

The effect of differential recruitment, non-response and non-recruitment on estimators for respondent-driven sampling

Amber Tomas

*Department of Statistics
The University of Oxford
Oxford, UK*

and

Krista J. Gile

*Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA, USA*

Abstract: Respondent-driven sampling is a widely-used network sampling technique, designed to sample from hard-to-reach populations. Estimation from the resulting samples is an area of active research, with software available to compute at least four estimators of a population proportion. Each estimator is claimed to address deficiencies in previous estimators, however those claims are often unsubstantiated. In this study we provide a simulation-based comparison of five existing estimators, focusing on sampling conditions which a recent estimator is designed to address. We find no estimator consistently out-performs all others, and highlight sampling conditions in which each is to be preferred.

Keywords and phrases: Respondent Driven Sampling, non-response.

Received January 2011.

1. Introduction

Traditional sampling and estimation begins with a list of all units in the population of interest, i.e. a *sampling frame*. Typically a random sample of units is selected from the sampling frame, and estimates are calculated based on the sample values and the probabilities of selection for the sampled units. However, often it may be of interest to obtain estimates of population quantities for *hard-to-reach populations* (sometimes referred to as *hidden populations*) - populations for which it is difficult or impossible to obtain a practicable sampling frame. *Respondent-driven sampling* (RDS) is a variant of link-tracing network sampling designed for sampling hard-to-reach human populations. It has been used most extensively in public health, especially among injecting drug users,

men who have sex with men, and sex workers, in dozens of countries around the world [15]. It has also been used in populations such as unregulated workers [2] and jazz musicians [12].

In RDS, sampling proceeds as follows: first, a small number of individuals are selected in the sample. Once in the sample, participants are given a number of coupons to distribute among others they know within the population, inviting them to join the sample. This process is repeated recursively until a large enough sample has been obtained.

In this paper we focus on the problem of using RDS to estimate a population proportion in a hard-to-reach population with two groups, which we refer to as *infected* and *uninfected*. Several estimators of population proportion have been developed for data collected through RDS, so it is important to understand the conditions under which each estimator is expected to perform well. We use a simulation study to compare the performance of five estimators under various conditions which may be encountered in practice. In particular, we focus on recruitment and referral characteristics which only the recently developed estimator presented in Heckathorn [11] is designed to address.

The original RDS paper [10] suggested using the sample proportion as an estimator of population proportion (we refer to this as the “Naïve” estimator), and showed that this estimator is unbiased under very strong assumptions about the sampling process. Subsequent papers [18, 19, 11, 3] have relaxed some of these assumptions and proposed several alternative estimators. We will refer to these estimators as the *Salganik-Heckathorn (SH) estimator* [18], the *Volz-Heckathorn (VH) estimator* [19], the *Heckathorn (H) estimator* [11], and the *Successive Sampling (SS) estimator* [3]. The SH- estimator has been implemented in the freely-available RDSAT software [20] for several years, and is widely used by researchers implementing RDS. The H-estimator has been implemented in the latest version of this software, RDSAT 6.0 Beta. There is an urgent need to better understand the properties of the Heckathorn estimator in practice, and in comparison to the other available estimators. The VH-, SS- and SH- estimators are implemented in the RDS **R** package [6].

Each of the estimators is easy to implement, and can be shown to behave well under the specific conditions used in its derivation. However, in practice, the assumptions used to derive each of the estimators will not be met, so there is substantial uncertainty about how the estimators behave in real-world applications. For example, most estimators are derived based on the assumption that sampling is with replacement or that seeds are selected randomly, whereas in practice these conditions almost never hold. Gile and Handcock [4] conducted a simulation study to study the VH- and SH-estimators under realistic sampling scenarios which violated these assumptions, and showed that such violations can significantly affect the estimates. Since this work was done the Heckathorn estimator has also been added to the RDSAT software and is already being used in practice. This estimator is designed to address violations of sampling assumptions that were not studied in [4].

In particular, all previous estimators assume that sampled individuals always respond, that they always recruit other individuals when requested, and that

they recruit uniformly at random from their acquaintances in the population¹. In practice, sampled individuals may not respond (non-response), may not always recruit available others (imperfect recruitment effectiveness) and may preferentially recruit people with particular characteristics (differential recruitment). In this paper we investigate the effect such deviations from the sampling assumptions have on the behaviour of the estimators. Thus, our comparison study is well-positioned to investigate the contributions of the H- estimator, which claims to adjust for differential recruitment and recruitment effectiveness [11].

Unfortunately, these features are difficult to distinguish in RDS data. If a coupon is not returned, for example, this could either indicate that a coupon was passed on but the recipient did not respond (non-response), or that the coupon was never passed on (imperfect recruitment effectiveness). Features of sampling are also often confounded with features of the network structure. For example, if infected participants recruit proportionally more infecteds to the sample than is the case for uninfected participants, this could either indicate that infected people tend to know proportionally more infected people (an example of *homophily* - the preferential tendency for individuals to have relationships with others similar to themselves), or that they preferentially recruit their infected counterparts (differential recruitment). For this reason, our simulation study is an invaluable tool for studying the impacts of population and sampling features not discernible from data.

Despite the difficulty in measuring these features, there is evidence of the presence of these sampling patterns in some reported studies. For example, rates of coupon return from the initial sample can be as low as 55% [16]. Rates of coupon return can also differ based on characteristics of the sample; one study [13] found that the rate of return of coupons given to HIV positive recruits was 69% greater than for those given to HIV negative recruits [12]. There is also anecdotal evidence to suggest that individuals tend to recruit others who have particular characteristics, rather than uniformly at random (differential recruitment). For example, people may recruit family members or friends they think would benefit most from the incentives for participating in the study [16]. In studies of HIV prevalence respondents may prefer to recruit acquaintances they suspect might benefit from the testing and possible treatment available to study participants.

In the remainder of this paper we first present an overview of the sampling variants we consider, followed by a discussion of the estimators we will compare, with a particular focus on the SH- and H- estimators. We then use a simulation study to investigate the effects of differential recruitment, imperfect recruitment effectiveness and non-response on the bias and variance of the estimators. In all these simulations the SH- and H- estimates were very similar to each other, so we extend the simulation study further to explore the conditions under which these estimates are likely to differ. Finally, we summarise the results and make some concluding remarks.

¹An individual is said to have been “recruited” if they receive a coupon inviting them to join the study.

2. Variations of sampling structure

The basic method used to select a respondent-driven sample is as follows: First, a small number of population members are selected as *seeds*, typically selected from among individuals known to the researcher. Each seed is given a number of *coupons*, each of which has a unique identifier, and is asked to pass on the coupons to other people they know within the population. When an individual has received a coupon, they are asked to report to a study centre where information of interest is collected by the researcher (such information is also collected from the seeds). A small monetary reward is often offered at this stage to encourage response. The responders are then themselves given coupons, and are asked to hand them on to others they know within the population, usually only to those who have not yet been recruited. In this manner, after the initial selection of seeds, the sampling is driven by the respondents. Those who report to the study centre are known to those who have already been selected, and recruitee-recruiter relationships can be determined from the identifiers of the coupons. The information available on which to base an estimate is therefore information collected from the respondents and from the recruitment patterns. Respondents are also asked how many people they know within the population of interest. This provides an estimate of *degree*, as described later.

In this paper, we focus on variations of three components of the sampling process:

- **Differential Recruitment:** Respondents are more likely to pass coupons to contacts with certain characteristics. We consider both the situation in which respondents pass coupons preferentially to others similar to themselves, as well as the case where all respondents preferentially pass coupons to contacts with certain characteristics.
- **Recruitment Effectiveness:** This refers to respondents' rates of coupon distribution. In particular, we study cases in which recruitment effectiveness varies by individual characteristics.
- **Non-response:** Some people receiving coupons choose not to participate in the study. We consider cases in which this tendency varies by individual characteristics.

For each of these conditions, we consider performance based on the characteristic “infection status,” as well as performance based on the characteristic “degree.” The specific parameters of our study are described more precisely in Sections 5 and 6.

3. Estimators of a population proportion

In this section, we briefly present the form of the five RDS estimators we consider in the simulation study.

In the following presentation of the estimators we use π_i to represent the true, unknown probability of selection for individual i . The set of individuals selected

in the sample is denoted by s , and s_g denotes the intersection of s and the group g . We assume that the response variable is binary, and refer to the levels as infected and uninfected. A is used to denote the set of all infected individuals, and B to denote the set of uninfected individuals. The population quantity of interest is the proportion of infected individuals, denoted P_A . Following standard sampling notation, we denote population totals by upper-case letters and sample quantities by lower-case letters. Hence $\Pi_g = \sum_{i \in g} \pi_i$, for example. The population size is denoted by N , and the sample size by n . We use d_i to denote the degree of individual i . The *degree* of an individual is considered to be the number of other individuals in the population to which that individual could potentially pass a coupon.

Most of the estimators described in this paper are functions of Hansen-Hurwitz [9] estimators, or, in the case of the SS estimator, of the closely related Horvitz-Thompson estimators. The *Hansen-Hurwitz estimator* of a population total Y is unbiased and given by [9]

$$\hat{Y} = \sum_{i \in s} \pi_i^{-1} y_i. \quad (1)$$

The *generalised Hansen-Hurwitz estimator* of a population mean, \hat{Y}/\hat{N} , is not unbiased but is asymptotically unbiased. We refer to any ratio of Hansen-Hurwitz estimators as a generalised Hansen-Hurwitz estimator.

3.1. Naïve estimator

The Naïve estimate [10] is equal to the sample proportion of infected individuals, i.e.

$$\widehat{P}_A^N = \frac{n_A}{n}. \quad (2)$$

If the true sampling probabilities are given by $\pi_i, i = 1, \dots, N$, then \widehat{P}_A^N is a generalised Hansen-Hurwitz estimator of

$$\frac{\Pi_A}{\Pi_A + \Pi_B} \quad (3)$$

(an explanation is given in Appendix A).

Equation (3) shows that if the sampling probabilities of all individuals are equal, then \widehat{P}_A^N is a generalised Hansen-Hurwitz estimator of N_A/N . Thus, under very restrictive assumptions, \widehat{P}_A^N is asymptotically unbiased for P_A ². If the true probabilities of selection for infected individuals are greater than those of uninfected individuals, then it can be seen from (3) that \widehat{P}_A^N will be biased upwards.

²We use the term asymptotically unbiased somewhat reluctantly, as the asymptotics require infinite population size which is often a poor approximation to the scenarios where RDS is used. However, since asymptotic arguments are used to justify the use of these estimators, we retain them here, noting that it is precisely the failure of assumptions necessary for these asymptotics that necessitates studies such as this one.

3.2. Volz-Heckathorn estimator

The Volz-Heckathorn (VH-) estimator [19] is given by

$$\widehat{P}_A^{\text{VH}} = \frac{\sum_{i \in s_A} 1/\tilde{d}_i}{\sum_{i \in s} 1/\tilde{d}_i}, \quad (4)$$

where \tilde{d}_i denotes the reported degree of individual i . Using the same approach as in Section 3.1, if the true selection probability of individual i is π_i , $i = 1, \dots, N$, it can be shown that $\widehat{P}_A^{\text{VH}}$ is a generalised H-H estimator of

$$\frac{\sum_{i \in A} \pi_i / \tilde{d}_i}{\sum_{i \in A} \pi_i / \tilde{d}_i + \sum_{i \in B} \pi_i / \tilde{d}_i}. \quad (5)$$

The VH-estimator is asymptotically unbiased for P_A if $\pi_i \propto \tilde{d}_i$ for all i , and under many circumstances has been shown to have smaller mean square error than the Naïve estimator [19, 4].

Similarly to the case for \widehat{P}_A^{N} , it can be seen from expression (5) that if the true probability of selection for infected individuals is larger than assumed, then $\widehat{P}_A^{\text{VH}}$ is likely to be biased upwards.

3.3. Successive sampling estimator

The Successive Sampling (SS-) estimator [3] is very similar in form to the VH-estimator. The difference is in the estimation of the sampling weights. Rather than assuming sampling probabilities proportional to degree as in $\widehat{P}_A^{\text{VH}}$, this estimator substitutes a function of the degree, based on approximating the sampling process as a successive sampling process. This process is without replacement, whereas the Naïve and VH- estimators are derived based on the assumption of with-replacement sampling (among others). Therefore, the key contribution of the SS- estimator over the other estimators is the relaxing of the assumption of with-replacement sampling. Its form is given by:

$$\widehat{P}_A^{\text{SS}} = \frac{\sum_{i \in s_A} 1/\tilde{\pi}_i}{\sum_{i \in s} 1/\tilde{\pi}_i}, \quad (6)$$

where $\tilde{\pi}_i$ denotes the estimated sampling probability of individual i , estimated according to the successive sampling procedure described in [3]. Again, using the approach in Section 3.1, if the true selection probability of individual i is π_i , $i = 1, \dots, N$, it can be shown that $\widehat{P}_A^{\text{SS}}$ is a generalised H-T estimator of

$$\frac{\sum_{i \in A} \pi_i / \tilde{\pi}_i}{\sum_{i \in A} \pi_i / \tilde{\pi}_i + \sum_{i \in B} \pi_i / \tilde{\pi}_i}. \quad (7)$$

Note that here the Horvitz-Thompson estimator [14] is used in place of the Hansen-Hurwitz estimator because of the without-replacement sampling as-

sumption³. Similarly to the case for \widehat{P}_A^N and \widehat{P}_A^{VH} , it can be seen from (7) that if the true probability of selection for infected individuals increases relative to uninfected individuals, \widehat{P}_A^{SS} will tend to increase.

This estimator also requires knowledge of the population size, N . The sensitivity of the SS- estimator to the value of N is considered in Gile [3]. For all simulations in this paper we assume that the true value of N is known, and compute estimates using the **R** package RDS [6].

3.4. Salganik-Heckathorn estimator

The Salganik-Heckathorn (SH-) estimator makes use of the fact that it is possible to measure the proportion of within-group and cross-group recruitments in the sample by keeping track of the barcodes of coupons distributed and returned.

The Salganik-Heckathorn estimator is given by

$$\widehat{P}_A^{SH} = \frac{\widehat{C}_{BA}}{\widehat{C}_{BA} + \widehat{C}_{AB} \frac{\widehat{D}_B}{\widehat{D}_A}}. \quad (8)$$

Here \widehat{C}_{AB} denotes the proportion of all individuals recruited by members of group A who are members of group B , and can be thought of as an estimate of the probability a randomly selected recruit in group A recruits from group B . \widehat{D}_A is an estimate of mean degree given by

$$\widehat{D}_g \stackrel{\text{def}}{=} \frac{n_g}{\sum_{i \in s_g} 1/\tilde{d}_i}, \quad (9)$$

and is a H-H estimator of $\sum_{i \in g} \pi_i / \sum_{i \in g} \pi_i / \tilde{d}_i$. Note that as the probability of selection of infected individuals increases relative to uninfected individuals, the ratio $\widehat{D}_B / \widehat{D}_A$ will tend to decrease and hence, from (8), \widehat{P}_A^{SH} will tend to increase.

Because of its more complicated form, and the stronger assumptions required to derive it, the SH- estimator is relatively difficult to study analytically compared to the Naïve- or VH-estimators.

3.5. Heckathorn estimator

This estimator is an extension of the SH-estimator, and was motivated by the need to control for biases introduced by differential recruitment and recruitment effectiveness [11]. Before introducing the form of the estimator, we need to consider the model used in its derivation: that of a Markov chain on degree groups. *Degree groups* are formed by partitioning the reported degrees into groups of

³Although this estimator has the same mathematical form, the difference is that the sampling probabilities used are without-replacement sampling probabilities, as opposed to the draw-wise selection probabilities used in the Hansen-Hurwitz estimator.

contiguous degrees. Then assuming there is only one coupon, perfect recruitment effectiveness and no non-response, the sampling process is treated as a Markov chain. Here, time is indexed by the wave of the sample, and the state of the chain in a given wave is the degree group of the sampled individual. The transition probability from group g to g' is the probability that a coupon will be passed from an individual in group g to someone in group g' . Quite strong assumptions about respondent behaviour are required for this model to be appropriate, notably assumptions which are stronger than those required for the sampling process to be modelled as a Markov chain on nodes.

The H- estimator is constructed by using an “adjusted” degree estimate \widehat{AD}_A in place of the estimates of mean degree in (8). The adjusted degree estimate is defined as

$$\widehat{AD}_A = \frac{\sum_{i \in s_A} \text{RCD}_i}{\sum_{i \in s_A} \left(\frac{1}{d_i} \text{RCD}_i \right)}. \quad (10)$$

RCD_i is called the “recruitment component of degree” for individual i , and is formed based on *degree* groups defined by the reported degree of the respondents. RCD_i is defined as the ratio of the estimated degree group equilibrium probability and the degree group sample proportion. That is,

$$\text{RCD}_i = \frac{\hat{E}_g}{n_g}, \text{ for } i \text{ in degree group } g, \quad (11)$$

where n_g is the number of respondents in degree group g and \hat{E}_g is the estimated equilibrium probability of being in degree group g . For all degree groups g, g' , the probability of transition for the Markov chain from g to g' is estimated by the proportion of recruitments from g to g' , $\widehat{C}_{gg'}$. The estimated equilibrium probability of being in group g , \hat{E}_g , is then calculated from the matrix of estimated transition probabilities [11, pg 171].

The Heckathorn estimator uses the adjusted degree estimates in place of the unadjusted degree estimates in the SH-estimator, which gives

$$\widehat{P}_A^{\text{H}} = \frac{\widehat{AD}_B \widehat{C}_{BA}}{\widehat{AD}_A \widehat{C}_{AB} + \widehat{AD}_B \widehat{C}_{BA}}. \quad (12)$$

If $\text{RCD}_i = 1$ for all i , then it can be seen from (10) that $\widehat{AD}_g = \widehat{D}_g$ for all g , in which case $\widehat{P}_A^{\text{H}} = \widehat{P}_A^{\text{SH}}$.

To define the degree groups we use the recommended method of Heckathorn [11], to mimic as closely as possible what we believe to be the default method implemented in RDSAT. In this method the user specifies a “mean cell size” n_c , which determines the “aggregation level” $AL = \sqrt{n/n_c}$. The range of degrees is then split into AL groups of approximately equal size. The default value is $n_c = 12$, which is the value we use for all simulations carried out in later sections.

It should be noted that if all recruits from group g only recruit others from group g , the estimated equilibrium probability \hat{E}_g will be 1. This means that

\widehat{AD}_A will be based only on individuals within the “absorbing” degree group, which leads to instability. However, this case did not arise for any of the simulations presented in this paper.

Comments on the Heckathorn estimator

\widehat{AD}_A can be seen as some kind of adjustment for divergence from equilibrium of the Markov chain on degree groups. Suppose all the assumptions that are required for the sampling process to be described as a Markov chain on degree groups hold. Then if the Markov chain started in equilibrium, n_g/n is an unbiased estimator of E_g for all degree groups g . Therefore, from (11) it can be seen that it is unlikely for the H- and SH- estimates to differ unless the Markov chain on degree groups does not start in equilibrium. This is likely to occur if seeds are chosen disproportionately from individuals of low or high degree, for example, or if only infected individuals are chosen as seeds and the degree distribution of infected and uninfected individuals differ. Further, we would expect to see larger differences between the H- and SH- estimates if the chain is out-of-equilibrium for a larger proportion of the sample, for example for smaller sample fractions or if convergence is slowed by homophily on degree-group.

The above reasoning is based on the assumption that the sampling process is a Markov chain on degree groups. This is extremely unlikely to ever occur in practice, but may be a good enough model to inform our investigation. We comment further on this in Section 7.

3.6. Summary

All the estimators considered substitute some estimate for the true sampling probabilities into a H-H or H-T estimator: the Naïve estimator assumes all π_i are equal, the VH- and SH- estimators assume $\pi_i \propto \tilde{d}_i$, and the SS- estimator uses a successive sampling estimate of the π_i . Although all the estimators considered in this paper can be shown to be asymptotically unbiased under certain conditions (see Neely [17] and Gile [3] for details), in practice these conditions will not hold. Of interest is how the estimators compare when the assumptions are violated. Differential recruitment, imperfect recruitment effectiveness and non-response are all sampling behaviours which violate the assumptions under which the estimators were derived. How much these types of sampling behaviour affect the estimates is likely to depend on how great the resulting difference is between the true and estimated sampling probabilities [17]. We investigate this via a simulation study.

4. Design of the study

In this section we present the simulation design used for all simulations presented in this paper. Because it is not possible to consider all configurations of

network and sampling parameters, we focus on the effect of varying the sampling parameters for several configurations of the network parameters. Parameter values are chosen to approximate real-life conditions, to be comparable to other simulations in the literature, and to allow us to illustrate qualitative trends. Where possible, parameters were chosen to match the characteristics of the pilot data from the CDC surveillance program [1, 4] or to represent a range of values observed in real data sets. Parameter values were also chosen to match those used in the simulation studies of Gile and Handcock [4]. In particular, because finite population effects are not the focus of this study, we use a fixed population size of $N = 1000$ to match the largest population size studied in Gile and Handcock [4]. We test the robustness of our results to the sample fraction by replicating the study with different sample sizes and highlighting the few cases in which the results differed.

The parameterisation of the sampling process relating to differential recruitment, recruitment effectiveness and non-response will be discussed in more detail in the relevant sections. Because it is not possible to directly estimate differential recruitment, non-response or recruitment effectiveness from a standard RDS, we chose values for these parameters that we felt covered a range of values which might be reasonably expected in practice. We also used some more extreme values to further investigate the qualitative effects of varying these parameters. The values of non-response and non-recruitment we consider correspond to rates of coupon return from roughly 40% to 100%.

The population is modelled as an undirected network, where each node represents one individual in the population⁴. An edge between two nodes i and i' means that there is a non-zero probability that i will recruit i' when given a coupon, and vice versa. The *neighbours* of a node i are all those nodes which have an edge connecting them to node i . The set of potential recruits of an individual i is therefore given by the neighbours of node i , and the number of neighbours is equivalent to the degree of i , d_i .

For all simulations the network consists of 1000 nodes. 200 of the nodes are infected, so $P_A = 0.2$. The overall mean degree of the network is equal to seven, and there is a moderate amount of homophily in all the networks. Specifically, we fix the relative probabilities of edges between two infected nodes and between an infected and an uninfected node, such that the former is five times as likely. In the case of 20% infected nodes and no differential activity (defined below), this implies the probability of an uninfected-uninfected edge is twice the probability of an uninfected-infected edge.

Differential activity (DA) is defined as the ratio of the mean degree of infected nodes to the mean degree of uninfected nodes, i.e. $\overline{D_A}/\overline{D_B}$. It is not unusual in practice for the subpopulations of interest to have differing levels of connectedness based on nodal characteristics. For example, in some communities with strong HIV support groups, HIV positive respondents may have systematically higher degrees than their HIV negative counterparts. Gile and Handcock [4] showed that differential activity can have a significant effect on the bias of es-

⁴From now on we refer to nodes and individuals interchangeably.

timators for RDS, so we consider several values for this parameter which are expected to cover a reasonable range of values based on data available to the authors. The differential activity parameter is varied throughout the simulations and takes values in $\{0.5, \mathbf{1}, 1.8\}$, where bold face is used to represent the default parameterisation.

Given a population represented by the network, we simulate a respondent-driven sampling process. In every simulation we draw ten seeds. Unless otherwise stated, seeds are selected at random with probability proportional to degree from all nodes. Sampling is done without replacement, and all respondents are given two coupons⁵.

We consider samples of size 200 and 500, but unless there are qualitative differences we only present the simulation results for samples of size 200.

For any given value of the population and sampling parameters, the simulations are run as follows:

1. Simulate 1000 networks. To allow for comparison, we used the same networks as were used in Gile and Handcock [4]. These networks were simulated from an exponential random graph model (ERGM) using the **R** package `statnet` [7, 8]. This approach allows for the mean degree, differential activity and homophily to be controlled by including the relevant statistics in the model.
2. For each network, simulate one respondent-driven sample. We therefore have 1000 samples in total.
3. For each sample, calculate the estimates \widehat{P}_A^N , \widehat{P}_A^{SS} , \widehat{P}_A^{VH} , \widehat{P}_A^{SH} and \widehat{P}_A^H . Recall from Sections 3.3 and 3.5 that for all simulations we calculate the SS-estimator assuming the true population size is known, and calculate the H-estimator using $n_c = 12$.

The mean of the 1000 estimates from any estimator is a measure of the expected value of that estimator under the given population and sampling parameters. The variance of the estimates is a measure of the sampling variation of that estimator in addition to how sensitive the estimator is to variations in population structure.

5. Differential recruitment

Whereas the sample design is determined by the survey organiser, differential recruitment is a property of respondent behaviour which is not possible for the survey organiser to observe directly. Therefore, it is important to consider how robust the estimators are to the presence of differential recruitment.

In this section we show that the bias of all the estimators is significantly affected by differential recruitment. The direction of this bias can be explained by considering how differential recruitment affects the probabilities of selection π_i .

⁵We repeated all simulations with four coupons rather than two, but found that this did not significantly alter the results.

We first define the types of differential recruitment which we will consider in this paper.

Definition 1 (Differential recruitment) *Suppose the proportion of neighbours of node i which belong to group g is equal to p_{gi} , for all $g \in G$. Then differential recruitment exists if and only if the probability i passes a given coupon to a neighbour from group g is not equal to p_{gi} for any $g \in G$ and any i .*

In other words, differential recruitment exists if a respondent is more (or less) likely to pass on a coupon to neighbours of some group than if he were choosing uniformly at random from his neighbours. We will consider two types of differential recruitment:

1. Within-group differential recruitment, when nodes preferentially recruit neighbours from within their own group, and
2. Between-group differential recruitment, when all nodes preferentially recruit nodes from a particular group.

For each type of differential recruitment we distinguish between infection-group differential recruitment and degree-group differential recruitment. It is useful to note that either type of differential recruitment can induce the other if the degree distributions of infected and uninfected nodes are not the same.

We first investigate the influence of within-group differential recruitment in some detail, then consider between-group differential recruitment.

5.1. Within-group differential recruitment

We model within-group differential recruitment with the following assumption:

Assumption 1 *Suppose that node i is a member of group g , and is currently recruiting. Then the probability that node i recruits a neighbour in group g is proportional to K_g , and the probability that node i recruits a neighbour from some other group $g' \neq g$ is proportional to 1.*

This is similar to the two-group model of Goel and Salganik [5], except that we allow the strength of differential recruitment to vary by group and do not assume that every individual has a non-zero probability of recruiting any other individual. Clearly if differential recruitment is greater for individuals in group g than for individuals in group g' , i.e. $K_g > K_{g'}$, then the sampling probabilities of group- g individuals will increase relative to those of group- g' individuals. If differential recruitment exists but is similar for both groups, i.e. $K_g \approx K_{g'}$, then the relative sampling probabilities of nodes in groups g and g' should not change dramatically.

The effect of within-degree group differential recruitment on the estimators of P_A can therefore be seen to depend on the distribution of degree among infection groups. If it is the same, then within-degree-group differential recruitment will not induce within-infection group differential recruitment, and hence will not cause additional bias when estimating P_A . If the distribution of degree is

not the same among infection groups, for example when there exists differential activity, then within-degree-group differential recruitment will induce within-infection-group differential recruitment. For example, suppose differential activity is equal to 1.8, so infected nodes have higher mean degree than uninfected nodes. If nodes of high-degree prefer to recruit other nodes of high-degree, this will induce differential recruitment by infection group, because infected nodes will preferentially recruit other high-degree infected nodes.

Note that it is not always possible to distinguish from the sample if differential recruitment exists, because its effect on the resulting sampling chain is similar to that of *homophily* - the proportion of within-group edges. For example, suppose the proportion of recruitments from group A to group B , $\widehat{C}_{AB} = 0.5$. One possibility is that 50% of the neighbours of group A nodes are in group B , and recruitments were made uniformly at random. Alternatively, it might be the case that 70% of the neighbours of group A nodes are in group B , but that group A nodes preferentially recruited other group A nodes. Thus, sampling probabilities for a given group will increase with either differential recruitment or homophily.

Based on the properties of the estimators considered in Section 3, we can predict whether the estimates will increase or decrease with increasing within-group differential recruitment. We would expect that as K_A increases relative to K_B , that all of the estimates will also increase.

Results

Results of the simulations for three levels of differential activity are shown in Figure 1. Note that the counts at the top of the boxplot denote the number of samples for which the estimate was equal to one. Counts at the bottom of the boxplot denote the number of samples for which an estimate could not be calculated⁶.

For all estimators and for all levels of differential activity, as K_A increases relative to K_B the estimates increase, as expected. It can also be seen that the VH-, SH-, H- and SS- estimates are significantly higher or lower than the sample proportion for differential activity levels of 0.5 and 1.8 respectively. The SS- and VH- estimators are generally less biased than the SH- and H- estimators, and the VH- and SS- estimators also have a lower variance than the SH- and H- estimators. Other patterns, which are expected from previous studies, are that the SS-, VH- and SH- estimators correct for differential activity (in contrast to the Naïve estimator) [4, 3], and that the SS- estimates lie between those of the Naïve and VH- estimators [3]. However, there is no evidence that any of the estimators are adjusting for differential recruitment, because the estimates increase or decrease by approximately the same amount as the sample proportion as the differential recruitment level is varied for a given level of differential activity.

⁶The SH- and H- estimators will return a value of 1 if recruitments were made from infection group A to B , but not from B to A . In every case that an estimate could not be computed, it was due to there being no recruitments at all from at least one of the infection groups.

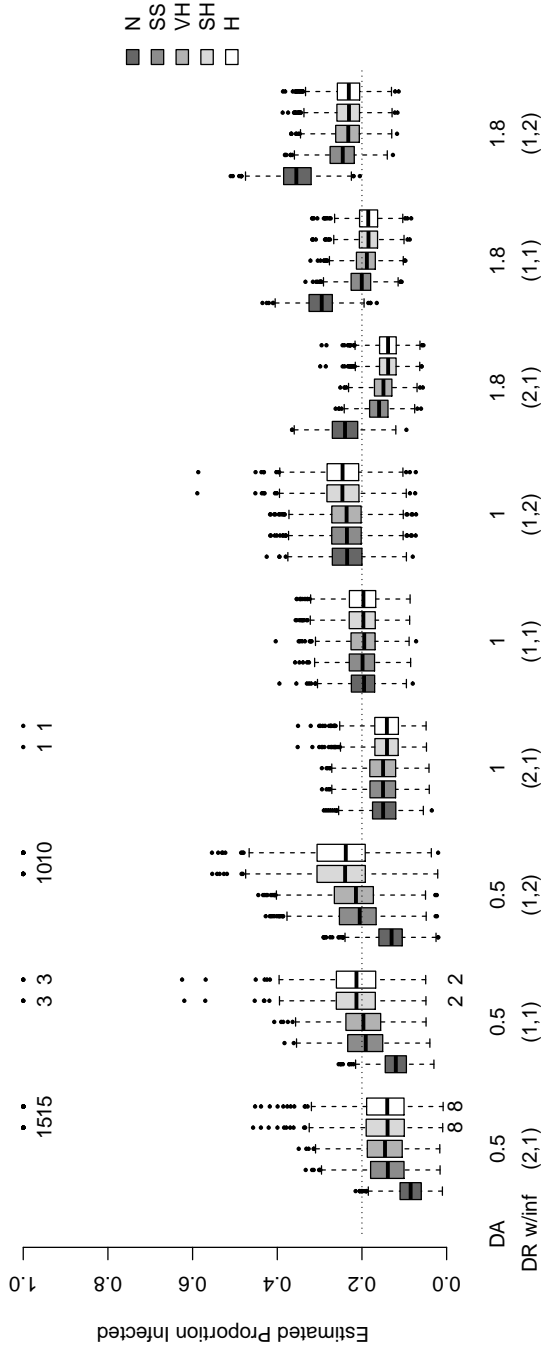


FIG 1. Simulation results for varying levels of within-infection-group differential recruitment ($DR_{w/Inf}$) and differential activity (DA). Differential recruitment is coded as (K_B, K_A) , where K_g denotes the relative probability of a within-group recruitment by individuals in group g .

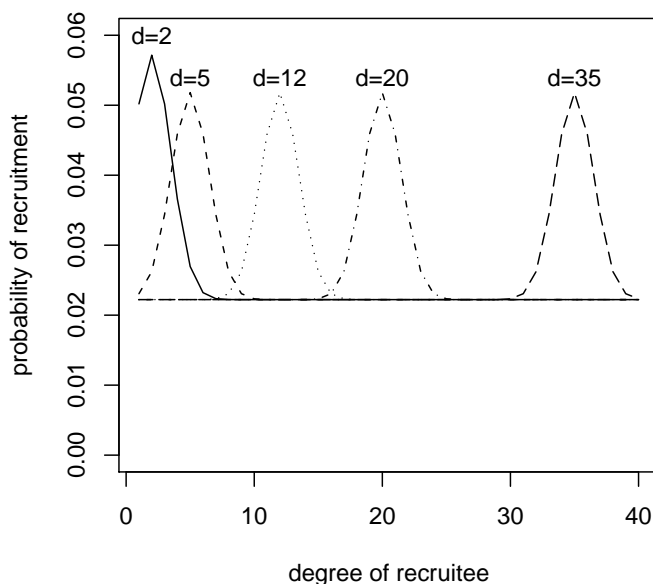


FIG 2. Relative probabilities that a node of degree d will recruit a node with degree shown on the x -axis for the simulations with within-degree-group differential recruitment, for several values of d .

There are no significant differences between the H- and SH- estimates for any of the parameter combinations⁷. We also ran simulations with $K_A = K_B \neq 1$ (results not shown), and these verified that the estimates are very similar to the case that $K_A = K_B = 1$, as expected.

In order to parameterise within-degree-group differential recruitment for our simulations, we assumed that all nodes preferentially recruit other nodes of similar degree to themselves. In this way, no matter how the boundaries of the degree groups are defined by the practitioner, within-degree-group differential recruitment will exist. Figure 2 shows the probability that a node of degree i will recruit a node of degree j , for all j and some i , for the simulations with within-degree-group differential recruitment.

The simulation results are shown in Figure 3. From this figure it can be seen that, for the parameterisation studied in this paper, the effect of within-degree-group differential recruitment on the bias of the estimates is small. In the absence of differential activity, within-degree group differential recruitment has no significant effect on the bias of the estimators. When there exists differential activity, then within-degree group differential recruitment reduces (DA = 0.5) or increases (DA = 1.8) the mean of each estimator. However, the effect of changing levels of differential recruitment is quite small. This can be

⁷All claims of statistical significance are based on paired t-tests of the null-hypothesis of no difference in mean at the 5% level of significance with Bonferroni correction for multiple comparisons.

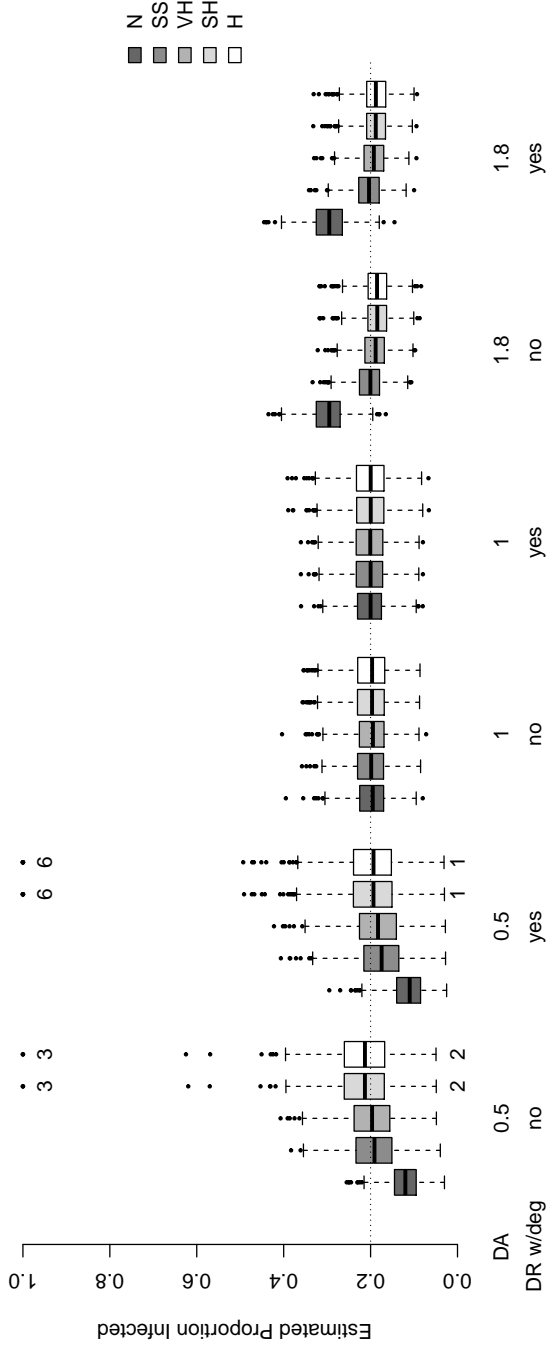


FIG 3. Simulation results for varying levels of differential activity (DA), with and without within-degree-group differential recruitment (DR w/deg).

explained by the fact that the induced levels of within-infection-group differential recruitment are likely to be similar for both infected and uninfected nodes, i.e. $K_A \approx K_B$. Hence, from the results of within-infection-group differential recruitment discussed above, the estimators will not show substantial additional bias.

Similarly to the case for within-infection-group differential recruitment, the variance of the SS-estimator is similar to that for the VH-estimator, and the variance of the SH- and H- estimators is slightly greater. There was only one significant difference between the means of the H- and SH- estimates: when there was both differential recruitment within degree group and differential activity equal to 1.8 (adjusted p-value 0.0018). In this case the mean estimate returned by the H- estimator was 3.7×10^{-4} greater than the mean for the SH- estimator.

5.2. Between-group differential recruitment

Suppose infected nodes are preferentially recruited by all nodes. In this case infected nodes will be more likely to be sampled than if there were no differential recruitment, so the true sampling probability of infected nodes will be higher than that assumed by the estimators. Therefore, following the arguments of Section 3, we would expect all the estimates to increase.

For our simulations, we parameterised differential recruitment between infection-group by the ratio of the probability a node will recruit an infected node relative to uninfected node. Possible values are $\{2, 1, 0.5\}$. Results of the simulations for three levels of differential activity are shown in Figure 4. This shows that for any fixed level of differential activity, as the probability an infected node will be recruited increases relative to the probability an uninfected node will be recruited, all of the estimates of P_A also increase. There is no evidence that any of the estimators considered correct for differential recruitment in this case, although the SS- and VH- estimators tend to be slightly less biased than the SH- and H- estimators. The mean of the estimates changes by about the same amount as for the sample proportion as the level of differential recruitment between infection-group is changed.

Now consider between-degree-group differential recruitment. If high-degree nodes are preferentially recruited by all nodes then this is likely to influence the estimates of P_A only if the occurrence of infection is greater or less in nodes of high-degree than in the general population. That is, we would expect between-degree-group differential recruitment to affect the estimates of P_A only if the degree distribution of infected nodes is different from the degree distribution of uninfected nodes, for example if there exists differential activity. In this case between-degree-group differential recruitment will induce between-infection-group differential recruitment, and we would expect the corresponding bias of the estimators.

For these simulations, we used three levels of differential recruitment: none, low degree nodes are preferentially recruited (“low”), and high-degree nodes are preferentially recruited (“high”). For the latter two cases, the relative probabilities that a node of any degree will recruit a node of degree d are illustrated in

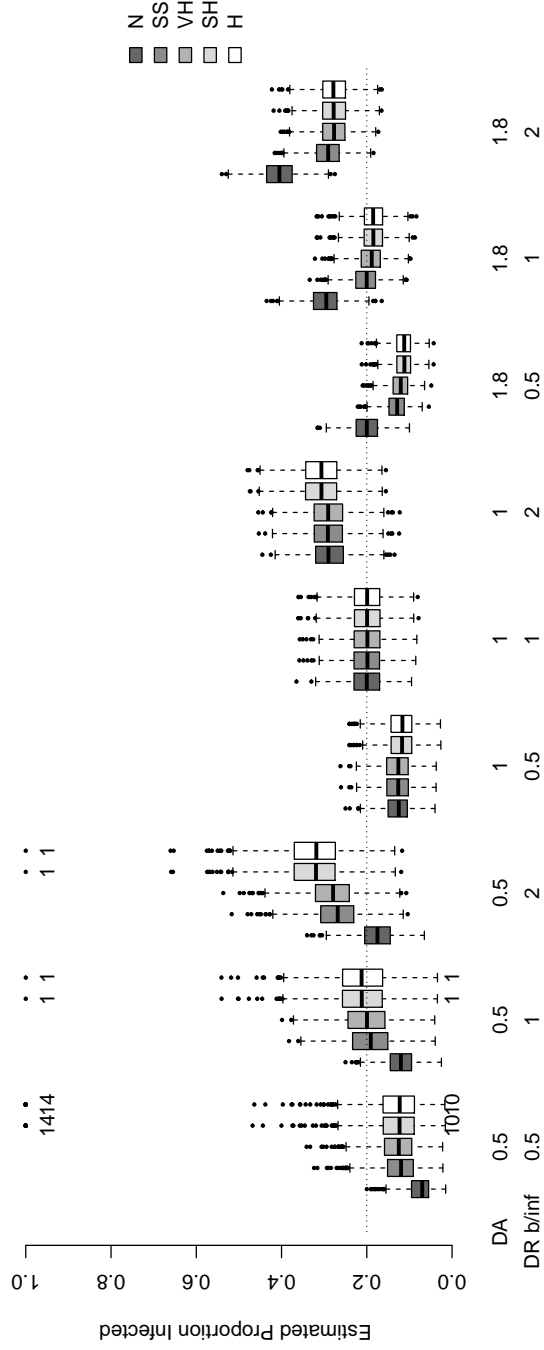


FIG 4. Simulation results for varying levels of between-infection-group differential recruitment ($DR_{b/inf}$) and differential activity (DA). Differential recruitment is coded as the relative probability an infected individual will be recruited, compared to an uninfected individual.

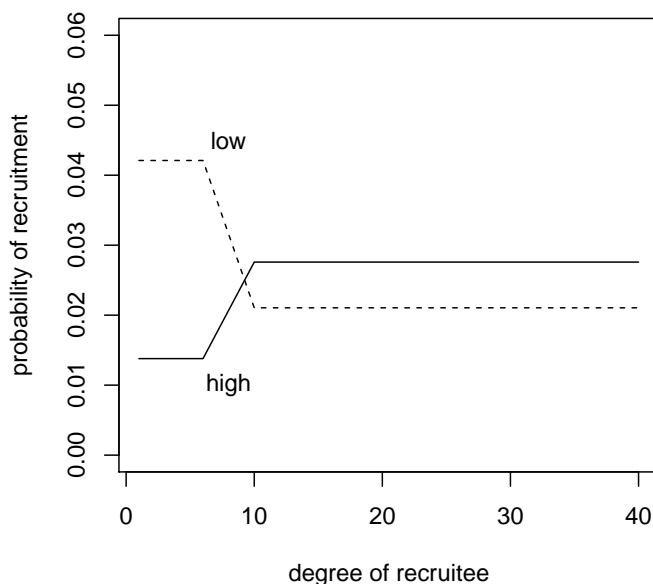


FIG 5. Relative probabilities of recruitment for nodes, for different levels of differential recruitment between degree-group. The dashed line corresponds to preferential recruitment of low-degree nodes, and the solid line to preferential recruitment of high degree nodes. As with all simulations, the mean degree of all nodes is equal to 7.

Figure 5. Results of the simulations for the three levels of between-degree-group differential recruitment and three levels of differential activity are shown in Figure 6. From this figure we can see that when low-degree nodes are preferentially recruited this results in an increase in the mean of the estimators if differential activity is equal to 0.5 and a decrease in mean of the estimators if differential activity is equal to 1.8. This is as we would expect based on the induced differential recruitment between infection groups. There is no evidence that any of the estimators correct for this type of differential recruitment.

6. Recruitment effectiveness and non-response

In this section we show that all estimators behave in a similar and predictable way in the presence of non-response. We also explore the behaviour of the estimators when there is imperfect recruitment effectiveness, and find that the SH- and H- estimators behave differently from the other estimators considered.

We define recruitment effectiveness and non-response as follows:

Definition 2 (Recruitment effectiveness) Denote by R_g the probability that a coupon given to an individual in group g is passed on to another candidate individual in the population, given that such an individual exists. Then R_g is the “recruitment effectiveness” of group g .

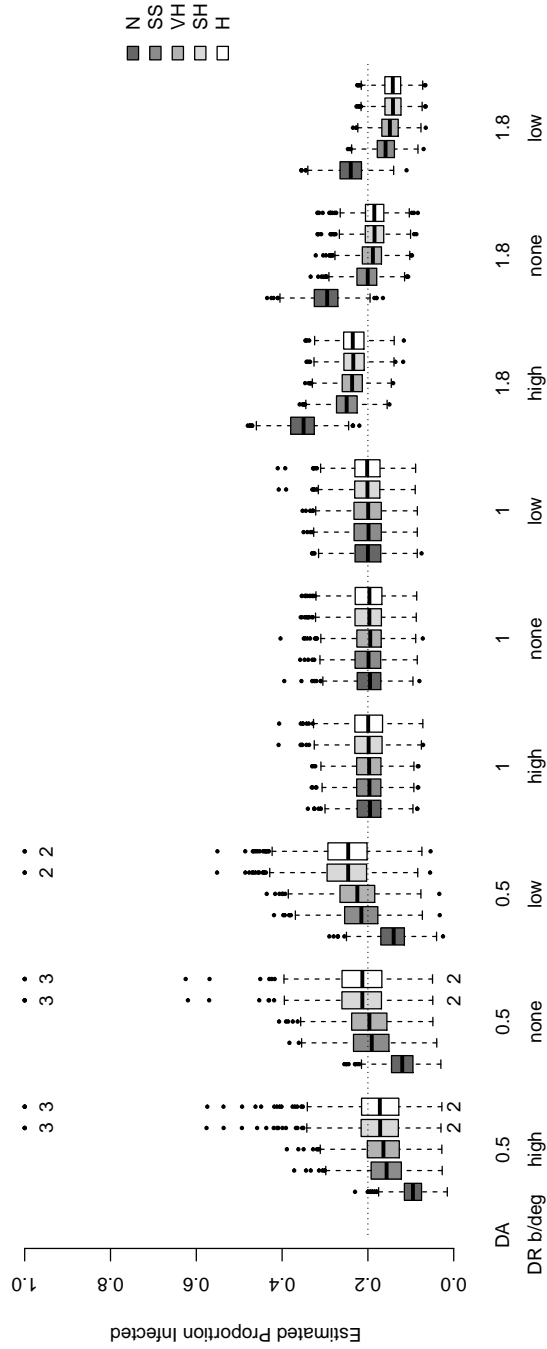


FIG 6. Simulation results for varying levels of between-degree-group differential recruitment ($DR\ b/deg$) and differential activity (DA). The labels indicate whether low ("low") or high ("high") degree nodes are preferentially recruited.

Definition 3 (Non-response) Denote by V_g the probability that an individual in group g reports to the study centre after having been given a coupon. Then if $V_g \neq 1$ we say there exists non-response, and refer to V_g as the “response rate” of group g .

Recruitment effectiveness and response rate are closely linked. If a coupon given to an individual is not returned, this could be either because that individual does not use the coupon to recruit anybody (imperfect recruitment effectiveness, $R_g < 1$), or because the person they recruit does not report back to the study centre (non-response, $V_g < 1$). It is not always possible to tell reliably for which reason a coupon was not returned, nor to directly estimate recruitment effectiveness and non-response rates.

In this section we consider how imperfect recruitment effectiveness and non-response are likely to affect the estimators, and carry out a simulation study to test these arguments.

6.1. Changes in the selection probabilities

Suppose that node $j \in g$ is given a coupon and there is probability $R_g < 1$ that he uses it to recruit another participant. If j does not pass on the coupon, then the children of j miss out on a chance to be recruited. Therefore, if $R_g < 1$ the children of nodes in group g will have a lower probability of selection than if R_g were equal to 1. Hence in general one would expect that if $R_g < R_{g'}$, nodes who have many parents in group g will have a lower probability of selection than nodes with many parents in group g' , everything else being equal. If this effect is not balanced between infection groups, and in the presence of homophily, it is likely to result in biased estimates of P_A . Because recruitment effectiveness does not affect the choice of recruit given that a coupon is passed on, changes in recruitment effectiveness would not be expected to affect the cross-group recruitment probabilities used in the SH- and H-estimators.

6.2. Additional effect of non-response

Rather than changing coupon-passing probabilities, non-response affects which coupons are returned. Hence rather than reflecting the population proportion of infected individuals, \widehat{P}_A will “estimate” the proportion of infected individuals among the population of respondents. Suppose we denote the set of respondents in the population by R . Then by definition 3, $V_g = N_{g \cap R}/N_g$ and so the population proportion of infected individuals can be written as

$$\frac{N_A}{N} = \frac{N_{A \cap R}}{N_{A \cap R} + N_{B \cap R} \frac{V_A}{V_B}} \quad (13)$$

(derivation given in Appendix B). If $V_A < V_B$, N_A/N will be greater than $N_{A \cap R}/N_R$. Thus if there are no other sources of bias, all the estimators will be biased downward. If $V_A > V_B$, all estimators will be biased upward.

Similarly, non-response will affect the estimated cross-group recruitment probabilities. This is because the estimates are based only on the recruitments to nodes who actually respond, which is a subset of all recruitments actually made. For example, if infected nodes have a lower response-rate than uninfected nodes, then the estimated probability of an uninfected node recruiting an infected node will be reduced compared to the case of perfect response.

6.3. Results

6.3.1. Recruitment effectiveness

For our simulations, recruitment effectiveness by infection group is parameterised as a vector (R_B, R_A) , where R_B and R_A are the probability that a coupon is passed on given that a candidate for recruitment exists, for uninfected and infected nodes respectively. We consider values $\{(0.6, 0.9), (\mathbf{1}, \mathbf{1}), (0.9, 0.6)\}$.

Recruitment effectiveness by degree group is parameterised in the form (α, β) , where nodes of degree 1–5 have relative probability α of recruiting, nodes of degree greater than 10 have relative probability β of recruiting, and the relative probability for nodes of degree 6–10 increases linearly from α to β . Possible values are $\{(0.5, 1), (\mathbf{1}, \mathbf{1}), (1, 0.5)\}$.

Results of the simulations for the different levels of recruitment effectiveness by infection group are shown in Figure 7. Figure 7 shows that, for all levels of differential activity, as the recruitment effectiveness of uninfected nodes increases relative to the recruitment effectiveness of infected nodes, the sample proportion of infected nodes decreases. This is as expected due to the population homophily. The VH-estimates and SS- estimates also decrease as the sample proportion decreases, demonstrating that they are biased by different levels of recruitment effectiveness. However, it is interesting to note that neither the SH- nor the H-estimates appear to be influenced to the same extent by recruitment effectiveness as the VH-, SS- or Naïve estimates. If there is no differential activity then the SH- and H- estimators appear to effectively control for different recruitment effectiveness within infection groups. In the presence of differential activity there is a slight bias in the opposite direction to the Naïve estimator, but this tends to be small relative to the bias of the other estimators. We also ran the simulation for a sample size of 500. In this case, shown in Figure 8, the H- and SH- estimators have a substantial bias in the opposite direction to the Naïve and VH- estimators. Thus the simulation results suggest that the SH- and H- estimates correct for bias due to differential recruitment in small sample fractions. Again, there is no significant difference between the means of the SH- and H- estimators.

In the case of different levels of recruitment effectiveness by degree-group, shown in Figure 9, a similar effect can be seen. In this case, if there is no differential activity then different levels of recruitment effectiveness by degree group does not bias the estimators of P_A , because the induced recruitment effectiveness by infection-group is the same for infected and uninfected groups.

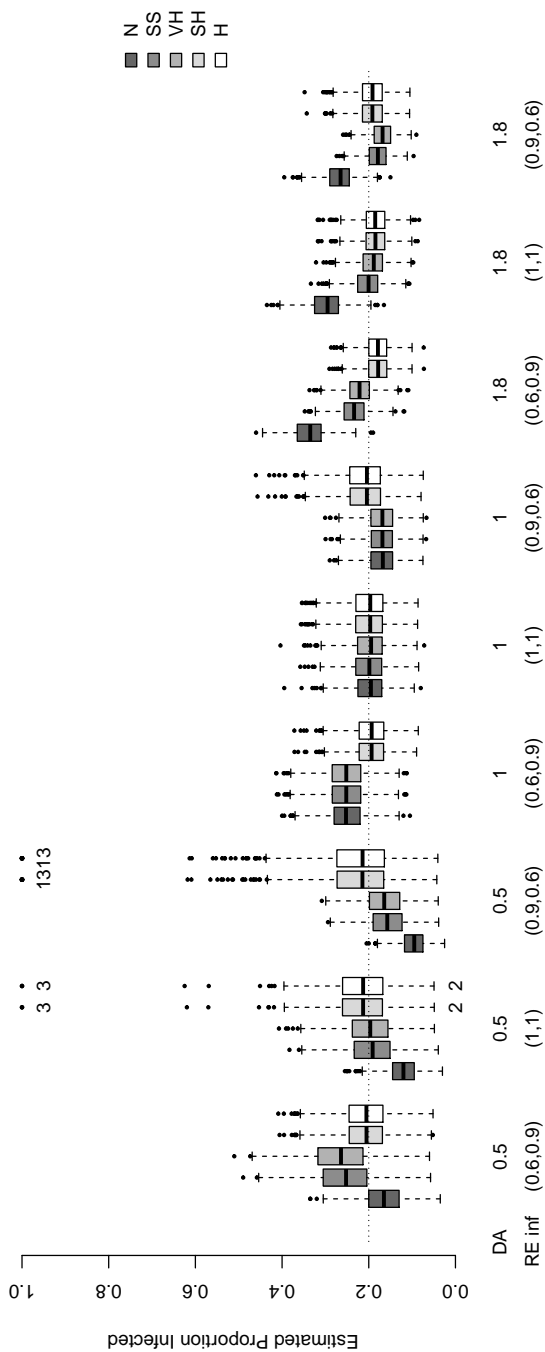


FIG 7. Simulation results for varying levels of recruitment effectiveness by infection group (RE_{inf}) and differential activity (DA). Recruitment effectiveness (the probability of passing-on a coupon) is coded as (R_B, R_A) , where R_g is the recruitment effectiveness of individuals in group g .

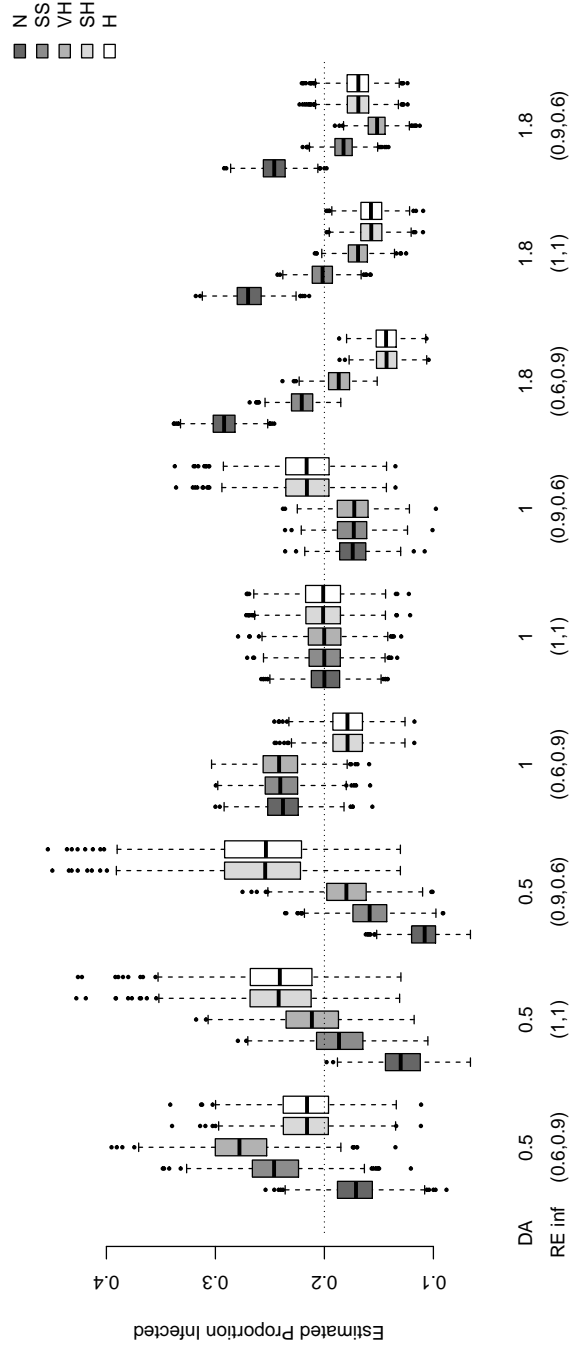


FIG 8. Simulation results for varying levels of recruitment effectiveness by infection group (RE_{inf}) and differential activity (DA), for samples of size 500 rather than 200. Recruitment effectiveness (the probability of passing-on a coupon) is coded as (R_B, R_A) , where R_g is the recruitment effectiveness of individuals in group g .

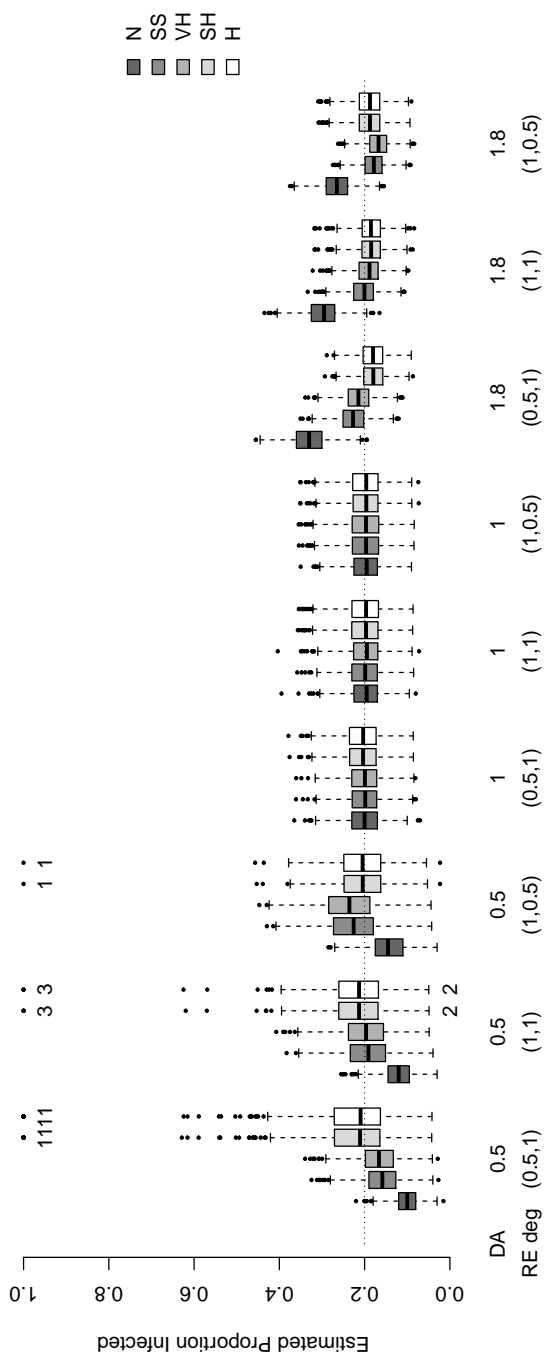


FIG 9. Simulation results for varying levels of recruitment effectiveness by degree group (RE deg) and differential activity (DA). Recruitment effectiveness (the probability of passing-on a coupon) is parameterised in the form (α, β) , where nodes of degree 1–5 have recruitment effectiveness α , nodes of degree greater than 10 have recruitment effectiveness β , and the recruitment effectiveness for nodes of degree 6–10 increases linearly from α to β .

For differential activity greater than 1, if low-degree nodes have higher recruitment rates than high-degree nodes (parameterisation $(1, 0.5)$), this means that on average uninfected nodes will have higher recruitment rates than infected nodes. Accordingly the sample proportion of infected individuals decreases as recruitment effectiveness by degree-group changes from $(0.5, 1)$ to $(1, 0.5)$ for differential activity equal to 1.8, and the reverse is true if differential activity is equal to 0.5. Therefore, the patterns of bias are similar to the case of recruitment effectiveness by infection group. The bias of the H- and SH- estimators increases accordingly for samples of size 500 (not shown).

The difference in mean between the SH- and H- estimators is statistically significant when recruitment effectiveness by degree-group is equal to $(0.5, 1)$ and differential activity is 0.5 or 1.8. In both of these cases the mean of the H-estimates was closer to the true value than the mean of the SH-estimates. However, from the simulation results of Figure 9, this difference may not be large enough to be of practical significance.

6.3.2. Non-response

Non-response by infection group is parameterised as a vector (V_B, V_A) , where V_B and V_A are the probability that a node responds given that they have been recruited, for uninfected and infected nodes respectively. Possible values are $\{(0.6, 0.9), (\mathbf{1}, \mathbf{1}), (0.9, 0.6)\}$. Non-response by degree group is parameterised in the same way as recruitment effectiveness by degree group. Possible values are $\{(0.5, 1), (\mathbf{1}, \mathbf{1}), (1, 0.5)\}$.

Results of the simulations for the different levels of non-response by infection group and degree group are shown in Figures 10 and 11 respectively. From Figure 10, as the response-rate decreases for infected nodes relative to uninfected nodes, i.e. the parameterisation changes from $(0.6, 0.9)$ to $(0.9, 0.6)$, the sample proportion of infected nodes decreases, as expected. Again, all estimators are biased by changes in the relative rates of non-response. As the sample proportion of infected individuals decreases, so too do the Naïve, SS-, VH-, SH- and H-estimates. There are no significant differences between the mean SH- and H-estimates.

For the case of non-response by degree-group, similarly to the case of recruitment effectiveness by degree-group, the results can be interpreted in terms of induced non-response by infection-group. Viewed in this way, the results shown in Figure 11 are consistent with those of Figure 10. There are significant differences between the SH- and H- estimates when there exists both non-response by degree group and differential activity, but again these differences don't appear to be large. None of the estimators seem to control for different rates of non-response between infected and uninfected groups.

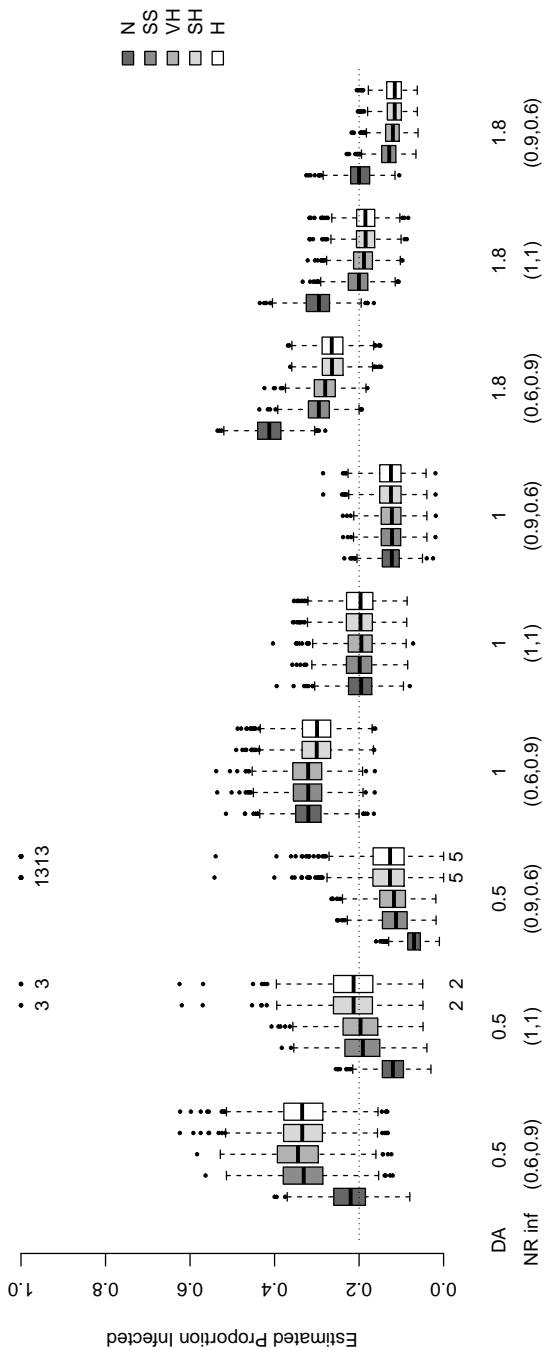


FIG 10. Simulation results for varying levels of non-response by infection group (NR inf) and differential activity (DA). Non-response is coded as (V_B, V_A) , where V_g is the response rate of individuals in group g .

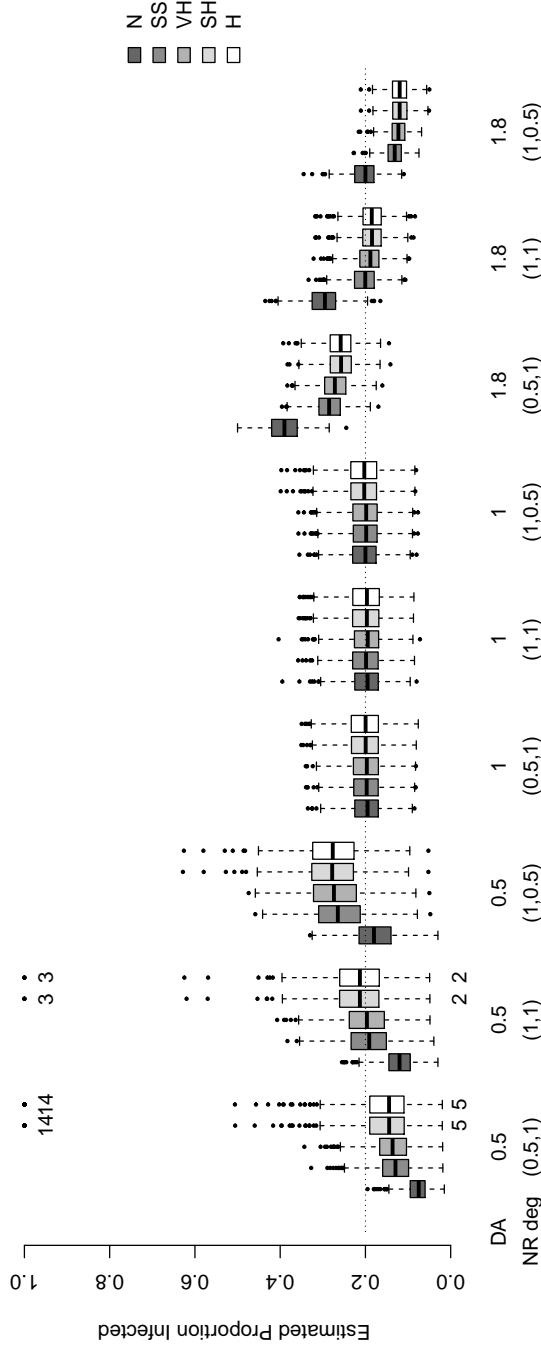


FIG 11. Simulation results for varying levels of non-response by degree group (NR deg) and differential activity (DA). Non-response is parameterised in the form (α, β) , where nodes of degree 1–5 have response rate α , nodes of degree greater than 10 have response rate β , and the response rate for nodes of degree 6–10 increases linearly from α to β .

7. Comparison of the Salganik-Heckathorn and Heckathorn estimators

In Section 3.5 we hypothesised that the SH- and H- estimates would differ if the Markov chain on degree groups was started out of equilibrium. Further, the more slowly the chain reaches equilibrium, the larger the expected difference between the SH- and H- estimates for a fixed sample size.

In the simulations presented in Sections 5 and 6 there were very few statistically significant differences between the mean SH- and H-estimates. Those differences which were significant always occurred when there was differential activity and differences in recruitment or response behaviour between degree groups. These are the cases where the equilibrium probabilities of the Markov chain on degree groups are likely to be furthest from the degree distribution of the seeds. However, in all simulations the differences between the mean (and median) estimates of the H- and SH- estimators were unlikely to be large enough to be of practical concern (all mean differences were less than 0.001), and the variances of the estimates were also nearly identical.

Heckathorn [11, pp. 185] admits that traditional RDS designs are likely to result in only “modest” differences between the two estimators. However, he theorises that

“...changes in research design that induce an association between degree and opportunities to recruit would produce much larger potential effects.”

A design that gave respondents with higher reported degree more coupons than respondents with low reported degree would have such an effect. Such a design can be viewed as equivalent to one where all nodes are given the same number of coupons but nodes of lower degree have much lower recruitment effectiveness. This case was considered in Section 6, Figure 9, and again produced only very small differences between the H- and SH- estimates. Similarly, a design where infected nodes are given more coupons than uninfected nodes is equivalent to the case we considered in Figure 7, where uninfected nodes have lower recruitment effectiveness than infected nodes. Again, only very small differences between the H- and SH- estimates were observed. We also re-ran the simulations using four coupons rather than two, but this did not produce any significant changes in the results.

To try and induce larger differences between the estimates we looked at the effect of increasing homophily (to slow convergence), preferentially selecting seeds from nodes of low or high degree (to start the chain out of equilibrium), and more extreme differences in recruitment effectiveness. The largest differences between the SH- and H- estimates arose when the seeds were selected from nodes of low degree. Compared to the effect of seed selection, increasing differential activity or homophily had little effect on the difference between the estimates. Simulation results for the case that ten seeds are chosen uniformly at random from the twenty nodes of lowest degree are shown in Figure 12. The observable differences between the H- and SH- estimates when seeds are selected from low-degree nodes are statistically significant. It is also very interesting to note that in

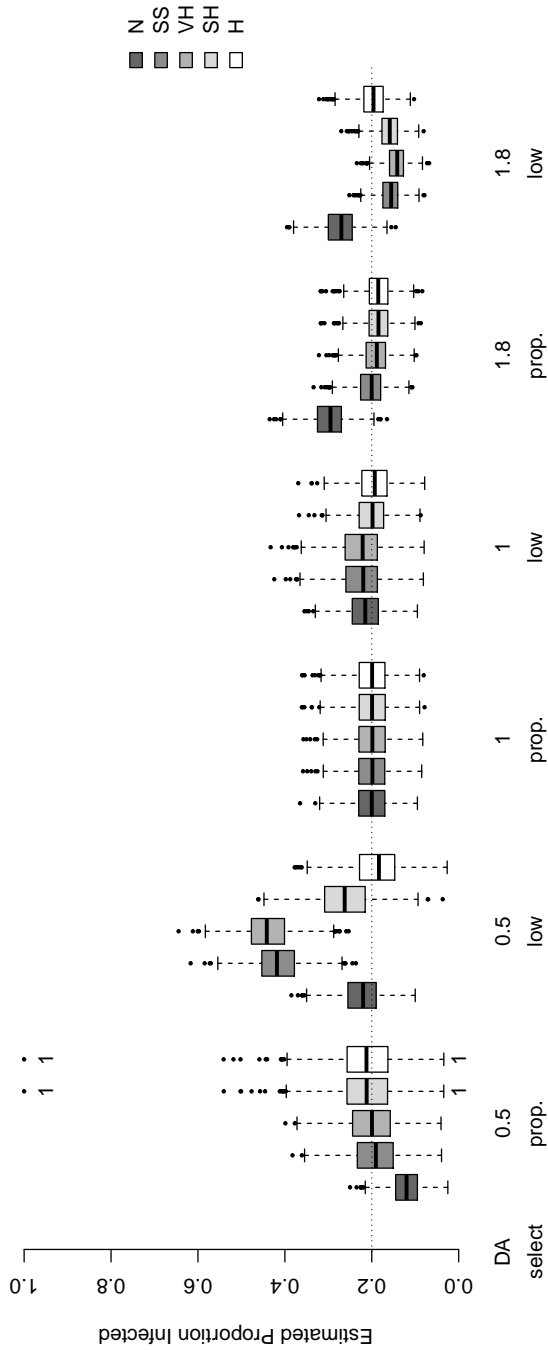


FIG 12. Box-plots of the estimates for different methods of selecting seeds (select), for several levels of differential activity (DA). Seeds are either selected with probability proportional to degree (prop.) or uniformly at random from the twenty nodes of lowest degree (low).

these cases the H-estimator appears to correct somewhat for the bias introduced by the seed selection mechanism.

To investigate further, we considered the differences in absolute deviations from 0.2 of the H- and SH- estimates. Figure 13 shows that on average the H-estimate is closer to the true value of 0.2 than the SH- estimate when seeds are chosen from nodes of either high or low degree. This observation is robust

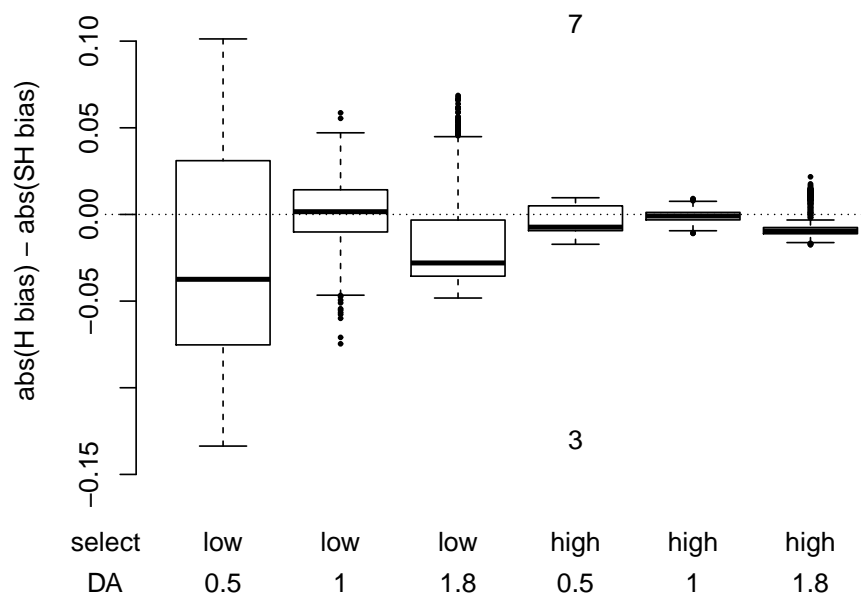


FIG 13. Box-plots of the difference in absolute deviations from 0.2 of the estimates returned by the SH- and H- estimators for different levels of differential activity (DA). Seeds are chosen either from among the nodes of lowest (select = low) or highest (select = high) degree.

to changes in sample-size. The simulation results therefore suggest that the H-estimator corrects for bias due to selecting seeds with degrees not representative of the theoretical equilibrium distribution on degree groups. If there are no other sources of bias, the H- estimator is likely to be less biased than the SH- estimator in this case. However, it should be noted that the seed selection scenarios which produced the observed differences in behaviour were quite extreme, and are unlikely to occur unless by design.

We also considered how the SH- and H- estimators behave when seeds are selected from nodes of low degree and, in addition, there exists differential recruitment, unbalanced recruitment effectiveness or non-response. In this case the biases due to seed selection and recruitment effectiveness or non-response behaved in an approximately additive fashion. Hence depending on the components of the additive bias, either the SH- or H- estimates could have the lower total bias.

In practice, a scenario as extreme as that illustrated in Figure 12 is unlikely to arise. More realistically, it could happen that there exists differential activity

and that seeds are selected only from infected individuals. If infected nodes have a lower or higher average degree than uninfected nodes, then this will induce the same effect as considered above. We ran the simulations for this case and, although the difference between the mean of the H- and SH- estimators was statistically significant for the cases with differential activity and seeds chosen only from infected nodes (and only those cases), the size of the differences (approximately 0.001) is unlikely to be of practical concern. This indicates that it is unlikely that these conditions will result in large differences between H- and SH- estimates in traditional respondent driven sampling designs.

8. Discussion

In this paper we compare the performance of several respondent driven sampling estimators under conditions of differential recruitment, recruitment effectiveness and non-response. Our specific findings are summarised in Table 1. Of particular importance is the performance of the estimator introduced in [11], which claims to adjust for differential recruitment and recruitment effectiveness.

We did not find any evidence that the H-estimator adjusts for differential recruitment or non-response. However, in the case of recruitment effectiveness, the resulting bias of the H-estimator and the SH- estimator is in the opposite direction to that of the Naïve, SS- and VH- estimators, and the size of this bias depends on the sampling fraction (see Figures 7 and 8). For the simulations with sample size 200, this resulted in the SH- and H- estimators effectively controlling for the bias due to imperfect recruitment effectiveness.

When seeds were randomly selected with probability proportional to degree, under all conditions of differential recruitment, recruitment effectiveness and non-response the estimates of the H-estimator were almost identical to those

TABLE 1
Summary of conditions under which tested respondent behaviours induce bias

Condition	Description	Estimator				
		N	SS	VH	SH	H
Differential Recruitment	Non-random coupon-passing among alters					
Within-Group	Different rates of recruitment to own group					
Infection	To own infection group	✓	✓	✓	✓	✓
Degree	To own degree group	†				
Between-Group	One group preferentially recruited					
Infection	Infection group preferred	✓	✓	✓	✓	✓
Degree	Degree group preferred	†	†	†	†	†
Recruitment Effectiveness	Differential rates of coupon-passing					
Infection Group	Infection group passes more coupons	✓	✓	✓	✓*	✓*
Degree Group	Degree group passes more coupons	†	†	†	†*	†*
Non-Response	Differential rates of coupon return					
Infection Group	Infection group returns more coupons	✓	✓	✓	✓	✓
Degree Group	Degree group returns more coupons	†	†	†	†	†
Low-degree Seeds	Seeds selected from low degree nodes	✓	✓	✓	†	

Key

- ✓ Estimator shows bias regardless of differential activity
- † Bias only in the presence of differential activity
- * Bias only with large sample fraction

of the SH-estimator. In Sections 5 (differential recruitment) and 6 (recruitment effectiveness and non-response), we found statistically significant differences between these two estimators in only 10 of 69 simulation cases (sample size 200). The magnitudes of the differences in these cases were small enough to be negligible for practical purposes.

This led us to explore, in Section 7, other conditions which might lead to differences between the H- and SH- estimates. Here, we found that we can induce differences between the two estimators when the following two conditions both hold:

1. The seeds are selected with probability far from the theoretical equilibrium probabilities of the Markov chain on degree groups; and
2. The degree distribution of infected and uninfected nodes is not the same (for example if there is differential activity).

Under these conditions the H-estimator was significantly less biased than the SH-estimator, indicating that the H-estimator may correct for the bias introduced by this type of seed selection. Specifically, we found the largest differences between the estimates of the H- and SH- estimators when seeds were selected from among the nodes of lowest degree. Smaller, but substantial differences were also apparent when selecting seeds from among nodes of the highest degree. It is possible this scenario may arise by design, where study designers try to recruit well-connected population members as seeds.

If there are no sources of bias other than unbalanced selection of seeds by degree, then the H- estimator is likely to have a smaller bias than the other estimators considered in this paper. However, if other sources of bias are present (which will almost always be the case), then these will interact to mean that the absolute bias of the H- estimator may not be the smallest of the estimators considered.

All estimators were subject to bias induced by differential recruitment, recruitment effectiveness and non-response. It is interesting, however, that when the differential recruitment and recruitment effectiveness acted only on degree groups, bias was only present in the estimators when the infection groups had different degree distributions (differential activity $\neq 1$). Without this condition, over or under representing degree groups and their sampling patterns in the estimators does not differentially affect the two groups. Across these simulations, for random selection of seeds the H- and SH- estimators tended to perform similarly in each case. The Naïve estimator (sample mean) tended to exhibit the largest bias, whereas the SS-, VH-, SH- and H- estimators showed reduced bias due to adjusting for differential activity [4]. For all cases the SS- estimates fell between the Naïve and VH- estimates. In general, apart from the conditions of seed selection mentioned above, none of the SS-, VH-, SH- or H- estimators were consistently less biased than any other. They also had similar variances, although variance of the VH- and SS- estimators appeared to be slightly less in many cases, and the estimates returned by the SS- and VH- estimators were not subject to the same instability as the SH- and H- estimates.

Overall, our findings, indicate that no estimator consistently out-performs

the others under all conditions. We find that with small sample fractions, the SH- and H- estimators correct for differential recruitment effectiveness across groups, which induces bias in the other estimators in the presence of homophily. This is an advantage, and therefore these estimators should be used whenever the sample fraction is small and differential recruitment effectiveness is suspected. However, these estimators also are also subject to more instability, and have larger variance and greater bias than the VH- or SS- estimators under many other sample conditions, both in this paper, and as described in [4]. This suggests that absent differential recruitment effectiveness, the VH- or SS- estimators are to be preferred. In the case of large sample fractions, the SS- estimator is clearly to be preferred [3]. Our results do not provide evidence that the H-estimator should be preferred over the SH-estimator. Although we found differences between the H- and SH- estimators in some extreme cases, we suspect these cases are rare enough so as to not justify the additional complexity of the H-estimator. The lack of clear preference among these estimators highlights the need for both new estimators, and also for new diagnostic tests aimed at identifying circumstances in which each estimator is to be preferred.

It is also important to keep in mind the limitations of this study. It was not possible to consider all possible combinations and levels of the parameterisations of the effects studied in this paper. We have tried to focus on a range of simulation conditions and parameter values most likely to highlight contributions of the H-estimator, and considered variations of many parameters at once to draw out potential interaction effects. Nevertheless, there may well be other population or sampling parameters whose levels will impact the results of our study. While we have tried to include any confounders we felt might affect our qualitative findings, any effect sizes should be understood as specific to the simulation conditions under study.

The code used to generate the networks, draw the samples and calculate the estimates is available publicly, or on request from the authors.

Appendix A: Derivation of Hansen-Hurwitz estimators

Consider the Hansen-Hurwitz estimator of Π_A ,

$$\begin{aligned}\widehat{\Pi}_A &= \sum_{i \in s_A} \pi_i^{-1} \pi_i \\ &= n_A.\end{aligned}$$

Thus, taking the ratio of Hansen-Hurwitz estimators for Π_A and $\Pi_A + \Pi_B$ shows that $n_A/(n_A + n_B)$ is a generalised Hansen-Hurwitz estimator of $\Pi_A/(\Pi_A + \Pi_B)$.

Appendix B: Non-response

By definition 3, $V_g = N_{g \cap R}/N_g$ and so the population proportion of infected individuals can be written as

$$\frac{N_A}{N} = \frac{V_B V_A N_A}{V_B V_A N_A + V_A V_B N_B}$$

$$\begin{aligned}
&= \frac{V_B N_{A \cap R}}{V_B N_{A \cap R} + V_A N_{B \cap R}} \\
&= \frac{N_{A \cap R}}{N_{A \cap R} + N_{B \cap R} \frac{V_A}{V_B}}
\end{aligned}$$

as claimed.

References

- [1] ABDUL-QUADER, A. S., HECKATHORN, D. D., MCKNIGHT, C., BRAMSON, H., C. NEMETH, K. G. and JARLAIS, D. C. D. (2006). Effectiveness of respondent-driven sampling for recruiting drug users in New York City: Findings from a pilot study. *Journal of Urban Health* **83** 459–476.
- [2] BERNHARDT, A., HECKATHORN, D., MILKMAN, R. and THEODORE, N. (2006). Documenting Unregulated Work: A Survey of Workplace Violations in New York City. *The Future of Work*.
- [3] GILE, K. J. (2011). Improved Inference for Respondent-Driven Sampling Data with Application to HIV Prevalence Estimation. *Journal of the American Statistical Association* **106** 135–146.
- [4] GILE, K. J. and HANDCOCK, M. S. (2010). Respondent-Driven Sampling: An Assessment of Current Methodology. *Sociological Methodology* **40** 285–327.
- [5] GOEL, S. and SALGANIK, M. J. (2009). Respondent-driven sampling as Markov chain Monte Carlo. *Statistics in Medicine* **28** 2202–2229.
- [6] HANDCOCK, M. S., GILE, K. J. and NEELY, W. W. (2009). **RDS**: R Functions for Respondent-Driven Sampling Hidden Population Methods Research Group <http://hpmrg.org/>, Seattle, WA R package version 0.10.
- [7] HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2003). **statnet**: Software Tools for the Statistical Modeling of Network Data Statnet Project <http://statnetproject.org/>, Seattle, WA R package version 2.0.
- [8] HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). **statnet**: Software tools for the representation, visualization, analysis and simulation of social network data. *Journal of Statistical Software* **24**. <http://www.jstatsoft.org/v24/i01/>.
- [9] HANSEN, M. H. and HURWITZ, W. N. (1943). On the Theory of Sampling from Finite Populations. *Annals of Mathematical Statistics* **14** 333–362.
- [10] HECKATHORN, D. D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems* **44** 174–199.
- [11] HECKATHORN, D. D. (2007). Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment. *Sociological Methodology* **37** 151–207.
- [12] HECKATHORN, D. D. and JEFFRI, J. (2001). Finding the beat: using respondent-driven sampling to study jazz musicians. *Poetics* **28** 307–329.

- [13] HECKATHORN, D. D., BROADHEAD, R. S., ANTHONY, D. L. and WEAKLIEM, D. L. (1999). Aids and Social Networks: HIV Prevention Through Network Mobilization. *Sociological Focus* **32** 307–329.
- [14] HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47** 663–685.
- [15] JOHNSTON, L. G., MALEKINEJAD, M., KENDALL, C., IUPPA, I. M. and RUTHERFORD, G. W. (2008). Implementation Challenges to Using Respondent-Driven Sampling Methodology for HIV Biological and Behavioral Surveillance: Field Experiences in International Settings. *AIDS and Behavior* **12** 131–141.
- [16] KOGAN, S. M., WEJNERT, C., FU CHEN, Y., BRODY, G. H. and SLATER, L. M. (2011). Respondent-driven sampling with hard-to-reach emerging adults: an introduction and case study with rural african americans. *Journal of Adolescent Research* **26** 30–60.
- [17] NEELY, W. W. (2009). Statistical Theory & Respondent-Driven Sampling (under review).
- [18] SALGANIK, M. J. and HECKATHORN, D. D. (2004). Sampling and Estimation in Hidden Populations using Respondent-driven Sampling. *Sociological Methodology* **34** 193–239.
- [19] VOLZ, E. and HECKATHORN, D. D. (2008). Probability Based Estimation Theory for Respondent Driven Sampling. *Journal of Official Statistics* **24** 79–97.
- [20] VOLZ, E., WEJNERT, C., DEGANI, I. and HECKATHORN, D. D. (2007). Respondent-Driven Sampling Analysis Tool (RDSAT) Version 5.6. www.respondentdrivensampling.org.