Generalized least squares can overcome the critical threshold in respondent-driven sampling

Sebastien Rocha,1 and Karl Roheb,1,2

a,1Department of Mathematics, University of Wisconsin, Madison, WI 53706; and bDepartment of Statistics, University of Wisconsin, Madison, WI 53706

Edited by Adrian E. Raftery, University of Washington, Seattle, WA, and approved August 21, 2018 (received for review April 21, 2017)

To sample marginalized and/or hard-to-reach populations, respondent-driven sampling (RDS) and similar techniques reach their participants via peer referral. Under a Markov model for RDS, previous research has shown that if the typical participant refers too many contacts, then the variance of common estimators does not decay like $O(n^{-1})$, where $n$ is the sample size. This implies that confidence intervals will be far wider than under a typical sampling design. Here we show that generalized least squares (GLS) can effectively reduce the variance of RDS estimates. In particular, a theoretical analysis indicates that the variance of the GLS estimator is $O(n^{-1})$. We then derive two classes of feasible GLS estimators. The first class is based upon a Degree Corrected Stochastic Blockmodel for the underlying social network. The second class is based upon a rank-two model. It might be of independent interest that in both model classes, the theoretical results show that it is possible to estimate the spectral properties of the population network from a random walk sample of the nodes. These theoretical results point the way to entirely different classes of estimators that account for the network structure beyond node degree. Diagnostic plots help to identify situations where feasible GLS estimators are more appropriate. The computational experiments show the potential benefits and also indicate that there is room to further develop these estimators in practical settings.

Respondent-driven sampling (RDS) is a popular network-based approach to sample marginalized and/or hard-to-reach populations (1). RDS has become particularly popular in HIV research because the populations most at risk for HIV (e.g., people who inject drugs, female sex workers, and men who have sex with men) cannot be sampled using conventional techniques. Several domestic and international institutions use RDS to quantify the prevalence of HIV in at-risk populations, including the Centers for Disease Control (CDC), the World Health Organization (WHO), and the Joint United Nations Program on HIV/AIDS (UNAIDS) (2). The most recent review of the literature in 2015 counted over 460 different RDS studies, in 69 different countries (3).

Because RDS collects samples from link-tracing the relationships in a social network, adjacent samples are dependent. In a simulation study, ref. 4 showed how this can lead to highly variable estimates. Under independent sampling, the variance of standard estimators decays like $O(n^{-1})$. This implies that a sample size of $4n$ will have a 50% smaller SE than a sample of size $n$. However, this does not necessarily hold for RDS. Under a Markov model, ref. 5 showed how the dependence induced by RDS can drastically inflate the variance of traditional estimators, making it decay at a rate slower than $O(n^{-1})$. This implies that reducing the sampling error by 50% can require far more than 4 times as many samples. This means that confidence intervals are much wider than under independent sampling. Using the covariance function derived in ref. 5, this paper studies the generalized least squares (GLS) estimator for RDS. Our theoretical analysis establishes that the variance of the GLS estimator is $O(n^{-1})$. We then derive a feasible GLS (fGLS) estimator based upon the Degree Corrected Stochastic Blockmodel (DC-SBM). Two alternative estimators are derived. These estimators first construct estimates about the spectral properties of the population social graph, which might be of independent interest. Our IGLS estimators easily accommodate any preliminary reweighting of the data to adjust for the sampling biases that occur in RDS (e.g., refs. 6 and 7). We study these estimators with simulations and propose a simple diagnostic plot to compare the different IGLS estimators.

A Simple Motivating Example

Fig. 1 uses a model studied in ref. 8. In this example, the population that we wish to sample is equally divided into two groups: HIV+ and HIV−. The seed participant is selected uniformly at random. Starting with the seed participant, every participant refers two additional participants (as in a complete binary tree). The participant refers a person that matches his or her own HIV status with probability $p$ and refers a person with the opposite status with probability $1−p$. Each referral is independent. Using this sample, we wish to estimate the proportion of the population that is HIV+ (i.e., 50%). Fig. 1 compares two estimators: (i) the sample proportion and (ii) the GLS estimator proposed in this paper.

Under this sampling with replacement model, the variances of both the sample proportion and the GLS estimator have closed form solutions (see ref. 8 and the proof of Theorem 2 in SI Appendix, Section S4). Fig. 1 gives the ratio of these formulas as a function of the sample size $n$. There are three lines, corresponding to $p = 0.6$, $p = 0.75$, and $p = 0.9$. In all cases, the lines are less than 1, indicating that the GLS estimator has a smaller variance than the sample proportion. Under this simulation model, if $p > 0.86$, then the variance of the sample proportion decays slower.

Significance

Respondent-driven sampling (RDS) is a popular technique to sample marginalized or hard-to-reach populations, where participants can refer multiple contacts into the sample. Using the sampled participants, we wish to estimate properties of the population, often the proportion of individuals that are HIV+. Because contacts often share the same HIV status, adjacent samples are dependent. As a result, RDS can lead to highly variable estimates of HIV prevalence. This paper studies an estimation technique for HIV prevalence that is based upon the classical idea of generalized least squares.
Under the stationary \((T, P)\)—walk on \(G\), the sample average of the \(Y_r\)s is an estimate of

\[
\mu = \mathbb{E}(Y_0) = \sum_i y(i)\pi_i.
\]

In general, \(\mu \neq \mu_{\text{true}}\) (where \(\mu_{\text{true}}\) was defined in Eq. 1), and the sample average must be adjusted with sampling weights to obtain an unbiased estimator of \(\mu_{\text{true}}\). Define

\[
y^w(i) = \frac{y(i)}{\pi_iN}, \quad Y^w_r = y^w(X_r).
\]

The sample average of the \(Y^w_r\)s is the inverse probability weighted (IPW) estimator; it is an unbiased estimator of \(\mu_{\text{true}}\) (10). However, the weights \(\pi_iN\) are unknown and must be estimated with additional information, as we describe next.

Under the \((T, P)\)—walk on \(G\),

\[
N\pi_i = \frac{N\text{deg}(i)}{\sum_j \text{deg}(j)} = \frac{\text{deg}(i)}{d},
\]

where \(d = N^{-1}\sum_j \text{deg}(j)\). The popular Volz–Heckathorn (VH) estimator replaces \(d\) with the harmonic mean of those degrees (6). Recall that \(T\) has \(n\) nodes and define

\[
H^{-1}_{\text{deg}}(\pi) = \frac{1}{n} \sum_{r \in T} \frac{1}{\text{deg}(X_r)}, \quad \tilde{\pi}_r = H^{-1}_{\text{deg}}(\text{deg}(i)), \quad y^\pi_r(i) = \frac{y(i)}{\tilde{\pi}_r}.
\]

and \(Y^\pi_r = y^\pi(X_r)\). The VH estimator is the sample average of the \(Y^\pi_r\)s, and it is an asymptotically unbiased estimator of \(\mu_{\text{true}}\) under the \((T, P)\)—walk on \(G\). [In practice, \text{deg}(i) is estimated by asking participants how many contacts they have. Recall that \text{deg}(i) = \sum_j w_{ij}. If the graph is weighted, then the \((T, P)\)—walk on \(G\) exhibits preferential recruitment (as discussed in SI Appendix, Section S1) and the number of contacts will not necessarily align with \text{deg}, making the estimator biased.]

**Remark**

The next section will drop the superscript \(\pi\) and \(\tilde{\pi}\) in \(Y^\pi_r\) and \(Y^\pi_r\). Using the \(Y_r\)s to construct the GLS estimator will lead to an unbiased estimator of \(\mu\). In practice, before doing any of the GLS computations, one could replace the \(Y_r\)s with \(Y^\pi_r\) or \(Y^\pi_r\) to estimate \(\mu_{\text{true}}\). The simulations in this paper use a reweighting that is similar to \(\tilde{\pi}\) but replaces \(H^{-1}_{\text{deg}}\) with a GLS estimator of \(\mathbb{E}(1/\text{deg}(X_r))\). In ref. 7, sampling weights are estimated under an alternative, non-Markovian model. These weights could also be used before doing GLS computations.

**GLS for RDS**

The GLS estimator is the weighted average of the \(Y_r\)s with smallest variance (11); that is, it is the solution \(g^*\) to

\[
\min_{g} \text{Var} \left( \sum_{r \in T} g_r Y_r \right) \text{ such that } \sum_{r \in T} g_r = 1.
\]

Because of the constraint that the weights \(g_r\) sum to 1, the linearity of expectation, and the fact that the \((T, P)\)—walk on \(G\) is stationary, the resulting estimator is an unbiased estimate of \(\mathbb{E}(Y_r)\). Define the covariance matrix \(\Sigma \in \mathbb{R}^{n \times n}\) as

\[
\Sigma_{\sigma, \tau} = \text{Cov}_{\text{RDS}}(Y_\sigma, Y_\tau),
\]

which is assumed to be nonsingular. It can be seen that the solution to Eq. 3 depends upon solving a system of equations
involving the covariance matrix—namely, that \( g^* = (x^T 1)^{-1} x^T \) where \( 2x = 1 \). (Throughout, we use the notation \( 1_M \) for the all-one vector of length \( M \). We drop the length when clear from context.) If \( Y \in \mathbb{R}^n \) is the vector of \( y \)'s, then the GLS estimator can be expressed as

\[
\hat{\mu}_{GLS} = (1^T \Sigma^{-1} 1)^{-1} 1^T \Sigma^{-1} Y. \tag{5}
\]

The rest of this section contains our main theoretical results, which study how

\[
\text{Var}_{RDS}(\hat{\mu}_{GLS}) = (1^T \Sigma^{-1} 1)^{-1} \tag{6}
\]
decays with the sample size.

**Main Result.** In our main result, we assume that \( T \) is a complete binary tree with \( n \) nodes, but we expect the result to hold for more general tree topologies.

**Theorem 1 (Main Result).** Let \( X_\tau : \tau \in T \) be sampled from the \((T, P)\)–walk on \( G \) for a fixed \( N \times N \) transition matrix \( P \) that is irreducible and reversible with respect to a stationary distribution \( \pi \). If \( T \) is a complete binary tree with \( n \) nodes, then the variance of the GLS estimator defined in Eq. 5 decays like \( O(1/n^2) \) as \( n \to \infty \).

The proof, which is contained in SI Appendix, Section S4, does not directly compute the variance of the GLS estimator. Instead, it proceeds by constructing an explicit linear estimator that does not directly compute the variance of the GLS estimator. The resulting estimator, \( \hat{\mu}_{GLS} \), is derived in this section. Two additional estimators based upon the rank-two assumption are derived in Chapter 3.

**Theorem 2.** Under a rank-two model,

\[
\text{Var} (\hat{\mu}_{GLS}) \to \frac{1 + \lambda}{1 - \lambda} \beta^2 \text{ as } n \to \infty. \tag{11}
\]

This proof follows from the fact that under a rank-two model, \( \Sigma^{-1} \) has a closed form expression (see SI Appendix, Section S6).

**Using RDS to Estimate the Spectral Properties of the Graph for fGLS.**

The fGLS estimator depends upon an estimated covariance matrix \( \Sigma \) (e.g., see ref. 13):

\[
\hat{\mu}_{fGLS}(\Sigma) = \Sigma^{-1} Y. \tag{12}
\]

With this notation, observe that \( \hat{\mu}_{fGLS} \approx \hat{\mu}_{GLS}(\Sigma) \).

In our setting, estimating \( \Sigma \) is equivalent to estimating \( \gamma(\cdot) \). We propose and compare several estimators for \( \gamma \). An estimator based upon the DC-SBM is derived in this section. Two additional estimators based upon the rank-two assumption are derived in SI Appendix, Section S3.

**Definition (DC-SBM).** Partition the \( N \) nodes into \( K \) blocks with \( z: \{1, 2, \ldots, N\} \to \{1, 2, \ldots, K\} \), and assign each node \( i \) a value \( \theta_i > 0 \) such that the \( \theta s \) sum to 1 within each block—that is,

\[
\sum_{i: z(i) = u} \theta_i = 1, \quad \text{for all } u \in \{1, \ldots, K\}. \tag{13}
\]

The block membership of node \( i \) is \( z(i) \), and the parameter \( \theta_i \) controls the degree heterogeneity within each block. Let \( B \) be a symmetric \( K \times K \) matrix such that \( B_{ab} \geq 0 \) for all \( a, b \in \{1, \ldots, K\} \). Under the DC-SBM:

\[
\mathbb{P}(\{i, j\} \in E) = \theta_i \theta_j B_{z(i), z(j)}
\]

for all pairs \( i, j = 1, 2, \ldots, N \) and each possible edge is independent.

In much of the previous literature on the DC-SBM, the full network is observed, and we wish to estimate the partition \( z \). In this paper, we presume that \( z \) is observed on the sampled nodes in the \((T, P)\)–walk on \( G \), and we wish to estimate the spectral properties of \( P \). This is reasonable in RDS because each participant takes a survey that records several salient demographic variables (e.g., gender, race, neighborhood, etc.). In practice, the block labels should be chosen such that they are highly autocorrelated from one referral to the next. Many RDS papers already report such statistics. For example, the original RDS paper (1)
presented four empirical transition matrices on four different demographic partitions (i.e., race, gender, drug preference, and location).

The derivations below condition on the block labels $z$; only the graph $G$ is random. Let $A \in \{0, 1\}^{N \times N}$ be the (random) adjacency matrix; $A_{ij} = 1$ if and only if $(i, j) \in E$. Define $G \in \mathbb{R}^{N \times N}$ as a diagonal matrix with $(i, i)$-th element $\sigma_{ij}$. Define $G = G^{-1} \sigma$ as a population version of $P$.

The inspiration for the following estimators is based on a population version of the chain and relies on three results. Define the matrix $\hat{Q} = \mathbb{E}(\tilde{Q})$ such that for any two blocks $u, v$,

\[
\hat{Q}_{uv} = \frac{1}{n} \times \text{number of referrals from block } u \text{ to block } v. \tag{14}
\]

Proposition 1 below shows that $\hat{Q}$ is an estimator of $B$ under a $(\bar{T}, \bar{\mathcal{P}})$—walk on $G$. Then, Proposition 2 shows that a normalized version of $B$ has spectral properties that match the spectral properties of $\mathcal{P}$. Finally, under the DC-SBM, if the smallest expected degree is growing fast enough, then $P$ converges to $\mathcal{P}$ in spectral norm (e.g., see ref. 16). So estimates of the spectral properties of $P$ are similar to the spectral properties of $\mathcal{P}$. With these facts in mind, we propose estimating the spectral properties of $P$ with the spectral properties of a normalized version of $\hat{Q}$. We let $Z \in \{0, 1\}^{N \times K}$ be such that $Z_{ij} = 1$ if and only if $z(i) = j$.

**Proposition 1.** If $\mathcal{P}$ is constructed from the DC-SBM and if $\hat{Q}$ is computed via a sample from the $(\bar{T}, \bar{\mathcal{P}})$—walk on $G$, then

\[
\mathbb{E}(\hat{Q}) = B/m,
\]

where $m = 1^T B 1$.

**Proposition 2.** Define $D_B \in \mathbb{R}^{K \times K}$ to be a diagonal matrix that contains the row sums of $B \in \mathbb{R}^{K \times K}$—that is, $D_B = \text{diag}(B 1_K)$—and define $B_\ell = D_B^{-1/2} B B_\ell D_B^{-1/2}$. Define $U$ and $\Lambda$ via the eigendecomposition, $B_\ell = U \Lambda U^T$. Define $\beta^* = (y, f^*_\pi)\pi^*$, where $\pi^*$ is the stationary distribution of $\mathcal{P}$. Then, (i) the nonzero eigenvalues of $B_\ell$ are identical to the nonzero eigenvalues of $\mathcal{P}$; (ii) the columns of $f^* = \sqrt{m} Z D_B^{-1/2} U$ are eigenvectors of $\mathcal{P}$; and (iii) if $X$ is sampled from $\pi^*$, then

\[
\beta^*_\ell = \mathbb{E}(y(X) f^*_\ell(X)), \quad \ell \leq K. \tag{16}
\]

The proofs of the propositions are given in SI Appendix, Section S5. We now introduce our estimator of $\Sigma$ and $\mu$.

**SBM-fGLS.** Using $\tilde{z} : T \to \{1, \ldots, K\}$ as an observed partition of the nodes (e.g., by demographic characteristics), the SBM estimator of $\Sigma$ is computed with the following steps. Each step uses a plug-in estimator using the previously derived formulas. After the statement of the algorithm, the steps are matched to the motivating equation.

For notational convenience, denote $Y_{\tau}, \tilde{z}(\tau)$, and $\text{deg}(\tau)$ as $y(X_{\tau}), \tilde{z}(X_{\tau})$, and $\text{deg}(X_{\tau})$ for each sampled individual $X_{\tau}$. Moreover, suppose a one-to-one mapping between the node set of $T$ and $\{1, \ldots, n\}$.

i) Compute $\hat{Q}$ via Eq. 14 using the block memberships $\tilde{z}(\tau)$.

Define $\hat{Q}^{(S)} = (\hat{Q} + \hat{Q}^T)/2$. This symmetrization ensures the eigenvalues are real-valued.

ii) Row and column normalize $\hat{Q}^{(S)}$, as $\hat{Q}_L = D^{-1/2}_Q \hat{Q}^{(S)} D^{-1/2}_Q$, where $D_Q = \text{diag}(\hat{Q} 1_K) \in \mathbb{R}^{K \times K}$.

iii) Take an eigendecomposition of $\hat{Q}_L = \tilde{U} \tilde{\Lambda} \tilde{U}^T$. \tag{17}

iv) Compute $\tilde{f} = Z D^{-1/2}_Q U$, where $Z \in \{0, 1\}^{n \times K}$ contains $Z_{ij} = 1$ if $\tilde{z}(i) = j$.

v) For $\ell = 1, \ldots, K$, compute $\tilde{\beta}_\ell = \frac{1}{Z} \sum_{\tau} Y_{\tau} \tilde{f}_\ell(\tau)$, where $\tilde{f}_\ell(\tau)$ is the $(\ell, \tau)$ element of $\tilde{f}$.

vi) Compute an estimate of the autocovariance function as $\gamma_{SBM}(d) = \frac{K}{\ell = 1} \beta^2_\ell \Lambda_{\ell \ell}$.

vii) Define $\hat{s}^2$ to be the sample variance of the $Y_{\tau}$. For $\sigma, \tau \in T$,

\[
\gamma_{SBM}(\sigma, \tau) = \begin{cases} \gamma_{SBM}(\sigma) + \hat{s}^2 & \text{if } \sigma \neq \tau, \\ \gamma_{SBM}(0) & \text{if } \sigma = \tau, \end{cases}
\]

where $\hat{s}^2$ provides for Tikhonov regularization in $(\hat{\Sigma}_{SBM})^{-1}$.

viii) Define $\hat{v} \in \mathbb{R}^n$ to solve the system of equations $\hat{\Sigma}_{SBM} \hat{v} = 1$.

ix) Estimate $\mathbb{E}(Y_{\tau})$ with $\sum_{\tau \in \gamma} Y_{\tau} / \sum_{\tau \in \gamma} \gamma_{SBM}(0)$.

Step i comes from Proposition 1. Steps iv and v come from Eqs. 15 and 16 in Proposition 2. Step vi comes from Eq. 9. In all of the plug-in formulas, it is unnecessary to estimate $m$ because we must only specify $\Sigma$ up to a constant of proportionality; this constant appears in both the numerator and denominator of $\mu_{G_{SBM}}$ in step ix.

**Simulations.** This section compares the SBM-fGLS estimator to the VH estimator via simulation. Each simulated sample is collected by tracing contacts in social graphs collected in the National Longitudinal Study of Adolescent Health (Add Health). In the 1994-95 school year, the Add Health study collected a nationally represented sample of adolescents in grades 7-12. The sample covers 84 pairs of middle and high schools in which students nominated up to five male and five female friends in their middle/high school network (17). In this analysis, all graphs are restricted to the largest connected component. SI Appendix, Section S1 performs a similar simulation on the Colorado Spring Project 90 network (18). These networks were previously studied in refs. 4 and 19. The simulation was performed without replacement on the directed edges; both of these settings are different from the model used in the theoretical results. Details of the simulation settings are given in SI Appendix.

Fig. 2 shows the RMSE for fGLS and VH estimators; $\text{RMSE} = \sqrt{\mathbb{E}(\hat{\mu} - \mu)^2}$, Overall, the SBM-fGLS estimators have a smaller RMSE. Each panel in Fig. 2 has one line with an asterisk. These lines correspond to the same school, which has both (i) a referral bottleneck between the white and black populations and (ii) a referral bottleneck between the high school and middle school. None of the fGLS estimators model both bottlenecks, yet they perform well.

The most difficult quantity to estimate, high school, also has one of the largest absolute reductions in RMSE. This is consistent with the broader pattern of the experiments and the theory: The VH estimator can have excessive error on quantities that are aligned with the community structures in the network. None of the fGLS estimators perform well in such cases. However, we must
be careful to extrapolate either the frequency or magnitude of
the fGLS improvement. These are highly empirical quantities.
Moreover, the social networks that are available to perform sim-
ulation experiments (Add Health and P90) are not necessarily
representative of the typical RDS population. More discussion
of the idiosyncrasies of these networks is given in SI Appendix,
Section S1.

Fig. 3 presents a diagnostic plot to evaluate the fGLS esti-
mators using only data that are observed in a single sam-
ple. This diagnostic plot was created from the first sim-
ulated sample taken on the school that has the asterisk
in Fig. 2.

The horizontal axis in Fig. 3 gives eigenvalue(s) of \( \Sigma \)
estimated by the fGLS technique. The vertical axis gives the plug-in
estimate for the RSE:

\[
RSE(\hat{\Sigma}) = \sqrt{\frac{\text{Var}(\hat{\mu}_{\text{GLS}}(\hat{\Sigma}))}{\text{Var}(\hat{\mu})}} = \sqrt{\left( \frac{1}{\text{Var}(\hat{\mu})} \right)} \left( \frac{1}{1} \right) \hat{\Sigma}^{1/2} \Sigma^{1/2} \cdot
\]

We should prefer the fGLS estimators that have a smaller ratio.
As is justified in more detail in SI Appendix, Section S6, estima-
tors with smaller RSE make reductions in the variance by taking
advantage of the dependencies. Notice how the fGLS estima-
tors have smaller ratios for the outcomes of black, white, and
high school. For these outcomes, fGLS significantly reduces the
RMSE in Fig. 2. It fails to identify the reduction in RMSE for the
outcome male. For Asian, Hispanic, and male, the ratio of SEs is
closer to 1.

**Summary**

This paper derives and studies GLS and fGLS estimators that
account for the covariance between samples in an RDS. Under
the Markov model where the covariance between samples is
known, Theorems 1 and 2 show that the variance of the GLS esti-
mator decays like \( O(n^{-1}) \). To estimate the covariance be-
 tween samples, we use the fact that the covariance between adja-
cent samples can be exactly specified in terms of the spectral
properties of the Markov transition matrix (5, 20–24). These
essential spectral properties of the network can be estimated
from the observed data under the DC-SBM and the rank-two
model.

Simulations shows in simulations on the Add Health networks
that the fGLS estimates typically have smaller RMSEs than VH
estimates. This simulation is performed under a more realistic
model than the models used in the technical results (Theorems
1 and 2 and Propositions 1 and 2). First, the RDS is simulated
on social graphs that were recorded in the Add Health study
(though not rank-two nor simulated from the DC-SBM). Second,
the sampling is without replacement. Third, the edges have not
been symmetrized. Despite these departures from the reversible
Markov model in the technical results, the estimators appear to
still perform well. This finding is empirical, and given that these
networks are not necessarily representative of the typical RDS
population, we must be careful to extrapolate this intuition to other scenarios.

The diagnostic plots in Fig. 3 help to determine whether the outcome of interest is correlated in the observed sample. For quantities that are correlated (e.g., race, ethnicity, and school), Fig. 2 shows that fGLS estimates significantly reduce the RMSE.

**SI Appendix.** Sections S2 and S3 present two additional simulations to investigate the role of (i) sample size, (ii) referral rates, (iii) alignment of the outcome \( y \) with the blocks \( z \), and (iv) preferential recruitment. In those simulations, when the outcome of interest correlates or aligns with the underlying structure of the graph and the referral rate is larger than the critical threshold identified in ref. 5, fGLS estimators can appreciably reduce the variability over previous estimators. In some simulations, the fGLS estimators have a smaller RMSE with 500 samples than the VH estimators have with 1,000 samples. While the fGLS estimators are derived under a Markov model, all simulations were performed under a without-replacement (i.e., non-Markovian) model.

Under the Markov model in Theorem 1 and under the simulations on the networks to which we have access, our results suggest effective ways (i) to diagnose strong dependence between samples and (ii) to alleviate such dependence. However, we must be careful in extrapolating specific values from the simulations (e.g., the amount that fGLS reduces the RMSE). The Add Health and P90 networks that are available to perform simulation experiments are not necessarily representative of the typical RDS population. The RMSE of the VH estimator and magnitude of the reduction in RMSE from fGLS are two highly empirical quantities that change between networks and outcomes.

**ACKNOWLEDGMENTS.** S.R. is supported by NSF Grants DMS-1149312 (CAREER), DMS-1614242, and CCF-1740707 (TRIPODS). K.R. is supported by NSF Grant DMS-1612456 and Army Research Office Grant W911NF-15-1-0423.