

• Growing interest in studying hard-to-reach, rare, elusive, hidden populations
  – HIV at-risk population: MSMs, Sex workers, IDUs
  – LGBT populations
  – Recent immigrants

• No clear and practical solution with probability sampling
  – High screening costs
  – Hesitant to be identified
Respondent Driven Sampling – 2

• Proposed by Heckathorn (1997, 2002)
• Popular usage in public health
• Exploits social networks among rare population members for sampling purposes
  – Sampled members also play a role of a recruiter
  – Incentivized recruitment from own network through coupons and this continues in waves/chains
  – Recruitment assumed to be random within each individual’s network and to follow memory-less Markov chain and reach equilibrium
    • Under these assumptions, unbiased estimators can be obtained after equilibrium using weights equal to the number of nodes for a subject’s recruiter.
Respondent Driven Sampling – 3

**WAVE 1**
- Seed 1
  - Recruit 1
  - Recruit 2
  - Recruit 3
- Seed 2
- Seed 3
- Seed S

**WAVE 2**
- Recruit 1
  - Recruit 1
  - Recruit 2
  - Recruit 3

**WAVE 3**
- Recruit 1
  - Recruit 1
  - Recruit 2
  - Recruit 3

**WAVE W**
- Recruit 1
- Recruit 2
- Recruit 3
- Recruit R-2
- Recruit R-1
- Recruit R

---

S. Lee

Workshop on Improving Health Research for Small Populations
Respondent Driven Sampling – 4

<table>
<thead>
<tr>
<th>WAVE 1</th>
<th>WAVE 2</th>
<th>WAVE 3</th>
<th>WAVE W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed 1</td>
<td>Recruit 1</td>
<td>Recruit 1</td>
<td>Recruit 1</td>
</tr>
<tr>
<td></td>
<td>Recruit 2</td>
<td>Recruit 2</td>
<td>Recruit 2</td>
</tr>
<tr>
<td></td>
<td>Recruit 3</td>
<td>Recruit 3</td>
<td>Recruit 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.......</td>
</tr>
</tbody>
</table>

Recruitment Chain:

Seed 1

Seed 2

Seed 3

Seed S

WAVE 1

WAVE 2

WAVE 3

WAVE W

Recruit R

Recruit R -1

Recruit R -2

Recruit 1

Recruit 2

Recruit 3

Recruit Coupon

S. Lee

Workshop on Improving Health Research for Small Populations
Network Sampling vs. RDS

Similar:
• Rely on social networks

Different:
• Network specification
  – NS: biological siblings, immediate family members
  – RDS: jazz musicians
• Who selects the sample
  – NS: researchers
  – RDS: study participants
Application 1:  
**Project PATH** (Positive Attitudes Towards Health)

Funded by the National Science Foundation (GRANT NUMBER SES-1461470)
PATH Data Collection

• Injection drug users in Southeast Michigan
• Phone screener
  → In-person screener + Main interview + ~3 Coupons
  → In-person follow-up interview
• Data collection sites
  – Detroit: Urban; Tues, Thur @ Detroit Center
  – Macomb: Suburban; Weds @ County PH Depart
  – St. Clair: Rural; Mon (+Weds) @ County PH Depart
  – 4 interviewers rotating between sites
• Field Period: 5/1/2017 – 10/31/2017
## Demographics

<table>
<thead>
<tr>
<th></th>
<th>Detroit</th>
<th>St Clair/ Macomb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (avg)</td>
<td>56 yrs</td>
<td>40 yrs</td>
</tr>
<tr>
<td>Age: &lt;30 years old</td>
<td>2%</td>
<td>32%</td>
</tr>
<tr>
<td>Male</td>
<td>68%</td>
<td>53%</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>11%</td>
<td>73%</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>81%</td>
<td>16%</td>
</tr>
<tr>
<td>Education: &lt;High School</td>
<td>32%</td>
<td>18%</td>
</tr>
<tr>
<td>Employed</td>
<td>8%</td>
<td>18%</td>
</tr>
<tr>
<td>Ever homeless past 12 mos</td>
<td>40%</td>
<td>56%</td>
</tr>
</tbody>
</table>
Substance Use

CURRENT SUBSTANCE USE

Most frequently used: Heroin
98% Detroit
78% St. Clair/Macomb
Application 2: Health and Life Study of Koreans (HLSK)

Funded by the National Science Foundation (GRANT NUMBER SES-1461470)
HLSK

• Targets foreign-born Korean American adults in
  – Los Angeles County
  – State of Michigan

• Web-RDS survey
  [http://sites.lsa.umich.edu/korean-healthlife-study/](http://sites.lsa.umich.edu/korean-healthlife-study/)
  – Unique number required for participation
  – Incentive payment through checks

• Target n=800 (currently ~600)

• Benchmarks from American Community Survey
HLSK Formative Research

• 3 rounds of focus group discussions
  – ~30 participants; 2 rounds in Korean and 1 in English
  – Discussion focused on
    • Web surveys
      → URL, Web site contents, etc.
    • Concept of RDS
    • Coupons
      → Up to 2 coupons
      → “Expire” in 2 weeks
    • Level of incentives
      → $20 for main, $5 for follow-up, $0 for recruitment
HLSK Data Collection

• Started with 12 seeds in LA in June 2016
• MI added in November 2016

• LA seeds (initially)
  – Recruited through referral
  – Balanced on gender, age, dominant language
  – In-person introduction about the study

→ It became clear the protocols would not work
  – Provide recruitment incentives
  – Add more seeds
HLSK Data Collection Progress

n=306
110 seeds
578 coupons

n=249
85 seeds
477 coupons
HLSK vs. ACS – 1

• American Community Survey 2011-2015 data
• HLSK sample estimates
  – Unweighted (UW)
  – RDS-I
  – Weighted: RDS-II
  – Weighted: Post-stratification (PS) by age, sex, educ
  – Weighted: RDS-II + PS
HLSK vs. ACS – 3

Benchmarks and Sample Estimates: LA (n=309)

Proportion

0.00  0.25  0.50  0.75  1.00

Age ≥ 30  Male  Educ ≥ College  Married  Worked past wk  Arrived ≥ 2000  US Citizen  High Eng Prof  ADL/ADL

ACS
HLSK vs. ACS – 4

Benchmarks and Sample Estimates: LA (n=309)
Summary
What did we learn?

• Non-cooperation is an issue for generating long chains (memorylessness unlikely)
• Had to improvise to make RDS “work”
• Sample size (hence, chain length) is a random variable affected by many (mostly unknown) factors
• Inferences limited
• YET, difficult-to sample groups can be recruited
  – E.g., highly-educated young recent immigrants
Where should we go?

• Non-cooperation is critical for
  – meeting theoretical assumptions (hence, inferences)
  – study design
  – replications of the same study

• Yet to be addressed in the literature and accounted for in inferences
Thank you

sungheel@umich.edu