Randomized Recruitment-Driven Sampling

Adam Visokay
University of Washington

Session on Special Populations

The Team

Rachel Heath

Laura Boudreau

Tyler McCormick

UNIVERSITY of WASHINGTON

IMPRS
for Population,
Health and Data Science
INTERNATIONAL MAX PLANCK
RESEARCH SCHOOL

Background

2013 Rana Plaza Collapse

Background

2013 Rana Plaza Collapse

Journal of Economic Behavior & Organization
Volume 219, March 2024, Pages 196-213

Research Paper

Migrants, experience, and working conditions in Bangladeshi garment factories ☆

Laura Boudreau [a], Rachel Heath [b], Tyler H. McCormick [c]

Journal of Development Economics
Volume 163, June 2023, 103107

Regular article

The effects of international scrutiny on manufacturing workers: Evidence from the Rana Plaza collapse in Bangladesh ☆

Laurent Bossavie [a], Yoonyoung Cho [a], Rachel Heath [b]

Journal of Political Economy > Volume 126, Number 4

PREVIOUS ARTICLE NEXT ARTICLE

Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories

Rachel Heath

COVID-19

How can we administer surveys in disaster settings?

World Health Organization

COVID-19

How can we administer surveys in disaster settings?

Remote access

Valid population inference

COVID-19

How can we administer surveys in disaster settings?

Remote access -> Cell phones

Valid population inference -> Network sampling

JOURNAL ARTICLE

Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations*

Douglas D. Heckathorn Author Notes

Social Problems, Volume 44, Issue 2, 1 May 1997, Pages 174–199,
https://doi.org/10.2307/3096941
Published: 31 July 2014

‘RDS’ was developed to access otherwise hard to reach (HTR) populations

Recruitment

Peer-to-peer

Dual incentive (physical voucher)

Unique identifiers to keep track of respondents

Estimation - Volz-Heckathorn (2008)

d , degree, number of connections

High degree, weighted ⇩

Low degree, weighted ⇧

$$\hat{\mu}_{\mathrm{VH}} = \frac{\sum_{i=1}^{n} \frac{x_i}{d_i}}{\sum_{i=1}^{n} \frac{1}{d_i}}$$

Remote access

Cell phones: Inghels et al. (2021)

Internet, cell phones: Pham et al. (2023)

Smart-phones: Bramley (2022)

Web-based with SMS/WhatsApp: McGowan et al. (2023)

RDS Assumptions

Sociological Methodology

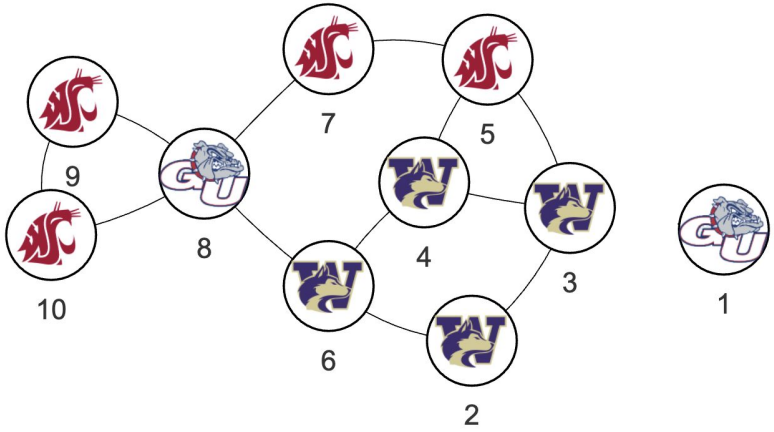
Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling

Matthew J. Salganik, Douglas D. Heckathorn

First published: 26 November 2004 | https://doi.org/10.1111/j.0081-1750.2004.00152.x |

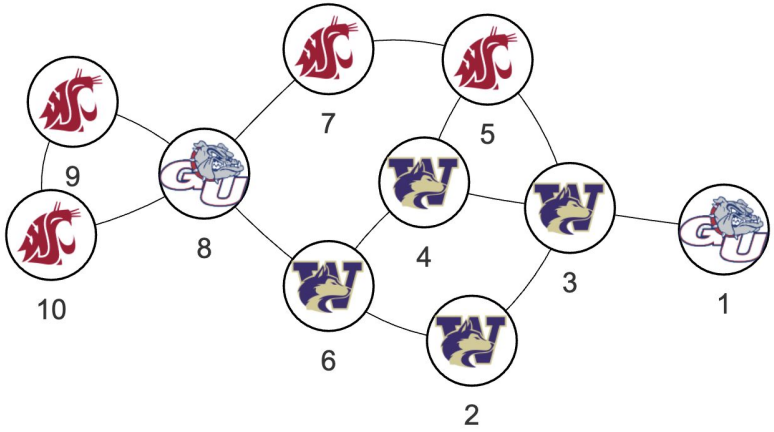
RDS Assumptions

1. Single Component



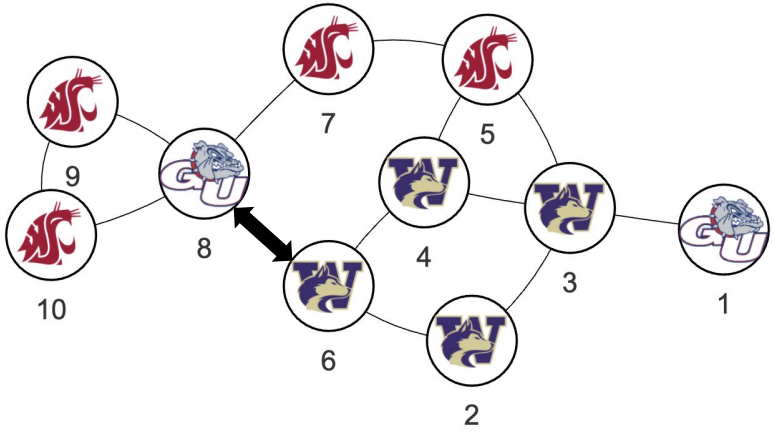
RDS Assumptions

1. Single Component



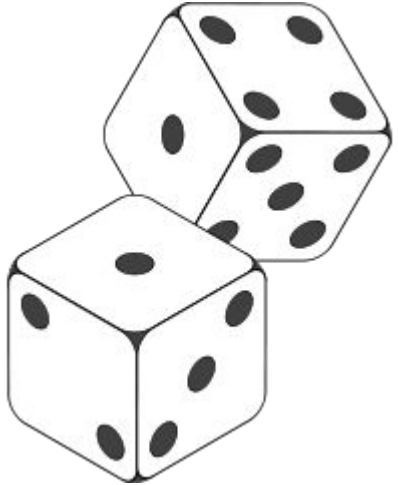
RDS Assumptions

1. Single Component
2. Reciprocity



RDS Assumptions

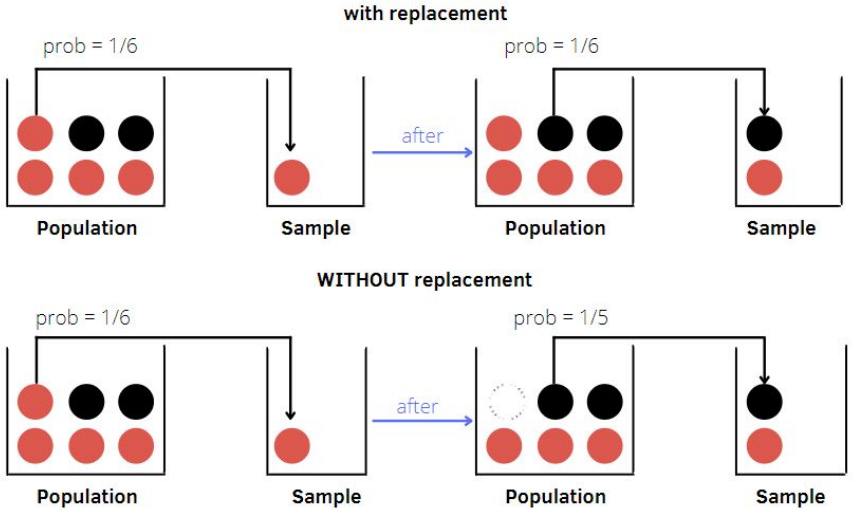
1. Single Component
2. Reciprocity
3. Random referrals



https://en.wikipedia.org/wiki/File:2-Dice-Icon.svg

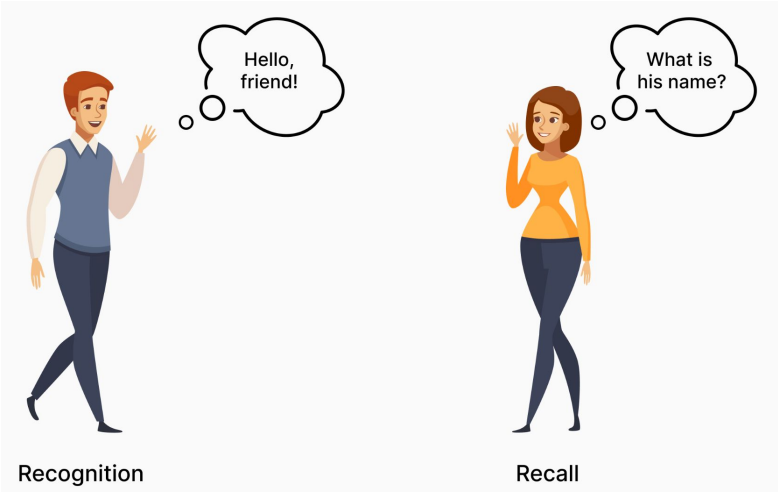
RDS Assumptions

1. Single Component
2. Reciprocity
3. Random referrals
4. Sampling with replacement



https://towardsdatascience.com/an-introduction-to-probability-sampling-methods-7a936e486b5/

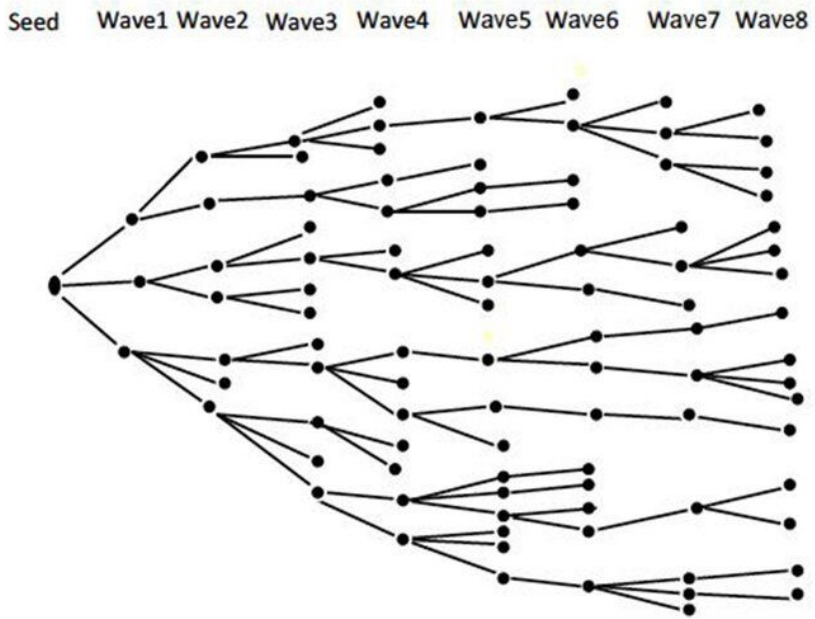
RDS Assumptions



1. Single Component
2. Reciprocity
3. Random referrals
4. Sampling with replacement
5. Degree (number of contacts) is accurately reported

https://think360studio.com/blog/recall-recognition

RDS Assumptions



1. Single Component
2. Reciprocity
3. Random referrals
4. Sampling with replacement
5. Degree (number of contacts) is accurately reported
6. Sufficient recruitment chains

Aysha Twoseİlgi and Bozdağİlgi Bozdağ (2019)

RDS Assumptions

1. Single Component
2. Reciprocity
3. Random referrals
4. Sampling with replacement
5. Degree (number of contacts) is accurately reported
6. Sufficient recruitment chains

https://en.wikipedia.org/wiki/File:2-Dice-Icon.svg

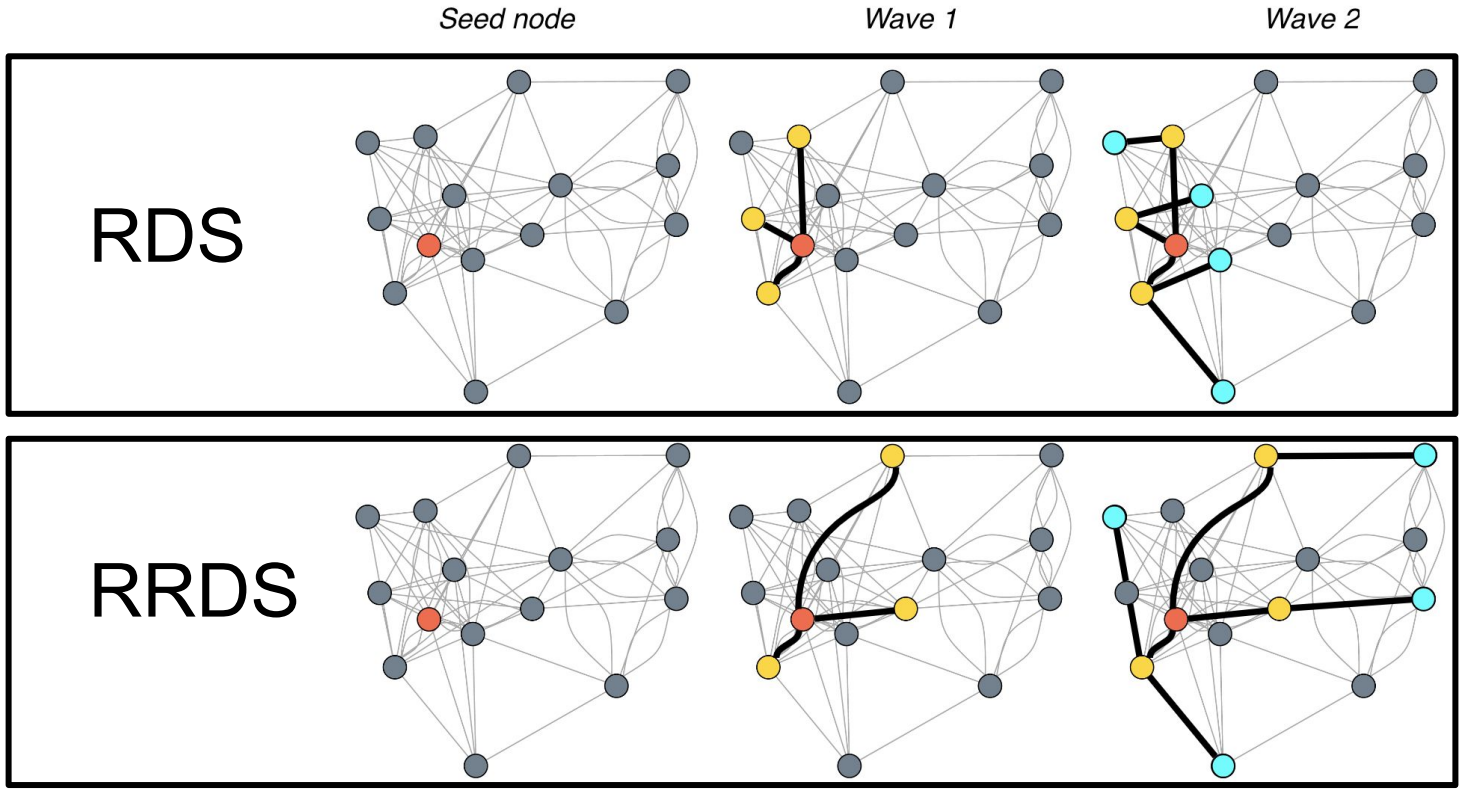
Randomized Recruitment-Driven Sampling (RRDS)

We formalize RRDS as a two-stage recruitment procedure. For each respondent i:

1. **Nomination stage:** Respondent i produces a list L_i of size m_i from their eligible network neighborhood N_i of size k_i (the respondent's degree).

2. **Recruitment stage:** The researcher samples s_i individuals uniformly at random from L_i.

Randomized Recruitment-Driven Sampling (RRDS)

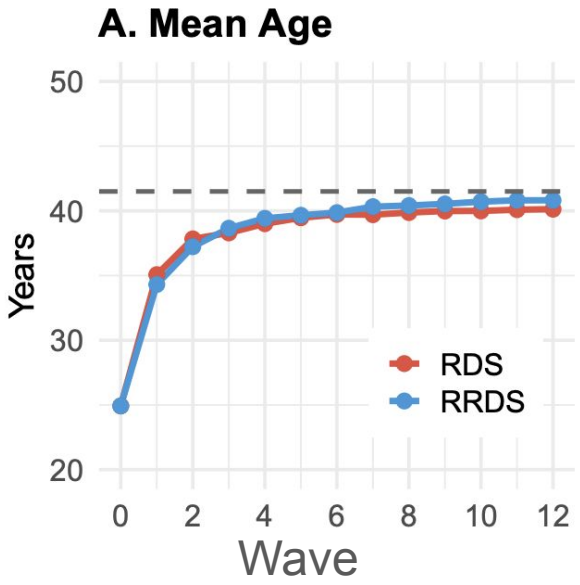


Simulation Study

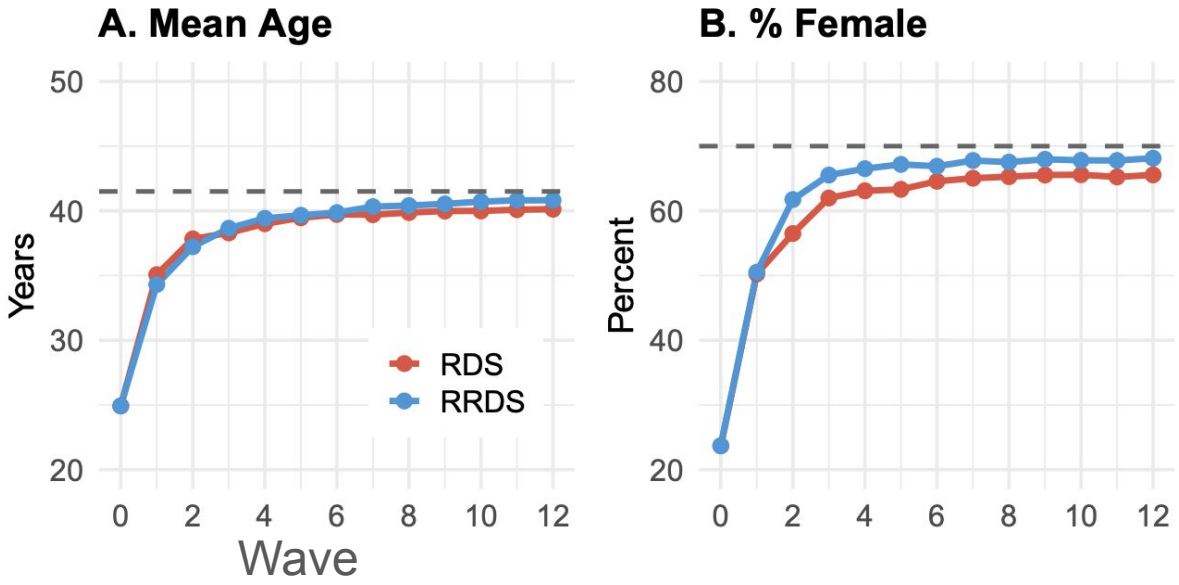
Generate a network of 10,000 nodes.

Parallel data collection comparing traditional RDS with RRDS.

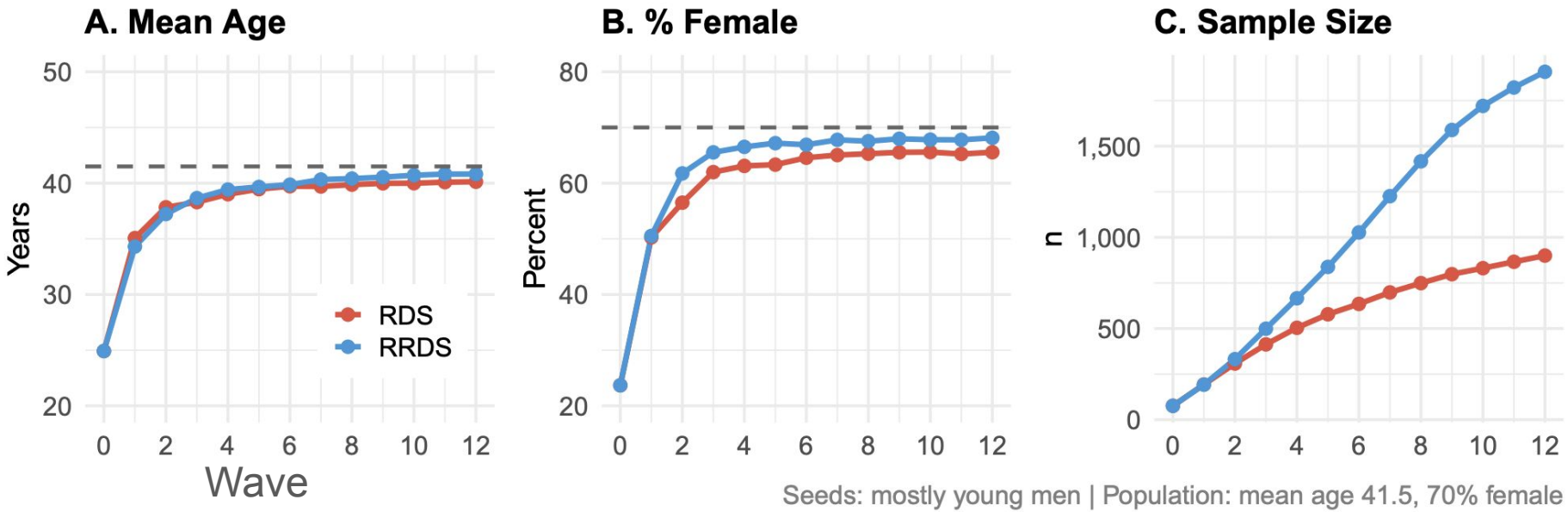
Simulation Study



Simulation Study



Simulation Study



Empirical Example: Garment workers in Bangladesh

2017 survey as baseline for validation.

Parallel RDS and RRDS arms.

Data collected between 2020-2021.

Elicit respondent’s call log, randomly select contacts.

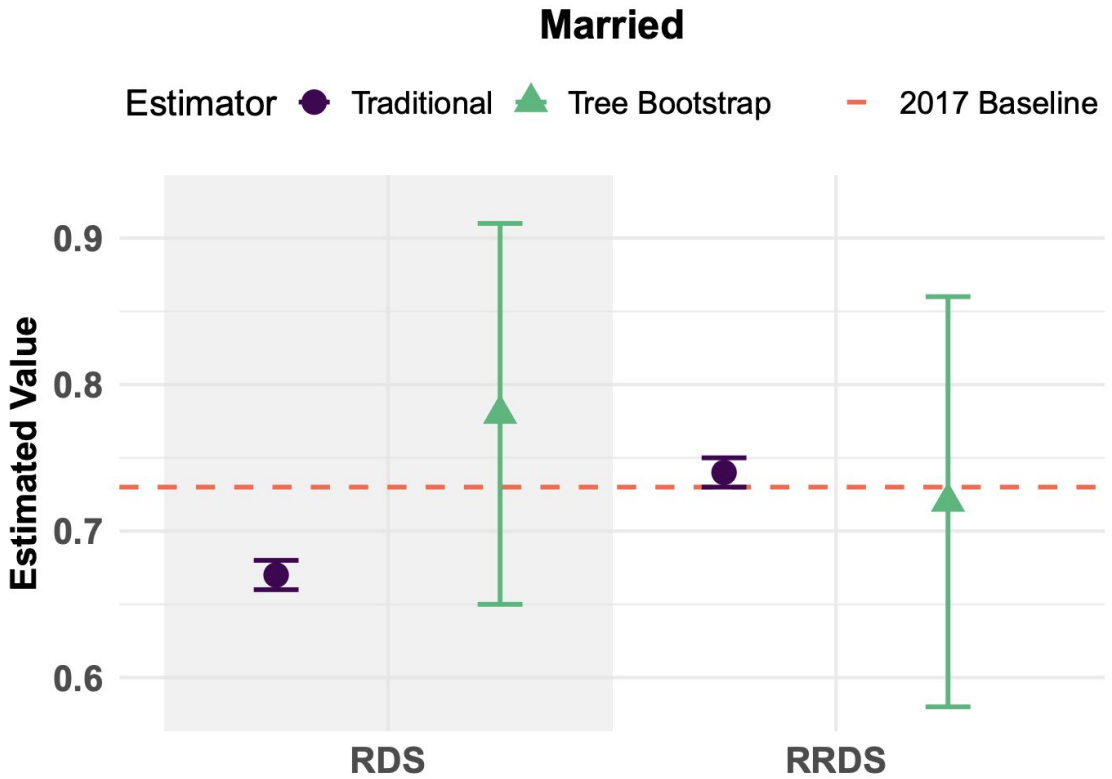
Empirical Example: Garment workers in Bangladesh

Women were far less likely to provide referral contacts than men.

Adapted data collection to follow all referrals from women, still insufficient data.

Focus our results on Male respondents.

Empirical Example: Garment workers in Bangladesh



Empirical Example: Garment workers in Bangladesh

Table 1: Comparison of Recruitment Methods and Estimators. The RRDS recruitment strategy with the Tree Bootstrap estimator produces the least biased estimates and has the best coverage (95% CI containing the baseline) of the underlying population.

Recruitment	**Estimator**	**Bias (RMSE)**	**CI Coverage (Count/6)**
RDS	Volz-Heckathorn	3.713	0/6
RDS	Tree Bootstrap	3.118	4/6
RRDS	Volz-Heckathorn	3.198	1/6
RRDS	Tree Bootstrap	**2.503**	**5/6**

Conclusion

Promising results, but contextual.

Wider use than just disasters / cell phones!

More experiments!

Acknowledgements:

Partial support for this research came from a Shanahan Endowment Fellowship and a Eunice Kennedy Shriver National Institute of Child Health and Human Development training grant, T32 HD101442, to the Center for Studies in Demography & Ecology at the University of Washington. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support for this research came from support to CSDE from the College of Arts & Sciences, the UW Provost, eSciences Institute, the Evans School of Public Policy & Governance, College of Built Environment, School of Public Health, the Foster School of Business, and the School of Social Work.

I also gratefully acknowledge the resources provided by the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS).

Thank you!

Adam Visokay

avisokay@uw.edu

avisokay.github.io