

Respondent Driven Sampling Design Considerations

Ai Rene Ong, Michael Elliott, Sunghee Lee

AAPOR 2022

Acknowledgement: This research was supported by NIH, grant number R01 AG060936 (PI: Lee)



INSTITUTE FOR SOCIAL RESEARCH

PROGRAM IN SURVEY AND DATA SCIENCE

UNIVERSITY OF MICHIGAN

Background (1)

- Respondent driven sampling (RDS) is commonly used to sample hard-to-reach populations
- In RDS, respondents are asked to invite a fixed number of people they know to participate in the study
 - Chain-referral process continues until target sample size achieved
- Despite its popularity, difficulties in the field have been reported and resulted in researchers needing to change their strategy on an ad-hoc basis

Background (2)

- Lacking standardized guidelines
 - RDS design (e.g., how many seeds?)
 - Implementation (e.g., what do coupons look like?)
- Lacking transparency
 - Recent (in 2015) development about what RDS studies should report; STROBE-RDS
 - Adherence to reporting standards is not 100%, hampering the quality and utility of systematic reviews to give recommendations on best practices for RDS design

Objectives

- As there is a lack of standardization and transparency on the reporting of RDS study design information, approaching the researchers to ask about the design of their study would fill in the gaps on what is reported about their RDS studies
- Current study: Uses data on a survey of RDS researchers on their RDS studies to explore the RDS design features that are associated with peer recruitment productivity

Data and Methods (1)

- Survey of RDS researchers who have published an article or have applied for a grant to collect RDS data between 2009-2020
 - Respondents were asked to report on the details of their RDS research design (e.g., number of seeds, sample size, incentive amount)
- Sampling frame was a database of RDS researchers built using a systematic search in PubMed, JSTOR, and Web of Science, and from the NIH and NSF database
 - Search terms: “respondent driven sampling” OR “respondent-driven sampling”
 - N = 344 unique researchers
- 122 researchers responded to the survey (RR=35.8%)

Data and Methods (2)

- As the sampling frame represents a census of RDS researchers, a finite population correction factor of $1-f = \sqrt{.64}$ has been applied to the standard errors during analysis
- RDS productivity was examined using a multivariate linear model, with peer recruitment productivity as a continuous outcome
- Interactions were explored where it is sensible

Data and Methods (3)

- RDS productivity measured as:

- $Overall\ productivity = \frac{Achieved\ sample\ size}{Target\ sample\ size}$

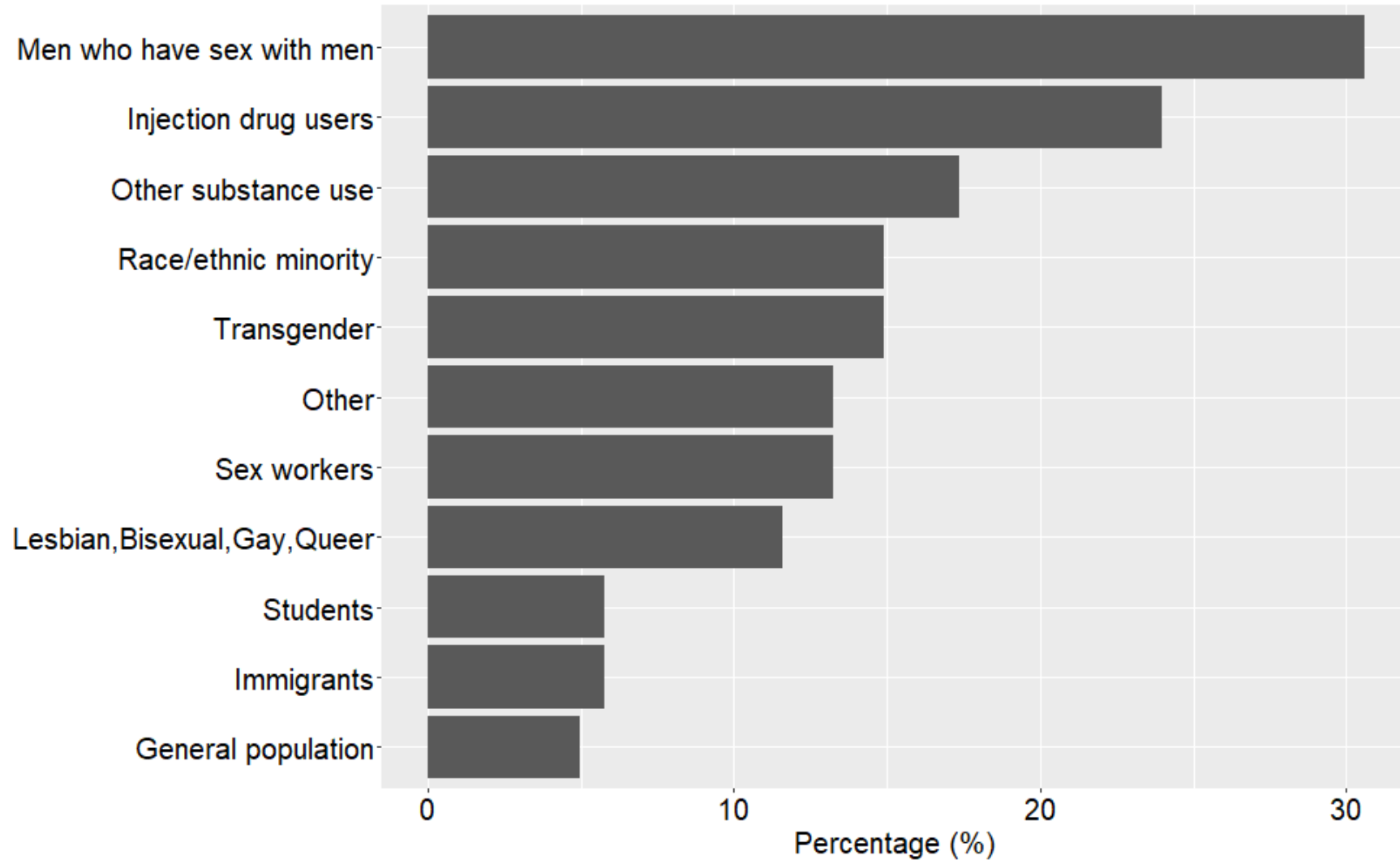
- $Seed\ productivity = \frac{Achieved\ sample\ size}{Final\ seed\ size}$

- Values lower than 1 for overall productivity indicate less than ideal productivity (i.e., did not achieve target sample size)
- Seed productivity measures the how efficient the seeds are in getting respondents on average
 - Due to very large and skewed spread of the values, the natural log of seed productivity is used instead

Data and Methods (4)

- Predictors of productivity considered:
 - Target population of the RDS study
 - Web mode (1 = Web, 0 = Not web)
 - Formative research conducted (1 = Yes, 0 = No)
 - Target population in the formative study (Target population only, target pop. and other informants, target pop. not included)
 - Type of formative research conducted (Focus group + In-depth interview + Field observation vs. Focus group + In-depth interview only vs. Field observation only vs. Other)
 - More than one type of recruitment instruction given (1 = Yes, 0 = No)
 - Incentive in USD
 - Duration of data collection in years
 - Location (In U.S. only vs. Outside of U.S. included)
 - Seed difference: Difference in the planned vs. achieved number of seeds

Target populations

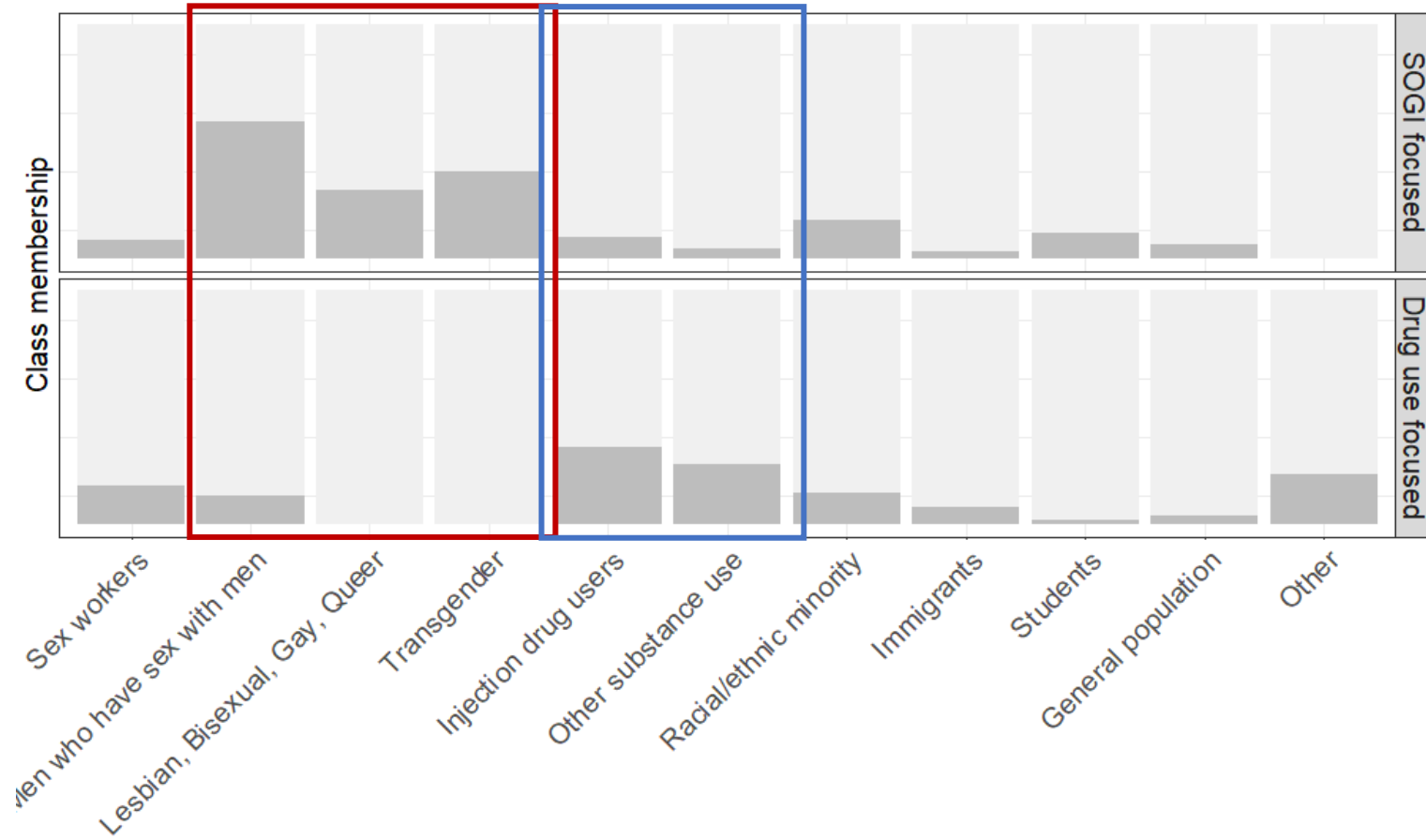


*Note: A single study may have more than one target population (e.g., injection drug users and other substance users)

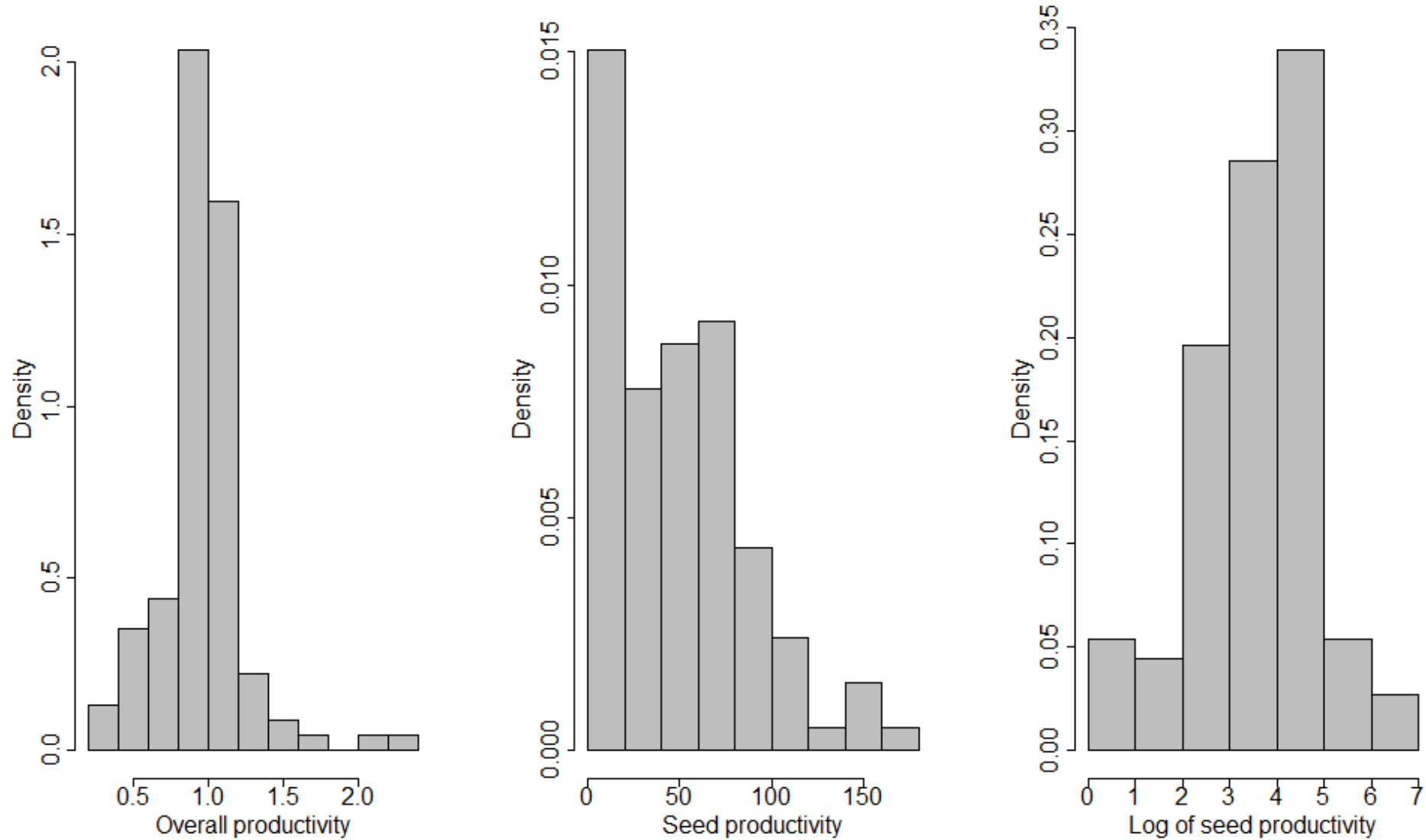
Data and Methods (5)

- A given RDS study may target multiple target populations
- Latent class analysis used to identify population grouping patterns
 - Two-class solution yielded the best fit (tested 2 to 6 classes)
 - Class 1: Sexual orientation and gender minority
 - Class 2: Substance use disorders and others (e.g., racial/ethnic minority)

Latent Class Analysis



Distribution of the outcomes



Results: Overall productivity

	Est (Std. Err)
Target pop – Drug use focused	.392 (.117)**
Web mode	-.204 (.079)*
Formative research sample (Ref: Target population only)	
- Target pop + other informants	.016 (.073)
- Target pop not included	.364 (.142)*
Formative research sample x Target population: Drug use focused (Ref: Target population only)	
- Target pop + other informants	-.424 (.122)**
- Target pop not included	-.344 (.131)*

*p < .05, **p < .01; Adjusted R² = 0.19

Note: Only the model with interactions shown here. Only 16 studies are web-RDS studies. Fitted this model after removing outliers and the direction of the estimates and strength of associations are the similar

	Est (Std. Err)
Location – U.S. only	-.025 (.049)
Duration of data collection (years)	.012 (.691)
Seed difference/100	-.019 (.042)
Amount in incentive (USD)/10	-.008 (.013)
More than one type of recruitment instruction given	.029 (.056)
Formative research conducted	.152 (.123)
Formative research type (Ref: Interview + focus group only)	
- Interview + focus group + field observation	-.027 (.061)
- Field observation only	-.482 (.241)*
- Other	-.180 (.120)

*p < .05, **p < .01; Adjusted R² = 0.19

Results: Seed productivity

	Est (Std. Err)
Target pop – Drug use focused	.286 (.169)
Location – U.S. only	-.537 (.212)*
Web mode	-.197 (.355)
Duration of data collection (years)	-.122 (.127)
Seed difference/100	-1.273 (.218)**
Amount in incentive (USD)/10	.004 (.056)
Formative research conducted	.083 (.656)
More than one type of recruitment instruction given	.080 (.170)
Formative research type (Ref: Interview + focus group only)	
- Interview + focus group + field observation	.145 (.399)
- Field observation only	-1.968 (1.069)
- Other	-1.082 (.648)
Formative research sample (Ref: Target population only)	
- Target pop + other informants	.017 (.192)
- Target pop not included	1.242 (.731)

*p < .05, **p < .01; Adjusted R² = 0.24

Summary

- Productivity in RDS studies administered over the web seems to be lower than those administered in other modes
 - Web-RDS use is reported much less frequently in the publications and grant proposals than other modes (in-person interview and telephone interviews), so this finding should be interpreted cautiously
- Not including the target population in the formative research seems to result in RDS studies with better overall productivity
 - This relationship is moderated by the type of target population
 - Lack of information on who are the other informants
- Location of the study is weakly associated with seed productivity
- Most of the measures tested are not associated with seed productivity
 - Chain length might be a better indicator of seed productivity than the current measure

Thank you

Contact: aireneo@umich.edu

Target population x Format research target interaction

