
IMPROVING VALIDITY IN WEB SURVEYS WITH HARD-TO-REACH TARGETS: ONLINE RDS METHODOLOGY

5-th Internet Survey Methods Workshop, The Hague, 29-31 August 2011

Aigul Mavletova, PhD

National Research University – Higher School of Economics

Postdoctoral Researcher, Bonn University

Sample design in web surveys

REFERENCE POPULATION=	SAMPLING
SURVEY SAMPLE	Researcher does not control the selection process. Voluntary participation.
WEB-SITE VISITORS	Random sample (exit-polls type)
SOCIAL GROUPS (Rare population)	Adaptive sampling
INTERNET USERS	Probability sampling based on offline sampling frame
POPULATION	Probability sampling of the population based on offline sampling frame

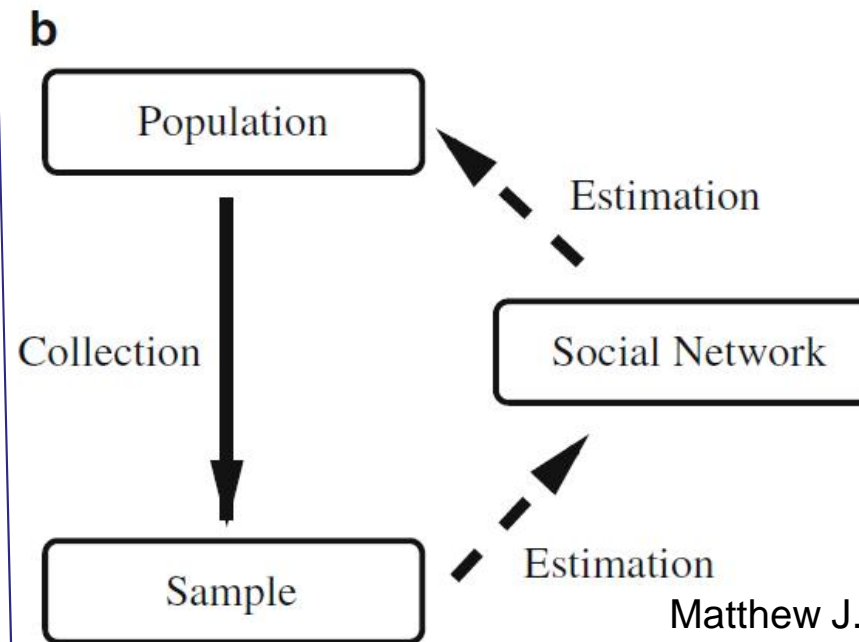
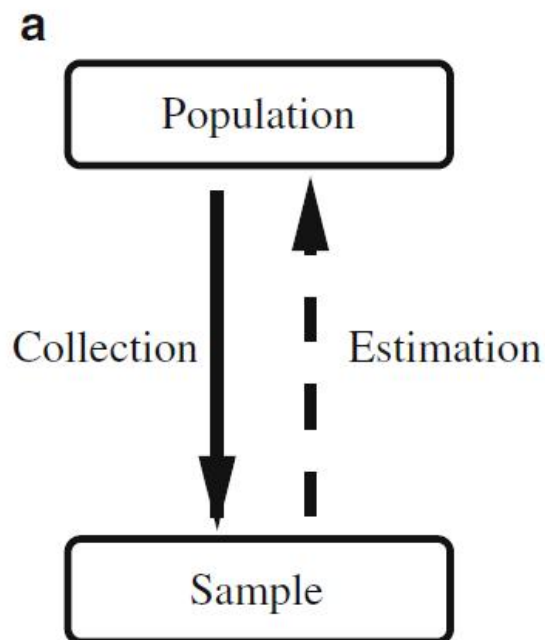
Respondent-driven sampling

- Methodological issue in the surveys with hard-to-access groups: hardly possible to draw random sample
- Researchers use network-based adaptive sampling
- Method which can be efficiently applied in Internet-based surveys for hard-to-reach target is a respondent-driven sampling

Respondent-driven sampling

BASIC IDEA:

Respondents are selected not from a sampling frame but from the participant's social networks. The estimation process should not be directly based on the sample but on social network estimates.



Matthew J. Salganik

BASIC PROCEDURES:

- Researcher selects seeds (initial respondents).
- Seeds recruit other participants.
- Quota on the number of the participants recruited.
- Incentives for participation and recruitment.

Online RDS

- D.Heckathorn, C.Weinert
- Target group: Cornell University students
- In 2004: 150 students (for 72 hours, max.incentive - \$55)
- In 2008: 369 students (for 6 weeks, max.incentive - \$25)

RDS online experiment: casino gamblers

- Goal : to test online RDS methodology and assess it's applicability towards studying hard-to-reach groups.
- Online survey of the casino gamblers who play in casinos or/and online casinos.
- Sample size: 99 respondents
- Coverage: Moscow
- Incentives: no
- Fieldwork: May-August 2009

Seeds	4
Sample size	99
Number of the waves	5
Number of recruits	6

RDS online experiment: casino gamblers

- 4 seeds in the survey.

The seeds were selected according to the following criteria:

- Have a good number of friends who play in casino.
- Are interested in the study and can involve other participants.
- Have different socio-demographic profile.

Basic RDS assumptions

1. Reciprocal connections between recruiter and respondent.

- Who sent the link to online questionnaire?
- Was the link sent by a friend, acquaintance or a stranger?

All ties in the survey were reciprocal.

Basic RDS assumptions

2. Peer recruitment is a random selection from the recruiter's network.

- How many individuals who gamble in casinos / online casinos you know? What % among them males, what – females?

-What percent among them play in “offline” casino, online casinos only, and what – both in “offline” and online casino?

Selection among peers was non-random.

Basic RDS assumptions

3. Respondents can accurately report their personal network size, defined as the number of acquaintances who fall within the target population.

Respondents *were not always able to calculate* how many of their friends play in casino, how many - only in online casinos, and how many – in both “offline” and online casino.

Basic RDS assumptions

4. Each respondent recruits a single peer.

This condition *is hardly feasible in any RDS study.*

Basic RDS assumptions

5. Respondents are linked by a network composed of a single component. In other words, each respondent can be recruited by a peer after a certain number of waves.

The assumption has not been met for those individuals who play in online casinos solely.

They are not familiar with the most of the players they "meet" online. While among "offline" casino players the social networks are larger and closer.

RDS estimates efficiency

Criteria validity: comparison the socio-demographic profile of the gamblers with the estimates of the Fund “Social Opinion” (FSO) (face-to-face survey with probability sampling, 2006).

2 differences between RDS and FSO estimates:

- (1) FSO did not include those gamblers who play in online casinos only.
- (2) RDS included only those who have Internet access

RDS estimates efficiency

Recruiter's gender	Males	Females	TOTAL
Males			
Number of the respondents	69	10	79
Selection probability (S)	87%	13%	100%
Adjusted number of the respondents	71.8	10.4	82.2
Females			
Number of the respondents	13	3	16
Selection probability (S)	81%	19%	100%
Adjusted number of the respondents	10.4	2.4	11.8
TOTAL:			
Number of the respondents	82	13	95
Total number of the respondents	84	15	99
Group proportion in sample	0.85	0.15	
Equilibrium proportion	0.87	0.13	
Sampling weight	0.87	1.76	
Degree component	0.85	1.97	
Recruitment component	1.02	0.89	
Mean degree (adjusted estimate)	11.5	5.0	
Estimates	0.73	0.27	

RDS estimates efficiency

	Under 35 y.o.	35+	TOTAL
Under 35 y.o.			
Number of the respondents	32	17	49
Selection probability (S)	65%	35%	100%
Adjusted number of the respondents	34.5	18.3	41.8
35+			
Number of the respondents	20	26	46
Selection probability (S)	43.5%	56.5%	100%
Adjusted number of the respondents	18.3	23.8	42.1
TOTAL:			
Number of the respondents	52	43	95
Total number of the respondents	54	45	99
Group proportion in sample	0.55	0.45	
Equilibrium proportion	0.56	0.44	
Sampling weight	1.28	0.66	
Degree component	1.26	0.68	
Recruitment component	1.02	0.98	
Mean degree (adjusted estimate)	7.56	14.37	
Homophilia indicator	-0.07	0.38	
Estimate	0.70	0.30	

RDS estimates efficiency

Estimates	Males	Females
Online RDS estimate	0.73	0.27
Fund “Social Opinion” estimate	0.73	0.27

Estimates	Up to 35 y.o.	35+
Online RDS estimate	0.70	0.30
Fund “Social Opinion” estimate	0.60	0.40

Advantages and limitations of online RDS methodology

ADVANTAGES

1. Access to hard-to-reach targets

2. Time and organizational costs are lower

3. Statistical inference about social group can be drawn

LIMITATIONS

1. Not high cooperation and confidentiality level

2. Organizational difficulties of contacting respondents, motivating them to participate in the survey

3. Basic methodological assumptions are not met

4. Noncoverage of those who do not have Internet access.