INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

Respondent-Driven Sampling Model Evaluation for Sampling without Replacement in Estimating Hidden Populations

Yakubu Anjikwi[a,b*], Danjuma Jibasen[b], Ikeme J. Dike[c] and Emmanuel Torsen[b]

[a]Department of Agricultural Economics, Faculty of Agriculture, University of Maiduguri, Borno, Nigeria
[b]Department of Statistics, Faculty of Physical Sciences, Modibbo Adama University, Yola, Adamawa, Nigeria
[c]Department of Operations Research, Faculty of Computing, Modibbo Adama University, Yola, Adamawa, Nigeria

ARTICLE INFO

Article history:
Received 15 February 2026
Received in revised form 10 March 2026
Accepted 18 March 2026

Keywords:
Respondent-Driven Sampling, Gender, Without Replacement, Profile

MSC 2020 Subject classification:
62D05

ABSTRACT

This study evaluates respondent-driven sampling (RDS) models that use sampling without replacement to estimate characteristics of hidden populations. Traditional RDS estimators, such as Salganik and Heckathorn (SH-RDS) and Volz and Heckathorn (VH-RDS), assume sampling with replacement and require many recruitment waves to reach statistical equilibrium, which is rarely achieved in practice. Most real-world RDS studies are conducted without replacement and use fewer waves, leading to biases like overrepresentation of highly connected individuals. Recent estimators, such as Gile's Successive Sampling (G-SS), address some limitations but still face challenges, including instability with large samples, broad confidence intervals, and inadequate handling of non-random recruitment and seed selection. To address these issues, a new estimator is proposed that incorporates sampling without replacement and strategic multiple-seed selection. Simulations and real-world data analysis (using the Project 90 dataset) demonstrate that estimator performance varies by sample size. For small samples $(n < 500)$*, SH-RDS and VH-RDS are most accurate for gender estimation. For larger samples* $(n > 500)$*, the proposed estimator is most efficient, with minimal variance. G-SS shows moderate, reliable performance, while the Naïve estimator becomes less reliable as the sample size increases. Analysis also reveals that the proposed estimator performs well for groups with higher connectivity, though variance remains high for the "Unemployed" group. Overall, the proposed estimator was recommended for large samples and complex networks, especially among hard-to-reach populations.*

1. Introduction

Hidden or hard-to-reach populations are typically small, marginalized groups that operate outside conventional data collection frameworks, rendering them difficult to identify, access, and study systematically. Such populations often face significant risks if their identities or behaviors become public, necessitating heightened considerations of safety and confidentiality. Examples include people who inject drugs (PWID), victims of human trafficking, men who have sex with men (MSM), individuals experiencing homelessness, and undocumented migrants. Social stigmatization and legal vulnerability commonly drive these groups to the margins of society, which both increases their need for privacy

* Corresponding author. Tel.: +2348036267312
E-mail address: y.anjikwi@gmail.com (Anjikwi, Y.)
https://doi.org/10.62054/ijdm/0301.16

and complicates efforts to collect reliable data (Sarah *et al.,* 2022).

Due to the absence of comprehensive sampling frames for such populations, researchers are compelled to rely on non-probability sampling methodologies. Common approaches include convenience sampling techniques, such as snowball sampling, whereby current participants recruit future subjects from their social circles. Other methods include time-location sampling, which targets specific venues and times likely to yield eligible participants, and respondent-driven sampling (RDS), which incentivizes peer recruitment. While these methods can reveal important characteristics and dynamics of hidden populations, they introduce significant challenges, particularly regarding sample representativeness and potential biases. Addressing these methodological challenges is essential for producing accurate and meaningful empirical insights.

The RDS has emerged as a particularly innovative and effective strategy for recruiting participants from populations lacking formal sampling frames. RDS leverages the social networks within hidden populations, allowing researchers to reach individuals who would otherwise remain inaccessible through conventional means. Its utility has been demonstrated in medical and public health research, where it has facilitated the recruitment of high-risk groups such as people who inject drugs, MSM, city musicians, and individuals experiencing homelessness (Card *et al.,* 2017; Heckathorn & Cameron, 2017; Lyons *et al.,* 2017; Sypsa *et al.,* 2017; White *et al.,* 2015).

Beyond the health sciences, RDS has also proven effective in connecting with migrant communities and other marginalized populations across diverse settings (Górny & Napierała, 2016; Keygnaert *et al.,* 2014). RDS's unique ability to generate statistically robust population estimates, coupled with its logistical feasibility, has led to its widespread global adoption (Johnston *et al.,* 2016). This methodological trend underscores the growing significance of RDS in expanding the empirical understanding of hidden populations and informing interventions tailored to their needs.

The RDS process begins with the identification and recruitment of initial participants, or "seeds," selected via convenience sampling from the broader target population. These seeds complete a baseline survey, often administered online for enhanced accessibility, and are then invited to recruit a limited number of peers from within their social networks. A coupon system is employed to track recruitment chains and to ensure that each participant discloses the size of their network within the target population. Through successive waves of peer recruitment, the sample expands and becomes progressively less dependent on the initial convenience sample. Importantly, RDS employs a dual incentive system, rewarding participants both for participating and for successfully recruiting others. This incentive structure is critical for maximizing recruitment efficiency and sample diversity (Gile & Handcock, 2010; White *et al.,* 2015).

RDS has been implemented in over 460 studies globally White et al. (2015), utilizing both traditional face-to-face and web-based (WebRDS) recruitment modalities. The effectiveness of WebRDS has been substantiated in research by Stromdahl *et al.* (2015), Logan et al. (2016), and Stein *et al.* (2014). A robust body of statistical models underpins RDS analysis, including the Naïve estimator (Heckathorn, 1997), RDS-HK1 (Heckathorn, 2002), SH-RDS (Salganik & Heckathorn, 2004), VH-RDS (Volz & Heckathorn, 2008), and G-SS (Gile, 2011).

All these estimators share a foundational assumption: recruitment occurs exclusively through existing social network ties within the hidden population. This means that recruitment is only possible between individuals with mutual awareness and social connection. Observing this principle is vital for the methodological integrity of RDS, particularly when studying populations inaccessible to traditional survey methods. Notably, the SH-RDS estimator is highly sensitive to unequal recruitment probabilities and the distribution of social ties within and between subgroups. While this can lead to underestimation of certain characteristics when recruitment is non-random, it also means that the SH-RDS estimator can outperform alternatives in cases with complex recruitment dynamics (Salganik & Heckathorn, 2004).

RDS is now the dominant methodology for sampling and studying hidden populations, including people who inject drugs, sex workers, MSM, and other marginalized groups inaccessible to probability-based approaches (Berndt, 2020). Despite its widespread adoption, RDS estimators, most notably SH-RDS and VH-RDS, are grounded in theoretical assumptions that are regularly violated in applied research. These estimators presuppose with-replacement sampling and the attainment of a stationary distribution over a large number of recruitment waves (Gile & Handcock, 2010). In practice, RDS operates without replacement and typically completes far fewer waves, resulting in systematic biases, such as the overrepresentation of individuals with large social networks. These biases undermine the validity of population inferences and compromise the utility of RDS in informing public health interventions (Avery & Rotondi, 2020).

Recent methodological developments by Naser *et al*. (2018) and Kim *et al*. (2026) have introduced without-replacement estimators that eliminate the need for stationarity and better reflect the sequential nature of RDS recruitment. While these advances have reduced bias, they are limited by instability in large sample sizes, vulnerability to non-random recruitment patterns, and insufficient protocols for seed selection and weighting, particularly in heterogeneous network structures (Spiller *et al.,* 2023; Crawford *et al*., 2018; Fellows, 2019; Avery *et al.,* 2021). Thus, there remains an urgent need for an RDS estimator that: (1) operates under realistic without-replacement assumptions; (2) maintains stability and precision in large samples and limited waves; (3) is robust to violations of recruitment and network assumptions; and (4) incorporates strategic seed selection and adaptive weighting mechanisms (Spiller et al., 2023).

This study responds to these gaps by developing and validating a novel RDS estimator that integrates multiple-seed selection and adaptive differential weighting, explicitly designed for reliable performance under typical field conditions involving large samples, limited recruitment waves, and heterogeneous network topologies. The estimator is tested through comprehensive simulations and validated using real-world datasets from Project 90, Facebook user pages, and populations of people who inject drugs, as referenced in recent RDS literature (Avery *et al.,* 2021; Spiller *et al.,* 2018; Abdesselam *et al.,* 2020). This study aims to develop a model for respondent-driven sampling that enables the sampling and estimation of hidden populations using a without-replacement sampling framework. The specific objectives are to: Develop an RDS estimator utilizing without-replacement sampling; Estimate the proportion of categorical variables (gender) in hidden populations, and evaluate the performance of the proposed RDS estimator.

2. Methods

This study introduced the existing RDS estimators, which were grounded in the theory of undirected social relation networks. It elucidated how these estimators were prone to bias when applied to populations characterized by a large sampling fraction. Subsequently, this work proposed an alternative estimator that incorporated sampling without replacement and strategic multiple-seed selection to mitigate the impact of biases found in the previous RDS estimator when sampling within hidden populations.

2.1 The Naïve Estimator.

The Naive estimator was proposed by Heckathorn (1997) and was simply the proportion of infected individuals found in the sample.

$$\hat{\mu}_A^N = \frac{n_a}{n} \tag{1}$$

where,

$\hat{\mu}_A^N$ was the population proportion, n_a was the sample size of group A, and n is the total sample.

Likewise, if the actual sampling probabilities were represented by $\pi_i, i = 1, \dots, N,$ then they could serve as a generalized Hansen-Hurwitz estimator of the target parameter.

$$\hat{\mu}_{NV} = \frac{\phi_A}{\phi_A + \phi_B} \tag{2}$$

where,

$\hat{\mu}_{NV}$ was the population proportion, ϕ_A was the number of recruits in group A, and was the number of recruits in group B.

Equation (2) indicated that when equal sampling probabilities were observed for individuals in both group A and group B, it acted as a generalized Hansen-Hurwitz estimator for the parameter of interest.

The Assumptions of the Naïve Estimator were:

i. Respondents recruited peers from their social contacts with equal probability.
ii. Sampling was done with replacement.
iii. The degree of respondents was normally distributed.
iv. The social network of the population was undirected. The population formed a connected network.

2.2 SH-RDS Estimator

The SH-RDS was developed by Salganik and Heckathorn (2004) for estimating population proportion using with replacement sampling, and the model was defined as

$$\hat{\mu}_{SH_A} = \frac{\hat{D}_B \cdot \hat{C}_{B,A}}{\hat{D}_A \cdot \hat{C}_{A,B} + \hat{D}_B \cdot \hat{C}_{B,A}} \tag{3}$$

for group A, and

$$\hat{\mu}_{SH_B} = \frac{\hat{D}_A \cdot \hat{C}_{A,B}}{\hat{D}_A \cdot \hat{C}_{A,B} + \hat{D}_B \hat{C}_{B,A}} \tag{4}$$

for group B,

where:

$$\hat{C}_{A,B} = \frac{R_{AB}}{R_{AA} + R_{AB}}$$

$$\hat{C}_{B,A} = \frac{R_{BA}}{R_{BB} + R_{BA}}$$

$$R_{AA} = \sum_{i \in A} d_i$$

$$\hat{D}_A = \frac{n_A}{\sum_{i=1}^{n_A} \frac{1}{d_i}}$$

$$\hat{D}_B = \frac{n_B}{\sum_{i=1}^{n_B} \frac{1}{d_i}}$$

$\hat{\mu}_{SH_A}$ was the estimate of the population proportion of group A,

$\hat{\mu}_{SH_B}$ was the estimate of the population proportion of group B,

$\hat{C}_{B,A}$ was the estimated probability of cross-sample recruitment from group B to group A

$\hat{C}_{A,B}$ was the estimated probability of cross-sample recruitment from group A to group B

R_{AA} was the total number of ties (edges) that contained a person in group A

$\hat{D}_A$ was the estimated mean degree of group A

$\widehat{D}_B$ was the estimated mean degree of group B

The Assumptions of the SH-RDS estimator are:

i. Respondents recruit peers from their social contacts with equal probability.
ii. Each recruitment consists of only one peer.
iii. Sampling is done with replacement.
iv. The degree of respondents is reported without error.
v. The social network of the population is undirected, and
vi. The population forms a connected network.

2.3 VH-RDS Estimator

The VH-RDS model was developed by Voltz and Heckathorn (2008) to estimate population proportion using with-replacement sampling, and the model was defined as

$$\hat{\mu}_{VH} = \left(\frac{n_A}{n}\right)\left(\frac{\widehat{D}_u}{\widehat{D}_{A_V}}\right) \tag{5}$$

The variance of VH-RDS was defined as

$$\widehat{V}_{HH}(\langle\hat{y}\rangle) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{\widehat{D}_u}{\hat{d}_i} - \langle\hat{y}\rangle\right) \tag{6}$$

where:

n_A was the sample size of group A, and n was the total sample.

$\hat{y}$ was an estimate of the mean value of y

y was an indicator function that a value $I_i(i) = \begin{cases} 1, & i \in A \\ 0, & otherwise \end{cases}$

$\widehat{D}_u$ was the estimate of the mean degree of the sample u

$$\widehat{D}_{A_V} = \frac{N_u}{\sum_{i=1}^{N_u}\frac{1}{d_i}} \tag{7}$$

$\widehat{D}_{A_V}$ was the estimate of mean degree of the sample group A

The Assumptions of the VH-RDS Estimator are:

i. Recruitment are random. When recruiting others, respondents select uniformly at random from their network.
ii. Each recruitment consists of only one peer.
iii. Sampling is done with replacement.
iv. Respondents accurately report their degree in the network.
v. Network connections are reciprocal.
vi. The population forms a connected network.
vii. Convergence. Recruitment is modeled as a Markov process (MP), where the state of the MP was the last individual recruited.

2.4 Gile's Successive Sampling (G-SS) Estimator

The G-SS model was developed by Gile (2011) for estimating population proportion characteristics using without replacement sampling. The inclusion probability associated with a specific unit was defined as

$$\tilde{\pi}_{SS} = \frac{U_i+1}{M+1} \tag{8}$$

where,

U_i was the number of times unit i was sampled in M trials, and the model was defined as

$$\hat{\mu}_{SS} = \frac{\sum_{j=1}^{N} \frac{S_j Z_j}{\hat{\pi}(d_j)}}{\sum_{j=1}^{N} \frac{S_j}{\hat{\pi}(d_j)}} \quad (9)$$

$\hat{\mu}_{SS}$ was the estimate of population proportion.

The assumptions of Gile's Successive Sampling (G-SS) Estimator are:

i. Recruitment is random. When recruiting others, respondents select uniformly at random from their network.
ii. Each recruitment consists of only one peer.
iii. Sampling are done without replacement.
iv. Respondents accurately report their degree in the network.
v. Network connections are reciprocal.
vi. The population forms a connected network.
vii. The population size N is known.

2.5 The Proposed RDS Estimator

The assumptions of the proposed model are:

i. Respondents recruit peers from their social contacts with equal probability (random).
ii. Each recruitment consists of only one peer (throughout the sampling period).
iii. Sampling are done without replacement.
iv. The degree of respondents reported has a negligible error.
v. The network are directed.
vi. The population forms a connected network.
vii. The population size N is unknown.

2.5.1 Estimation of inclusion probability

The idea of probability proportional to size without replacement (PPSWOR) was extended to an RDS estimator to sample a hidden population (Naser et al., 2018). Since a node was recruited into an RDS sample with a probability proportional to its degree,

let $s = \{i_1 i_2, \ldots, i_n\}$ be the RDS sample and $\lambda_{i_{proposed}}$ be the proposed inclusion probability, then,

$$\lambda_{i_{proposed}} = P(i \in S) \quad (10)$$

Therefore, the proposed inclusion probability can be modelled using the (Naser *et al.,* 2018) approach as

$$\lambda_{i_{proposed}} = \sum_{s \in S} P(s) \times I(i \in s) \quad (11)$$

where,

$P(s)$ was the probability of selecting sample s

$I(i \in s)$ was an indicator function (1 if i ∈ s, 0 otherwise)

For the RDS sample without replacement, $P(s)$ can be modelled as:

$$P(s) = P(i_1) \prod_{k=2}^{n} P(i_k | i_{k-1})$$

where,

$P(i_1)$ was the probability of selecting the seed node i_1(often not random, or usually 1).

$P(i_k|i_{k-1})$ was the probability that node i_k was recruited by node i_{k-1}:

$$P(i_k|i_{k-1}) = \frac{D_{i_k}}{\sum_{j\in N(i_{k-1})-\{i_1 i_2,\dots,i_{k-1}\}} D_j} \tag{12}$$

where,

$N(i_{k-1})$ the set of neighbors of i_{k-1}, not yet recruited

Substitute P(s) into equation (11)

$$\lambda_{i_{proposed}} \approx \sum_{s\in S}\left(\prod_{k=2}^{n} \frac{D_{i_k}}{\sum_{j\in N(i_{k-1})-\{i_1 i_2,\dots,i_{k-1}\}} D_j}\right) \times I(i \in s) \tag{13}$$

Simplify the expression in the equation

$$\lambda_{i_{proposed}} \approx \frac{D_i}{(\sum_{j\in \cup} D_j)} \tag{14}$$

where,

$I(i \in s)$ was equal to 1 if i belonged to S

$\cup$ was the total number of nodes.

$\lambda_{i_{proposed}}$ can be approximated as

$$\lambda_{i_{proposed}} \approx \frac{D_i}{E} \tag{15}$$

where,

E was the total number of edges in the network.

$$E \approx \sum_{j\in \cup} D_j \tag{16}$$

2.5.2 Estimation of mean degree ($\boldsymbol{D_{proposed}}$)

The proposed mean degree ($D_{proposed}$) of a node i can be estimated as:

Let d_i be the degree of node i, since $\pi_{proposed_i}$ was the inclusion probability of node i as (approximated as $\frac{D_i}{(E)}$), then the estimate of the mean degree of node i was given as

$$\widehat{D}_{proposed} = \frac{\left(\sum_{i\in S}\frac{d_i}{\lambda_{i_{proposed}}}\right)}{\sum_{i\in S}\frac{1}{\lambda_{i_{proposed}}}} \tag{17}$$

Substitute $\lambda_{i_{proposed}}$ into equation (17)

$$\widehat{D}_{proposed} = \frac{\left(\sum_{i\in S}\frac{d_i}{\frac{D_i}{(E)}}\right)}{\sum_{i\in S}\frac{1}{\frac{D_i}{(E)}}} \tag{18}$$

Simplify equation (18)

$$\widehat{D}_{proposed} = \left(\frac{E\sum_{i\in S}\frac{d_i}{D_i}}{E\sum_{i\in S}\frac{1}{D_i}}\right) \tag{19}$$

Cancel out the constant E

$$\widehat{D}_{proposed} = \frac{\left(\sum_{j \in S} \frac{d_i}{D_i}\right)}{\left(\sum_{j \in S} \frac{1}{D_i}\right)} \tag{20}$$

2.5.3 Estimation of cross-group edges ($S_{g_a g_b}$)

Let $S_{g_a g_b}$ be the number of cross-group edges, that is, from group g_a to group g_b,

Then the probability of a cross-group edge being reported can be modelled as:

$$\text{P(edge (i, j)was reported)} = \frac{D_i}{E} \times \left(\frac{1}{D_i}\right) + \frac{D_j}{E} \times \left(\frac{1}{D_j}\right) \tag{21}$$

where,

$i \in g_a$ and $j \in g_b$ (or vice versa)

D_i was the degree of node i

D_j was the degree of node j

E was the total number of edges in the network

Equation (21) can be simplified as:

$$\text{P(edge (i, j)is reported)} = \frac{1}{E} + \frac{1}{E}$$

$$= \frac{1}{E} \tag{22}$$

Estimate $S_{g_a g_b}$ using the reported cross-group edges.

$$E\left[S_{g_a g_b}\right] = \sum_{\{i \in g_a, j \in g_b\}} P(edge(i, j)\ \ is\ reported)$$

$$= \frac{S_{g_a g_b}}{E} \tag{23}$$

Use the RDS data to estimate $S_{g_a g_b}$

$$S_{g_a g_b} = E \times \frac{number\ of\ g_a \rightarrow g_b\ edges\ reported}{\left(\sum_{i \in s} \frac{1}{D_i}\right)}$$

where s was the RDS sample.

E can be approximated as defined as in (16)

2.5.4 Estimation of population proportion

The proposed RDS estimator was estimated by using the idea of two ratios of Horvitz-Thompson (1952) estimator given by:

$$\hat{P}_{HT} = \frac{\left(\sum_{i=1}^{n} \frac{y_i}{\pi_i}\right)}{\left(\sum_{i=1}^{n} \frac{x_1}{\pi_i}\right)} \tag{24}$$

where,

y_i= value of the variable of interest for unit (i) (e.g., presence/absence of a trait)

(x_i) = value of the auxiliary variable for unit (i) (e.g., usually 1 if estimating a proportion, or some other covariate)

π_i= inclusion probability for node i

Building on this general framework, the proposed RDS estimator was specified for the case of estimating the proportion of individuals with a particular trait in an RDS network setting. Specifically, if $S_{g_a g_b}$ denotes the total number of observed links (edges) from group g_a to group g_b, λ_1 and λ_0 are the respective inclusion probabilities for individuals with and without the trait, D_1 and D_0 are their respective degrees, the estimator was modelled as:

$$\widehat{RDS}_{proposed} = \frac{\left(\frac{S_{g_a g_b}}{\lambda_1}\right)}{\left(\frac{S_{g_a g_b}}{\lambda_1}+\frac{S_{g_b g_a}}{\lambda_0}\right)} \tag{25}$$

$$= \frac{\left(\frac{S_{g_a g_b}}{\frac{D_1}{(\mathrm{E})}}\right)}{\left(\frac{S_{g_a g_b}}{\frac{D_1}{(\mathrm{E})}}+\frac{S_{g_b g_a}}{\frac{D_1}{(\mathrm{E})}}\right)}$$

$$= \frac{\left(E*\frac{S_{g_a g_b}}{D_1}\right)}{\left(E*\frac{S_{g_a g_b}}{D_1}+E*\frac{S_{g_b g_a}}{D_0}\right)}$$

$$= \frac{\left(\frac{S_{g_a g_b}}{D_1}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)}$$

$$= \frac{\left(D_0*S_{g_a g_b}\right)}{\left(D_0*S_{g_a g_b}+D_1*S_{g_a g_b}\right)} \tag{26}$$

Thus, equation (26) represents a specific application of the general Horvitz-Thompson ratios estimator (24), where the numerator and denominator are constructed using the observed cross-group links between two groups and their associated inclusion probabilities. This formulation allows for the estimation of the proportion of group (g_a) in the population, accounting for the network structure and sampling design inherent to RDS.

2.5.5 Estimation of Variance of Proposed Model

Let's estimate the variance of $\widehat{RDS}_{proposed}$

$$\widehat{RDS}_{proposed} = \frac{\left(\frac{S_{g_a g_b}}{D_1}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)} \tag{27}$$

where,

$S_{g_a g_b}$ was the number of cross-group edges from group g_a to group g_b.

As discussed by Beutner (2024) and Wikipedia (2024), the variance of the proposed estimator

$(\widehat{RDS}_{proposed})$ can be estimated as:

$$Var\left(\widehat{RDS}_{roposed}\right) \approx \left(\frac{\partial \widehat{RDS}_{proposed}}{\partial S_{g_a g_b}}\right)^2 var\left(S_{g_a g_b}\right) + \left(\frac{\partial \widehat{RDS}_{proposed}}{\partial S_{g_b g_a}}\right)^2 var\left(S_{g_b g_a}\right)$$

$$+2\left(\frac{\partial\widehat{RDS}_{proposed}}{\partial S_{g_a g_b}}\right)\left(\frac{\partial\widehat{RDS}_{proposed}}{\partial S_{g_b g_a}}\right)Cov(S_{g_a g_b}, S_{g_b g_a}) \quad (28)$$

Compute the partial derivatives.

Let

$$\frac{\partial\widehat{RDS}_{proposed}}{\partial S_{g_a g_b}} = \frac{\left(\frac{1}{D_1}\right)\left(\frac{S_{g_b g_a}}{D_0}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)^2} \quad (29)$$

$$\frac{\partial\widehat{RDS}_{proposed}}{\partial S_{g_b g_a}} = -\frac{\left(\frac{S_{g_a g_b}}{D_1}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)^2\left(\frac{1}{D_0}\right)} \quad (30)$$

$$var\left(S_{g_a g_b}\right) \approx \frac{S_{g_a g_b}(1-\lambda_1)}{\lambda_1} \quad (31)$$

$$var\left(S_{g_b g_a}\right) \approx \frac{S_{g_b g_a}(1-\lambda_0)}{\lambda_0} \quad (32)$$

$$Cov(C_{ab}, C_{ba}) = E[(C_{ab} - E[C_{ba}])(C_{ba} - E[C_{ab}])] \quad (33)$$

Substitute the partial derivatives Equation (29, 30, 31, 32, and 33 into 28)

$$Var\left(\widehat{RDS}_{proposed}\right) \approx \left(\frac{\left(\frac{1}{D_1}\right)\left(\frac{S_{g_b g_a}}{D_0}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)^2}\right)^2 \frac{S_{g_a g_b}(1-\lambda_1)}{\lambda_1} + \left(-\frac{\left(\frac{S_{g_a g_b}}{D_1}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)^2\left(\frac{1}{D_0}\right)}\right)^2 \frac{S_{g_b g_a}(1-\lambda_0)}{\lambda_0} +$$

$$2\left(\frac{\left(\frac{1}{D_1}\right)\left(\frac{S_{g_b g_a}}{D_0}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)^2}\right)\left(-\frac{\left(\frac{S_{g_a g_b}}{D_1}\right)}{\left(\frac{S_{g_a g_b}}{D_1}+\frac{S_{g_b g_a}}{D_0}\right)^2\left(\frac{1}{D_0}\right)}\right)E[(C_{ab} - E[C_{ba}])(C_{ba} - E[C_{ab}])] \quad (34)$$

2.5.6 Estimation of Bias of Proposed Model

The bias of $\widehat{RDS}_{proposed}$ is the difference between the expected value and the true population parameter:

$$Bias\left(\widehat{RDS}_{proposed}\right) \approx E\left[\widehat{RDS}_{proposed}\right] - RDS_{proposed} \quad (35)$$

2.6 Simulation

To assess the robustness of the proposed estimator, $RDS_{proposed}$, the study simulated and calculated the outcome variable of gender for a series of samples($n = 500, n = 1000, n = 1500, n = 2000$ *and* $n > 2000$ drawn from the overall population of $10{,}000$. For each sample, the study meticulously computed the proportions of male and female participants, along with important statistical metrics such as the variance and bias. Initially, 8 key recruits, known as "seeds," were selected. Each of the 8 seeds was tasked with 2 coupons (a unique code that allowed researchers to track recruitment) to recruit their associates, friends, relatives, and acquaintances. These individuals then recruited their associates and so on, until the desired sample size was reached. That is, 6 waves were attained to obtain $n = 500$, 7 waves for $n = 1000$, 9 waves for $n = 1500$, 10 waves for $n = 2000$, and 12 waves for $n > 2000$. The recruitment tree was presented in Fig. 1, while Project 90 data was used as real-world data to validate the model.

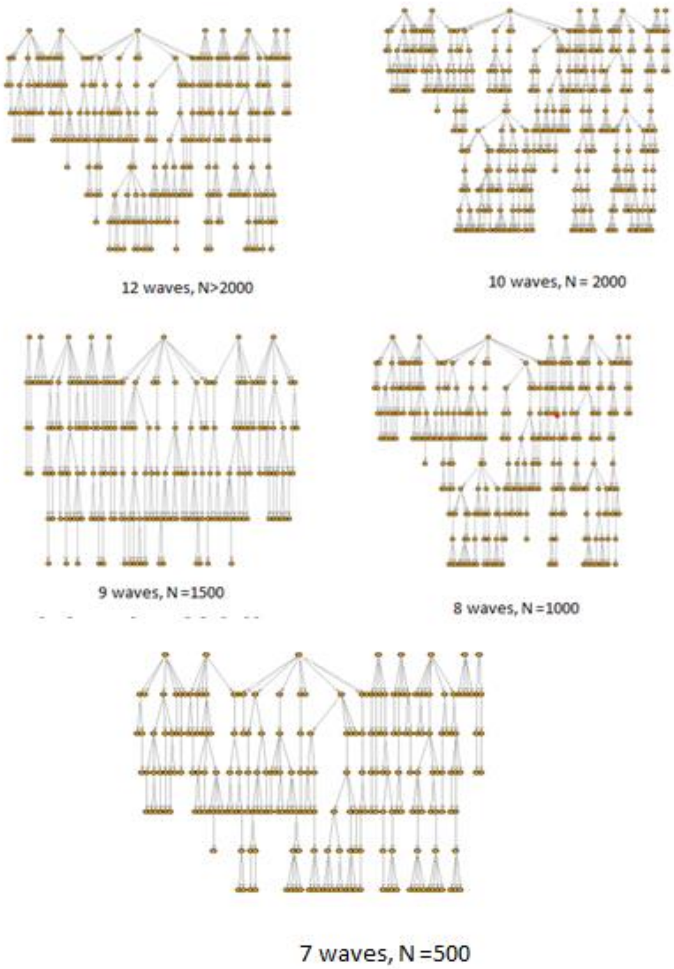

Figure 1: RDS recruitment Tree

3. Results and Discussion

3.1 Simulation Results for RDS Estimators

This section presents a detailed overview of five simulated populations, each of which was carefully designed to share essential characteristics related to population dynamics and random variables. These characteristics are aligned with the specified target values to ensure consistency across all populations. The study maintains average values for critical metrics such as network degree, the number of waves, and recruits per seed, establishing a robust framework for comparison. Additionally, it estimates population proportions and provides detailed calculations of the variance and bias for various estimators, including the Naïve, $RDS_{proposed}$, SH-RDS, VH-RDS, and G-SS estimators. The estimation procedures guarantee that the findings are both reliable and meaningful. The results are presented in Table 1.

Table 1: Estimates of gender proportion and Variance($\boldsymbol{\sigma^2}$) for proposed and existing RDS estimators

Samples	Gender	Naïve		$\widehat{RDS}_{proposed}$		SH-RDS		VH-RDS		G-SS	
		$\hat{\mu}_{NV}$	σ^2	$\hat{\mu}_{proposed}$	σ^2	$\hat{\mu}_{SH}$	σ^2	$\hat{\mu}_{VH}$	σ^2	$\hat{\mu}_{SS}$	σ^2
500	Female	0.3596	0.00022	0.3033	0.00067	0.3909	0.00013	0.3803	0.00014	0.3059	0.00050
	Male	0.6404		0.6967		0.6091		0.6197		0.6941	
1000	Female	0.264	0.00295	0.3877	0.00011	0.3001	0.00071	0.3195	0.00063	0.3595	0.00031
	Male	0.736		0.6123		0.6999		0.6805		0.6405	
1500	Female	0.3396	0.00950	0.3453	0.00005	0.3549	0.00115	0.3663	0.00113	0.3096	0.00062
	Male	0.6604		0.6547		0.6451		0.6337		0.6904	
2000	Female	0.3293	0.0116	0.3473	0.00003	0.3146	0.00216	0.3582	0.00117	0.3501	0.00041
	Male	0.6707		0.6527		0.6854		0.6418		0.6499	
>2000	Female	0.3191	0.0246	0.3573	0.00000	0.3032	0.00316	0.3449	0.00300	0.3508	0.00023
	Male	0.6809		0.6427		0.6968		0.6551		0.6492	

The results in Table 1 indicate that the proportion of females is slightly lower than that of males, with the Naïve estimator consistently showing lower female estimates compared to SH-RDS and VH-RDS. Notably, the $RDS_{proposed}$ and G-SS estimators yield similar values for both genders, indicating potential robustness in these methods. Variance estimates for females are generally higher with the Naïve estimator, especially as sample size increases (>2000, σ^2 = 0.0246), while the $RDS_{proposed}$ and G-SS methods maintain very low variances, suggesting increased precision. The SH-RDS and VH-RDS estimators occasionally display higher means for females at certain sample sizes, but their variances remain low, indicating stable estimates.

This finding mirror patterns observed in prior research, where RDS estimators, especially SH-RDS and VH-RDS, offer greater stability and lower variance in proportion estimates when compared to the Naïve approach (Volz & Heckathorn, 2008; Gile & Handcock, 2010). The convergence of estimates from the proposed RDS and G-SS methods supports recent literature advocating for Giles successive sampling as a reliable approach when sampling fractions are moderate (Gile, 2011). The observed gender proportion stability across increasing sample sizes further aligns with findings that larger samples can reduce estimator bias and variance, enhancing representativeness (Salganik & Heckathorn, 2004).

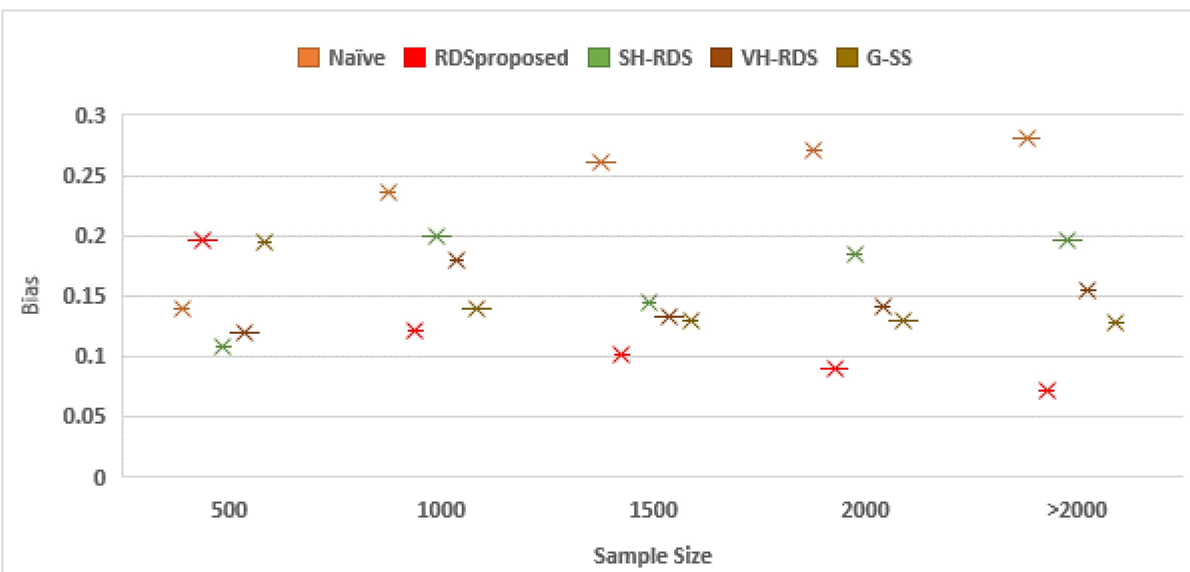

Figure 2: Simulated Bias estimate of Gender Proportion for $RDSproposed$ and Existing RDS estimators.

Figure 2 indicates that the bias estimates across different sample sizes and estimators (Naïve, $RDS_{proposed}$, SH-RDS, VH-RDS, and G-SS) reveal notable patterns. For the Naïve estimator, bias increases steadily with larger sample sizes, peaking at 0.2809 for samples greater than 2000. In contrast, the proposed RDS estimator shows a consistent decrease in bias as sample size increases, dropping from 0.1967 (n=500) to 0.072 (n>2000), indicating improved estimator accuracy with larger samples. The SH-RDS and VH-RDS estimators exhibit fluctuating bias across sample sizes, with no clear trend of reduction. G-SS bias decreases from 0.1941 at n=500 to 0.1292 at n>2000, but the reduction is modest and levels off at larger sample sizes. Overall, the $RDS_{proposed}$ estimator demonstrates the most substantial improvement in bias with increasing sample size, while the Naïve estimator's performance actually worsens.

This finding align with previous research highlighting the limitations of the Naïve estimator in RDS studies, especially its vulnerability to bias due to network effects and non-random recruitment (Salganik & Heckathorn, 2004; Gile & Handcock, 2010). The observed reduction in bias with increasing sample size for the $RDS_{proposed}$ estimator is consistent with literature suggesting that more sophisticated estimators, which adjust for network structure and recruitment patterns, yield more accurate population estimates as sample size grows (Gile, 2011). The irregular bias trends for SH-RDS, VH-RDS, and G-SS are also reported in empirical studies, indicating that while these estimators can outperform the Naïve approach, their effectiveness may be sensitive to specific sample compositions and underlying network characteristics (Volz & Heckathorn, 2008).

3.2 Application of Real-life Datasets

3.2.1 Estimating Project 90 Data

This data was drawn from the Project 90 dataset (https://opr.princeton.edu/archive/p90/), which examines hard-to-reach populations, including drug cooks, homeless individuals, sex workers, drug dealers, and unemployed persons, along with a "Gender" category. In previous analyses (Avery et al., 2021), the focus primarily rested on evaluating the single largest connected network within this dataset. However, this study takes a broader approach by incorporating all individuals (n=5,475) in the dataset to ensure a more comprehensive analysis. The analysis focuses on the estimated population proportions ($\hat{\mu}_{proposed}$) and variances (σ²) for different status groups across four degree categories (Equally, Moderate, Very High, Very Low): the results are presented in Table 2

Table 2: Estimating the Degree Distribution and Variance of Project 90 Data Population Characteristics

Degree	Status	**Naïve**		$\widehat{RDS}_{proposed}$		**SH-RDS**		**VH-RDS**		**G-SS**	
		$\hat{\mu}_{NV}$	σ^2	$\hat{\mu}_{proposed}$	σ^2	$\hat{\mu}_{SH}$	σ^2	$\hat{\mu}_{VH}$	σ^2	$\hat{\mu}_{SS}$	σ^2
Equally	drug cook	0.01	0.04	0.01	0.06	0.015	0.02	0.015	0.02	0.01	0.04
	Homeless	0.01	0.04	0.01	0.04	0.01	0.04	0.01	0.02	0.02	0.06
	Retired	0.04	0.04	0.03	0.08	0.01	0.04	0.01	0.02	0.04	0.02
	Thief	0.02	0.04	0.03	0.014	0.03	0.02	0.03	0.02	0.03	0.06
	Pimp	0.03	0.04	0.02	0.08	0.03	0.04	0.03	0.04	0.02	0.06
	Housewife	0.04	0.06	0.07	0.01	0.07	0.04	0.07	0.02	0.05	0.06
	Disabled	0.04	0.06	0.03	0.08	0.04	0.02	0.04	0.04	0.05	0.06
	sex work client	0.07	0.06	0.09	0.012	0.09	0.08	0.09	0.06	0.08	0.08
	drug dealer	0.06	0.06	0.04	0.08	0.045	0.02	0.045	0.06	0.05	0.04
	Gender	0.44	0.06	0.45	0.1	0.45	0.02	0.45	0.04	0.42	0.04
	sex worker	0.07	0.06	0.06	0.1	0.06	0.02	0.06	0.06	0.05	0.04
	Unemployed	0.17	0.06	0.18	0.04	0.15	0.22	0.15	0.02	0.18	0.22
Moderate	drug cook	0.01	0.06	0.01	0.04	0.015	0.02	0.015	0.02	0.01	0.02
	Homeless	0.01	0.06	0.01	0.04	0.01	0.04	0.01	0.04	0.02	0.04
	Retired	0.039	0.06	0.029	0.04	0.01	0.04	0.01	0.04	0.04	0.02
	Thief	0.02	0.1	0.03	0.04	0.03	0.04	0.03	0.1	0.03	0.02
	Pimp	0.03	0.06	0.02	0.06	0.03	0.04	0.03	0.04	0.02	0.04

	Housewife	0.039	0.04	0.07	0.04	0.07	0.04	0.07	0.04	0.05	0.02
	Disabled	0.04	0.12	0.03	0.03	0.04	0.08	0.04	0.06	0.05	0.04
	sex work client	0.07	0.18	0.07	0.04	0.09	0.06	0.09	0.06	0.08	0.04
	drug dealer	0.056	0.18	0.06	0.02	0.045	0.06	0.045	0.08	0.048	0.01
	Gender	0.44	0.18	0.413	0.04	0.45	0.08	0.45	0.04	0.43	0.04
	sex worker	0.069	0.18	0.08	0.01	0.06	0.12	0.06	0.06	0.046	0.04
	Unemployed	0.18	0.22	0.178	0.01	0.149	0.26	0.149	0.28	0.176	0.02
Very high	drug cook	0.01	0.22	0.01	0.04	0.015	0.06	0.015	0.06	0.01	0.04
	Homeless	0.01	0.22	0.01	0.02	0.01	0.08	0.01	0.04	0.02	0.04
	Retired	0.039	0.22	0.029	0.02	0.01	0.04	0.01	0.06	0.039	0.04
	Thief	0.02	0.22	0.03	0.02	0.03	0.08	0.03	0.06	0.029	0.04
	Pimp	0.029	0.22	0.019	0.02	0.03	0.1	0.03	0.04	0.02	0.06
	Housewife	0.039	0.22	0.07	0.02	0.07	0.04	0.07	0.06	0.05	0.04
	Disabled	0.039	0.22	0.029	0.02	0.039	0.08	0.039	0.08	0.05	0.06
	sex work client	0.07	0.22	0.072	0.02	0.019	0.12	0.089	0.06	0.082	0.04
	drug dealer	0.055	0.28	0.062	0.02	0.045	0.1	0.045	0.06	0.048	0.08
	Gender	0.43	0.28	0.413	0.02	0.449	0.06	0.449	0.04	0.429	0.04
	sex worker	0.069	0.28	0.08	0.04	0.064	0.12	0.064	0.06	0.046	0.04
	Unemployed	0.19	0.28	0.178	0.02	0.149	0.3	0.149	0.286	0.176	0.04
Very low	drug cook	0.008	0.46	0.008	0.04	0.013	0.01	0.013	0.04	0.009	0.02
	Homeless	0.009	0.46	0.009	0.04	0.009	0.04	0.009	0.04	0.017	0.06
	Retired	0.034	0.48	0.025	0.04	0.009	0.04	0.009	0.02	0.035	0.04
	Thief	0.017	0.5	0.025	0.04	0.026	0.01	0.026	0.02	0.026	0.04
	Pimp	0.026	0.5	0.017	0.06	0.026	0.04	0.026	0.04	0.018	0.04
	Housewife	0.035	0.52	0.062	0.04	0.062	0.02	0.062	0.04	0.044	0.04
	Disabled	0.034	0.56	0.026	0.06	0.045	0.02	0.045	0.02	0.044	0.06
	sex work client	0.062	0.56	0.064	0.04	0.079	0.02	0.079	0.02	0.073	0.06
	drug dealer	0.049	0.56	0.055	0.08	0.05	0.02	0.05	0.06	0.043	0.06
	Gender	0.5	0.58	0.48	0.04	0.49	0.01	0.49	0.02	0.49	0.06
	sex worker	0.061	0.620	0.071	0.020	0.057	0.040	0.057	0.020	0.041	0.040

The results in Table 2 demonstrate that for most subgroups, the $RDS_{proposed}$ estimator and G-SS yielded similar proportion, but variances tended to be higher with the naïve estimator, especially in groups with “Very Low” connectivity (e.g., “Housewife”, “Disabled”, “Gender” categories). For higher degree populations (“Unemployed”, “Sex work client”), all estimators produced relatively consistent estimates, but some, such as the SH-RDS, occasionally produced lower variances compared to others. This suggests that the estimator choice can meaningfully impact the precision and stability of population proportion estimates, particularly in subgroups with skewed or heterogeneous degree distributions.

This finding was consistent with previous studies indicating that naïve estimators often overestimate variance, especially in sparse networks or when degree distributions are highly skewed (Gile & Handcock, 2010; Salganik & Heckathorn, 2004). The improved stability of variance estimates with SH-RDS and VH-RDS aligns with simulation studies demonstrating their robustness to outliers and recruitment bottlenecks (Volz & Heckathorn, 2008). Furthermore, similarity between the $RDS_{proposed}$ estimator and G-SS was observed; this supports recent advancements suggesting that Gile’s successive sampling can approximate population-level degree measures when sampling fractions are non-negligible (Gile, 2011). However, minor discrepancies in the estimates for categories such as “Retired” and “Pimp” highlight the sensitivity of certain estimators to the underlying recruitment structure and population heterogeneity, as documented in empirical network studies (Johnston et al., 2016).

4. Conclusion

The $RDS_{proposed}$ estimator effectively addresses major limitations of existing RDS methods by enabling sampling without replacement, using multiple seeds, and applying adaptive weighting, making it especially suitable for large, complex, and highly connected populations. The simulation results showed that SH-RDS and VH-RDS perform best in small samples, but the $RDS_{proposed}$ estimator outperforms all others as the sample size and network complexity increase. While G-SS remains moderately reliable, the Naïve estimator becomes less dependable with larger samples. Although the $RDS_{proposed}$ estimator offers significant improvements over traditional RDS estimators, especially in large and complex network samples, some practical and methodological limitations remain. These results should be interpreted in the context of several limitations. The estimators' assumptions, such as random mixing, accurate degree reporting, and independence of recruitment, may not be fully met in real-world settings, possibly inflating or deflating observed bias. The accuracy of degree estimation is contingent upon truthful self-reporting, which may be affected by recall bias or social desirability, particularly for stigmatized statuses. Second, sample sizes within some subgroups ("drug cook", "pimp") may be small, leading to unstable variance estimates and wide confidence intervals. Third, the estimators assume a certain degree of random recruitment and network connectivity, which may not hold in populations with tightly clustered or fragmented social networks. Finally, the generalizability of these findings is limited by the specific context of Project 90 and may not extend to populations with different network structures or recruitment dynamics.

Acknowledgements
The authors are grateful for the reviewer's assessment, feedback, and recommendations, which have significantly enhanced the quality of this work.

REFERENCES

Abdesselam, K., Yousfi, H., & Sari, T. (2020). Respondent-driven sampling for hard-to-reach populations: A review of methodological advances and applications. *International Journal of Social Research Methodology*, 23(6), 745–759.

Avery, L., & Rotondi, N. K. (2020). Respondent-driven sampling: A critical review of recent literature. *Current Epidemiology Reports*, 7(2), 61–74.

Avery, L., McAuley, A., & Rotondi, N. K. (2021). Performance of respondent-driven sampling estimators under different recruitment patterns: A case study using the Project 90 dataset. *Journal of Urban Health*, 98(3), 392–406.

Berndt, C. (2020). Respondent-driven sampling and the social networks of hard-to-reach populations. *Social Networks*, 62, 43–52. https://doi.org/10.1016/j.socnet.2020.04.002

Beutner, E. (2024). Delta method, asymptotic distribution. *WIREs Computational Statistics, 16*(1), e1634. https://doi.org/10.1002/wics.1634

Card, K. G., Lachowsky, N. J., Cui, Z., Shurgold, S., Gislason, M., Forrest, J. I., & Hogg, R. S. (2017). A comparison of respondent-driven sampling and venue-based sampling of gay and bisexual men in Vancouver, Canada. *BMC Medical Research Methodology*, 17, Article 63. https://doi.org/10.1186/s12874-017-0335-3

Crawford, F. W., Wu, J., & Heimer, R. (2018). Hidden population size estimation from respondent-driven sampling: A network approach. *Journal of the American Statistical Association*, 113(521), 240–251.

Fellows, I. E. (2019). RDS Analyst: An innovative tool for the analysis of respondent-driven sampling data. *Field Methods*, 31(2), 207–217.

Gile, K. J. (2011). Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *Journal of the American Statistical Association*, 106(493), 135–146. https://doi.org/10.1198/jasa.2011.ap09475

Gile, K. J., & Handcock, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology*, 40(1), 285–327. https://doi.org/10.1111/j.1467-9531.2010.01223.x

Kim D, Gile KJ, Mathers B, Mirandola M, Gios L, Toskin I, Sabin, K. M. (2026) Comparing snowball sampling and RDS: A methodology and case study. PLoS One 21(1). https://doi. org/10.1371/journal.pone.0331666

Górny, A., & Napierała, J. (2016). Comparing the effectiveness of respondent-driven sampling and snowball sampling in studying migrant populations. *International Journal of Social Research Methodology*, 19(6), 645–661.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44(2), 174–199.

Heckathorn, D. D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*, 49(1), 11–34.

Heckathorn, D. D., & Cameron, C. J. (2017). Network sampling: Respondent-driven sampling, link-tracing sampling, and snowball sampling. In P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *SAGE Research Methods Foundations*. SAGE Publications.

Horvitz, D. G. & Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *Journal of the American Statistical Association*, 47(260), 663–685.

Johnston, L. G., Sabin, K., Hien, M. T., & Huong, P. T. (2016). Assessment of respondent-driven sampling for recruiting female sex workers in two Vietnamese cities: Reaching the unseen sex worker. *Journal of Urban Health*, 93(5), 899–911.

Keygnaert, I., Dias, S. F., Degomme, O., Deville, W., Kennedy, P., Kovats, A., ... & Temmerman, M. (2014). Sexual and reproductive health of migrants: Does the EU care? *Health Policy*, 114(2–3), 215–225. https://doi.org/10.1016/j.healthpol.2013.10.007

Logan, J. E., Hall, J., & Kresnow, M. (2016). Conducting web-based respondent-driven sampling among men who have sex with men in the United States: Implications for HIV surveillance. *JMIR Public Health and Surveillance*, 2(2), e22.

Lyons, C., Ketende, S., Diouf, D., Simplice, A., Maman, S., & Wirtz, A. L. (2017). Potential impact of integrated stigma mitigation interventions in improving HIV/AIDS service delivery and uptake for key populations in Senegal. *Journal of the International AIDS Society*, 20(1), 21418.

Naser, A. Y., et al. (2018). Probability proportional to size without replacement for hidden population estimation: Methodology and application. *Statistics in Medicine*, 37(16), 2430–2441. https://doi.org/10.1002/sim.7649

Salganik, M. J., & Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, 34(1), 193–240. https://doi.org/10.1111/j.0081-1750.2004.00152.x

Sarah, R., · Michelle, A. D., · Jean C. D., · Yea-Hung, C., & Meghan, D. M. (2022). Respondent-Driven Sampling: A Sampling Method for Hard-to-Reach Populations and Beyond. *Current Epidemiology Reports* (2022) 9:38–47.

Spiller, M. W., Gile, K. J., Handcock, M. S., & Mar, C. M. (2018). Evaluating respondent-driven sampling estimators for hidden populations: The role of network degree. *Social Networks*, 55, 29–38.

Spiller, M. W., Gile, K. J., Handcock, M. S., & Mar, C. M. (2023). Advances in respondent-driven sampling diagnostics: Network structure and estimator performance. *Social Networks*, 74, 45–63. https://doi.org/10.1016/j.socnet.2023.03.003

Stein, M. D., Caviness, C. M., Anderson, B. J., Hebert, M., & Clarke, J. (2014). Respondent-driven sampling for the recruitment of injection drug users in rural and urban settings. *Substance Use & Misuse*, 49(3), 315–324. https://doi.org/10.3109/10826084.2013.841246

Stromdahl, S., Hickman, M., Lundgren, T., & Wiessing, L. (2015). A systematic review of web-based respondent-driven sampling: Implications for future research. *Addiction*, 110(1), 4–17. https://doi.org/10.1111/add.12750

Sypsa, V., Paraskevis, D., Malliori, M., Nikolopoulos, G., Kantzanou, M., Psichogiou, M., ... & Hatzakis, A. (2017). Homelessness and other risk factors for HIV infection in Athens, Greece: A case–control study during an outbreak among people who inject drugs. *AIDS and Behavior*, 21(4), 1066–1073.

Volz, E., & Heckathorn, D. D. (2008). Probability-based estimation theory for respondent-driven sampling. *Journal of Official Statistics*, 24(1), 79–97.

White, R. G., Hakim, A. J., Salganik, M. J., Abdala, N., Latkin, C., Spiller, M., ... & Johnston, L. G. (2015). Strengthening the reporting of observational studies in epidemiology for respondent-driven sampling studies: 'STROBE-RDS' statement. *Journal of Clinical Epidemiology*, 68(12), 1463–1471. https://doi.org/10.1016/j.jclinepi.2015.04.002

Wikipedia contributors. (2024, June 14). Delta method. In *Wikipedia, The Free Encyclopedia*. Retrieved September 23, 2025, from https://en.wikipedia.org/wiki/Delta_method