UC Berkeley
UC Berkeley Previously Published Works

Title
Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations.

Permalink
https://escholarship.org/uc/item/0ff7q2br

Journal
Sociological methodology, 46(1)

ISSN
0081-1750

Authors
Feehan, Dennis M
Salganik, Matthew J

Publication Date
2016-08-01

DOI
10.1177/0081175016665425

Peer reviewed

ॐ 5 ॐ

GENERALIZING THE NETWORK SCALE-UP METHOD: A NEW ESTIMATOR FOR THE SIZE OF HIDDEN POPULATIONS

*Dennis M. Feehan**
Matthew J. Salganik[†]

Abstract

The network scale-up method enables researchers to estimate the sizes of hidden populations, such as drug injectors and sex workers, using sampled social network data. The basic scale-up estimator offers advantages over other size estimation techniques, but it depends on problematic modeling assumptions. The authors propose a new generalized scale-up estimator that can be used in settings with nonrandom social mixing and imperfect awareness about membership in the hidden population. In addition, the new estimator can be used when data are collected via complex sample designs and from incomplete sampling frames. However, the generalized scale-up estimator also requires data from two samples: one from the frame population and one from the hidden population. In some situations these data from the hidden population can be collected by adding a small number of questions to already planned studies. For other situations, the authors develop interpretable adjustment factors that can be applied to the basic scale-up estimator.

*University of California, Berkeley, CA, USA
[†]Princeton Univerity, Princeton, NJ, USA

Corresponding Author:
Dennis M. Feehan, University of California, Berkeley, Department of Demography, 2232
Piedmont Avenue, Berkeley, CA 94720, USA
Email: feehan@berkeley.edu

The authors conclude with practical recommendations for the design and analysis of future studies.

Keywords

hidden populations, social networks, sampling, network scale-up method

1. INTRODUCTION

Many important problems in social science, public health, and public policy require estimates of the sizes of hidden populations. For example, in HIV/AIDS research, estimates of the size of the most at-risk populations—drug injectors, female sex workers, and men who have sex with men—are critical for understanding and controlling the spread of the epidemic. However, researchers and policymakers are unsatisfied with the ability of current statistical methods to provide these estimates (Joint United Nations Programme on HIV/AIDS 2010). We address this problem by improving the network scale-up method, a promising approach to size estimation. Our results are immediately applicable in many substantive domains in which size estimation is challenging, and the framework we develop advances the understanding of sampling in networks more generally.

The core insight behind the network scale-up method is that ordinary people have embedded within their personal networks information that can be used to estimate the sizes of hidden populations, if that information can be properly collected, aggregated, and adjusted (Bernard et al. 1989, 2010). In a typical scale-up survey, randomly sampled adults are asked about the number of connections they have to people in a hidden population (e.g., "How many people do you know who inject drugs?") and a series of similar questions about groups of known size (e.g., "How many widowers do you know?" "How many doctors do you know?"). Responses to these questions are called *aggregate relational data* (McCormick et al. 2012).

To produce size estimates from aggregate relational data, previous researchers have begun with the *basic scale-up model*, which makes three important assumptions: (1) Social ties are formed completely at random (i.e., random mixing), (2) respondents are perfectly aware of the characteristics of their alters, and (3) respondents are able to provide accurate answers to survey questions about their personal networks. From the basic scale-up model, Killworth, McCarty, et al. (1998)

derived the *basic scale-up estimator*. This estimator, which is widely used in practice, has two main components. For the first component, the aggregate relational data about the hidden population are used to estimate the number of connections that respondents have to the hidden population. For the second component, the aggregate relational data about the groups of known size are used to estimate the number of connections that respondents have in total. For example, a researcher might estimate that members of her sample have 5,000 connections to people who inject drugs and 100,000 connections in total. The basic scale-up estimator combines these pieces of information to estimate that 5 percent $(5,000/100,000)$ of the population injects drugs. This estimate is a sample proportion, but rather than being taken over the respondents, as would be typical in survey research, the proportion is taken over the respondents' alters. Researchers who desire absolute size estimates multiply the alter sample proportion by the size of the entire population, which is assumed to be known (or estimated using some other method).

Unfortunately, the three assumptions underlying the basic scale-up model have all been shown to be problematic. Scale-up researchers call violations of the random mixing assumption *barrier effects* (Killworth et al. 2006; Maltiel et al. 2015; Zheng, Salganik, and Gelman 2006), they call violations of the perfect awareness assumption *transmission error* (Killworth et al. 2006; Maltiel et al. 2015; Salganik, Mello, et al. 2011; Shelley et al. 1995, 2006), and they call violations of the respondent accuracy assumption *recall error* (Killworth et al. 2003, 2006; Maltiel et al. 2015; McCormick and Zheng 2007).

In this paper, we develop a new approach to producing size estimates from aggregate relational data. Rather than depending on the basic scale-up model or its variants (e.g., Maltiel et al. 2015), we use a simple identity to derive a series of new estimators. Our new approach reveals that one of the two main components of the basic scale-up estimator is problematic. Therefore, we propose a new estimator—the *generalized scale-up estimator*—that combines the aggregate relational data traditionally used in scale-up studies with similar data collected from the hidden population. Collecting data from the hidden population is a major departure from current scale-up practice, but we believe that it enables a more principled approach to estimation. For researchers who are not able to collect data from the hidden population, we propose a series of adjustment factors that highlight the possible biases of the basic scale-up estimator. Ultimately, researchers must balance the

trade-offs between the basic scale-up estimator, generalized scale-up estimator, and other size estimation techniques on the basis of the specific features of their research setting.

In Section 2, we derive the generalized scale-up estimator, and we describe the data collection procedures needed to use it. In Section 3, we compare the generalized and basic scale-up approaches analytically and with simulations; our comparison leads us to propose a decomposition that separates the difference between the two approaches into three measurable and substantively meaningful factors (equation 15). In Section 4, we make practical recommendations for the design and analysis of future scale-up studies, and in Section 5, we conclude with a discussion of the steps that follow. Appendices A to G in the online journal provide technical details and supporting arguments.

2. THE GENERALIZED SCALE-UP ESTIMATOR

The generalized scale-up estimator can be derived from a simple accounting identity that requires no assumptions about the underlying social network structure in the population. Figure 1 helps illustrate the derivation, which was inspired by earlier research on multiplicity estimation (Sirken 1970) and indirect sampling (Lavallée 2007). Consider a population of seven people, two of whom are drug injectors (Figure 1a). In this population, two people are connected by a directed edge $i \rightarrow j$ if person i would count person j as a drug injector when answering the question "How many drug injectors do you know?" Whenever $i \rightarrow j$, we say that i makes an *out-report* about j and that j receives an *in-report* from i.[1]

Each person can be viewed as both a source of out-reports and a recipient of in-reports, and in order to emphasize this point, Figure 1b shows the population with each person represented twice: on the left as a sender of out-reports and on the right as a receiver of in-reports. This visual representation highlights the following identity:

$$\text{total out-reports} = \text{total in-reports.} \tag{1}$$

Despite its simplicity, the identity in equation (1) turns out to be very useful because it leads directly to the new estimator that we propose.

In order to derive an estimator from equation (1), we must define some notation. Let U be the entire population, and let $H \subset U$ be the hidden population. Furthermore, let $y_{i,H}$ be the total number of
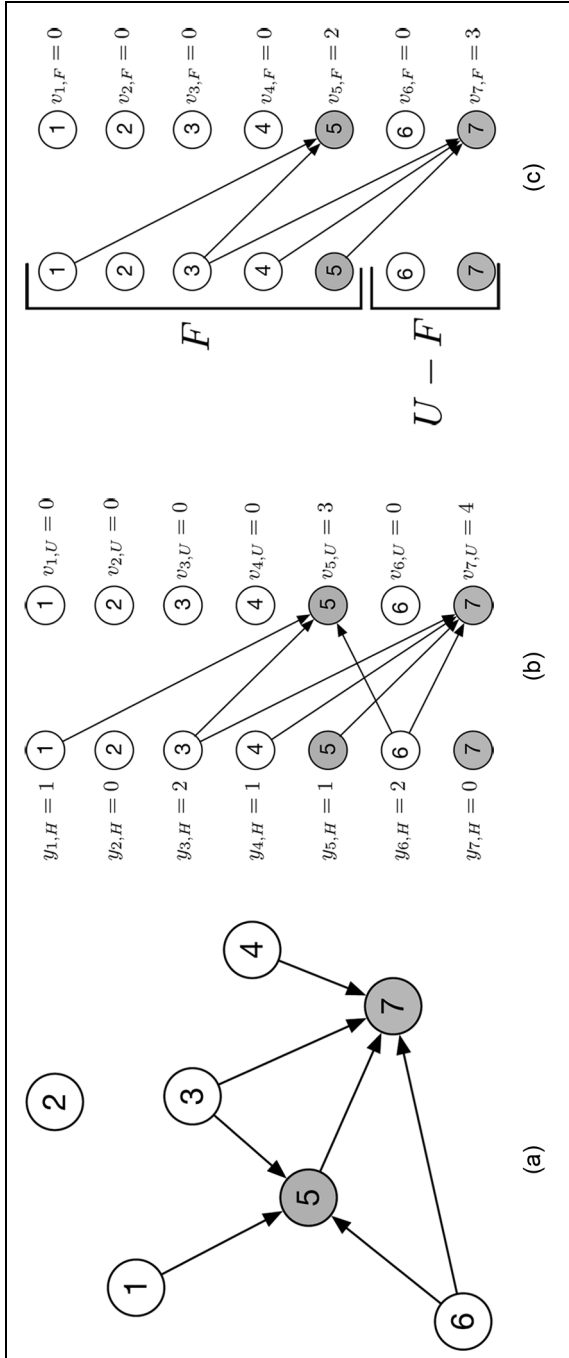
Figure 1. Illustration of the derivation of the generalized scale-up estimator. (a) A population of seven people, two of whom are drug injectors (shown in gray). A directed edge $i \rightarrow j$ indicates that i counts j as a drug injector when answering the question "How many drug injectors do you know?" (b) The same population, redrawn so that each person now appears twice: as a source of out-reports, on the left, and as a recipient of in-reports, on the right. This arrangement shows that total out-reports and total in-reports must be equal. (c) The same population again, but now some of the people are in the frame population F and some are not. In real scale-up studies, we can only learn about out-reports from the frame population.

157

out-reports from person i (i.e., person i's answer to the question "How many drug injectors do you know?"). For example, Figure 1b shows that person 5 would report knowing one drug injector, so $y_{5,H} = 1$. Let $v_{i,U}$ be the total number of in-reports to i if everyone in U is interviewed; that is, $v_{i,U}$ is the *visibility* of person i to people in U. For example, Figure 1b shows that person 5 would be reported as a drug injector by three people so $v_{5,U} = 3$. Because total out-reports must equal total in-reports, it must be the case that

$$y_{U,H} = v_{U,U}, \tag{2}$$

where $y_{U,H} = \sum_{i \in U} y_{i,H}$ and $v_{U,U} = \sum_{i \in U} v_{i,U}$. Multiplying both sides of equation (2) by N_H, the number of people in the hidden population, and then rearranging terms, we get

$$N_H = \frac{y_{U,H}}{v_{U,U}/N_H}. \tag{3}$$

Equation (3) is an expression for the size of the hidden population that does not depend on any assumptions about network structure or reporting accuracy; it is just a different way of expressing the identity that the total number of out-reports must equal the total number of in-reports. If we could estimate the two terms on the right side of equation (3)—one term related to out-reports ($y_{U,H}$) and one term related to in-reports ($v_{U,U}/N_H$)—then we could estimate N_H.

However, in order to make the identity in equation (3) useful in practice we need to modify it to account for an important logistical requirement of survey research. In real scale-up studies, researchers do not sample from the entire population U, but instead they sample from a subset of U called the frame population, F. For example, in almost all scale-up studies the frame population has been adults (but note that our mathematical results hold for any frame population). In standard survey research, restricting interviews to a frame population does not cause problems because inference is being made about the frame population. In other words, when respondents report about themselves, it is clear to which group inferences apply. However, with the scale-up method, respondents report about others, so the group that inferences are being made about is not necessarily the same as the group that is being interviewed. As we show in Section 4.2, failure to consider this fact requires the introduction of an awkward adjustment factor that had previously

gone unnoticed. Here, we avoid this awkward adjustment factor by deriving an identity explicitly in terms of the frame population. Restricting our attention to out-reports coming from people in the frame population, it must be the case that

$$N_H = \frac{y_{F,H}}{v_{U,F}/N_H},$$ (4)

where $y_{F,H} = \sum_{i \in F} y_{i,H}$ and $v_{U,F} = \sum_{i \in U} v_{i,F}$. The only difference between equation (3) and equation (4) is that equation (4) restricts out-reports and in-reports to come from people in the frame population (Figure 1c). The identity in equation (4) is extremely general: it does not depend on any assumptions about the relationship between the entire population U, the frame population F, and the hidden population H. For example, it holds if no members of the hidden population are in the frame population, if there are barrier effects, and if there are transmission errors. Thus, if we could estimate the two terms on the right side of equation (4)—one term related to out-reports ($y_{F,H}$) and one term related to in-reports ($v_{U,F}/N_H$)—then we could estimate N_H under very general conditions.

Unfortunately, despite repeated attempts, we were unable to develop a practical method for estimating the term related to in-reports ($v_{U,F}/N_H$). However, if we make an assumption about respondents' reporting behavior, then we can re-express equation (4) as an identity made up of quantities that we can actually estimate. Specifically, if we assume that the out-reports from people in the frame population only include people in the hidden population, then it must be the case that the visibility of everyone not in the hidden population is 0: $v_{i,F} = 0$ for all $i \notin H$. In this case, we can rewrite equation (4) as

$$N_H = \frac{y_{F,H}}{v_{H,F}/N_H} = \frac{y_{F,H}}{\bar{v}_{H,F}} \quad \text{if } v_{i,F} = 0 \text{ for all } i \notin H,$$ (5)

where $\bar{v}_{H,F} = v_{H,F}/N_H$.

To understand the reporting assumption substantively, consider the two possible types of reporting errors: false positives and false negatives. Previous scale-up research on transmission error focused on the problem of false negatives, where a respondent is connected to a member of the hidden population but does not report this, possibly because she is not aware that the person she is connected to is in the hidden population (Bernard et al. 2010). Because hidden populations such as drug

injectors are often stigmatized, it is reasonable to suspect that false nega-
tives will be a serious problem for the scale-up method. Fortunately,
equation (5) holds even if there are false negative reporting errors.
However, false positives—which do not seem to have been considered
previously in the scale-up literature—are also possible. For example, a
respondent who is not connected to any drug injectors might report that
one of her acquaintances is a drug injector. These false positive reports
are not accounted for in the identity in equation (5) and the estimators
that we derive subsequently. If false positive reports exist, they will intro-
duce a positive bias into estimates from the generalized scale-up estima-
tor. Therefore, in Appendix A in the online journal we (1) formally
define an interpretable measure of false positive reports, the *precision of
out-reports*; (2) analytically show the bias in size estimates as a function
of the precisions of out-reports; and (3) discuss two research designs that
could enable researchers to estimate the precision of out-reports.

2.1. *Estimating N_H from Sampled Data*

Equation (5) relates our quantity of interest, the size of the hidden popu-
lation (N_H), to two other quantities: the total number of out-reports from
the frame population ($y_{F,H}$) and the average number of in-reports in the
hidden population ($\bar{v}_{H,F}$). We now show how to estimate $y_{F,H}$ with a
probability sample from the frame population and $\bar{v}_{H,F}$ with a relative
probability sample from the hidden population.

The total number of out-reports ($y_{F,H}$) can be estimated from respon-
dents' reported number of connections to the hidden population,

$$\widehat{y}_{F,H} = \sum_{i \in s_F} \frac{y_{i,H}}{\pi_i}, \tag{6}$$

where s_F denotes the sample, $y_{i,H}$ denotes the reported number of con-
nections between i and H, and π_i is i's probability of inclusion from a
conventional probability sampling design from the frame population.
Because $\widehat{y}_{F,H}$ is a standard Horvitz-Thompson estimator, it is consistent
and unbiased as long as all members of F have a positive probability of
inclusion under the sampling design (Sarndal, Swensson, and Wretman
1992); for a more formal statement, see Result B.1. This estimator
depends only on an assumption about the sampling design for the frame
population, and in Table D.2, we show the sensitivity of our estimator
to violations of this assumption.

Estimating the average number of in-reports for the hidden population $(\bar{v}_{H,F})$ is more complicated. First, it will usually be impossible to obtain a conventional probability sample from the hidden population. As we show below, however, estimating $\bar{v}_{H,F}$ requires only a relative probability sampling design in which hidden population members have a nonzero probability of inclusion and respondents' probabilities of inclusion are known up to a constant of proportionality, $c\pi_i$ (see Appendix C.1 in the online journal for a more precise definition). Of course, even selecting a relative probability sample from a hidden population can be difficult.

A second problem arises because we do not expect respondents to be able to easily and accurately answer direct questions about their visibility ($v_{i,F}$). That is, we do not expect respondents to be able to answer questions such as "How many people on the sampling frame would include you when reporting a count of the number of drug injectors that they know?" Instead, we propose asking hidden population members a series of questions about their connections to certain groups and their visibility to those groups. For example, each sampled hidden population respondent could be asked "How many widowers do you know?" and then "How many of these widowers are aware that you inject drugs?" This question pattern can be repeated for many groups (e.g., widowers, doctors, bus drivers). We call data with this structure *enriched aggregate relational data* to emphasize its similarity to the aggregate relational data that is familiar to scale-up researchers. An interviewing procedure called the *game of contacts* enables researchers to collect enriched aggregated relational data, even in realistic field settings (Salganik, Mello, et al. 2011; Maghsoudi et al. 2014).

Given a relative probability sampling design and enriched aggregate relational data, we can now formalize our proposed estimator for $\bar{v}_{H,F}$. Let $A_1, A_2, \ldots, A_J$, be the set of groups about which we collect enriched aggregate relational data (e.g., widowers, doctors). Here, to keep the notation simple, we assume that these groups are all contained in the frame population, so that $A_j \subset F$ for all j; in Appendix C.4 in the online journal, we extend the results to groups that do not meet this criterion. Let $\mathcal{A}$ be the concatenation of these groups, which we call the *probe alters*. For example, if A_1 is widowers and A_2 is doctors, then the probe alters $\mathcal{A}$ is the collection of all widowers and all doctors, with doctors who are widowers included twice. Also, let $\tilde{v}_{i,A_j}$ be respondent i's report about her visibility to people in A_j and let v_{i,A_j} be respondents i's actual visibility to people in A_j (i.e., the number of times that this respondent

would be reported about if everyone in A_j was asked about their connections to the hidden population).

The estimator for $\bar{v}_{H,F}$ is

$$\widehat{\bar{v}}_{H,F} = \frac{N_F}{N_\mathcal{A}} \frac{\sum_{i \in s_H} \sum_j \widetilde{v}_{i,} A_j / (c\pi_i)}{\sum_{i \in s_H} 1 / (c\pi_i)}, \qquad (7)$$

where $N_\mathcal{A}$ is the number of probe alters, c is the constant of proportionality from the relative probability sample, and s_H is a relative probability sample of the hidden population. Equation (7) is a standard weighted sample mean (Sarndal et al. 1992, Section 5.7) multiplied by a constant, $N_F / N_\mathcal{A}$. Result C.2 shows that this estimator is consistent and *essentially unbiased*,[2] when three conditions are satisfied: one about the design of the survey, one about reporting behavior, and one about sampling from the hidden population.

The first condition underlying the estimator in equation (7) is related to the design of the survey, and we call it the *probe alter condition*. This condition describes the required relationship between the visibility of the hidden population to the probe alters and the visibility of the hidden population to the frame population:

$$\frac{v_{H,\mathcal{A}}}{N_\mathcal{A}} = \frac{v_{H,F}}{N_F}, \qquad (8)$$

where $v_{H,\mathcal{A}}$ is the total visibility of the hidden population to the probe alters, $v_{H,F}$ is the total visibility of the hidden population to the frame population, $N_\mathcal{A}$ is the number of probe alters, and N_F is the number of people in the frame population. In words, equation (8) says that the rate at which the hidden population is visible to the probe alters must be the same as the rate at which the hidden population is visible to the frame population. For example, in a study to estimate the number of drug injectors in a city, drug treatment counselors would be a poor choice for membership in the probe alters because drug injectors are probably more visible to drug treatment counselors than to typical members of the frame population. On the other hand, postal workers would probably be a reasonable choice for membership in the probe alters because drug injectors are probably about as visible to postal workers as they are to typical members of the frame population. Additional results about the probe alter condition are presented in the online appendices: (1) Result C.3 presents three other algebraically equivalent formulations of probe

alter condition, some of which offer additional intuition; (2) Result C.4 provides a method to empirically test the probe alter condition; and (3) Table D.1 quantifies the bias introduced when the probe alter condition is not satisfied.

The second condition underlying the estimator $\widehat{\bar{v}}_{H,F}$ (equation 7) is related to reporting behavior, and we call it *accurate aggregate reports about visibility*:

$$\tilde{v}_{H,\mathcal{A}} = v_{H,\mathcal{A}}, \tag{9}$$

where $\tilde{v}_{H,\mathcal{A}}$ is the total reported visibility of members of the hidden population to the probe alters ($\sum_{i \in H} \sum_{A_j \in \mathcal{A}} \tilde{v}_{i,A_j}$) and $v_{H,\mathcal{A}}$ is the total actual visibility of members of the hidden population to the probe alters ($\sum_{i \in H} \sum_{A_j \in \mathcal{A}} v_{i,A_j}$). In words, equation (9) says that hidden population members must be correct in their reports about their visibility to probe alters in aggregate, but equation (9) does not require the stronger condition that each individual report be accurate. In practice, we expect that there are two main ways that there might not be accurate aggregate reports about visibility. First, hidden population members might not be accurate in their assessments of what others know about them. For example, research on the "illusion of transparency" suggests that people tend to overestimate how much others know about them (Gilovich, Savitsky, and Medvec 1998). Second, although we propose asking hidden population members what other people know about them (e.g., "How many of these widowers know that you are a drug injector?") what actually matters for the estimator is what other people would report about them (e.g., "How many of these widowers would include you when reporting a count of the number of drug injectors that they know?"). In cases in which the hidden population is extremely stigmatized, some respondents to the scale-up survey might conceal the fact that they are connected to people whom they know to be in the hidden population, and if this were to occur, it would lead to a difference between the information that we collect ($\tilde{v}_{i,\mathcal{A}}$) and the information that we want ($v_{i,\mathcal{A}}$). Unfortunately, there is currently no empirical evidence about the possible magnitude of these two problems in the context of scale-up studies. However, Table D.1 quantifies the bias introduced into estimates if the accurate aggregate reports about visibility condition is not satisfied.

Finally, the third condition underlying the estimator $\widehat{\overline{v}}_{H,F}$ (equation 7) is that researchers have a relative probability sample from the hidden population. Currently the most widely used method for drawing relative probability samples from hidden populations is respondent-driven sampling (Heckathorn 1997); see Volz and Heckathorn (2008) for a set of conditions under which respondent-driven sampling leads to a relative probability sample. Although respondent-driven sampling has been used in hundreds of studies around the world (White et al. 2015), there is active debate about the characteristics of samples that it yields (Bengtsson and Thorson 2010; Gile and Handcock 2010, 2015; Gile, Johnston, and Salganik 2015; Goel and Salganik 2010; Heimer 2005; Li and Rohe 2015; McCreesh et al. 2012; Mills et al. 2012; Rohe 2015; Rudolph et al. 2013; Salganik 2012; Scott 2008; Yamanis et al. 2013). If other methods for sampling from hidden populations are demonstrated to be better than respondent-driven sampling (e.g., see Karon and Wejnert 2012; Kurant, Markopoulou, and Thiran 2011; Mouw and Verdery 2012), then researchers should consider these methods when using the generalized scale-up estimator. Furthermore, researchers can use Table D.2 to quantify the bias that results if the condition requiring a relative probability sample is not satisfied.

To recap, using two different data collection procedures—one with the frame population and one with the hidden population—we can estimate the two components of the expression for N_H given in equation (5). The estimator for the numerator $(\widehat{y}_{F,H})$ depends on an assumption about the ability to select a probability sample from the frame population (see Result B.1), and the estimator for the denominator $(\widehat{\overline{v}}_{H,F})$ depends on assumptions about survey construction, reporting behavior, and the ability to select a relative probability sample from the hidden population (see Result C.2).

We can combine these component estimators to form the *generalized scale-up estimator*:

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\overline{v}}_{H,F}}. \qquad (10)$$

Result C.8 proves that the generalized scale-up estimator will be consistent and essentially unbiased if (1) the estimator for the numerator $(\widehat{y}_{F,H})$ is consistent and essentially unbiased, (2) the estimator for the

denominator ($\widehat{\bar{v}}_{H,F}$) is consistent and essentially unbiased, and (3) there are no false positive reports.

One attractive feature of the generalized scale-up estimator (equation 10) is that it is a combination of standard survey estimators. This structure enabled us to derive very general sensitivity results about the impact of violations of assumptions, either individually or jointly. We return to the issue of assumptions and sensitivity analysis when discussing recommendations for practice (Section 4).

3. COMPARISON BETWEEN THE GENERALIZED AND BASIC SCALE-UP APPROACHES

In Section 2, we derived the generalized network scale-up estimator by using an identity relating in-reports and out-reports as the basis for a design-based estimator. The approach we followed differs from previous scale-up studies, which have posited the basic scale-up model and derived estimators conditional on that model. In this section, we compare these two different approaches from a design-based perspective.

We begin our comparison by reviewing the basic scale-up model, which was used in most of the studies listed in Table 1. To review this model, we need to define another quantity: we call $d_{i,U}$ person i's degree, the number of undirected network connections she has to everyone in U.

The basic scale-up model assumes that each person's connections are formed independently, that reporting is perfect, and that visibility is perfect (Killworth, McCarty, et al. 1998). Together, these three assumptions lead to the probabilistic model:

$$y_{i,A_j} = d_{i,A_j} \sim \text{Binomial}\left(d_{i,U}, \frac{N_{A_j}}{N}\right), \tag{11}$$

for all i in U and for any group A_j. In words, this model suggests that the number of connections from a person i to members of a group A_j is the result of a series of $d_{i,U}$ independent random draws, where the probability of each edge being connected to A_j is $\frac{N_{A_j}}{N}$.

The basic scale-up model leads to what we call the basic scale-up estimator:

$$\widehat{N}_H = \frac{\sum_{i \in s_F} y_{i,H}}{\sum_{i \in s_F} \widehat{d}_{i,U}} \times N, \tag{12}$$

Table 1. Network Scale-up Studies That Have Been Completed

Hidden Population(s)	Location	Citation
Mortality in earthquake	Mexico City, Mexico	Bernard et al. (1989)
Rape victims	Mexico City, Mexico	Bernard et al. (1991)
HIV prevalence, rape, and homelessness	United States	Killworth, McCarty, et al. (1998)
Heroin use	14 U.S. cities	Kadushin et al. (2006)
Choking incidents in children	Italy	Snidero et al. (2007, 2009, 2012)
Groups most at risk for HIV/AIDS	Ukraine	Paniotto et al. (2009)
Heavy drug users	Curitiba, Brazil	Salganik, Fazito, et al. (2011)
Groups most at risk for HIV/AIDS	Kerman, Iran	Shokoohi, Baneshi, and Haghdoost (2012)
Men who have sex with men	Japan	Ezoe et al. (2012)
Groups most at risk for HIV/AIDS	Almaty, Kazakhstan	Scutelniciuc (2012a)
Groups most at risk for HIV/AIDS	Moldova	Scutelniciuc (2012b)
Groups most at risk for HIV/AIDS	Thailand	Aramrattan and Kanato (2012)
Groups most at risk for HIV/AIDS	Rwanda	Rwanda Biomedical Center (2012)
Groups most at risk for HIV/AIDS	Chongqing, China	Guo et al. (2013)
Groups most at risk for HIV/AIDS	Tabriz, Iran	Khounigh et al. (2014)
Men who have sex with men	Taiyuan, China	Jing et al. (2014)
Drug and alcohol users	Kerman, Iran	Sheikhzadeh et al. (2014)
Men who have sex with men	Shanghai, China	Wang et al. (2015)
Men who have sex with men	Tbilisi, Georgia	Sulaberidze et al. (2016)
Illicit drug users	Iran	Nikfarjam et al. (2016)

where $\widehat{d}_{i,U}$ is the estimated degree of respondent i from the known population method (Killworth, Johnsen, et al. 1998). Killworth, McCarty, et al. (1998) showed that equation (12) is the maximum likelihood estimator for N_H under the basic scale-up model, conditional on the additional assumption that $d_{i,U}$ is known for each $i \in s_F$.

Given this background, we can now compare the basic and generalized scale-up approaches by comparing their estimands; that is, we compare the quantities that they produce in the case of a census with perfectly observed degrees. The basic scale-up estimand can be written

$$\widehat{N}_H = \frac{y_{F,H}}{d_{F,U}} \times N = \frac{y_{F,H}}{\bar{d}_{U,F}}, \tag{13}$$

where $d_{F,U} = \sum_{i \in F} d_{i,U}$ and $\bar{d}_{U,F} = d_{U,F}/N = d_{F,U}/N$. Furthermore, as shown in Section 2, the generalized scale-up estimand is

$$\widehat{N}_H = \frac{y_{F,H}}{\bar{v}_{H,F}}. \tag{14}$$

Comparing equations (13) and (14) reveals that both estimands have the same numerator, but they have different denominators. The network reporting identity from Section 2 (total out-reports = total in-reports) shows that the appropriate way to adjust the out-reports is based on in-reports, as in the generalized scale-up approach. However, the basic scale-up approach instead adjusts out-reports with the degree of respondents. Although using the degree of respondents cleverly avoids any data collection from the hidden population, our results reveal that it will be correct only under a very specific special case ($\bar{d}_{U,F} = \bar{v}_{H,F}$).

To further clarify the relationship between the basic and generalized scale-up approaches, we propose a decomposition that separates the difference between the two estimands into three measurable and substantively meaningful *adjustment factors*:

$$N_H = \underbrace{\left(\frac{y_{F,H}}{\bar{d}_{U,F}}\right)}_{\substack{\text{basic} \\ \text{scale-up}}} \times \underbrace{\frac{1}{\bar{d}_{F,F}/\bar{d}_{U,F}}}_{\substack{\text{frame ratio} \\ \phi_F}} \times \underbrace{\frac{1}{\bar{d}_{H,F}/\bar{d}_{F,F}}}_{\substack{\text{degree ratio} \\ \delta_F}} \times \underbrace{\frac{1}{\bar{v}_{H,F}/\bar{d}_{H,F}}}_{\substack{\text{true positive rate} \\ \tau_F}} = \underbrace{\left(\frac{y_{F,H}}{\bar{v}_{H,F}}\right)}_{\substack{\text{generalized} \\ \text{scale-up}}}. \tag{15}$$

$$\underbrace{\phantom{\frac{1}{\bar{d}_{F,F}/\bar{d}_{U,F}} \times \frac{1}{\bar{d}_{H,F}/\bar{d}_{F,F}} \times \frac{1}{\bar{v}_{H,F}/\bar{d}_{H,F}}}}_{\text{adjustment factors}}$$

The decomposition shows that when the product of the adjustment factors is 1, the two estimands are both correct. However, when the

product of the adjustment factors is not 1, then the generalized scale-up estimand is correct but the basic scale-up estimand is incorrect. We now describe each of the three adjustment factors in turn.

First, we define the frame ratio, ϕ_F, to be

$$\phi_F = \frac{\text{avg \# connections from a member of } F \text{ to the rest of } F}{\text{avg \# connections from a member of } U \text{ to } F} = \frac{\bar{d}_{F,F}}{\bar{d}_{U,F}}. \quad (16)$$

ϕ_F can range from 0 to infinity, and in most practical situations we expect that ϕ_F will be greater than 1. Result B.6 shows that we can make consistent and essentially unbiased estimates of ϕ_F from a sample of F.[3]

Next, we define the degree ratio δ_F to be

$$\delta_F = \frac{\text{avg \# connections from a member of } H \text{ to } F}{\text{avg \# connections from a member of } F \text{ to the rest of } F} = \frac{\bar{d}_{H,F}}{\bar{d}_{F,F}}. \quad (17)$$

δ_F ranges from 0 to infinity, and it is less than 1 when the hidden population members have, on average, fewer connections to the frame population than frame population members. Result C.6 shows that we can make consistent and essentially unbiased estimates of δ_F from samples of F and H.

Finally, we define the true positive rate, τ_F, to be

$$\tau_F = \frac{\text{\# in-reports to } H \text{ from } F}{\text{\# edges connecting } H \text{ and } F} = \frac{v_{H,F}}{d_{H,F}} = \frac{\bar{v}_{H,F}}{\bar{d}_{H,F}}. \quad (18)$$

τ_F relates network degree to network reports.[4] τ_F ranges from 0, if none of the edges are correctly reported, to 1 if all of the edges are reported. Substantively, the more stigmatized the hidden population, the closer we would expect τ_F to be to 0. Result C.7 shows that we can make consistent and essentially unbiased estimates of τ_F from a sample of H.

Furthermore, the decomposition in equation (15) can be used to derive an expression for the bias in the basic scale-up estimator when we have a census and degrees are known:

$$\text{bias}\left(\widehat{N}_H^{\text{basic}}\right) \equiv \widehat{N}_H^{\text{basic}} - N_H \quad (19)$$

$$= \widehat{N}_H^{\text{basic}} \left[1 - \frac{1}{\phi_F \delta_F \tau_F} \right]. \quad (20)$$

The comparison between the basic and generalized scale-up approaches leads to two main conclusions. First, the estimand of the basic scale-up approach is correct only in one particular situation: when the product of the three adjustment factors is 1. The estimand of generalized scale-up approach, in contrast, is correct more generally. Second, as equation 15 shows, if the adjustment factors are known (or have been estimated), then they can be used to improve basic scale-up estimates.

3.1. *Illustrative Simulation*

To illustrate our comparison between the basic and generalized scale-up approaches, we conducted a series of simulation studies. The simulations were not meant to be a realistic model of a scale-up study, but rather, they were designed to clearly illustrate our analytic results. More specifically, the simulation investigated the performance of the estimators as three important quantities vary: (1) the size of the frame population F, relative to the size of the entire population U; (2) the extent to which people's network connections are not formed completely at random; and (3) the accuracy of reporting, as captured by the true positive rate τ_F (see equation 18).[5]

As described in detail in Appendix G in the online journal, we created populations of $5,000$ people with different proportions of the population on the sampling frame (p_F). Next, we connected the people with a social network created by a stochastic block model (Wasserman and Faust 1994; White, Boorman, and Breiger 1976) in which the randomness of the mixing was controlled by a parameter ρ such that $\rho = 1$ is equivalent to random mixing (i.e., an Erdos-Reyni random graph) and the mixing becomes more nonrandom as $\rho \to 0$. Then, for each combination of parameters, we drew 10 populations, and within each of these populations, we simulated 500 surveys. For each survey, we drew a probability sample of 500 people from the frame population, a relative probability sample of 30 people from the hidden population, and simulated responses with a specific level of reporting accuracy (τ_F). Finally, we used these reports and the appropriate sampling weights to calculate the basic and generalized scale-up estimates.

Figure 2 shows that the simulations support our analytic results. First, the simulations show that the generalized scale-up estimator is unbiased even in the presence of incomplete sampling frames, nonrandom mixing, and imperfect reporting. Second, they show that the basic scale-up
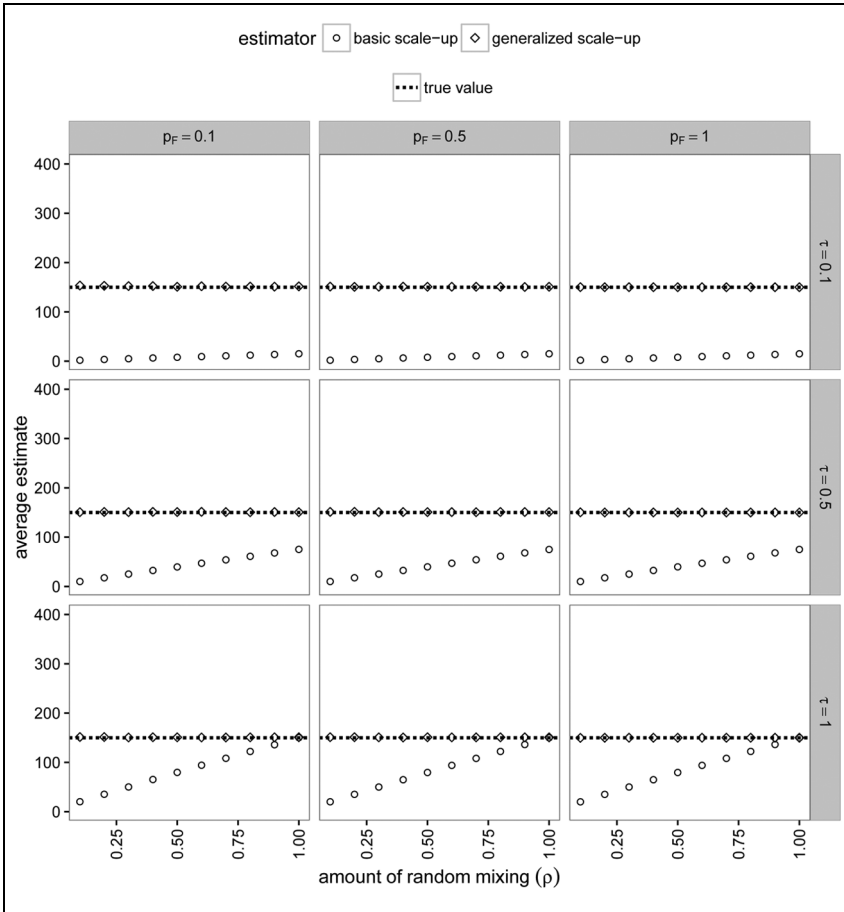
Figure 2. Estimated size of the hidden population for the generalized and basic scale-up estimators. Each panel shows how the two estimators change as the amount of random mixing is varied from low ($\rho = 0.1$; members of the hidden population are relatively unlikely to form contacts with nonmembers) to high ($\rho = 1$; members of the hidden population form contacts independent of other people's hidden population membership). The columns show results for different sizes of the frame population, from small (left column, $p_F = 0.1$), to large (right column, $p_F = 1$). The rows show results for different levels of reporting accuracy, from a small amount of true positives (top row, $\tau_F = 0.1$), to perfect reporting (bottom row, $\tau_F = 1$). For example, looking at the middle of the center panel, when $p_F = 0.5$, $\tau_F = 0.5$, and $\rho = 0.5$, we see that the average basic scale-up estimate is about 50, while the average generalized scale-up estimate is 150 (the true value). The generalized scale-up estimator is unbiased for all parameter combinations, while the basic scale-up estimator is only unbiased for certain special cases (e.g., when $\rho = 1$, $\tau_F = 1$, and $p_F = 1$). Full details of the simulation are presented in Appendix G in the online journal.
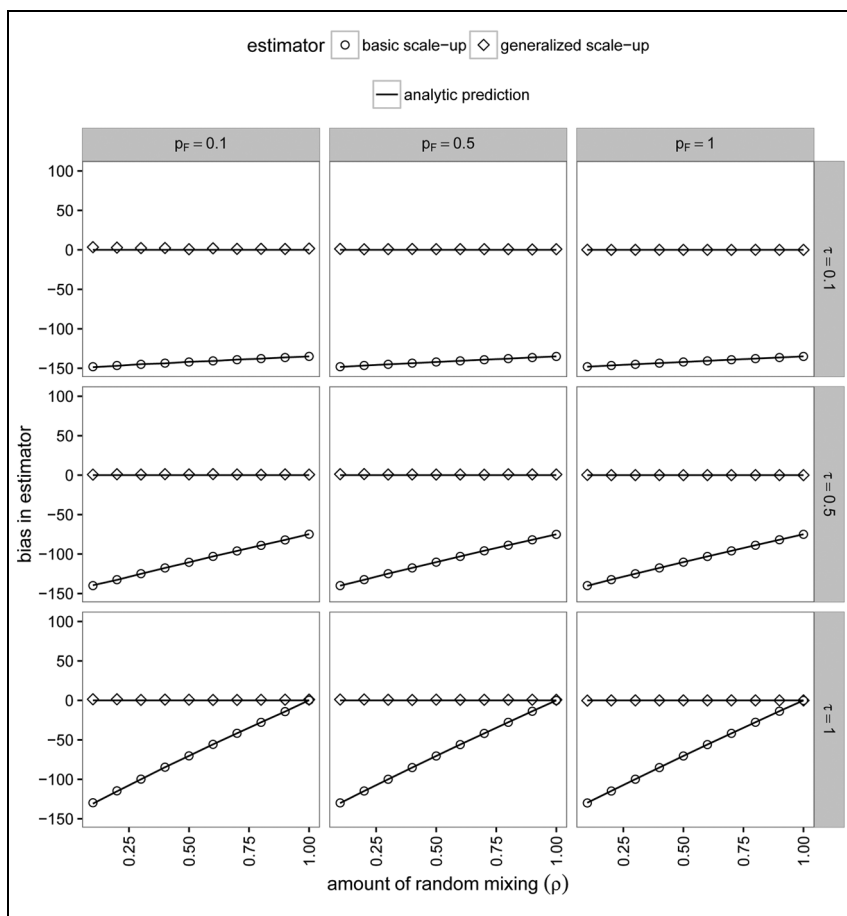
Figure 3. Bias (open circles and diamonds) and predicted bias (solid lines) in the basic scale-up estimates and generalized scale-up estimates for the same parameter configurations depicted in Figure 2. Our analytical results (equation 20) accurately predict the bias observed in our simulation study.

estimator is unbiased in a much smaller set of situations. More concretely, the basic scale-up estimator is unbiased in situations in which the basic scale-up model holds—when everyone is in the frame population ($p_F = 1$), there is random mixing ($\rho = 1$), and respondents' reports are perfect ($\tau_F = 1$).[6] Furthermore, Figure 3 illustrates that our analytic approach (equation 3) can correctly predict the bias of the basic scale-up estimator.

4. RECOMMENDATIONS FOR PRACTICE

The results in Sections 2 and 3 lead us to recommend a major departure from current scale-up practice. In addition to collecting a sample from the frame population, we recommend that researchers consider collecting a sample from the hidden population so that they can use the generalized scale-up estimator. As our results clarify, researchers using the scale-up method face a decision: they can collect data from the hidden population or they can make assumptions about the adjustment factors described in Section 3. The appropriate decision depends on a number of factors, but we think that two are most important: (1) the difficulty of sampling from the hidden population and (2) the availability of high-quality estimates of the adjustment factors in Section 3. For example, if it is particularly difficult to sample from a specific hidden population and high-quality estimates of the adjustment factors are already available, then a basic scale-up estimator may be appropriate. If, however, it is possible to sample from the hidden population and there are no high-quality estimates of adjustment factors, then the generalized scale-up estimator may be appropriate. Many realistic situations will be somewhere between these two extremes, and the trade-offs must be weighed on a case-by-case basis.

To aid researchers deciding between basic and generalized scale-up approaches, we collected the conditions needed for consistent and essentially unbiased estimates into Table 2; formal proofs of these results are presented in Online Appendices B and C. We find it helpful to group these conditions into four broad categories: sampling, survey construction, network structure, and reporting behavior.

A review of the conditions in Table 2 necessarily raises practical concerns. In situations in which researchers are trying to make estimates about real hidden populations, they probably will not know how close they are to meeting these conditions. Therefore, researchers may wonder how their estimates will be affected by violations of these assumptions, both individually (e.g., "How would my estimates be affected if there was a problem with the survey construction?") and jointly (e.g., "How would my estimate be affected if there was a problem with my survey construction and reporting behavior?"). To address this concern, in Appendix D in the online journal, we develop a framework for sensitivity analysis that shows researchers exactly how estimates will be affected by violations of all assumptions, either individually or jointly. Table 3 summarizes the results of our sensitivity framework.

Table 2. Summary of the Conditions Needed for the Generalized and Modified Basic Network Scale-up Estimators, and Their Components, to Produce Estimates That Are Consistent and Essentially Unbiased

Quantity	Conditions Required	Condition Type	Result
Reported connections to H ($\hat{y}_{F,H}$)	1. Probability sample from F	Sampling	B1
Average personal network size of F ($\hat{\bar{d}}_{F,F}$)	1. Probability sample from F 2. Groups of known size total is accurate N_A 3. Probe alter condition ($\bar{d}_{A,F} = \bar{d}_{F,F}$) 4. Accurate reporting condition ($y_{F,A} = d_{F,A}$)	Sampling Survey construction Survey construction Reporting behavior	B3
Average visibility of H ($\hat{\bar{v}}_{H,F}$)	1. Relative probability sample from H 2. Groups of known size total is accurate $N_{A_H \cap F}$ 3. Probe alter condition ($\frac{v_{H,A \cap F}}{N_{A \cap F}} = \frac{v_{H,F}}{N_F}$) 4. Accurate aggregate reports about visibility ($\bar{v}_{H,A_H \cap F} = v_{H,A_H \cap F}$)	Sampling Survey construction Survey construction Reporting behavior	C2
Generalized scale-up ($\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{\bar{v}}_{H,F}}$)	1. Conditions needed for $\hat{y}_{F,H}$ 2. Conditions needed for $\hat{\bar{v}}_{H,F}$ 3. No false positive reports about connections to $H(\eta_F = 1)$	Sampling Sampling, survey construction, Reporting behavior Reporting behavior	C8
Modified basic scale-up ($\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{\bar{d}}_{F,F}}$)	1. Conditions needed for $\hat{y}_{F,H}$ 2. Condition needed for $\hat{\bar{d}}_{F,F}$ 3. No false positive reports about connections to $H(\eta_F = 1)$ 4. Members of H and members of F have same average personal network size ($\delta_F = 1$) 5. No false negative reports about connections to $H(\tau_F = 1)$	Sampling Sampling, survey construction Reporting behavior Reporting behavior Network structure Reporting behavior	Sections 2 and 3

Note: This table uses the version of the basic scale-up estimator we recommend in Section 4.2.

173

Table 3. Analytical Expressions Researchers Can Use to Perform Sensitivity Analysis for Estimates Made Using Scale-up Estimators

Quantity	Conditions Required	Adjusted Estimand for Sensitivity Analysis
Generalized scale-up $\left(\widehat{N}_H = \dfrac{\widehat{y}_{F,H}}{\widehat{\overline{v}}_{H,F}}\right)$	1. Probability sample from F with accurate weights ($K_{F_2} = 0$ and $\overline{\in}_F = 1$) 2. Relative probability sample from H with accurate weights ($K_H = 0$) 3. Conditions needed for $\widehat{\overline{v}}_{H,F}\left(\widehat{\overline{v}}_{F,H} = k\overline{v}_{F,H}\right)$ 4. No false positive reports about connections to $H(\eta_F = 1)$	$\widehat{N}_H \cdot \dfrac{(1+K_H)}{\overline{\in}_F(1+K_{F_2})} \cdot k \cdot \eta_F \rightsquigarrow N_H$
Modified basic scale-up $\left(\widehat{N}_H = \dfrac{\widehat{y}_{F,H}}{\widehat{\overline{d}}_{F,F}}\right)$	1. Probability sample from F with accurate weights for $y_{F,H}(K_{F_2} = 0)$ 2. Probability sample from F with accurate weights for $y_{F,A}(K_{F_1} = 0)$ 3. Condition needed for $\widehat{\overline{d}}_{F,F}\left(\widehat{\overline{d}}_{F,F} = k\overline{d}_{,F}\right)$ 4. No false positive reports about connections to $H(\eta_F = 1)$ 5. Members of H and members of F have same average personal network size ($\delta_F = 1$) 6. No false negative reports about connections to $H(\tau_F = 1)$	$\widehat{N}_H \cdot \dfrac{(1+K_{F_1})}{(1+K_{F_2})} \cdot k \cdot \dfrac{\eta_F}{\delta_F \tau_F} \rightsquigarrow N_H$

Note: See Appendix D in the online journal for more details. K_{F_1}, K_{F_2}, and K_H are indices that reflect how imperfect the sampling weights researchers use to make estimates are; when these K values are 0, the weights are exactly correct; the farther they are from 0, the more imperfect the weights are. (We use the symbol $\rightsquigarrow$ as a shorthand for "is consistent and essentially unbiased for.")

Another problem researchers face in practice is putting appropriate confidence intervals around estimates. The procedure currently used in scale-up studies was proposed by Killworth, McCarty, et al. (1998), but it has a number of conceptual problems, and in practice, it produces intervals that are anticonservative (i.e., the actual coverage rate is lower than the desired coverage rate). Both of these problems—theoretical and empirical—do not seem to be widely appreciated in the scale-up literature. Therefore, instead of the current procedure, we recommend that researchers use the rescaled bootstrap procedure (Rao, Wu, and Yue 1992; Rao and Wu 1988; Rust and Rao 1996), which has strong theoretical foundations, does not depend on the basic scale-up model, can handle both simple and complex sample designs, and can be used for both the basic scale-up estimator and the generalized scale-up estimator. In Appendix F in the online journal, we review the current scale-up confidence interval procedure and the rescaled bootstrap, highlighting the conceptual advantages of the rescaled bootstrap. Furthermore, we show that the rescaled bootstrap produces slightly better confidence intervals in three real scale-up data sets: one collected via simple random sampling (McCarty et al. 2001) and two collected via complex sample designs (Salganik, Fazito, et al. 2011; Rwanda Biomedical Center 2012). Finally, and somewhat disappointingly, our results show that none of the confidence interval procedures work very well in an absolute sense, a finding that highlights an important problem for future research.

We now provide more specific guidance for researchers based on the data they decide to collect. In Section 4.1 we present recommendations for researchers who collect a sample from both the frame population, F, and the hidden population, H; in Section 4.2, we present recommendations for researchers who only select a sample from the frame population.

4.1. *Estimation with Samples from F and H*

We recommend that researchers who have samples from F and H use a generalized scale-up estimator to produce estimates of N_H (see Section 2):

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\overline{v}}_{H,F}}. \tag{21}$$

For researchers using the generalized scale-up estimator, we have three specific recommendations. Of all the conditions needed for consistent and essentially unbiased estimation, the ones most under the control of the researcher are those related to survey construction, so we recommend that researchers focus on these during the study design phase. In particular, we recommend that the probe alters be designed so that the rate at which the hidden population is visible to the probe alters is the same as the rate at which the hidden population is visible to the frame population (see Result C.2 for a more formal statement, and see Section C5 for more advice about choosing probe alters). Second, when presenting estimates, we recommend that researchers use the results in Table 3 to also present sensitivity analyses highlighting how the estimates may be affefcted by assumptions that are particularly problematic in their setting. Finally, we recommend that researchers produce confidence intervals around their estimate using the rescaled bootstrap procedure, keeping in mind that this will likely produce intervals that are anticonservative.

We also have three additional recommendations that will facilitate the cumulation of knowledge about the scale-up method. First, although the generalized scale-up estimator does not require aggregate relational data from the frame population about groups of known size, we recommend that researchers collect these data so that the basic and generalized estimators can be compared. Second, we recommend that researchers publish estimates of δ_F and τ_F, although these quantities play no role in the generalized scale-up estimator (Figure 4). As a body of evidence about these adjustment factors accumulates (e.g., Salganik, Fazito, et al. 2011; Maghsoudi et al. 2014), studies that are not able to collect a sample from the hidden population will have an empirical foundation for adjusting basic scale-up estimates, either by borrowing values directly from the literature or by using published values as the basis for priors in a Bayesian model. Finally, we recommend that researchers design their data collections—both from the frame population and the hidden population—so that size estimates from the generalized scale-up method can be compared with estimates from other methods (e.g., see Salganik, Fazito, et al. 2011a). For example, if respondent-driven sampling is used to sample from the hidden population, then researchers could use methods that estimate the size of a hidden population from recruitment patterns in the respondent-driven sampling data (Berchenko, Rosenblatt,
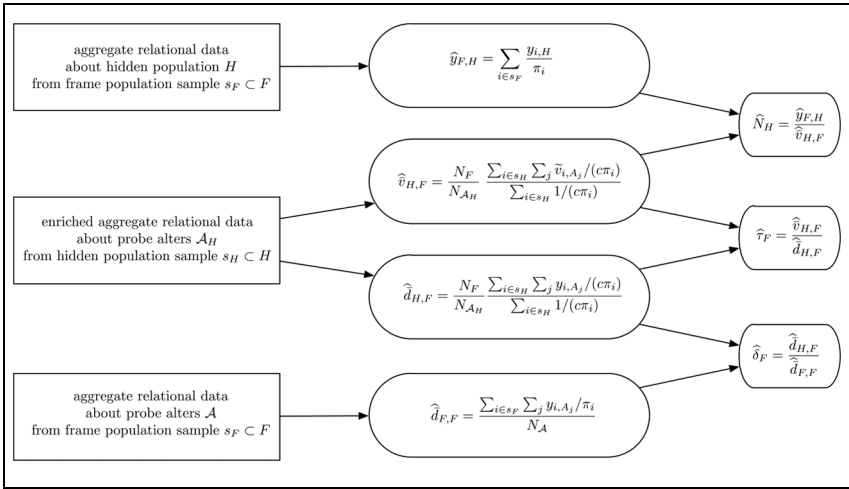
Figure 4. Recommended schematic of inputs and outputs for a study using the generalized scale-up estimator. We recommend that researchers produce size estimates using the generalized scale-up estimator and that researchers produce estimates of the adjustment factors δ_F and τ_F to aid other researchers.

and Frost 2013; Crawford, Wu, and Heimer 2015; Handcock, Gile, and Mar 2014, 2015; Johnston et al. 2015; Wesson et al. 2015).

4.2. *Estimation with Only a Sample from* F

If researchers cannot collect a sample from the hidden population, we have three recommendations. First, we recommend two simple changes to the basic scale-up estimator that remove the need to adjust for the frame ratio, ϕ_F. Recall, that the basic scale-up estimator that has been used in previous studies (see Section 3) is

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,U}} \times N = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,U}\big/ N}. \qquad (22)$$

Instead of equation 22, we suggest a new estimator, called the modified basic scale-up estimator, that more directly deals with the fact that researchers sample from the frame population F (typically adults), and not from the entire population U (adults and children):

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,F}} \times N_F = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,F} / N_F} . \tag{23}$$

There are two differences between the modified basic scale-up estimator (equation 23) and the basic scale-up estimator (equation 22). First, we recommend that researchers estimate $\widehat{d}_{F,F}$ (i.e., the total number of connections between adults and adults) rather than $\widehat{d}_{F,U}$ (i.e., the total number of connections between adults and everyone). To do so, researchers should design the probe alters for the frame population so that they have similar personal networks to the frame population; in Appendix B.4 in the online journal, we define this requirement formally, and in Section B.4.1 we provide guidance for choosing the probe alters. Second, we recommend that researchers use N_F rather than N.[7] These two simple changes remove the need to adjust for the frame ratio ϕ_F, and thereby eliminate an assumption about an unmeasured quantity. An improved version of the basic scale-up estimator would then be

$$\widehat{N}_H = \underbrace{\frac{\widehat{y}_{F,H}}{\left(\widehat{d}_{F,F} / N_F\right)}}_{\text{modified basic scale-up}} \times \underbrace{\frac{1}{\widehat{\delta}_F} \times \frac{1}{\widehat{\tau}_F}}_{\text{adjustment factors}} . \tag{24}$$

Our second recommendation is that researchers using the modified basic scale-up estimator (equation 23) perform a sensitivity analysis using the results in Table 3. In particular, we think that researchers should be explicit about the values that they assume for the adjustment factors δ_F and τ_F. Our third recommendation is that researchers construct confidence intervals using the rescaled bootstrap procedure, while explicitly accounting for the fact that there is uncertainty around the assumed adjustment factors and bearing in mind that this procedure will likely produce intervals that are anticonservative.

5. CONCLUSION AND NEXT STEPS

In this paper, we developed the generalized network scale-up estimator. This new estimator improves upon earlier scale-up estimators in several ways: it enables researchers to use the scale-up method in populations with nonrandom social mixing and imperfect awareness about membership in the hidden population, and it accommodates data collection with complex sample designs and incomplete sampling frames. We also

compared the generalized and basic scale-up estimators, leading us to introduce a framework that makes the design-based assumptions of the basic scale-up estimator precise. Finally, researchers who use either the basic or generalized scale-up estimator can use our results to assess the sensitivity of their size estimates to assumptions.

The approach we followed to derive the generalized scale-up estimator has three elements, and these elements may prove useful in other problems related to sampling in networks. First, we distinguished between the network of reports and the network of relationships. Second, using the network of reports, we derived a simple identity that permitted us to develop a design-based estimator free of any assumptions about the structure of the network of relationships. Third, we combined data from different types of samples. Together, these three elements may help other researchers in other situations derive relatively simple, design-based estimators that are an important complement to complex, model-based techniques.

Although the generalized scale-up estimator has many attractive features, it also requires that researchers obtain two different samples, one from the frame population and one from the hidden population. In cases in which studies of the hidden population are already planned (e.g., the behavioral surveillance studies of the groups most at risk for HIV/ AIDS), the necessary information for the generalized scale-up estimator could be collected at little additional cost by appending a modest number of questions to existing questionnaires. In cases in which these studies are not already planned, researchers can either collect their own data from the hidden population, or they can use the modified basic scale-up estimator and borrow estimated adjustment factors from other published studies.

The generalized scale-up estimator, like all estimators, depends on a number of assumptions, and we think three of them will be most problematic in practice. First, the estimator depends on the assumption that there are no false positive reports, which is unlikely to be true in all situations. Although we have derived an estimator that works even in the presence of false positive reports (Appendix A in the online journal), we were not able to design a practical data collection procedure that would allow us to estimate one of the terms it requires. Second, the generalized scale-up estimator depends on the assumption that hidden population members have accurate aggregate awareness about visibility (equation 9). That is, researchers have to assume that hidden population

respondents can accurately report whether or not their alters would report them, and we expect this assumption will be difficult to check in most situations. Third, the generalized scale-up estimator depends on having a relative probability sample from the hidden population. Unfortunately, we cannot eliminate any of these assumptions, but we have stated them clearly and we have derived the sensitivity of the estimates to violations of these assumptions, individually and jointly.

Our results and their limitations highlight several directions for further work, in terms of both of improved modeling and improved data collection. We think the most important direction for future modeling is developing estimators in a Bayesian framework, and a recent paper by Maltiel et al. (2015) offers some promising steps in this direction. We see two main advantages of the Bayesian approach in this setting. First, a Bayesian approach would allow researchers to propagate the uncertainty they have about the many assumptions involved in scale-up estimates, whereas our current approach captures only uncertainty introduced by sampling. Furthermore, as more empirical studies produce estimates of the adjustment factors (τ_F and δ_F), a Bayesian framework would permit researchers to borrow values from other studies in a principled way. In terms of future directions for data collection, researchers need practical techniques for estimating the rate of false positive reporting. These estimates, combined with the estimator in Appendix A in the online journal, would permit the relaxation of one of the most important remaining assumptions made by all scale-up studies to date. We hope that the framework introduced in this paper will provide a basis for these and other developments.

Notes

1. Throughout the paper, we consider only the case ion which i never reports j more than once.
2. We use the term *essentially unbiased* because equation (7) is not, strictly speaking, unbiased; the ratio of two unbiased estimators is not itself unbiased. However, a large literature confirms that the biases caused by the nonlinear form of ratio estimators are typically insignificant relative to other sources of error in estimate (e.g., Sarndal et al. 1992, chap. 5). Unfortunately, many of the estimators we propose are actually ratios of ratios, sometimes called "compound ratio estimators" or "double ratio estimators." In Appendix E in the online journal, we demonstrate that the bias caused the nonlinear form of our estimators is not a practical cause for concern.
3. Note that because $\bar{d}_{U,F} = (N_F/N)\bar{d}_{F,U}$, an equivalent expression for the frame ratio is $\phi_F = \frac{\bar{d}_{F,F}}{\bar{d}_{F,U}(N_F/N)} = \frac{\bar{d}_{F,F}}{\bar{d}_{F,U}}\frac{N}{N_F}$.
4. Note that the fact that in-reports must equal out-reports means that τ_F can also be defined as $\tau_F = \frac{\text{\# reported edges from } F \text{ actually connected to } H}{\text{\# edges connecting } F \text{ and } H} = \frac{y_{F,H}^+}{d_{F,H}}$. Here we have written $y_{F,H}^+$ to mean the true positive reports among the $y_{F,H}$; see Appendix A in the online journal for a detailed explanation.
5. Computer code to perform the simulations was written in R (R Core Team 2014) and used the following packages: devtools (Wickham and Chang 2013), functional (Danenberg 2013), ggplot2 (Wickham 2009), igraph (Csardi and Nepusz 2006), networkreporting (Feehan and Salganik 2014), plyr (Wickham 2011), sampling (Tillé and Matei 2015), and stringr (Wickham 2012).
6. In addition to the settings in which the basic scale-up model holds, the basic scale-up estimator can also be unbiased when its different biases cancel (e.g., when the product of the adjustment factors is 1).
7. In some cases this difference between N_F and N can be substantial. For example, if F is adults, then in many developing countries, $N \approx 2N_F$.

References

Aramrattan, A., and M. Kanato. 2012. "Network Scale-up Method: Application in Thailand." Presented at Consultation on Estimating Population Sizes through Household Surveys: Successes and Challenges, March 28–30, New York.

Bengtsson, L., and A. Thorson. 2010. "Global HIV Surveillance among MSM: Is Risk Behavior Seriously Underestimated?" *AIDS* 24(15):2301–303.

Berchenko, Y., J. Rosenblatt, and S.D.W. Frost. 2013. "Modeling and Analysing Respondent Driven Sampling as a Counting Process." arXiv:1304.3505 [stat].

Bernard, H. R., T. Hallett, A. Iovita, E. C. Johnsen, R. Lyerla, C. McCarty, M. Mahy, M. J. Salganik, T. Saliuk, O. Scutelniciuc, G. A. Shelley, P. Sirinirund, S. Weir, and D. F. Stroup. 2010. "Counting Hard-to-count Populations: The Network Scale-up Method for Public Health." *Sexually Transmitted Infections* 86(Suppl. 2):ii11–15.

Bernard, H. R., E. C. Johnsen, P. D. Killworth, and S. Robinson. 1989. "Estimating the Size of an Average Personal Network and of an Event Subpopulation." Pp. 159–75 in *The Small World*, edited by M. Kochen. Norwood, NJ: Ablex.

Bernard, H. R., E. C. Johnsen, P. Killworth, and S. Robinson. 1991. "Estimating the Size of an Average Personal Network and of an Event Subpopulation: Some Empirical Results." *Social Science Research* 20(2):109–21.

Crawford, F. W., J. Wu, and R. Heimer. 2015. "Hidden Population Size Estimation from Respondent-driven Sampling: A Network Approach." arXiv:1504.08349 [stat].

Csardi, G., and T. Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal, Complex Systems*:1695.

Danenberg, P. 2013. "Functional: Curry, Compose, and Other Higher-order Functions." R package version 0.4.

Ezoe, S., T. Morooka, T. Noda, M. L. Sabin, and S. Koike. 2012. "Population Size Estimation of Men Who Have Sex with Men through the Network Scale-up Method in Japan." *PLoS ONE* 7(1):e31184.

Feehan, D. M., and M. J. Salganik. 2014. "The Network Reporting Package." R package version 0.1.1.

Gile, K. J., and M. S. Handcock. 2010. "Respondent-driven Sampling: An Assessment of Current Methodology.'' Pp. 285–327 in *Sociological Methodology*, vol. 40, edited by Tim Futing Liao. Hoboken, NJ: Wiley-Blackwell.

Gile, K. J., and M. S. Handcock.2015. "Network Model-assisted Inference from Respondent-driven Sampling Data." *Journal of the Royal Statistical Society, Series A: Statistics in Society* 178(3):619–39.

Gile, K. J., L. G. Johnston, and M. J. Salganik. 2015. "Diagnostics for Respondent-driven Sampling." *Journal of the Royal Statistical Society, Series A: Statistics in Society* 178(1):241–69.

Gilovich, T., K. Savitsky, and V. H. Medvec. 1998. "The Illusion of Transparency: Biased Assessments of Others' Ability to Read One's Emotional States." *Journal of Personality and Social Psychology* 75(2):332–46.

Goel, S., and M. J. Salganik. 2010. "Assessing Respondent-driven Sampling." *Proceedings of the National Academy of Science* 107(15):6743–47.

Guo, W., S. Bao, W. Lin, G. Wu, W. Zhang, W. Hladik, A. Abdul-Quader, M. Bulterys, S. Fuller, and L. Wang. 2013. "Estimating the Size of HIV Key Affected Populations in Chongqing, China, Using the Network Scale-up Method." *PLoS ONE* 8(8): e71796.

Handcock, M. S., K. J. Gile, and C. M. Mar. 2014. "Estimating Hidden Population Size Using Respondent-driven Sampling Data." *Electronic Journal of Statistics* 8(1): 1491–1521.

Handcock, M. S., K. J. Gile, and C. M. Mar. 2015. "Estimating the Size of Populations at High Risk for HIV Using Respondent-driven Sampling Data." *Biometrics* 71(1): 258–66.

Heckathorn, D. D. 1997. "Respondent-driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44(2):174–99.

Heimer, R. 2005. "Critical Issues and Further Questions about Respondent-driven Sampling: Comment on Ramirez-Valles, et al." *AIDS and Behavior* 9(4):403–8.

Jing, L., C. Qu, H. Yu, T. Wang, and Y. Cui. 2014. "Estimating the Sizes of Populations at High Risk for HIV: A Comparison Study." *PLoS ONE* 9(4):e95601.

Johnston, L. G., K. R. McLaughlin, H. El Rhinila, A. Latifi, A. Toufik, A. Bennani, K. Alami, B. Elomari, and M. S. Handcock. 2015. "Estimating the Size of Hidden Populations Using Respondent-driven Sampling Data: Case Examples from Morocco." *Epidemiology* 26(6):846–52.

Joint United Nations Programme on HIV/AIDS. 2010. "Guidelines on Estimating the Size of Populations Most at Risk to HIV." Geneva, Switzerland: UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance.

Kadushin, C., P. D. Killworth, H. R. Bernard, and A. A. Beveridge. 2006. "Scale-up Methods as Applied to Estimates of Heroin Use." *Journal of Drug Issues* 36(2): 417–40.

Karon, J., and C. Wejnert. 2012. "Statistical Methods for the Analysis of Time-location Sampling Data." *Journal of Urban Health* 89(3):565–86.

Khounigh, A. J., A. A. Haghdoost, S. SalariLak, A. H. Zeinalzadeh, R. Yousefi-Farkhad, M. Mohammadzadeh, and K. Holakouie-Naieni. 2014. "Size Estimation of Most-at-risk Groups of HIV/AIDS Using Network Scale-up in Tabriz, Iran." *Journal of Clinical Research and Governance* 3(1):21–26.

Killworth, P. D., E. C. Johnsen, C. McCarty, G. A. Shelley, and H. Bernard. 1998. "A Social Network Approach to Estimating Seroprevalence in the United States." *Social Networks* 20(1):23–50.

Killworth, P. D., C. McCarty, H. R. Bernard, G. A. Shelley, and E. C. Johnsen. 1998. "Estimation of Seroprevalence, Rape, and Homelessness in the United States Using a Social Network Approach." *Evaluation Review* 22(2):289–308.

Killworth, P. D., C. McCarty, H. R. Bernard, E. C. Johnsen, J. Domini, and G. A. Shelly. 2003. "Two Interpretations of Reports of Knowledge of Subpopulation Sizes." *Social Networks* 25(2):141–60.

Killworth, P. D., C. McCarty, E. C. Johnsen, H. R. Bernard, and G. A. Shelley. 2006. "Investigating the Variation of Personal Network Size under Unknown Error Conditions." *Sociological Methods and Research* 35(1):84–112.

Kurant, M., A. Markopoulou, and P. Thiran. 2011. "Towards Unbiased BFS Sampling." *IEEE Journal on Selected Areas in Communications* 29(9):1799–1809.

Lavallée, P. 2007. *Indirect Sampling*. New York: Springer.

Li, X., and K. Rohe. 2015. "Central Limit Theorems for Network Driven Sampling." arXiv:1509.04704 [math, stat].

Maghsoudi, A., M. R. Baneshi, M. Neydavoodi, and A. Haghdoost. 2014. "Network Scale-up Correction Factors for Population Size Estimation of People Who Inject Drugs and Female Sex Workers in Iran." *PLoS ONE* 9(11):e110917.

Maltiel, R., A. E. Raftery, T. H. McCormick, and A. Baraff. 2015. "Estimating Population Size Using the Network Scale Up Method." *Annals of Applied Statistics* 9(3):1247–77.

McCarty, C., P. D. Killworth, H. R. Bernard, E. Johnsen, and G. A. Shelley. 2001. "Comparing Two Methods for Estimating Network Size." *Human Organization* 60(1):28–39.

McCormick, T., R. He, E. Kolaczyk, and T. Zheng. 2012. "Surveying Hard-to-reach Groups through Sampled Respondents in a Social Network." *Statistics in Biosciences* 4(1):177–95.

McCormick, T. H., and T. Zheng. 2007. "Adjusting for Recall Bias in 'How Many X's Do You Know?' Surveys." *Proceedings of the Joint Statistical Meetings*, Salt Lake City, UT.

McCreesh, N., S. Frost, J. Seeley, J. Katongole, M. N. Tarsh, R. Ndunguse, F. Jichi, N. L. Lunel, D. Maher, L. G. Johnston, et al. 2012. "Evaluation of Respondent-driven Sampling." *Epidemiology* 23(1):138.

Mills, H. L., C. Colijn, P. Vickerman, D. Leslie, V. Hope, and M. Hickman. 2012. "Respondent Driven Sampling and Community Structure in a Population of Injecting Drug Users, Bristol, UK." *Drug and Alcohol Dependence* 126(3):324–32.

Mouw, T., and A. M. Verdery. 2012. "Network Sampling with Memory: A Proposal for More Efficient Sampling from Social Networks." Pp. 206–56 in *Sociological Methodology*, vol. 42, edited by Tim Futing Liao. Thousand Oaks, CA: Sage.

Nikfarjam, A., M. Shokoohi, A. Shahesmaeili, A. A. Haghdoost, M. R. Baneshi, S. Haji-Maghsoudi, A. Rastegari, A. A. Nasehi, N. Memaryan, and T. Tarjoman. 2016. "National Population Size Estimation of Illicit Drug Users through the Network Scale-up Method in 2013 in Iran." *International Journal of Drug Policy* 31:147–52.

Paniotto, V., T. Petrenko, V. Kupriyanov, and O. Pakhok. 2009. "Estimating the Size of Populations with High Risk for HIV Using the Network Scale-up Method." Technical Report. Kiev, Ukraine: Kiev Internation Institute of Sociology.

R Core Team. 2014. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing.

Rao, J., C. Wu, and K. Yue. 1992. "Some Recent Work on Resampling Methods for Complex Surveys." *Survey Methodology* 18(2):209–17.

Rao, J. N., and C.F.J. Wu. 1988. "Resampling Inference with Complex Survey Data." *Journal of the American Statistical Association* 83(401):231–41.

Rohe, K. 2015. "Network Driven Sampling: A Critical Threshold for Design Effects." arXiv:1505.05461 [math, stat].

Rudolph, A. E., C. M. Fuller, and C. Latkin. 2013. "The Importance of Measuring and Accounting for Potential Biases in Respondent-driven Samples." *AIDS and Behavior* 17(6):2244–52.

Rust, K., and J. Rao. 1996. "Variance Estimation for Complex Surveys Using Replication Techniques." *Statistical Methods in Medical Research* 5(3):283–310.

Rwanda Biomedical Center. 2012. "Estimating the Size of Key Populations at Higher Risk of HIV through a Household Survey (ESPHS) Rwanda 2011." Technical Report. Calverton, MD: RBC/IHDPC, SPF, UNAIDS, and ICF International.

Salganik, M. J. 2012. "Commentary: Respondent-driven Sampling in the Real World." *Epidemiology* 23(1):148–50.

Salganik, M. J., D. Fazito, N. Bertoni, A. H. Abdo, M. B. Mello, and F. I. Bastos. 2011. "Assessing Network Scale-up Estimates for Groups Most at Risk of HIV/AIDS:

Evidence from a Multiple-method Study of Heavy Drug Users in Curitiba, Brazil." *American Journal of Epidemiology* 174(10):1190–96.

Salganik, M. J., M. B. Mello, A. H. Abdo, N. Bertoni, D. Fazito, and F. I. Bastos. 2011. "The Game of Contacts: Estimating the Social Visibility of Groups." *Social Networks* 33(1):70–78.

Sarndal, C. E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer.

Scott, G. 2008. "'They Got Their Program, and I Got Mine': A Cautionary Tale Concerning the Ethical Implications of Using Respondent-driven Sampling to Study Injection Drug Users." *International Journal of Drug Policy* 19(1):42–51.

Scutelniciuc, O. 2012a. "Network Scale-up Method Experiences: Republic of Kazakhstan." Presented at Consultation on Estimating Population Sizes through Household Surveys: Successes and Challenges, March 28–30, New York.

Scutelniciuc, O. 2012b. "Network Scale-up Method Experiences: Republic of Moldova." Presented at Consultation on Estimating Population Sizes through Household Surveys: Successes and Challenges, March 28–30, New York.

Sheikhzadeh, K., M. R. Baneshi, M. Afshari, and A. A. Haghdoost. 2014. "Comparing Direct, Network Scale-up, and Proxy Respondent Methods in Estimating Risky Behaviors among Collegians." *Journal of Substance Use* 21(1):9–13.

Shelley, G. A., H. R. Bernard, P. Killworth, E. Johnsen, and C. McCarty. 1995. "Who Knows Your HIV Status? What HIV+ Patients and Their Network Members Know about Each Other." *Social Networks* 17(3–4):189–217.

Shelley, G. A., P. D. Killworth, H. R. Bernard, C. McCarty, E. C. Johnsen, and R. E. Rice. 2006. "Who Knows Your HIV Status II? Information Propagation within Social Networks of Seropositive People." *Human Organization* 65(4):430–44.

Shokoohi, M., M. R. Baneshi, and A. A. Haghdoost. 2012. "Size Estimation of Groups at High Risk of HIV/AIDS Using Network Scale Up in Kerman, Iran." *International Journal of Preventive Medicine* 3(7):471–76.

Sirken, M. G. 1970. "Household Surveys with Multiplicity." *Journal of the American Statistical Association* 65(329):257–66.

Snidero, S., B. Morra, R. Corradetti, and D. Gregori 2007. "Use of the Scale-up Methods in Injury Prevention Research: An Empirical Assessment to the Case of Choking in Children." *Social Networks* 29(4):527–38.

Snidero, S., N. Soriani, I. Baldi, F. Zobec, P. Berchialla, and D. Gregori. 2012. "Scale-up Approach in CATI Surveys for Estimating the Number of Foreign Body Injuries in the Aero-digestive Tract in Children." *International Journal of Environmental Research and Public Health* 9(11):4056–67.

Snidero, S., F. Zobec, P. Berchialla, R. Corradetti, and D. Gregori. 2009. "Question Order and Interviewer Effects in CATI Scale-up Surveys." *Sociological Methods and Research* 38(2):287–305.

Sulaberidze, L., A. Mirzazadeh, I. Chikovani, N. Shengelia, N. Tsereteli, and G. Gotsadze. 2016. "Population Size Estimation of Men Who Have Sex with Men in Tbilisi, Georgia: Multiple Methods and Triangulation of Findings." *PLoS ONE* 11(2):e0147413.

Tillé, Y., and A. Matei. 2015. "Sampling: Survey Sampling." R package version 2.7.

Volz, E., and D. D. Heckathorn. 2008. "Probability-based Estimation Theory for Respondent-driven Sampling." *Journal of Official Statistics* 24(1):79–97.

Wang, J., Y. Yang, W. Zhao, H. Su, Y. Zhao, Y. Chen, T. Zhang, and T. Zhang. 2015. "Application of Network Scale Up Method in the Estimation of Population Size for Men Who Have Sex with Men in Shanghai, China." *PLoS ONE* 10(11):e0143118.

Wasserman, S., and K. Faust. 1994. *Social Network Analysis*. New York: Cambridge University Press.

Wesson, P., M. S. Handcock, W. McFarland, and H. F. Raymond. 2015. "If You Are Not Counted, You Don't Count: Estimating the Number of African-American Men Who Have Sex with Men in San Francisco Using a Novel Bayesian Approach." *Journal of Urban Health* 92(6):1052–64.

White, H. C., S. A. Boorman, and R. L. Breiger. 1976. "Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions." *American Journal of Sociology* 81(4):730–80.

White, R. G., A. J. Hakim, M. J. Salganik, M. W. Spiller, L. G. Johnston, L. R. Kerr, C. Kendall, A. Drake, D. Wilson, K. Orroth, et al. 2015. "Strengthening the Reporting of Observational Studies in Epidemiology for Respondent-driven Sampling Studies: 'STROBE-RDS' Statement." *Journal of Clinical Epidemiology* 68(12):1463–71.

Wickham, H. 2009. "ggplot2: Elegant Graphics for Data Analysis." New York: Springer.

Wickham, H. 2011. "The Split-apply-combine Strategy for Data Analysis." *Journal of Statistical Software* 40(1):1–29.

Wickham, H. 2012. "stringr: Make It Easier to Work with Strings." R package version 0.6.2.

Wickham, H., and W. Chang. 2013. "devtools: Tools to Make Developing R Code Easier." R Package Version 1.4.1.

Yamanis, T. J., M. G. Merli, W. W. Neely, F. F. Tian, J. Moody, X. Tu, and E. Gao. 2013. "An Empirical Analysis of the Impact of Recruitment Patterns on RDS Estimates among a Socially Ordered Population of Female Sex Workers in China." *Sociological Methods and Research* 42(3):392–425.

Zheng, T., M. J. Salganik, and A. Gelman. 2006. "How Many People Do You Know in Prison? Using Overdispersion in Count Data to Estimate Social Structure in Networks." *Journal of the American Statistical Association* 101(474):409–23.

Author Biographies

Dennis M. Feehan is an assistant professor of demography at the University of California, Berkeley. His research lies at the intersection of demography, social networks, and quantitative methodology.

Matthew J. Salganik is a professor of sociology at Princeton University, and he is affiliated with several of Princeton's interdisciplinary research centers: the Office for Population Research, the Center for Information Technology Policy, the Center for Health and Wellbeing, and the Center for Statistics and Machine Learning. His research interests include social networks and computational social science.

Online Appendixes

A   Estimation in the presence of false positive reports

In the main text, we follow all previous scale-up studies to date in assuming that there are never any false positive reports. In this appendix, we generalize our analysis to the situation where false positive reports are possible.

In Section 2, Equation 5, we discussed false positive reports in terms of in-reports: we explained that if there are no false positive reports, then $v_{i,F} = 0$ for all $i \notin H$. In this appendix, we will re-orient the analysis and focus on how false positives affect out-reports. Each individual i's out-reports can be divided into two groups: true positives, which actually connect to the hidden population $(y_{i,H}^+)$; and false positives, which do not connect to the hidden population $(y_{i,H}^-)$. Therefore,

$$y_{i,H} = y_{i,H}^+ + y_{i,H}^-. \tag{A.1}$$

We can also define the aggregate quantities $y_{F,H}^+ = \sum_{i \in F} y_{i,H}^+$ and $y_{F,H}^- = \sum_{i \in F} y_{i,H}^-$, so that

$$y_{F,H} = y_{F,H}^+ + y_{F,H}^-. \tag{A.2}$$

Because the total number of true-positive out-reports must equal the total number of true-positive in-reports, it is the case that

$$y_{F,H}^+ = v_{H,F} \tag{A.3}$$

where $y_{F,H}^+$ is the total number of true-positive out-reports and $v_{H,F}$ is the total number of true positive in-reports. Dividing both sides by $v_{H,F}$, and then multiplying

both sides by N_H produces

$$N_H = \frac{y_{F,H}^+}{\bar{v}_{H,F}}. \tag{A.4}$$

In the main text, we introduce a strategy for estimating $\bar{v}_{H,F}$. If there was also a strategy for estimating $y_{F,H}^+$, then we could use Equation A.4 to estimate N_H, even if some reports are false positives. Unfortunately, we cannot typically estimate $y_{F,H}^+$ directly from F, since any attempt to do so would learn about $y_{F,H}$ instead. Therefore, we propose that researchers collect information about $y_{F,H}$ and then estimate an adjustment factor that relates $y_{F,H}$ to $y_{F,H}^+$. This approach leads us to introduce a new quantity called the *precision of out-reports*, η_F:

$$\eta_F = \frac{y_{F,H}^+}{y_{F,H}}. \tag{A.5}$$

The precision is useful because it relates the observed out-reports, $y_{F,H}$ to the true positive out-reports, $y_{F,H}^+$. It varies from 0, when none of the out-reports are true positives, to 1, when the out-reports are perfect. The precision allows us to derive an identity that relates out-reports to N_H:

$$N_H = \frac{\eta_F \, y_{F,H}}{\bar{v}_{H,F}}. \tag{A.6}$$

Equation A.6 then suggests the estimator:

$$\widehat{N}_H = \frac{\widehat{\eta}_F \, \widehat{y}_{F,H}}{\widehat{v}_{H,F}}. \tag{A.7}$$

If we could find a consistent and essentially unbiased estimator for η_F, then we could use Equation A.7 to form a consistent and essentially unbiased estimator for N_H,

even in the presence of false positive reports.

Unfortunately, we are not aware of a practical strategy for estimating the precision of out-reports. The most direct approach would be to interview each alter that a respondent reports as being in the hidden population. In other words, if a respondent reports knowing 3 drug injectors, researchers could try to interview these three people and see if they are actually drug injectors. Killworth et al. (2006) attempted a version of this procedure, which they called an "alter-chasing" study, but they later abandoned it because of the numerous logistical challenges that arose; see also Laumann (1969) for a related attempt. A second possible approach would be to conduct a census of a networked population where respondents are asked about themselves and specific people to whom they are connected. For example, Goel et al. (2010) collected responses about the political attitudes of thousands of interconnected people on Facebook, including respondents' attitudes as well as their beliefs about specific alters' attitudes. For a subset of respondents, they could compare i's belief about j's attitude with j's report of her own attitude in order to measure the precision. Unfortunately, we think it would be difficult to include a sufficiently large number of members of a stigmatized hidden population in this type of study.

We expect that the measurement of the precision of out-reports will pose a major challenge for future scale-up research, and we hope that practical solutions to this problem can be found. For the time being, we recommend that researchers show the impact that different values of the precision of out-reports would have on size estimates (Equation A.7).

B   Estimates with a sample from F

In this appendix, we present the full results for all of the estimators that require a sample from the frame population. First, we describe the general requirements that our sampling design for F must satisfy (Section B.1). Then we describe how to estimate the total number of out-reports, $y_{F,H}$ (Section B.2). Next we turn to some background material on multisets (Section B.3), which is needed for the following section on the known population method for estimating network degree (Section B.4). Finally, we present an estimator for the frame ratio, ϕ_F, which makes use of the known population method results (Section B.5).

B.1   Requirements for sampling designs from F

We follow Sarndal et al. (1992)'s definition of a probability sampling design, which we repeat here for convenience. Suppose that we have a set of possible samples $\{s_1, \ldots, s_j, \ldots, s_{\max}\}$, with each $s_j \subset F$. Furthermore, suppose $p(s_j)$ gives the probability of selection for each possible sample s_j. If we select a sample s_F at random using a process that will produce each possible sample s_j with probability $p(s_j)$, and if every element $i \in F$ has a nonzero probability of inclusion $\pi_i > 0$, then we will say that we have selected a *probability sample* and we call $p(\cdot)$ the *sampling design*.

B.2   Estimating the total number of out-reports, $y_{F,H}$

If we have a probability sample from the frame then estimating the total number of out-reports is a straightforward application of a standard survey estimator.

Result B.1 *Suppose we have a sample s_F taken from the frame population using a probability sampling design with probabilities of inclusion given by π_i (Sec. B.1). Then*

the estimator given by

$$\widehat{y}_{F,H} = \sum_{i \in s_F} y_{i,H}/\pi_i \tag{B.1}$$

is consistent and unbiased for $y_{F,H}$.

Proof: This follows from the fact that Equation B.1 is a Horvitz-Thompson estimator (Sarndal et al., 1992, Section 2.8). ∎

B.3 Reporting about multisets

Appendix B.4 and Appendix C both describe strategies that involve asking respondents to answer questions about their network alters in specific groups. In this section, we develop the notation and some basic properties of responses generated this way; these properties will be then be used in the subsequent sections.

Suppose we have several groups $A_1, \ldots, A_J$ with $A_j \subset U$ for all j, and also a frame population F of potential interviewees. (Note that we do not require $A_j \subset F$.) Imagine concatenating all of the people in populations $A_1, \ldots, A_J$ together, repeating each individual once for each population she is in. The result, which we call the *probe alters*, $\mathcal{A}$, is a multiset. The size of $\mathcal{A}$ is $N_\mathcal{A} = \sum_j N_{A_j}$.

Let y_{i,A_j} be the number of members of group A_j that respondent i reports having among the members of her personal network. We also write $y_{i,\mathcal{A}} = \sum_j y_{i,A_j}$ for the sum of the responses for individual i across all of $A_1, \ldots, A_J$, and $y_{F,\mathcal{A}} = \sum_{i \in F} \sum_j y_{i,A_j}$ to denote the total number of reports from F to $\mathcal{A}$. Similarly, we write $d_{i,\mathcal{A}} = \sum_j d_{i,A_j}$ for the sum of the network connections from individual i to each $A_1, \ldots, A_J$, and $d_{F,\mathcal{A}} = \sum_{i \in F} \sum_j d_{i,A_j}$ for the total of the individual $d_{i,\mathcal{A}}$ taken over all i. As always, we will write averages with respect to the first subscript so that, for example, $\bar{d}_{\mathcal{A},F} =$

$d_{\mathcal{A},F}/N_{\mathcal{A}}$.

We now derive a property of estimation under multisets that will be useful later on. Roughly, this property says that we can estimate the total number of reports from the entire frame population to the entire multiset of probe alters using only a sample from the frame population with known probabilities of inclusion (Section B.1). While this property might seem intuitive, we state it formally for two reasons. First, by stating it explicitly, we show that this property is very general: it does not require any assumptions about the contact pattern between the frame population and probe alters, nor does it require any assumptions about the probe alters. Second, it will turn out to be useful in several later proofs, and so we state it for compactness.

Property B.2 *Suppose we have a sample s_F from F taken using a probability sampling design with probabilities of inclusion π_i (Section B.1). Then*

$$\widehat{y}_{F,\mathcal{A}} = \sum_{i \in s_F} y_{i,\mathcal{A}}/\pi_i \tag{B.2}$$

is a consistent and unbiased estimator for $y_{F,\mathcal{A}}$.

Proof: If we define $a_i = \sum_j y_{i,A_j}$, the sum of the responses to each A_j for individual i, then we can write our estimator as

$$\widehat{y}_{F,\mathcal{A}} = \sum_{i \in s_F} a_i/\pi_i. \tag{B.3}$$

This is a Horvitz-Thompson esimator (see, e.g., Sarndal et al., 1992, chap. 2); it is unbiased and consistent for the total $\sum_{i \in F} a_i = y_{F,\mathcal{A}}$. ∎

B.4 Network degree and the known population method for estimating $\bar{d}_{F,F}$, $\bar{d}_{F,U}$, and $\bar{d}_{U,F}$

In order to conduct a scale-up study, we need a definition of the network that we will ask respondents to tell us about; that is, we need to define what it will mean for two members of the population to be connected by an edge. To date, most scale-up studies have used slight variations of the same definition: the respondent is told that she should consider someone a member of her network if she "knows" the person, where to know someone means (i) you know her and she knows you; (ii) you have been in contact in the past 2 years; and, (iii), if needed, you could get in touch with her (Bernard et al., 2010). Of course, many other definitions are possible, and an investigation of this issue is a matter for future study. The only restriction on the tie definition we impose here is that it be reciprocal; that is, the definition must imply that if the respondent is connected to someone, then that person is also connected to the respondent.

For a particular definition of a network tie an individual i's degree, $d_{i,U}$ may not be very easy to directly observe, even if the network is conceptually well-defined. For the basic scale-up estimator, the most commonly used technique for estimating respondents' network sizes is called the known population method (Killworth et al., 1998a; Bernard et al., 2010).[8] The known population method is based on the idea that we can estimate a respondent's network size by asking how many connections she has to a number of different groups whose sizes are known. The more connections a respondent reports to these groups, the larger we estimate her network to be. Current

---

[8]There are other techniques for estimating personal network size, including the summation method (McCarty et al., 2001; Bernard et al., 2010), which could be used in conjunction with many of our results. We focus on the known population method here because it is relatively easy to work with from a statistical perspective, and also because there is some evidence that it works better in practice (Salganik et al., 2011a; Rwanda Biomedical Center, 2012)

standard practice is to ask a respondent about her connections to approximately 20 groups of known size in order to estimate her degree (Bernard et al., 2010), although the exact number of groups used has no impact on the bias of the estimates as we show in Results B.3 and B.4.

The known population estimator was originally introduced to estimate the personal network size of each respondent individually (Killworth et al., 1998a), but in Sections 3 and 4.2 we showed that for the scale-up method the quantity of interest is actually the average number of connections from a member of the frame population F to the rest of the frame population F ($\bar{d}_{F,F}$), or the average number of connections from a member of the entire population U to the frame population F ($\bar{d}_{U,F}$).[9] This is fortunate, because it is easier to estimate an average degree over all respondents than it is to estimate the individual degree for each respondent.

B.4.1   Guidance for choosing the probe alters, $\mathcal{A}$

Result B.3, below, shows that the known population estimator will produce consistent and unbiased estimates of average network degree if (i) $y_{F,\mathcal{A}} = d_{F,\mathcal{A}}$ (*reporting condition*); and (ii) $\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}$ (*probe alter condition*). Stating these conditions precisely enables us to provide guidance about how the groups of known size $(A_1, A_2, \ldots A_J)$ should be selected such that the probe alters $\mathcal{A}$ will enable consistent and unbiased estimates.

First, the reporting condition ($y_{F,\mathcal{A}} = d_{F,\mathcal{A}}$) in Result B.3 shows that researchers should select probe alters such that reporting will be accurate in aggregate. One way to make the reporting condition more likely to hold is to select groups that are unlikely to suffer from transmission error (Shelley et al., 1995, 2006; Killworth

---

[9]Although we have framed our discussion here in terms of $\bar{d}_{F,F}$, the same ideas apply to $\bar{d}_{U,F}$ and $\bar{d}_{F,U}$.

et al., 2006; Salganik et al., 2011b; Maltiel et al., 2015). Another way to make the reporting condition more likely to hold is to avoid selecting groups that may lead to recall error (Killworth et al., 2003; Zheng et al., 2006; McCormick and Zheng, 2007; McCormick et al., 2010; Maltiel et al., 2015). That is, previous work suggests that respondents seem to under-report the number of connections they have to large groups, although the precise mechanism behind this pattern is unclear (Killworth et al., 2003). Researchers who have data that may include recall error can consider some of the empirically-calibrated adjustments that have been used in earlier studies (Zheng et al., 2006; McCormick and Zheng, 2007; McCormick et al., 2010; Maltiel et al., 2015).

Second, the probe alter condition ($\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}$) in Result B.3 shows that researchers should select groups to be typical of F in terms of their connections to F. In most applied situations, we expect that F will consist of adults, so that researchers should choose groups of known size that are composed of adults, or that are typical of adults in terms of their connections to adults. Further, when trying to choose groups that satisfy the probe alter condition, it is useful to understand how connections from the individual known populations to the frame ($\bar{d}_{A_1,F}, \ldots, \bar{d}_{A_J,F}$) aggregate up into connections from the probe alters to the frame ($\bar{d}_{\mathcal{A},F}$). Basic algebraic manipulation shows that the probe alter condition can be written as:

$$\frac{\sum_j \bar{d}_{A_j,F} \, N_{A_j}}{\sum_j N_{A_j}} = \bar{d}_{F,F}. \tag{B.4}$$

Equation B.4 reveals that the probe alter condition requires that $\bar{d}_{F,F}$ is equal to a weighted average of the average number of connections between each individual known population A_j and the frame population F ($\bar{d}_{A_j,F}$). The weights are given by the size of each known population, N_{A_j}. The simplest way that this could be satisfied

A9

is if $\bar{d}_{A_j,F} = \bar{d}_{F,F}$ for every known population A_j. If this is not true, then the probe alter condition can still hold as long as groups for which $\bar{d}_{A_j,F}$ is too high are offset by other groups for which $\bar{d}_{A_{j'},F}$ is too low.

In practice it may be difficult to determine if the reporting condition and probe alter condition will be satisfied. Therefore, we recommend that researchers assess the sensitivity of their size estimates using the procedures described in Online Appendix D. Further, we note that in many realistic situations, N_{A_j} might not be known exactly. Fortunately, researchers only need to know $\sum_j N_{A_j}$, and they can assess the sensitivity of their estimates to errors in the size of known populations using the procedures described in Online Appendix D.

B.4.2 The known population estimators

Given that background about selecting the probe alters, we present the formal results for the known population estimators for $\bar{d}_{F,F}$, $\bar{d}_{U,F}$, and $\bar{d}_{F,U}$.

Result B.3 *Suppose we have a sample s_F taken from the frame population using a probability sampling design with probabilities of inclusion given by π_i (see Section B.1). Suppose also that we have a multiset of known populations, $\mathcal{A}$. Then the known population estimator given by*

$$\widehat{\bar{d}}_{F,F} = \frac{\sum_{i \in s_F} \sum_j y_{i,A_j}/\pi_i}{N_{\mathcal{A}}} \tag{B.5}$$

is consistent and unbiased for $\bar{d}_{F,F}$ if

$$y_{F,\mathcal{A}} = d_{F,\mathcal{A}}, \qquad \textit{(reporting condition)} \tag{B.6}$$

A10

and if

$$\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}. \qquad \text{(probe alter condition)} \qquad \text{(B.7)}$$

Proof: By Property B.2, we know that our estimator is unbiased and consistent for $y_{F,\mathcal{A}}/N_{\mathcal{A}}$. By the reporting condition in Equation B.6, this means it is unbiased and consistent for $d_{F,\mathcal{A}}/N_{\mathcal{A}}$. Then, by the probe alter condition in Equation B.7, it is also unbiased and consistent for $\bar{d}_{F,F}$. ∎

Result B.4 *Suppose we have a sample s_F taken from the frame population using a probability sampling design with probabilities of inclusion given by π_i (see Section B.1). Suppose also that we have a multiset of known populations, $\mathcal{A}$. Then the known population estimator given by*

$$\widehat{\bar{d}}_{U,F} = \frac{\sum_{i \in s_F} \sum_j y_{i,A_j}/\pi_i}{N_{\mathcal{A}}} \qquad \text{(B.8)}$$

is consistent and unbiased for $\bar{d}_{U,F}$ if

$$y_{F,\mathcal{A}} = d_{F,\mathcal{A}}, \qquad \text{(reporting condition)} \qquad \text{(B.9)}$$

and if

$$\bar{d}_{\mathcal{A},F} = \bar{d}_{U,F}. \qquad \text{(probe alter condition)} \qquad \text{(B.10)}$$

Proof: By Property B.2, we know that our estimator is unbiased and consistent for $y_{F,\mathcal{A}}/N_{\mathcal{A}}$. By the reporting condition in Equation B.9, this means it is unbiased and consistent for $d_{F,\mathcal{A}}/N_{\mathcal{A}}$. Then, by the probe alter condition in Equation B.10, it is also unbiased and consistent for $\bar{d}_{U,F}$. ∎

A11

Since $\bar{d}_{F,U} = \frac{N}{N_F}\bar{d}_{U,F}$, as a direct consequence of Result B.4 we have the following corollary.

Corollary B.5 *If the conditions described in Result B.4 hold,*

$$\widehat{\bar{d}}_{F,U} = \widehat{\bar{d}}_{U,F} \; \frac{N}{N_F} \tag{B.11}$$

is consistent and unbiased for $\bar{d}_{F,U}$.

B.5   Estimating the frame ratio, ϕ_F

Given our estimator of $\bar{d}_{F,F}$ (Result B.3) and our estimator of $\bar{d}_{U,F}$ (Result B.4), we can estimate the frame ratio, ϕ_F.

Result B.6 *The estimator*

$$\widehat{\phi}_F = \frac{\widehat{\bar{d}}_{F,F}}{\widehat{\bar{d}}_{U,F}} \tag{B.12}$$

is consistent and essentially unbiased for ϕ_F if $\widehat{\bar{d}}_{F,F}$ is consistent and essentially unbiased for $\bar{d}_{F,F}$ and $\widehat{\bar{d}}_{U,F}$ is consistent and essentially unbiased for $\bar{d}_{U,F}$.

Proof: This follows from the properties of a ratio estimator (Sarndal et al., 1992, chap. 5). ■

More concretely, combining the estimator for $\bar{d}_{F,F}$ (Result B.3) and the estimator for $\bar{d}_{U,F}$ (Result B.4), and assuming that we have known populations $\mathcal{A}_{F_1}$ for $\bar{d}_{F,F}$, and $\mathcal{A}_{F_2}$ for $\bar{d}_{U,F}$, we obtain

$$\widehat{\phi}_F = \frac{N_{\mathcal{A}_{F_2}}}{N_{\mathcal{A}_{F_1}}} \; \frac{\sum_{i \in s_F} \sum_{A_j \in \mathcal{A}_{F_1}} y_{i,A_j}/\pi_i}{\sum_{i \in s_F} \sum_{A_k \in \mathcal{A}_{F_2}} y_{i,A_k}/\pi_i}. \tag{B.13}$$

In our discussion of $\widehat{\bar{d}}_{F,F}$ (Result B.3) and $\widehat{\bar{d}}_{U,F}$ (Result B.4), we concluded that we want the known populations $\mathcal{A}_{F_1}$ used for $\widehat{\bar{d}}_{F,F}$ to be typical of members of F in their connections to F. An analogous argument shows that we want the known populations $\mathcal{A}_{F_2}$ used for $\widehat{\bar{d}}_{U,F}$ to be typical of members of U in their connections to F. In general, we expect that it will not be appealing to assume that F and U are similar to each other in terms of their connections to F meaning that, unfortunately, it will not make sense to use the same set of known populations for $\widehat{\bar{d}}_{F,F}$ and $\widehat{\bar{d}}_{U,F}$. If researchers wish to estimate ϕ_F directly, one approach would be to choose $\mathcal{A}_{F_2}$ to be typical of U in such a way that some of the individual known populations are more typical of F, while others more typical of $U - F$. The multiset formed from only the ones that are more typical of F could then be our choice for $\mathcal{A}_{F_1}$. In this case, researchers would also want $\frac{N_{\mathcal{A}_{F_1}}}{N_{\mathcal{A}_{F_2}}} \approx \frac{N_F}{N}$. This complication is one of the reasons we recommend in Section 4 that future scale-up studies estimate $\bar{d}_{F,F}$ directly, thus avoiding the need to estimate ϕ_F entirely.

C   Estimates with samples from F and H

In this appendix, we present the full results for all of the estimators that require a sample from the hidden population. Section C.1 defines the general requirements that our sampling design for H must satisfy. Section C.2 describes a flexible data collection procedure called the game of contacts. Section C.3 introduces some background material on estimation using questions about multisets and presents an estimator for $\bar{v}_{H,F}$, the average number of in-reports among the members of the hidden population. Section C.5 gives some guidance about how to choose the probe alters for the known population method. Section C.6 presents estimators for the two adjustment factors introduced in Section 3: the degree ratio, δ_F, and the true positive rate, τ_F. Finally,

Section C.7 presents formal results for four different estimators for N_H.

C.1 Requirements for sampling designs from H

For the results that involve a sample from the hidden population s_H, we do not need a probability sample (Appendix B); instead, we need a weaker type of design. We require that every element $i \in H$ have a nonzero probability of selection $\pi_i > 0$, and that we can determine the probability of selection up to a constant factor c; that is, we only need to know $c\pi_i$. We are not aware of any existing name for this situation, so we will call it a *relative probability sample*. Because of the challenges involved in sampling hard-to-reach populations, the two most likely sampling designs for s_H will probably be time-location sampling (Karon and Wejnert, 2012) and respondent-driven sampling (Heckathorn, 1997). A relative probability sample allows us to use weighted sample means to estimate averages, but not totals. See Sarndal et al. (1992, Section 5.7) for more details on weighted sample means, also sometimes called Hájek estimators, which is what we use to estimate averages from a sample of hidden population members.

C.2 Data collection

In order to make estimates about the hidden population's visibility to the frame population, researchers will need to collect what we call *enriched aggregate relational data* from each respondent, and a procedure called the *game of contacts* has produced promising results from a study of heavy drug users in Brazil (Salganik et al., 2011b). In the main text, we assumed that the groups in the probe alters $A_1, \ldots, A_J$ were all contained in the frame population ($A_j \subset F$ for all j). However, the estimators in this Online Appendix are more general because they allow for the possibility that some of

the groups $A_1, \ldots A_J$ may not be contained entirely in F. For example, if the frame population is adults, then this flexibility enables researchers to use groups based on names, such as Michael, even though not all people named Michael are adults.

In order to allow for this flexibility, we need to introduce some new notation: let $A_1 \cap F, A_2 \cap F, \ldots, A_J \cap F$ be the intersection of these groups and the frame population, and let $\mathcal{A} \cap F$ be the concatenation of these intersected groups. For example, if the frame population is adults, A_1 is people named Michael, and A_2 is doctors, then $A_1 \cap F$ is adults named Michael, $A_2 \cap F$ is adult doctors, and $\mathcal{A} \cap F$ is the collection of all adult Michaels and all adult doctors, with adult doctors named Michael included twice. (In the special case discussed in the main text, $A_1 \cap F, \ldots A_J \cap F = A_1, \ldots, A_J$.)

The data collection begins with a relative probability sample (Section C.1) from the hidden population. For a set of groups, $A_1, A_2, \ldots A_J$, each respondent in the hidden population is asked, "How many people do you know in group A_j?" We call the response y_{i,A_j}. Next for each of the y_{i,A_j} alters, the respondent picks up a token and places it on a game board like the one in Figure C.1. From the location of the tokens on the board, the researcher can record whether each alter is in the frame population (or not) and whether the alter is aware that the respondent is in the hidden population (or not) (Table C.2). This process is then repeated until the respondent has been asked about all groups.

If all of the probe alters are in the frame population, then the process is much easier for respondents and the game board can be modified to collect alternative information. If all of the probe alters are not in the frame population, then it is important for the researcher to define the frame population as clearly as possible. If the respondents are not able to correctly indicate whether the alters are in the frame population or not, it could lead to biased estimates of $\bar{v}_{H,F}$. For more on the operational implementation of this procedure, see Salganik et al. (2011b).

Adult &	Adult &
Knows that I inject drugs	Does not know that I inject drugs
Child &	Child &
Knows that I inject drugs	Does not know that I inject drugs

Figure C.1: Example of a game board that could be used in the game of contacts interviewing procedure if the hidden population was people who inject drugs and the frame was made up of adults. This board is a variation of the board used in Salganik et al. (2011b).

	aware	not aware	total
frame population	$\widetilde{v}_{i,A_j \cap F}$	$\widetilde{h}_{i,A_j \cap F}$	$y_{i,A_j \cap F}$
not frame population	$\widetilde{v}_{i,A_j \cap (U-F)}$	$\widetilde{h}_{i,A_j \cap (U-F)}$	$y_{i,A_j \cap (U-F)}$
total	$\widetilde{v}_{i,A_j}$	$\widetilde{h}_{i,A_j}$	y_{i,A_j}

Table C.1: Responses collected during the game of contacts for each respondent i and each group A_j. We use $\widetilde{}$ to indicate reported values. For example, $\widetilde{v}_{i,A_j}$ is the respondent's reported visibility to people in A_j and v_{i,A_j} is respondent's actual visiblility to people in A_j. Also, using this notational convention, it is the case that $y_{i,A_j} = \widetilde{d}_{i,A_j}$, but we have written y_{i,A_j} in order to be consistent with the rest of the paper.

C.3 Estimation using aggregated relational data from the hidden population

In this section, we follow Section B.3 and present another useful property about estimates made using aggregate relational data from the hidden population. Roughly, this property says that we can estimate the average number of reports from the entire hidden population to the probe alters using only a relative probability sample from the hidden population (Section C.1). Similar to Property B.2, the result we present below does not require any assumptions about the contact pattern between the hidden population and the probe alters, nor about the probe alters themselves.

Property C.1 *Suppose we have a sample s_H from H taken using a relative probability design, allowing us to compute the relative probabilities of inclusion $c\pi_i$ for all sampled elements (Sec. C.1). Then*

$$\widehat{\bar{y}}_{H,\mathcal{A}} = \frac{\sum_{i \in s_H} y_{i,\mathcal{A}}/(c\pi_i)}{\sum_{i \in s_H} 1/(c\pi_i)} \tag{C.1}$$

is a consistent and essentially unbiased estimator for $\bar{y}_{H,\mathcal{A}} = y_{H,\mathcal{A}}/N_H$.

Proof: Note that the c in the relative probabilities of inclusion $c\pi_i$ cancel, so that

$$\widehat{\bar{y}}_{H,\mathcal{A}} = \frac{\sum_{i \in s_H} y_{i,\mathcal{A}}/(\pi_i)}{\sum_{i \in s_H} 1/(\pi_i)}. \tag{C.2}$$

If we define $a_i = \sum_j y_{i,A_j}$, the sum of the responses to each A_j for individual i, then we can write our estimator as

$$\widehat{\bar{y}}_{H,\mathcal{A}} = \frac{\sum_{i \in s_H} a_i/\pi_i}{\sum_{i \in s_H} 1/\pi_i}. \tag{C.3}$$

Now we have a standard weighted mean estimator (e.g. Sarndal et al., 1992, chap. 5); it is consistent and essentially unbiased for the average $\frac{1}{N_H} \sum_{i \in H} a_i = y_{H,\mathcal{A}}/N_H$. ∎

C.4 Estimating the average visibility, $\bar{v}_{H,F}$

Given the data collection procedure described in Sec. C.2, we can estimate the average visibility ($\bar{v}_{H,F}$) as long as three conditions are satisfied: one about reporting, one about the visibility of the hidden population to the probe alters, and one about sampling.

Result C.2 *Assume that we have a sample s_H taken from the hidden population using a relative probability design with relative probabilities of inclusion $c\pi_i$ for all sampled elements (Sec. C.1). Then*

$$\widehat{\bar{v}}_{H,F} = \frac{N_F}{N_{\mathcal{A} \cap F}} \frac{\sum_{i \in s_H} \sum_j \widetilde{v}_{i,A_j \cap F}/(c\pi_i)}{\sum_{i \in s_H} 1/(c\pi_i)} \tag{C.4}$$

is consistent and essentially unbiased for $\bar{v}_{H,F}$ if

$$\widetilde{v}_{H,\mathcal{A} \cap F} = v_{H,\mathcal{A} \cap F}, \qquad \text{(reporting condition)} \tag{C.5}$$

and

$$\frac{v_{H,\mathcal{A} \cap F}}{N_{\mathcal{A} \cap F}} = \frac{v_{H,F}}{N_F}. \qquad \text{(probe alter condition)} \tag{C.6}$$

Proof: Property C.1 holds for estimating $\bar{\bar{v}}_{F,\mathcal{A} \cap F}$ from $\widetilde{v}_{i,\mathcal{A} \cap F}$, just as it holds for estimating $\bar{y}_{H,\mathcal{A} \cap F}$ from $y_{i,\mathcal{A} \cap F}$. Applying Property C.1 here, we conclude that the

estimator is consistent and essentially unbiased for

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\bar{\widetilde{v}}_{H,\mathcal{A}\cap F} = \frac{N_F}{N_{\mathcal{A}\cap F}}\frac{\widetilde{v}_{H,\mathcal{A}\cap F}}{N_H}. \tag{C.7}$$

Next, by applying the reporting condition in Equation C.5 we can conclude that

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\frac{\widetilde{v}_{H,\mathcal{A}\cap F}}{N_H} = \frac{N_F}{N_{\mathcal{A}\cap F}}\frac{v_{H,\mathcal{A}\cap F}}{N_H}. \tag{C.8}$$

Finally, by applying the probe alter condition in Equation C.6 and rearranging terms, we conclude that

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\frac{v_{H,\mathcal{A}\cap F}}{N_H} = \frac{N_F}{N_H}\frac{v_{H,F}}{N_F} \tag{C.9}$$

$$= \bar{v}_{H,F} \tag{C.10}$$

∎

Note that Result C.2 requires us to know the size of the probe alters in the frame population, $N_{\mathcal{A}\cap F}$. In some cases, this may not be readily available, but it may be reasonable to assume that

$$N_{\mathcal{A}\cap F} = \frac{N_F}{N}\, N_{\mathcal{A}}. \tag{C.11}$$

Furthermore, if $\mathcal{A}$ is chosen so that all of its members are in F, then $N_{\mathcal{A}\cap F} = N_{\mathcal{A}}$ and $v_{i,A_j\cap F} = v_{i,A_j}$. In this situation, we do not need to specifically ask respondents about connections to $\mathcal{A}\cap F$; we can just ask about connections to $\mathcal{A}$.

The reporting condition required for Result C.5 states that the hidden population's total reported visibility from the probe alters on the frame must be correct. This might not be the case, if for example, respondents systematically over-estimate how

much others know about them (see e.g., Gilovich et al. (1998)).

The required condition for the probe alters is slightly more complex. It needs to be the case that the rate at which the hidden population is visible to the probe alters is the same as the rate at which the hidden population is visible to the frame population. There are several equivalent ways of stating this condition, as we show in a moment. First, we need to define two new quantities: the individual-level true positive rate and the average of the individual-level true positive rates.

Definition 1 *We define the individual-level true positive rate for respondent $i \in F$ to be*

$$\tau_i = \frac{v_{H,i}}{d_{i,H}}, \tag{C.12}$$

where $v_{H,i} = \sum_{j \in H} v_{j,i}$.

Definition 2 *We define the average of the individual true positive rates over a set F of respondents as*

$$\overline{\tau}_F = \frac{1}{N_F} \sum_{i \in F} \tau_i. \tag{C.13}$$

In general, $\overline{\tau}_F \neq \tau_F$. To see this, note that while $\overline{\tau}_F$ is the average of the individual-level true positive rates with each individual weighted equally, τ_F can be written as the weighted average of the individual true positive rates, with the weights given by each individual's degree. We can see the exact relationship between the two by expressing τ_F in terms of the τ_i:

$$\tau_F = \frac{\sum_{i \in F} \tau_i\, d_{i,H}}{\sum_{i \in F} d_{i,H}}, \tag{C.14}$$

since multiplying each τ_i by $d_{i,H}$ and summing is the same as summing the $v_{H,i}$.

Result C.3 *The following conditions are all equivalent.*

(i) $\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{v_{H,F}}{N_F}$

(ii) $\tau_{\mathcal{A}\cap F} \; \bar{d}_{\mathcal{A}\cap F,H} = \tau_F \; \bar{d}_{F,H}$

(iii) $\bar{\tau}_{\mathcal{A}\cap F} \; \bar{d}_{\mathcal{A}\cap F,H} + cov_{\mathcal{A}\cap F}(\tau_i, d_{i,H}) = \bar{\tau}_F \; \bar{d}_{F,H} + cov_F(\tau_i, d_{i,H})$

(iv) $\bar{y}_{F,H}^+ = \frac{\sum_j \bar{y}_{\mathcal{A}_j \cap F,H}^+ \; N_{\mathcal{A}_j \cap F}}{\sum_j N_{\mathcal{A}_j \cap F}},$

where cov_F is the finite-population covariance taken over the set F.[10]

    Proof: First, we show that

$$\tau_{\mathcal{A}\cap F} \; \bar{d}_{\mathcal{A}\cap F,H} = \tau_F \; \bar{d}_{F,H} \iff \frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{v_{H,F}}{N_F}. \tag{C.15}$$

By definition, $\tau_F \; \bar{d}_{F,H} = (v_{H,F}/d_{F,H}) \times (d_{F,H}/N_F) = v_{H,F}/N_F$. The same argument demonstrates that $\tau_{\mathcal{A}\cap F} \; \bar{d}_{\mathcal{A}\cap F,H} = v_{H,\mathcal{A}\cap F}/N_{\mathcal{A}}$. We conclude that $(i) \iff (ii)$.

    Next, we show that (ii) is equivalent to (iii). We can use the relationship between τ_F and the τ_i, Equation C.14, to deduce that

$$\tau_F \; d_{F,H} = \sum_{i \in F} \tau_i \; d_{i,H} = N_F \; [\bar{\tau}_F \; \bar{d}_{F,H} + cov_F(\tau_i, d_{i,H})]. \tag{C.16}$$

Dividing the left-most and right-most sides by N_F, we conclude that

$$\tau_F \; \bar{d}_{F,H} = \bar{\tau}_F \; \bar{d}_{F,H} + cov_F(\tau_i, d_{i,H}). \tag{C.17}$$

---

[10]We define the finite-population covariance to have a denominator of N_F; this differs from some other authors, who define the finite-population covariance to have $N_F - 1$ in the denominator.

The same argument shows that

$$\bar{d}_{\mathcal{A}\cap F,H}\ \tau_{\mathcal{A}\cap F} = \overline{\tau}_{\mathcal{A}\cap F}\ \bar{d}_{\mathcal{A}\cap F,H} + \mathrm{cov}_{\mathcal{A}\cap F}(\tau_i, d_{i,H}). \tag{C.18}$$

So we conclude that $(ii) \Longleftrightarrow (iii)$.

Finally, we show that (iv) is equivalent to (i). In Appendix A, showed that $y_{F,H}^+ = v_{H,F}$ (Equation A.3). Dividing both sides by N_F, we have $\bar{y}_{F,H}^+ = v_{H,F}/N_H$, which is the right-hand side of the identity in (i). Similarly, starting with the left-hand side of the identity in (i), we have

$$\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{\sum_j v_{H,A_j\cap F}}{\sum_j N_{A_j\cap F}} = \frac{\sum_j y_{A_j\cap F,H}^+}{\sum_j N_{A_j\cap F}} = \frac{\sum_j \bar{y}_{A_j\cap F,H}^+\ N_{A_j\cap F}}{\sum_j N_{A_j\cap F}}. \tag{C.19}$$

So we conclude that $(i) \Longleftrightarrow (iv)$.

Since $(i) \Longleftrightarrow (ii)$ and $(ii) \Longleftrightarrow (iii)$, it follows that $(i) \Longleftrightarrow (iii)$. Furthermore, since $(i) \Longleftrightarrow (iv)$, it follows that (iv) is equivalent to (ii) and (iii). ∎

Result C.3 shows that the probe alter condition can be expressed in many equivalent ways. One of these alternate expressions is especially useful because it leads to an empirical check of the probe alter condition that future scale-up studies can implement. This empirical check is a direct consequence of Result C.4, below. Intuitively, Result C.4 and the empirical check are a consequence of the identity in Equation 1, which says that in-reports from the perspective of H are also out-reports from the perspective of F.

Result C.4 *Suppose that the precision of out-reports from the frame population is the same as the precision of the out-reports from $\mathcal{A} \cap F$:*

$$\frac{y_{F,H}^+}{y_{F,H}} = \frac{y_{\mathcal{A}\cap F,H}^+}{y_{\mathcal{A}\cap F,H}} \tag{C.20}$$

A22

Then the probe alter condition (C.6) is satisfied if and only if

$$\bar{y}_{F,H} = \bar{y}_{\mathcal{A}\cap F,H}.$$
(C.21)

Proof: First, note that, by Result C.3, the probe alter condition is equivalent to

$$\bar{y}_{F,H}^{+} = \frac{\sum_{j} \bar{y}_{A_j\cap F,H}^{+} \, N_{A_j\cap F}}{\sum_{j} N_{A_j\cap F}}.$$
(C.22)

Since $\bar{y}_{A_j\cap F,H}^{+} = y_{A_j\cap F,H}^{+}/N_{A_j\cap F}$ for all j, the right-hand side of Equation C.22 is equal to $\bar{y}_{\mathcal{A}\cap F,H}^{+}$, meaning that the probe alter condition is also equivalent to

$$\bar{y}_{F,H}^{+} = \bar{y}_{\mathcal{A}\cap F,H}^{+}.$$
(C.23)

Second, note that the assumption in Equation C.20 can be re-written as

$$\frac{\bar{y}_{F,H}^{+}}{\bar{y}_{F,H}} = \frac{\bar{y}_{\mathcal{A}\cap F,H}^{+}}{\bar{y}_{\mathcal{A}\cap F,H}},$$
(C.24)

by multiplying the left-hand side by $\frac{N_F}{N_F}$ and the right-hand side by $\frac{N_{\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}}$. So we are left with the task of showing that if Equation C.24 is true, then Equation C.23 is satisfied if and only if Equation C.21 is satisfied. But this is the case, since Equation C.23 equates the numerators of the two fractions in Equation C.24 and Equation C.21 equates the denominators of the two fractions in Equation C.24. Two fractions that are equal will have equal numerators if and only if they have equal denominators. (Formally, if $a/b = c/d$ then $a = c$ if and only if $b = d$.) ∎

The implication of Result C.4 is that if (i) researchers design the probe alters so that the frame population sample s_F can be used to estimate $\bar{y}_{\mathcal{A}\cap F,H}$; and (ii) researchers assume that the precision of out-reports from the frame population is the

A23

same as the precision of out-reports from $\mathcal{A} \cap F$, then they can evaluate how well the probe alter condition is satisfied empirically by comparing $\widehat{\bar{y}}_{F,H}$ and $\widehat{\bar{y}}_{\mathcal{A} \cap F,H}$.

Finally, we can foresee four practical problems that might arise when researchers try to estimate $\bar{v}_{H,F}$. First, researchers might not be able to choose the probe alters to satisfy the probe alter condition (Equation C.6) because of limited information about the true visibility of the hidden population with respect to different social groups. A second problem might arise if researchers are not able to choose the probe alters to satisfy the reporting condition (Equation C.5) because of limited information about the hidden population's awareness about visibility. A third problem might arise due to errors in administrative records that would cause researchers to have incorrect information about the size of the multiset of probe alters on the frame ($N_{\mathcal{A} \cap F}$). Finally, a fourth problem might arise due to errors in the sampling method researchers use. Fortunately, as we show in Online Appendix D (Result D.6), it is possible to quantify the effect of these problems on the resulting estimates. In some cases they can cancel out, but in other cases they magnify each other.

C.5 Guidance for choosing the probe alters for the game of contacts, $\mathcal{A}$

Turning the results in Online Appendix C into easy to follow steps for selecting the probe alters for the game of contacts is an open and important research problem. Here, we briefly offer three recommendations for selecting the probe alters for the game of contacts. We realize that these recommendations may be difficult to follow exactly in practice. Therefore, we also discuss the sensitivity of the estimators to errors in the construction of the probe alters. Finally, we discuss one type of data that should be collected from the frame population in order to help the researchers

evaluate their choice of probe alters for the game of contacts.

First, we recommend that probe alters for the game of contacts be in the frame population. For example, if the frame population is adults, we recommend that all members of the probe alters be adults. This choice will simplify the data collection task in the game of contacts, and for all the advice listed below, we assume that it has been followed. If it is not possible, researchers can still use the more general procedures developed in this Online Appendix.

Second, we recommend that the probe alters be selected such that the probe alter condition in Result C.2 is satisfied. That is, the probe alters as a whole should be typical of the frame population in the following way: it should be the case that the rate at which the hidden population is visible to the probe alters is the same as the rate at which the hidden population is visible to the frame population ($\frac{v_{H,\mathcal{A}}}{N_{\mathcal{A}}} = \frac{v_{H,F}}{N_F}$). For example, in a study to estimate the number of drug injectors in a city, drug treatment counselors would be a poor choice for membership in the probe alters because drug injectors are probably more visible to drug treatment counselors than to typical members of the frame population. On the other hand, postal workers would probably be a reasonable choice for membership in the probe alters because drug injectors are probably about as visible to postal workers as they are to typical members of the frame population.

Third, we recommend that the probe alters be selected so that the reporting condition in Result C.2 is satisfied ($\widetilde{v}_{H,\mathcal{A}} = v_{H,\mathcal{A}}$). One way to help ensure that this condition holds is to avoid selecting large groups that may cause recall error (Killworth et al., 2003; Zheng et al., 2006; McCormick and Zheng, 2007; McCormick et al., 2010; Maltiel et al., 2015). In practice it might be difficult to meet each of these three conditions exactly, therefore we recommend a sensitivity analysis using the results in Online Appendix D.

Finally, the choice of probe alters for the game of contacts also has two implications for the design of the survey of the frame population. First, if researchers wish to estimate the degree ratio, δ_F, then they should design the probe alters $\mathcal{A}$ so that they can be asked of both members of the hidden population sample and members of the frame population sample (see Result C.6). Second, if researchers wish to test the probe alter condition using the approach in Result C.4, then additional information needs to be collected from each member of the frame population sample. For example, if one group in the probe alters for the game of contacts is postal workers, then members of the frame population sample should be asked if they are postal workers.

C.6   Term-by-term: δ_F and τ_F

In this section we describe how to estimate two adjustment factors: the degree ratio,

$$\delta_F = \frac{\bar{d}_{H,F}}{\bar{\bar{d}}_{F,F}} \tag{C.25}$$

and the true positive rate,

$$\tau_F = \frac{\bar{v}_{H,F}}{\bar{\bar{d}}_{H,F}}. \tag{C.26}$$

Estimating the degree ratio requires information from the survey of the hidden population and the survey of the frame population, while estimating the true positive rate only requires information from the survey of the hidden population (Fig. C.2). As Equations C.25 and C.26 make clear, both adjustment factors involve $\bar{d}_{H,F}$ so we first present an estimator for that quantity.

Result C.5 *Suppose we have a sample s_H taken from the hidden population using a relative probability sampling design with relative probabilities of inclusion denoted $c\pi_i$*
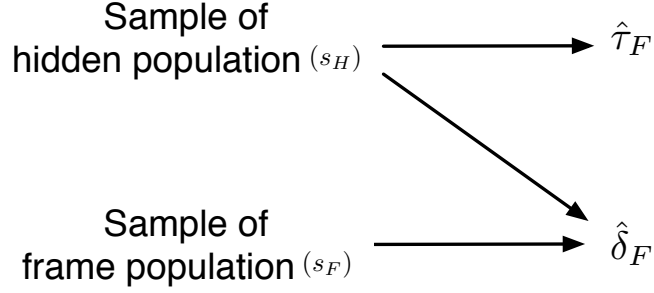
Figure C.2: We estimate the true positive rate $\hat{\tau}_F$ using data from the survey of the hidden population, and we estimate the degree ratio $\hat{\delta}_F$ using the sample of the hidden population and the sample of the frame population.

(Sec C.1). Then the estimator given by

$$\widehat{\bar{d}}_{H,F} = \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{\sum_{i\in s_H}\sum_j y_{i,(A_j\cap F)}/(c\pi_i)}{\sum_{i\in s_H} 1/(c\pi_i)} \tag{C.27}$$

is consistent and essentially unbiased for $\bar{d}_{H,F}$ *if:*

$$y_{H,\mathcal{A}\cap F} = d_{H,\mathcal{A}\cap F}, \qquad \text{(reporting condition)} \qquad \text{(C.28)}$$

and

$$\bar{d}_{\mathcal{A}\cap F,H} = \bar{d}_{F,H}. \qquad \text{(probe alter condition)} \qquad \text{(C.29)}$$

Proof: From Property C.1, we can see that our estimator is consistent and essentially unbiased for

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\frac{y_{H,\mathcal{A}\cap F}}{N_H} = \frac{N_F}{N_H}\frac{y_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}}. \tag{C.30}$$

Under the reporting condition (Equation C.28) this becomes

$$\frac{N_F}{N_H} \frac{y_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{N_F}{N_H} \frac{d_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} \tag{C.31}$$

Finally, applying the probe alter condition in Equation C.29, we have

$$\frac{N_F}{N_H} \frac{d_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{N_F}{N_H} \frac{d_{F,H}}{N_F} \tag{C.32}$$

$$= \bar{d}_{H,F}. \tag{C.33}$$

∎

Result C.5 requires that reports are, in total, correct (Equation C.28). Like Result C.2, Result C.5 also requires us to know the size of the probe alters on the frame, $N_{\mathcal{A}\cap F}$. In some cases, this may not be readily available, but it may be reasonable to assume that

$$N_{\mathcal{A}\cap F} = \frac{N_F}{N} \, N_{\mathcal{A}}. \tag{C.34}$$

Furthermore, if $\mathcal{A}$ is chosen so that all of its members are in F, then $N_{\mathcal{A}\cap F} = N_{\mathcal{A}}$ and $y_{i,A_j\cap F} = y_{i,A_j}$. In this situation, we do not need to specifically ask respondents about connections to $\mathcal{A}\cap F$; we can just ask about connections to $\mathcal{A}$. Result C.5 also requires a specific rate of connectivity between the probe alters and the hidden population (Equation C.29). We discussed some of the consequences of these assumption in the main text, where we made recommendations for practice (Section 4).

C.6.1 Estimating the degree ratio, δ_F

We can combine our estimator for $\bar{d}_{H,F}$ (Result C.5) and our estimator for $\bar{d}_{F,F}$ (Result B.3), to estimate the degree ratio, δ_F.

Result C.6 *The estimator*

$$\widehat{\delta}_F = \frac{\widehat{\bar{d}}_{H,F}}{\widehat{\bar{d}}_{F,F}} \tag{C.35}$$

is consistent and essentially unbiased for δ_F if $\widehat{\bar{d}}_{H,F}$ is consistent and essentially unbiased for $\bar{d}_{H,F}$ and $\widehat{\bar{d}}_{F,F}$ is consistent and essentially unbiased for $\bar{d}_{F,F}$.

Proof: This follows from the properties of a compound ratio estimator (Online Appendix E). ∎

More concretely, combing the estimators in Result C.5 and Result B.3, results in an estimator for $\widehat{\delta}_F$ with the following form:

$$\widehat{\delta}_F = \frac{\frac{N_F}{N_{\mathcal{A}_H \cap F}} \frac{\sum_{i \in s_H} \sum_{A_j \in \mathcal{A}_H} y_{i,(A_j \cap F)}/(c\pi_i^H)}{\sum_{i \in s_H} 1/(c\pi_i^H)}}{\frac{1}{N_{\mathcal{A}_F}} \sum_{i \in s_F} \sum_{A_k \in \mathcal{A}_F} y_{i,A_k}/\pi_i^F}. \tag{C.36}$$

If the probe alters for the frame population and the hidden population are the same, so that $\mathcal{A}_H = \mathcal{A}_F = \mathcal{A}$, and if the probe alters are randomly distributed in the frame population in the sense that

$$N_{\mathcal{A} \cap F} = N_{\mathcal{A}} \frac{N_F}{N}, \tag{C.37}$$

then we can reduce the constants in front of Equation C.36 to

$$\frac{\frac{N_F}{N_{\mathcal{A} \cap F}}}{\frac{1}{N_{\mathcal{A}}}} = \frac{\frac{N}{N_{\mathcal{A}}}}{\frac{1}{N_{\mathcal{A}}}} = N. \tag{C.38}$$

A29

In other words, when the probe alters for the frame and hidden population are the same, and when the probe alters are randomly distributed in the frame population, all of the factors involving the size of $\mathcal{A}$ drop out. This fact allows researchers to use groups defined by first names (e.g., people named Michael) in the probe alters $\mathcal{A}$, even if the size of these groups is not known, as long as it is reasonable to assume that $\mathcal{A}$ satisfies Equation C.37 (c.f., Salganik et al. (2011a)).

C.6.2 Estimating the true positive rate, τ_F

We can combine our estimator for $\bar{v}_{H,F}$ (Result C.2) and our estimator for $\bar{d}_{H,F}$ (Result C.5) to estimate the true positive rate τ_F.

Result C.7 *The estimator*

$$\widehat{\tau}_F = \frac{\widehat{\bar{v}}_{H,F}}{\widehat{\bar{d}}_{H,F}} \tag{C.39}$$

is consistent and essentially unbiased for τ_F if $\widehat{\bar{v}}_{H,F}$ is a consistent and essentially unbiased estimator of $\bar{v}_{H,F}$ and if $\widehat{\bar{d}}_{H,F}$ is a consistent and essentially unbiased estimator of $\bar{d}_{H,F}$.

Proof: This follows directly from the properties of a compound ratio estimator (Online Appendix E). ∎

More concretely, combing the estimator in Result C.2 and Result C.5 yields an estimator for $\widehat{\tau}_F$ with the following form:

$$\widehat{\tau}_F = \frac{\sum_{i \in s_H} \widetilde{v}_{i,\mathcal{A}_H}/(c\pi_i)}{\sum_{i \in s_H} y_{i,\mathcal{A}_H}/(c\pi_i)}. \tag{C.40}$$

All of the factors involving the size of $\mathcal{A}$ drop out of Equation C.40. This fact allows researchers to use groups defined by first names (e.g., people named Michael) in the

probe alters $\mathcal{A}$, even if the size of these groups is not known (c.f., Salganik et al. (2011b)).

C.7 Estimating the size of the hidden population, N_H

We now make use of all of the results for the individual terms we derived above to present four different estimators for the size of the hidden population, N_H.

Result C.8 *The generalized scale-up estimator given by*

$$\widehat{N}_H = \frac{\widehat{\overline{y}}_{F,H}}{\widehat{\overline{v}}_{H,F}} \tag{C.41}$$

is consistent and essentially unbiased for N_H if there are no false positive reports, if $\widehat{\overline{y}}_{F,H}$ is consistent and unbiased for $\overline{y}_{F,H}$, and if $\widehat{\overline{v}}_{H,F}$ is consistent and essentially unbiased for $\overline{v}_{H,F}$.

Proof: From the properties of a compound ratio estimator, we know that our estimator is consistent and essentially unbiased for $\overline{y}_{F,H}/\overline{v}_{H,F}$ (Appendix E). By the argument in the main text given in Section 2, leading to Equation 5, this quantity is equal to N_H. ∎

Result C.9 *The adjusted basic scale-up estimator given by*

$$\widehat{N}_H = \frac{\widehat{\overline{y}}_{F,H}}{\widehat{\overline{d}}_{U,F}} \; \frac{1}{\widehat{\phi}_F} \; \frac{1}{\widehat{\delta}_F} \; \frac{1}{\widehat{\tau}_F} \tag{C.42}$$

is consistent and essentially unbiased for N_H if there are no false positive reports, and if each of the individual estimators is consistent and essentially unbiased.

Proof: From the results in Online Appendix E, we know that this compound ratio estimator will be consistent and essentially unbiased for $\overline{y}_{F,H}/(\overline{d}_{U,F} \; \phi_F \; \delta_F \; \tau_F)$. The

denominator is $\bar{v}_{H,F}$ by construction, leaving us with $y_{F,H}/\bar{v}_{H,F}$. By the argument in the main text given in Section 2, leading to Equation 5, this quantity is equal to N_H. ∎

Result C.10 *The adjusted scale-up estimator*

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{F,F}} \frac{1}{\widehat{\delta}_F} \frac{1}{\widehat{\tau}_F} \tag{C.43}$$

is consistent and essentially unbiased for N_H if there are no false positives, and if each of the individual estimators is consistent and essentially unbiased.

Proof: From the results in Online Appendix E, we know that this compound ratio estimator will be consistent and essentially unbiased for $y_{F,H}/(\bar{d}_{F,F}\, \delta_F\, \tau_F)$. The denominator is $\bar{v}_{H,F}$ by construction, leaving us with $y_{F,H}/\bar{v}_{H,F}$. By the argument in the main text given in Section 2, leading to Equation 5, this quantity is equal to N_H. ∎

Result C.11 *The adjusted scale-up estimator*

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{F,F}} \frac{1}{\widehat{\delta}_F} \frac{1}{\widehat{\tau}_F} \widehat{\eta}_F \tag{C.44}$$

is consistent and essentially unbiased for N_H if each of the individual estimators is consistent and essentially unbiased.

Proof: From the results in Online Appendix E, we know that this compound ratio estimator will be consistent and essentially unbiased for $(y_{F,H}\, \eta_F)/(\bar{d}_{F,F}\, \delta_F\, \tau_F)$. The numerator is $y_{F,H}^{+}$ by construction and the product of the denominators is $\bar{v}_{H,F}$ by construction, leaving us with $y_{F,H}^{+}/\bar{v}_{H,F}$. By the argument in Online Appendix A this quantity is equal to N_H. ∎

D Sensitivity analysis

All of the estimators that we propose require that specific conditions hold in order to produce consistent and essentially unbiased estimates. These conditions can be divided into four groups: (1) survey construction, (2) reporting behavior, (3) network structure, and (4) sampling. In many practical settings, we expect that researchers may not be confident that these conditions hold perfectly. Therefore, in this appendix, we derive results that enable researchers to assess the sensitivity of their estimates to violations of all four types of conditions. First, in Section D.1, we develop a results that help researchers assess sensitivity to survey construction, reporting, and network structure; then, in Section D.2, we turn to results that help researchers assess sensitivity to sampling problems. Finally, in Section D.3, we combine all of the sensitivity results to derive expressions that enable researchers to conduct sensitivity analyses that simultaneously account for all of the conditions.

D.1 Sensitivity to non-sampling conditions: survey construction, reporting behavior, and network structure

Most estimators that we consider depend on conditions related to survey construction (for example, choosing the probe alters for the known population method) and to reporting (for example, the assumption that respondents make accurate aggregate reports about the probe alters); furthermore, the basic scale-up estimator is sensitive to conditions about network structure (for example, the relative size of hidden population and frame population members' personal networks). In this section, we develop sensitivity results for these nonsampling conditions. First, Result D.1 shows how one of these estimators ($\widehat{v}_{H,F}$) is impacted by violations of the conditions it depends upon. Next, using Result D.1 as a template, Table D.1 provides similar expressions for all

of the estimators we discuss in the main text.

Result D.1 *Suppose that $\widehat{N}_{\mathcal{A}\cap F}$, the researcher's estimate of $N_{\mathcal{A}\cap F}$, is incorrect, so that $\widehat{N}_{\mathcal{A}\cap F} = c_1 \cdot N_{\mathcal{A}\cap F}$. Suppose also that the reporting condition (Equation C.5) of Result C.2 is incorrect, so that $\tilde{v}_{H,\mathcal{A}\cap F} = c_2 \cdot v_{H,\mathcal{A}\cap F}$. Finally, suppose that the probe alter condition is incorrect, so that $\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = c_3 \cdot \frac{v_{H,F}}{N_F}$. Call the estimator under these imperfect conditions $\widehat{\tilde{v}}^{\star}_{H,F}$. Then $\widehat{\tilde{v}}^{\star}_{H,F}$ is consistent and essentially unbiased for $\frac{c_3\, c_2}{c_1} \bar{v}_{H,F}$ instead of $\bar{v}_{H,F}$.*

    Proof: Under the assumptions listed above, we can write the new estimator as

$$\widehat{\tilde{v}}^{\star}_{F,H} = \frac{1}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{\sum_{i\in s_H} \sum_j \tilde{v}_{i,A_j\cap F}/(c\pi_i)}{\sum_{i\in s_H} 1/(c\pi_i)}. \tag{D.1}$$

We follow the same steps as the proof of Result C.2, but each time we use one of our assumptions, the associated error is carried with it. So our estimator $\widehat{\tilde{v}}^{\star}_{F,H}$ is consistent and essentially unbiased for

$$\frac{1}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{\tilde{v}_{H,\mathcal{A}\cap F}}{N_H} = \frac{c_2}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{v_{H,\mathcal{A}\cap F}}{N_H} = \frac{c_3\, c_2}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{v_{H,F}}{N_H}. \tag{D.2}$$

In words, the estimand is now incorrect by $\frac{c_3\, c_2}{c_1}$. Since $\widehat{\tilde{v}}_{F,H}$ is consistent and essentialy unbiased for $\bar{v}_{F,H}$, we conclude that $\widehat{\tilde{v}}^{\star}_{F,H}$ is consistent and essentially unbiased for $\frac{c_3\, c_2}{c_1} \bar{v}_{F,H}$. Note that if the assumptions needed for Result C.2 hold, then $c_1 = 1$, $c_2 = 1$, and $c_3 = 1$, giving us the original result. ∎

    Table D.1 shows results analogous to Result D.1 for all of the estimators we propose. We do not prove each one individually, since the derivations all follow the pattern of Result D.1 very closely. Researchers who wish to understand the how their estimates are affected by the assumptions they make can use Table D.1 to conduct a sensitivity analysis. Note that any problems with the sampling design could result in

problems with the estimates that are not captured by the results in Table D.1. These sampling problems are the subject of the next section.

D.2 Sensitivity to sampling problems

All of the estimators we discuss throughout this paper rely upon assumptions about the sampling procedure that researchers use to obtain their data. In this section, we develop sensitivity results that enable researchers to assess how violations of these sampling assumptions will impact the resulting estimates. First, we investigate the sensitivity of the estimator $\widehat{y}_{F,H}$ from a probability sample (Online Appendix B.1), and, next, we investigate the estimator $\widehat{\bar{\bar{v}}}_{H,\mathcal{A}\cap F}$ from relative probability sample (Online Appendix C.1).

For both estimators, we investigate how estimates are affected by differences between the inclusion probabilities that researchers use to analyze their data and the true inclusion probabilities that come from the sampling mechanism. These problems could arise if the sampling design is not perfectly executed, or if there is a problem with the information underlying the sampling design.

D.2.1 Probability samples

First, we must define *imperfect sampling weights*.

Imperfect sampling weights. Suppose a researcher obtains a probability sample s_F from the frame population F (Online Appendix B.1). Let I_i be the random variable that assumes the value 1 when unit $i \in F$ is included in the sample s_F, and 0 otherwise. Let $\pi_i = \mathbb{E}[I_i]$ be the true probability of inclusion for unit $i \in F$, and let $w_i = \frac{1}{\pi_i}$ be the corresponding design weight for unit i. We say that researchers have *imperfect sampling weights* when researchers use imperfect estimates of the inclusion

Estimator	Imperfect assumptions	Effective estimand
$\widehat{\bar{d}}_{F,F}$ (Result B.3)	(i) $\widehat{N}_{\mathcal{A}} = c_1 \, N_{\mathcal{A}}$	$\frac{c_2 \, c_3}{c_1} \, \bar{d}_{F,F}$
	(ii) $\bar{d}_{\mathcal{A},F} = c_2 \, \bar{d}_{F,F}$	
	(iii) $y_{F,\mathcal{A}} = c_3 \, d_{F,\mathcal{A}}$	
$\widehat{\bar{d}}_{U,F}$ (Result B.4)	(i) $\widehat{N}_{\mathcal{A}} = c_1 \, N_{\mathcal{A}}$	$\frac{c_2 \, c_3}{c_1} \, \bar{d}_{U,F}$
	(ii) $\bar{d}_{\mathcal{A},F} = c_2 \, \bar{d}_{U,F}$	
	(iii) $y_{F,\mathcal{A}} = c_3 \, d_{F,\mathcal{A}}$	
$\widehat{\phi}_F$ (Result B.6)	(i) $\widehat{\bar{d}}_{F,F} \rightsquigarrow c_1 \, \bar{d}_{F,F}$	$\frac{c_1}{c_2} \, \phi_F$
	(ii) $\widehat{\bar{d}}_{U,F} \rightsquigarrow c_2 \, \bar{d}_{U,F}$	
$\widehat{\bar{v}}_{H,F}$ (Result C.2)	(i) $\widehat{N}_{\mathcal{A} \cap F} = c_1 \, N_{\mathcal{A} \cap F}$	$\frac{c_3 \, c_2}{c_1} \, \bar{v}_{H,F}$
	(ii) $\tilde{v}_{H,\mathcal{A} \cap F} = c_2 \, v_{H,\mathcal{A} \cap F}$	
	(iii) $\frac{v_{H,\mathcal{A} \cap F}}{N_{\mathcal{A} \cap F}} = c_3 \, \frac{v_{H,F}}{N_F}$	
$\widehat{\delta}_F$ (Result C.6)	(i) $\widehat{\bar{d}}_{H,F} \rightsquigarrow c_1 \, \bar{d}_{H,F}$	$\frac{c_1}{c_2} \, \delta_F$
	(ii) $\widehat{\bar{d}}_{F,F} \rightsquigarrow c_2 \, \bar{d}_{F,F}$	
$\widehat{\tau}_F$ (Result C.7)	(i) $\widehat{\bar{v}}_{H,F} \rightsquigarrow c_1 \, \bar{v}_{H,F}$	$\frac{c_1}{c_2} \, \tau_F$
	(ii) $\widehat{\bar{d}}_{H,F} \rightsquigarrow c_2 \, \bar{d}_{H,F}$	
$\widehat{N}_H$ (Result C.8)	(i) $\widehat{\bar{v}}_{H,F} \rightsquigarrow c_1 \, \bar{v}_{H,F}$	$\frac{1}{c_1} \, N_H$
$\widehat{N}_H$ (Result C.10)	(i) $\widehat{\bar{d}}_{F,F} \rightsquigarrow c_1 \, \bar{d}_{F,F}$	$\frac{1}{c_1 \, c_2 \, c_3} \, N_H$
	(ii) $\widehat{\delta}_F \rightsquigarrow c_2 \, \delta_F$	
	(iii) $\widehat{\tau}_F \rightsquigarrow c_3 \, \tau_F$	

Table D.1: Sensitivity of estimators to nonsampling assumptions. The first column lists the most important estimators we discuss in the main text and appendixes. The consistency and approximate unbiasedness of each estimator relies upon nonsampling conditions being satisfied. These conditions are given in the second column, with a modification: we add a constant to each condition; if the constant is 1, then the original condition is satisfied. The estimand is then effectively changed to the quantity listed in the third column. (NB: we use the symbol $\rightsquigarrow$ as a shorthand for 'is consistent and essentially unbiased for'.) For example, the first row shows $\widehat{\bar{d}}_{F,F}$ and the three conditions that the estimator in Result B.3 relies upon. Suppose that the first and third hold, so that $c_1 = 1$ and $c_3 = 1$, but that the second does not; instead, the probe alters $\mathcal{A}$ have been chosen so that $\bar{d}_{\mathcal{A},F} = 1.1 \, \bar{d}_{F,F}$. Then $c_2 = 1.1$. Looking at the third column, we can see that our estimator will then be consistent and essentially unbiased for $1.1 \times \bar{d}_{F,F}$ instead of $\bar{d}_{F,F}$.

probabilities π_i' and the corresponding design weights $w_i' = \frac{1}{\pi_i'}$. Note that we assume that both the true and the imperfect weights satisfy $\pi_i > 0$ and $\pi_i' > 0$ for all i.

The first result, Result D.2, concerns researchers who obtain a probability sample, but who estimate $y_{F,H}$ imperfect sampling weights.

Result D.2 shows the impact that imperfect sampling weights have on estimates of $y_{F,H}$ from a probability sample.

Result D.2 *Suppose researchers have obtained a probability sample s_F, but that they have imperfect sampling weights. Call the imperfect sampling weights $w_i' = \frac{1}{\pi_i'}$, call the true weights $w_i = \frac{1}{\pi_i}$, and define $\epsilon_i = \frac{w_i'}{w_i} = \frac{\pi_i}{\pi_i'}$. Call $\widehat{y}_{F,H}' = \sum_{i \in s_F} y_{i,H} w_i'$ the estimator for $y_{F,H}$ using the imperfect weights. Then*

$$bias[\widehat{y}_{F,H}'] = N_F[\bar{y}_{F,H}(\bar{\epsilon} - 1) + cov_F(y_{i,H}, \epsilon_i)]. \tag{D.3}$$

where $\bar{\epsilon} = \frac{1}{N_F} \sum_{i \in F} \epsilon_i$, and $cov_F(\cdot, \cdot)$ is the finite population unit covariance.

Proof: We can write the bias in the estimator $\widehat{y}_{F,H}'$ as

$$bias[\widehat{y}_{F,H}'] = \mathbb{E}[\widehat{y}_{F,H}'] - y_{F,H} \tag{D.4}$$

$$= \sum_{i \in F} w_i' \mathbb{E}[I_i] y_{i,H} - \sum_{i \in F} y_{i,H} \tag{D.5}$$

$$= \sum_{i \in F} \frac{\pi_i}{\pi_i'} y_{i,H} - \sum_{i \in F} y_{i,H} \tag{D.6}$$

$$= \sum_{i \in F} y_{i,H} (\epsilon_i - 1). \tag{D.7}$$

Now, recall that, for any a_i and b_i,

$$\sum_{i \in F} a_i \, b_i = N_F \left[\bar{a}\bar{b} + cov_F(a_i, b_i) \right], \tag{D.8}$$

A37

where $\bar{a}$ and $\bar{b}$ are the mean values of a and b, and $\text{cov}_F(a_i, b_i)$ is the finite population unit covariance between a_i and b_i. Applying this fact to Equation D.7, we have

$$\text{bias}[\widehat{y}'_{F,H}] = \sum_{i \in F} y_{i,H}(\epsilon_i - 1) \tag{D.9}$$

$$= N_F \left[\bar{y}_{F,H}(\overline{\epsilon - 1}) + \text{cov}_F(y_{i,H}, \epsilon_i - 1) \right], \tag{D.10}$$

$$= N_F \left[\bar{y}_{F,H}(\bar{\epsilon} - 1) + \text{cov}_F(y_{i,H}, \epsilon_i) \right]. \tag{D.11}$$

$$\blacksquare$$

In order to further understand Result D.2, it is helpful to use the identity

$$\text{cov}_F(y_{i,H}, \epsilon_i) = \text{cor}_F(y_{i,H}, \epsilon_i) \, \text{sd}_F(y_{i,H}) \, \text{sd}_F(\epsilon_i), \tag{D.12}$$

where $\text{sd}_F(\cdot)$ is the unit finite-population standard deviation, and $\text{cor}_F(y_{i,H}, \epsilon_i)$ is the correlation between the $y_{i,H}$ and the ϵ_i. Substituting this identity into Equation D.3 yields

$$\text{bias}[\widehat{y}'_{F,H}] = N_F \left[\bar{y}_{F,H}(\bar{\epsilon} - 1) + \text{cor}_F(y_{i,H}, \epsilon_i) \, \text{sd}_F(y_{i,H}) \, \text{sd}_F(\epsilon_i) \right]. \tag{D.13}$$

Equation D.13 provides a qualitative understanding for when errors in the weights will be more or less problematic. Several of the terms will typically be beyond the researcher's control: N_F, $\bar{y}_{F,H}$, and $\text{sd}_F(y_{i,H})$ are all properties of the population being studied. The remaining terms, however, are related to errors in the weights. The $\bar{\epsilon} - 1$ term says that the bias will be minimized when $\frac{\pi_i}{\pi'_i}$ is close to 1 for all i. The $\text{sd}_F(\epsilon_i)$ term says that the bias will be reduced when the $\frac{\pi_i}{\pi'_i}$ values have low variance—i.e., when deviations from the correct weight value do not vary between units. And, finally, the $\text{cor}_F(y_{i,H}, \epsilon_i)$ term says that bias is lower in absolute value

when errors in the weights are not related to the quantity being measured.

As we will see, it will be helpful to re-express Result D.2 in one additional way. This re-expression highlights the similarities between several of the sensitivity results we derive in this section. This final version of Result D.2 relies upon a quantity, K_F, which serves as an index for the amount of error in the weights. First, note that $\mathrm{sd}_F(\epsilon_i) = \bar{\epsilon}\, \mathrm{cv}_F(\epsilon_i)$, where $\mathrm{cv}(\epsilon_i)$ is the coefficient of variation (i.e., the standard deviation divided by the mean), and, likewise, $\mathrm{sd}_F(y_{i,H}) = \bar{y}_{F,H}\, \mathrm{cv}_F(y_{i,H})$. Now, define the index $K_F = \mathrm{cor}_F(y_{i,H})\mathrm{cv}_F(y_{i,H})\mathrm{cv}_F(\epsilon_i)$. K_F can be positive, negative, or zero. When the weights are exactly correct (i.e., $\pi'_i = \pi_i$ for all i), $K_F = 0$; on the other hand, when there are large errors in the weights, K_F will be far from 0.[11]

Using K_F enables us to re-write Equation D.13 as

$$\mathrm{bias}[\widehat{y}'_{F,H}] = \mathbb{E}[\widehat{y}'_{F,H}] - y_{F,H} = N_F\left[\bar{y}_{F,H}(\bar{\epsilon} - 1) + \bar{y}_{F,H}\, \bar{\epsilon}\, K_F\right] \tag{D.14}$$

$$\iff \mathbb{E}[\widehat{y}'_{F,H}] = y_{F,H} + y_{F,H}(\bar{\epsilon} - 1) + y_{F,H}\, \bar{\epsilon}\, K_F \tag{D.15}$$

$$= y_{F,H}\, \bar{\epsilon}\, (1 + K_F) \tag{D.16}$$

Therefore, Result D.2 directly implies Corollary D.3.

Corollary D.3 *From Result D.2, we also have*

$$\widehat{y}'_{F,H} \to y_{F,H} \cdot \bar{\epsilon} \cdot (1 + K_F), \tag{D.17}$$

where $\to$ means 'is consistent and unbiased for,' and $K_F = \mathrm{cor}_F(y_{i,H}, \epsilon_i)\mathrm{cv}_F(y_{i,H})\mathrm{cv}_F(\epsilon_i)$.

---

[11] K_F is similar to the identity in Equation D.12, except that it involves the coefficient of variation instead of the standard deviation. This is convenient, because the coefficient of variation is unitless, making K_F unitless (i.e., it does not depend on the scale of the particular quantity being estimated).

D.2.2 Relative probability samples

We now turn to the estimator for the average visibility of hidden population members $(\bar{v}_{H,F})$. This estimator turns out to be more complex than the estimator we investigated in the previous section. In order to derive complete sensitivity results for the estimator $\widehat{\bar{v}}_{H,F}$, it is useful to first understand the sensitivity of the estimator for the average reported visibility of hidden population members to the probe alters, $\bar{\bar{v}}_{H,\mathcal{A}\cap F}$ (see Online Appendix C.4). $\widehat{\bar{\bar{v}}}_{H,\mathcal{A}\cap F}$ turns out to be the only part of estimating $\bar{v}_{H,F}$ that is sensitive to imperfections in sampling.

Since visibility will typically be estimated from a relative probability sample, Result D.4 concerns researchers who obtain a relative probability sample but make estimates of $\bar{\bar{v}}_{H,\mathcal{A}\cap F}$ using what we call *imperfect relative sampling weights*. We define imperfect relative sampling weights precisely in the next paragraph, and then we present Result D.4.

Imperfect relative sampling weights. Suppose a researcher obtains a relative probability sample s_H from a population H (Online Appendix C.1). Let I_i be the random variable that assumes the value 1 when unit $i \in H$ is included in the sample s_H, and 0 otherwise, and let $\pi_i = \mathbb{E}[I_i]$. We say that researchers have *imperfect relative sampling weights* when the true π_i are not known and, instead, researchers use imperfect estimates of the relative inclusion probabilities $c'\pi_i'$, where c' is some unknown constant, and the corresponding imperfect relative probability design weights $w_i' = \frac{1}{c'\pi_i'}$. Note that we assume that both the true and the imperfect weights satisfy $\pi_i > 0$ and $\pi_i' > 0$ for all i.

Result D.4 *Suppose researchers have obtained a relative probability sample s_H, but that the researchers have imperfect relative sampling weights. Call the imperfect sam-*

pling weights $w_i' = \frac{1}{c'\pi_i'}$, and define $\epsilon_i = \frac{\pi_i}{\pi_i'}$. Call the estimator for $\bar{\tilde{v}}_{H,\mathcal{A}\cap F}$ (the reported visibilities; see Section C.2) using the imperfect relative sampling weights $\bar{\widehat{\tilde{v}}}'_{H,\mathcal{A}\cap F}$:

$$\bar{\widehat{\tilde{v}}}'_{H,\mathcal{A}\cap F} = \frac{\sum_{i\in s_H}\sum_j \tilde{v}_{i,A_j\cap F}/(c'\pi_i')}{\sum_{i\in s_H}1/(c'\pi_i')}. \tag{D.18}$$

Then

$$bias(\bar{\widehat{\tilde{v}}}'_{H,\mathcal{A}\cap F}) = \underbrace{\frac{cov_H(\tilde{v}_{i,\mathcal{A}\cap F},\epsilon_i)}{\bar{\epsilon}}}_{\text{bias from incorrect weights}} - \underbrace{\frac{cov(\bar{\widehat{\tilde{v}}}'_{H,\mathcal{A}\cap F},\widehat{N}_H')}{N_H'}}_{\text{bias from ratio estimator}}, \tag{D.19}$$

where $\bar{\epsilon} = \frac{1}{N_H}\sum_{i\in H}\epsilon_i$; $\widehat{N}_H' = \sum_{i\in s_H}w_i'$; $N_H' = \frac{1}{c'}\sum_{i\in H}\epsilon_i$; $cov(\cdot)$ is the covariance taken with respect to the sampling distribution; and $cov_H(\cdot)$ is the finite population unit covariance among hidden population members.

Proof: The classic result of Hartley and Ross (1954) (see also Sarndal et al., 1992, Result 5.6.1) shows that the expected value of the estimator in Equation D.18 is

$$\mathbb{E}[\bar{\widehat{\tilde{v}}}'_{H,\mathcal{A}\cap F}] = \frac{\mathbb{E}[\sum_{i\in s_H}w_i'\tilde{v}_{i,\mathcal{A}\cap F}]}{\mathbb{E}[\sum_{i\in s_H}w_i']} - \frac{cov(\bar{\widehat{\tilde{v}}}'_{H,\mathcal{A}\cap F},\widehat{N}_H')}{\mathbb{E}[\sum_{i\in s_H}w_i']}, \tag{D.20}$$

where the covariance is taken with respect to the sampling distribution. Now, note that

$$\mathbb{E}[\sum_{i\in s_H}w_i'] = \mathbb{E}[\sum_{i\in H}I_iw_i'] = \mathbb{E}[\sum_{i\in H}I_i\frac{1}{c'\pi_i'}] = \sum_{i\in H}\frac{\pi_i}{c'\pi_i'} = \frac{1}{c'}\sum_{i\in H}\epsilon_i = N_H'. \tag{D.21}$$

Therefore, we substitute N'_H for the denominator of the second term of Equation D.20, which produces

$$\mathbb{E}[\widehat{\tilde{v}}'_{H,\mathcal{A}\cap F}] = \frac{\mathbb{E}[\sum_{i\in s_H} w'_i \tilde{v}_{i,\mathcal{A}\cap F}]}{\mathbb{E}[\sum_{i\in s_H} w'_i]} - \frac{\text{cov}(\widehat{\tilde{v}}'_{H,\mathcal{A}\cap F}, \widehat{N}'_H)}{N'_H}. \tag{D.22}$$

We do not substitute N'_H for the denominator of the first term, because we will now see that we can instead produce a simpler expression.

The remainder of the proof focuses on the first term. Note that

$$\mathbb{E}[\sum_{i\in s_H} w'_i \tilde{v}_{i,\mathcal{A}\cap F}] = \mathbb{E}[\sum_{i\in H} I_i w'_i \tilde{v}_{i,\mathcal{A}\cap F}] = \mathbb{E}[\sum_{i\in H} I_i \frac{1}{c'\pi'_i} \tilde{v}_{i,\mathcal{A}\cap F}] = \sum_{i\in H} \frac{\pi_i}{c'\pi'_i} \tilde{v}_{i,\mathcal{A}\cap F} = \frac{1}{c'} \sum_{i\in H} \epsilon_i \tilde{v}_{i,\mathcal{A}\cap F}, \tag{D.23}$$

and also that

$$\mathbb{E}[\sum_{i\in s_H} w'_i] = \mathbb{E}[\sum_{i\in H} I_i w'_i] = \mathbb{E}[\sum_{i\in H} I_i \frac{1}{c'\pi'_i}] = \sum_{i\in H} \frac{\pi_i}{c'\pi'_i} = \frac{1}{c'} \sum_{i\in H} \epsilon_i. \tag{D.24}$$

A42

The bias of the estimator in Equation D.18 is therefore

$$\text{bias}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}) = \mathbb{E}[\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}] - \bar{\tilde{v}}_{H,\mathcal{A}\cap F} \tag{D.25}$$

$$= \frac{\sum_{i\in H}\epsilon_i\tilde{v}_{i,\mathcal{A}\cap F}}{\sum_{i\in H}\epsilon_i} - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F},\widehat{N}'_H)}{N'_H} - \frac{\sum_{i\in H}\tilde{v}_{i,\mathcal{A}\cap F}}{N_H} \tag{D.26}$$

$$= \left(\frac{\sum_{i\in H}\epsilon_i\tilde{v}_{i,\mathcal{A}\cap F}}{\sum_{i\in H}\epsilon_i} - \frac{\sum_{i\in H}\tilde{v}_{i,\mathcal{A}\cap F}}{N_H}\right) - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F},\widehat{N}'_H)}{N'_H} \tag{D.27}$$

$$= \left(\frac{\sum_{i\in H}\epsilon_i\tilde{v}_{i,\mathcal{A}\cap F} - \frac{1}{N_H}\sum_{i\in H}\tilde{v}_{i,\mathcal{A}\cap F}\sum_{i\in H}\epsilon_i}{\sum_{i\in H}\epsilon_i}\right) - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F},\widehat{N}'_H)}{N'_H} \tag{D.28}$$

$$= \left(\frac{\text{cov}_H(\tilde{v}_{i,\mathcal{A}\cap F},\epsilon_i)}{\bar{\epsilon}}\right) - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F},\widehat{N}'_H)}{N'_H}, \tag{D.29}$$

where $\text{cov}_H(\cdot,\cdot)$ is the finite-population unit variance among hidden population members.

∎

Result D.4 shows that the bias in the estimator $\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}$ with imperfect relative probability weights is the sum of two terms: one term that arises due to intrinsic bias in any ratio estimator, and one term that arises due to differences between the imperfect weights and the true weights. A large literature shows that, in many practical situations, the intrinsic bias in a ratio estimator will tend to be very small (see, for example, Online Appendix E and also Sarndal et al. (1992, Chap. 5)). When this intrinsic ratio bias is negligible, Result D.4 shows that the bias in the estimator for $\bar{\tilde{v}}_{H,\mathcal{A}\cap F}$ with imperfect weights can be approximated by

$$\text{bias}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}) \approx \frac{\text{cov}_H(\tilde{v}_{i,\mathcal{A}\cap F},\epsilon_i)}{\bar{\epsilon}}. \tag{D.30}$$

Similar to the discussion of Result D.2, we can obtain additional insight into Equation D.30 by using the fact that $\text{cov}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i) = \text{cor}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)\,\text{sd}_H(\tilde{v}_{i,\mathcal{A}\cap F})\,\text{sd}_H(\epsilon_i)$, where $\text{sd}_H(\cdot)$ is the unit finite-population standard deviation, and $\text{cor}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)$ is the correlation between the y_i and ϵ_i. Substituting this identity into Equation D.30 yields

$$\text{bias}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}) \approx \text{cor}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)\,\text{sd}_H(\tilde{v}_{i,\mathcal{A}\cap F})\frac{\text{sd}_H(\epsilon_i)}{\bar{\epsilon}}. \tag{D.31}$$

Equation D.31 provides a qualitative understanding of factors contributing to bias due to imperfect relative sampling weights. One term, $\text{sd}_H(\tilde{v}_{i,\mathcal{A}\cap F})$, is a property of the population being studied and will typically be beyond the researcher's control. The other two terms are related to errors in the weights: first, the factor $\frac{\text{sd}_H(\epsilon_i)}{\bar{\epsilon}}$ is the coefficient of variation in the ϵ_i; it will be minimized when the standard deviation of the ϵ_i is small, relative to the mean; that is, it will be minimized when the errors in the weights are uniform. Second, the magnitude of $\text{cor}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)$ will be minimized when there is no relationship between the imperfections in the weights, ϵ_i, and the quantity of interest, $\tilde{v}_{i,\mathcal{A}\cap F}$.

Next, note that $\text{sd}_H(\tilde{v}_{i,\mathcal{A}\cap F}) = \bar{\tilde{v}}_{H,\mathcal{A}\cap F}\,\text{cv}_H(\tilde{v}_{i,\mathcal{A}\cap F})$, where $\text{cv}_H(\tilde{v}_{i,\mathcal{A}\cap F})$ is the coefficient of variation. Equation D.31 can therefore be re-arranged to yield

$$\text{bias}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}) \approx \bar{\tilde{v}}_{H,\mathcal{A}\cap F}\,K_H, \tag{D.32}$$

where we have defined $K_H = \text{cor}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)\text{cv}_H(\tilde{v}_{i,\mathcal{A}\cap F})\text{cv}_H(\epsilon_i)$ as an index for the amount of error in the imperfect weights.

Using the index K_H helps to clarify the meaning of the ϵ_i in Result D.4. It may seem unintuitive to define $\epsilon_i = \frac{\pi_i}{\pi'_i}$, since the result assumes that neither π_i or

π'_i is known. But, we note that the K_H in Expression D.32 is not impacted if ϵ_i are multiplied by a constant. Therefore, if researchers find it more natural to work with a version of ϵ_i that involves multiplying all of the π'_i or π_i by a constant, then Result D.4 still applies. For example, imagine that a researcher has sampled from the hidden population using respondent-driven sampling, and then makes estimates under the assumption that respondents' inclusion probabilities are proportional to their degrees ($\pi'_i \propto d_i$). This researcher might wonder how her estimate would be impacted if this sampling assumption was incorrect ($\pi'_i \not\propto d_i$). In this case, the researcher could then make the necessary assumptions and calculate K_H assuming that, for example, ($\pi'_i \propto d_i^0$), or ($\pi'_i \propto d_i^2$).

Finally, since $\mathbb{E}[\widehat{\bar{\tilde{v}}}_{H,\mathcal{A}\cap F}] = \text{bias}(\widehat{\bar{\tilde{v}}}_{H,\mathcal{A}\cap F}) + \bar{\tilde{v}}_{H,\mathcal{A}\cap F}$, we can conclude that

$$\mathbb{E}[\widehat{\bar{\tilde{v}}}_{H,\mathcal{A}\cap F}] \approx \bar{\tilde{v}}_{H,\mathcal{A}\cap F}(1 + K_H). \tag{D.33}$$

Therefore, Result D.2 directly implies Corollary D.3.

Corollary D.5 *From Result D.2, we also have*

$$\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F} \rightsquigarrow \bar{\tilde{v}}_{H,\mathcal{A}\cap F}(1 + K_H), \tag{D.34}$$

where $\rightsquigarrow$ means 'is consistent and essentially unbiased for,' and $K_H = cor_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)\, cv_H(\tilde{v}_{i,\mathcal{A}\cap F})\, cv_H(\epsilon_i)$ is an index for the amount of error in the imperfect relative sampling weights.

D.2.3 Summary and results for all estimators

Table D.2 uses K, the index for the magnitude of errors introduced by imperfect weights, to summarize the results of our investigation into the impact that imperfect sampling weights will have on three quantities that play a central role in the estimators

Quantity	Relevant results	Effective estimand under imperfect sampling
$\widehat{y}'_{F,\mathcal{A}} = \sum_{i \in s_F} y_{i,\mathcal{A}}/\pi'_i$	(i) $\widehat{\bar{d}}_{F,F}$ (Result B.3)	$y_{F,\mathcal{A}} \cdot \bar{\epsilon} \cdot [1 + K_{F_1}]$
	(ii) $\widehat{\bar{d}}_{U,F}$ (Result B.4)	
	(iii) $\widehat{\phi}_F$ (Result B.6)	
	(iv) $\widehat{\delta}_F$ (Result C.6)	
$\widehat{y}'_{F,H} = \sum_{i \in s_F} y_{i,H}/\pi'_i$	(i) $\widehat{y}_{F,H}$ (Result B.1)	$y_{F,H} \cdot \bar{\epsilon} \cdot [1 + K_{F_2}]$
$\widetilde{\widehat{\bar{v}}}'_{H,\mathcal{A}\cap F} = \frac{\sum_{i \in s_H} \tilde{v}_{i,\mathcal{A}\cap F}/(c'\pi'_i)}{\sum_{i \in s_H} 1/(c'\pi'_i)}$	(i) $\widehat{\bar{v}}_{H,F}$ (Result C.2)	$\bar{\tilde{v}}_{H,\mathcal{A}\cap F} \cdot [1 + K_H]$

Table D.2: Summary of estimators' sensitivity to imperfect sampling. Here, s_F is a probability sample, s_H is a relative probability sample, and the Ks are indices for the magnitude of errors in the imperfect weights; $K_{F_1} = \mathrm{cor}_F(\epsilon_i, y_{i,\mathcal{A}}) \, \mathrm{cv}_F(\epsilon_i) \, \mathrm{cv}_F(y_{i,\mathcal{A}})$; $K_{F_2} = \mathrm{cor}_F(\epsilon_i, y_{i,H}) \, \mathrm{cv}_F(\epsilon_i) \, \mathrm{cv}_F(y_{i,H})$; and $K_H = \mathrm{cor}_H(\epsilon_i, \tilde{v}_{i,\mathcal{A}\cap F}) \, \mathrm{cv}_H(\epsilon_i) \, \mathrm{cv}_H(\tilde{v}_{i,\mathcal{A}\cap F})$. When the weights are exactly correct, each K is equal to 0.

we consider throughout this paper: $\widehat{y}'_{F,\mathcal{A}}$, $\widehat{y}'_{F,H}$, and $\widetilde{\widehat{\bar{v}}}'_{H,\mathcal{A}\cap F}$. The results in Table D.2 show how the magnitude of the index K is directly related to the bias that results from imperfect sampling weights.

D.3   Combined sensitivity results

We now combine our analysis of sensitivity to reporting, network structure, and survey construction (Section D.2.1) and sensitivity to sampling problems (Section D.2.2) to derive results that describe the sensitivity of the generalized and the modified basic scale-up estimator to all of the conditions they rely upon. Roughly, what we show below is that the results about estimators' sensitivity to nonsampling conditions (such as survey construction and reporting) and results about estimators' sensitivity to sampling conditions combine naturally.

D.3.1 Generalized scale-up

In this section, we derive an expression for the sensitivity of the generalized scale-up estimator to all of the conditions it relies upon. First, we derive a combined sensitivity result for $\widehat{\bar{v}}_{H,F}$ (Result D.6). We then make use of the combined sensitivity result for $\widehat{\bar{v}}_{H,F}$ to derive a combined sensitivity result for the generalized scale-up estimator (Result D.7 and Corollary D.8).

Result D.6 *Suppose researchers have obtained a relative probability sample s_H to estimate $\bar{v}_{H,F}$, but that the researchers have imperfect relative sampling weights. Call the imperfect relative sampling weights $w_i'^H = \frac{1}{c' \pi_i'^H}$, call the true probabilities of inclusion π_i, and define $\epsilon_i^H = \frac{\pi_i^H}{\pi_i'^H}$. Call the estimator for $\bar{\bar{v}}_{H,\mathcal{A}\cap F}$ using the imperfect relative sampling weights $\widehat{\bar{\bar{v}}}'_{H,\mathcal{A}\cap F}$.*

Suppose also that the researcher's estimate of $N_{\mathcal{A}\cap F}$ is incorrect, so that $\widehat{N}_{\mathcal{A}\cap F} = c_1 \cdot N_{\mathcal{A}\cap F}$. Suppose that the reporting condition (Equation C.5) of Result C.2 is incorrect, so that $\tilde{v}_{H,\mathcal{A}\cap F} = c_2 \cdot v_{H,\mathcal{A}\cap F}$. Finally, suppose that the probe alter condition is incorrect, so that $\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = c_3 \cdot \frac{v_{H,F}}{N_F}$. Call the estimator for $\bar{v}_{H,F}$ under these imperfect conditions $\widehat{\bar{v}}'^{\star}_{H,F}$.

Then

$$\widehat{\bar{v}}'^{\star}_{H,F} \rightsquigarrow \bar{v}_{H,F} \; \frac{c_3 \, c_2}{c_1} \; (1 + K_H) \tag{D.35}$$

where $\rightsquigarrow$ means 'is consistent and essentially unbiased for', and $K_H = cor_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i^H) cv_H(\tilde{v}_{i,\mathcal{A}\cap F}) cv_H(\epsilon_i^H)$

Proof: First, we note that Corollary D.5 shows that

$$\widehat{\bar{\bar{v}}}'_{H,\mathcal{A}\cap F} \rightsquigarrow \bar{\bar{v}}_{H,\mathcal{A}\cap F}(1 + K_H) = \frac{\tilde{v}_{H,\mathcal{A}\cap F}}{N_H}(1 + K_H). \tag{D.36}$$

The remainder of the proof follows the argument from Results D.1 and C.2 very closely. Under the assumptions listed above, we can write the imperfect estimator $\widehat{\tilde{v}}_{H,F}^{\prime\star}$ as

$$\widehat{\tilde{v}}_{H,F}^{\prime\star} = \frac{1}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \widehat{\tilde{v}}_{H,\mathcal{A}\cap F}^{\prime} \tag{D.37}$$

We follow the same steps as the proof of Results C.2, but each time we use one of our assumptions, the associated error is carried with it. So our estimator $\widehat{\tilde{v}}_{H,F}^{\prime\star}$ is consistent and essentially unbiased for

$$\widehat{\tilde{v}}_{H,F}^{\prime\star} \rightsquigarrow (1 + K_H)\frac{1}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{\tilde{v}_{H,\mathcal{A}\cap F}}{N_H} \tag{D.38}$$

$$= (1 + K_H)\frac{c_2}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{v_{H,\mathcal{A}\cap F}}{N_H} \tag{D.39}$$

$$= (1 + K_H)\frac{c_3\, c_2}{c_1} \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{v_{H,F}}{N_H}. \tag{D.40}$$

In words, the estimand is now incorrect by $(1 + K_H)\frac{c_3\, c_2}{c_1}$. Since $\widehat{\tilde{v}}_{H,F}$ is consistent and essentialy unbiased for $\bar{v}_{H,F}$, we conclude that $\widehat{\tilde{v}}_{H,F}^{\prime\star}$ is consistent and essentially unbiased for $(1 + K_H)\frac{c_3\, c_2}{c_1}\bar{v}_{H,F}$. Note that if the conditions needed for Result C.2 hold, then $c_1 = 1$, $c_2 = 1$, $c_3 = 1$, and $K_H = 0$, then we are left with our original result for $\widehat{\tilde{v}}_{H,F}$ (Result C.2). ∎

Result D.7 *Suppose researchers have obtained a probability sample s_F to estimate $y_{F,H}$, but that the researchers have imperfect sampling weights. Call the imperfect sampling weights $w_i^{\prime F} = \frac{1}{\pi_i^{\prime F}}$, call the true weights $w_i^F = \frac{1}{c\pi_i^F}$, and define $\epsilon_i^F = \frac{\pi_i^F}{\pi_i^{\prime F}} = \frac{w_i^{\prime F}}{w_i^F}$. Call the estimator for $y_{F,H}$ under these imperfect conditions $y_{F,H}^{\prime}$.*

Suppose also researchers have also obtained a relative probability sample s_H to estimate $\bar{v}_{H,F}$ but that the researchers have imperfect relative sampling weights. Call the

imperfect relative sampling weights $w_i'^H = \frac{1}{c'\pi_i'^H}$, *call the true probabilities of inclusion*

π_i, *and define* $\epsilon_i^H = \frac{\pi_i^H}{\pi_i'^H}$. *Suppose also that the researcher's estimate of* $N_{\mathcal{A}\cap F}$ *is incorrect, so that* $\widehat{N}_{\mathcal{A}\cap F} = c_1 \cdot N_{\mathcal{A}\cap F}$. *Suppose that the reporting condition (Equation C.5) of Result C.2 is incorrect, so that* $\tilde{v}_{H,\mathcal{A}\cap F} = c_2 \cdot v_{H,\mathcal{A}\cap F}$. *Finally, suppose that the probe alter condition is incorrect, so that* $\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = c_3 \cdot \frac{v_{H,F}}{N_F}$. *Call the estimator for* $\bar{v}_{H,F}$ *under these imperfect conditions* $\widehat{\bar{v}}_{H,F}'^{\star}$.

Finally, suppose that there are false positive reports, so that $y_{F,H}^+ = \eta_F y_{F,H}$. *Let the generalized scale-up estimator for* N_H *in this situation be* $\widehat{N}_H'^{\star} = \frac{y_{F,H}'}{\widehat{\bar{v}}_{H,F}'^{\star}}$. *Then*

$$\widehat{N}_H'^{\star} \rightsquigarrow \frac{\bar{\epsilon}^F(1+K_{F_1})}{1+K_H} \frac{c_1}{c_3 \, c_2} \frac{1}{\eta_F} N_H, \tag{D.41}$$

where $\rightsquigarrow$ *means 'is consistent and essentially unbiased for';* $\bar{\epsilon}^F = \frac{1}{N_F}\sum_{i\in F}\epsilon_i^F$; $K_H = cor_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i^H)cv_H(\tilde{v}_{i,\mathcal{A}\cap F})cv_H(\epsilon_i^H)$; *and* $K_{F_1} = cor_F(y_{i,H}, \epsilon_i^F)cv_F(y_{i,H})cv_F(\epsilon_i^F)$.

Proof: The generalized scale-up estimator is formed from a ratio of estimators, one in the numerator ($\widehat{y}_{F,H}$) and one in the denominator ($\widehat{\bar{v}}_{H,F}$). We have already derived results for each of the numerator and the denominator separately; our approach will therefore be to combine them. We must account for the fact that, in addition to the assumptions required for the estimator of the numerator and the denominator, the generalized scale-up estimator also requires the additional condition that there are no false positive reports.

We begin with the denominator, $\widehat{\bar{v}}_{H,F}$. Result D.6 shows that

$$\widehat{\bar{v}}_{H,F}'^{\star} \rightsquigarrow \bar{v}_{H,F} \frac{c_3 \, c_2}{c_1}(1+K_H), \tag{D.42}$$

where $K_H = cor_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i^H)cv(\tilde{v}_{i,\mathcal{A}\cap F})cv(\epsilon_i^H)$. Thus, Expression D.42 shows the sensitivity of the denominator of the generalized scale-up estimator to violations of

all of the conditions it relies upon.

Turning now to the numerator of the generalized scale-up estimator, Corollary D.3 shows that

$$\widehat{y}'_{F,H} \rightsquigarrow y_{F,H} \cdot \bar{\epsilon} \cdot (1 + K_{F_1}), \tag{D.43}$$

where $K_{F_1} = \mathrm{cor}_F(y_{i,H}, \epsilon_i^F)\mathrm{cv}_F(y_{i,H})\mathrm{cv}_F(\epsilon_i^F)$. Thus, Expression D.43 shows sensitivity of the numerator of the generalized scale-up estimator to violations of all of the conditions it relies upon.

Using the fact that a ratio estimator is consistent and essentially unbiased for the ratio of the estimand of its numerator and denominator (see Online Appendix E and Sarndal et al. (1992, chap. 5)), we therefore have

$$\widehat{N}_H'^\star \rightsquigarrow \frac{\bar{\epsilon}^F(1 + K_{F_1})}{1 + K_H} \frac{c_1}{c_3}\frac{y_{F,H}}{c_2}\frac{y_{F,H}}{\bar{v}_{H,F}}. \tag{D.44}$$

Finally, by definition we have $y_{F,H} = y_{F,H}^+/\eta_F$, which we can substitute into Expression D.44 to produce

$$\widehat{N}_H'^\star \rightsquigarrow \frac{\bar{\epsilon}^F(1 + K_{F_1})}{1 + K_H} \frac{c_1}{c_3}\frac{c_1}{c_2}\frac{y_{F,H}^+}{\eta_F}\frac{y_{F,H}^+}{\bar{v}_{H,F}}. \tag{D.45}$$

By the argument in Section 2 and Appendix A, $N_H = y_{F,H}^+/\bar{v}_{H,F}$. Substituting N_H for $y_{F,H}^+/\bar{v}_{H,F}$ in the expression above completes the proof. ■

Corollary D.8 *From Result D.7, it follows that, for the generalized scale-up estima-*

tor,

$$\widehat{N}_H'^\star \cdot \underbrace{\frac{1 + K_H}{\bar{\epsilon}^F(1 + K_{F_1})}}_{\substack{sampling \\ conditions}} \cdot \underbrace{\frac{c_3 \ c_2}{c_1}}_{\substack{visibility \\ estimator \\ conditions}} \cdot \underbrace{\eta_F}_{\substack{no \ false \\ positives \\ condition}} \rightsquigarrow N_H. \qquad (D.46)$$

Researchers who wish to conduct a sensitivity analysis for estimates made using the generalized scale-up method can therefore (1) assume values or ranges of values for K_H, $\bar{\epsilon}^F$, K_{F_1}, c_1, c_2, c_3, and η_F and (2) use Corollary D.8 to determine the resulting values of N_H. Thus, researchers can use this approach to explore the sensitivity of their estimates to all of the assumptions they had to make.

D.3.2 Modified basic scale-up

In this section, we develop an expression for the sensitivity of the modified basic scale-up estimator to all of the conditions it relies upon. First, we derive a combined sensitivity result for $\widehat{\bar{d}}_{F,F}$ (Result D.9). We then make use of the combined sensitivity result for $\widehat{\bar{d}}_{F,F}$ to derive a combined sensitivity result for the modified basic scale-up estimator (Result D.10 and Corollary D.11).

Result D.9 *Suppose researchers have obtained a probability sample s_F to estimate $\bar{d}_{F,F}$; however, suppose that the researchers have imperfect sampling weights. Call the imperfect sampling weights $w_i'^F = \frac{1}{\pi_i'^F}$, call the true weights $w_i^F = \frac{1}{c\pi_i^F}$, and define $\epsilon_i^F = \frac{\pi_i^F}{\pi_i'^F}$. Let the estimator for $y_{F,\mathcal{A}}$ using these imperfect weights be $\widehat{y}_{F,\mathcal{A}}'$.*

Suppose also that researchers have chosen a set of probe alters $\mathcal{A}$ in order to use the known population method (Result B.3). However, suppose that the researcher's estimate of $N_{\mathcal{A}}$ is incorrect, so that $\widehat{N}_{\mathcal{A}} = c_1 \cdot N_{\mathcal{A}}$. Suppose also that the reporting condition (Equation B.6) of Result B.3 is incorrect, so that $y_{F,\mathcal{A}} = c_2 \cdot d_{F,\mathcal{A}}$. Finally,

suppose that the probe alter condition (Equation B.7) of Result B.3 is incorrect, so that $\bar{d}_{\mathcal{A},F} = c_3 \cdot \bar{d}_{F,F}$. Call the estimator for $\bar{d}_{F,F}$ under these imperfect conditions $\widehat{\tilde{d}}_{F,F}^{\prime\star}$.

Let the known population estimator for $\bar{d}_{F,F}$ (Result B.3) under these imperfect conditions be $\widehat{\tilde{d}}_{F,F}^{\prime\star}$. Then

$$\widehat{\tilde{d}}_{F,F}^{\prime\star} \rightarrow \bar{\epsilon}^F (1 + K_{F_2}) \cdot \frac{c_2\, c_3}{c_1} \cdot \bar{d}_{F,F}, \tag{D.47}$$

where $\rightarrow$ means 'is consistent and unbiased for', and $K_{F_2} = cor_F(y_{i,\mathcal{A}}, \epsilon_i^F) cv_F(y_{i,\mathcal{A}}) cv_F(\epsilon_i^F)$.

Proof: Under the assumptions above, we can write the imperfect estimator $\widehat{\tilde{d}}_{F,F}^{\prime\star}$ as

$$\widehat{\tilde{d}}_{F,F}^{\prime\star} = \frac{1}{c_1} \cdot \frac{\widehat{y}_{F,\mathcal{A}}^{\prime}}{N_{\mathcal{A}}} \tag{D.48}$$

Using the exact same argument as Result D.2 and Corollary D.3, we have

$$\widehat{y}_{F,\mathcal{A}}^{\prime} \rightarrow \bar{\epsilon}^F (1 + K_{F_2}) \cdot y_{F,\mathcal{A}}. \tag{D.49}$$

Applying this to the imperfect estimator $\widehat{\tilde{d}}_{F,F}^{\prime\star}$, we have

$$\widehat{\tilde{d}}_{F,F}^{\prime\star} \rightarrow \bar{\epsilon}^F (1 + K_{F_2}) \cdot \frac{1}{c_1} \cdot \frac{y_{F,\mathcal{A}}}{N_{\mathcal{A}}} = \bar{\epsilon}^F (1 + K_{F_2}) \cdot \frac{1}{c_1} \cdot \bar{y}_{F,\mathcal{A}}. \tag{D.50}$$

We will obtain the rest of the result by following the argument of Result B.3 closely, but carrying the errors from the conditions that are not met through with each step.

First, by assumption, $\bar{y}_{F,\mathcal{A}} = c_2 \bar{d}_{F,\mathcal{A}}$, yielding

$$\widetilde{\widehat{d}}{}^{\star}_{F,F} \to \bar{\epsilon}^F (1 + K_{F_2}) \cdot \frac{c_2}{c_1} \cdot \bar{d}_{F,\mathcal{A}}. \tag{D.51}$$

Next, again by assumption, $\bar{d}_{F,\mathcal{A}} = c_3 \bar{d}_{F,F}$, so we have

$$\widetilde{\widehat{d}}{}^{\star}_{F,F} \to \bar{\epsilon}^F (1 + K_{F_2}) \cdot \frac{c_2\, c_3}{c_1} \cdot \bar{d}_{F,F}, \tag{D.52}$$

which is our result. $\blacksquare$

Result D.10 *Suppose researchers have obtained a probability sample s_F to estimate $y_{F,H}$ and $\bar{d}_{F,F}$ in order to produce estimates from the modified basic scale-up method. However, suppose that the researchers have imperfect sampling weights. Call the imperfect sampling weights $w'^F_i = \frac{1}{\pi'^F_i}$, call the true weights $w^F_i = \frac{1}{c\pi^F_i}$, and define $\epsilon^F_i = \frac{\pi^F_i}{\pi'^F_i} = \frac{w'^F_i}{w^F_i}$. Let the estimator for $y_{F,H}$ using these imperfect weights be $y'_{F,H}$.*

Suppose also that researchers have chosen a set of probe alters $\mathcal{A}$ in order to use the known population method (Result B.3). However, suppose that the researcher's estimate of $N_{\mathcal{A}}$ is incorrect, so that $\widehat{N}_{\mathcal{A}} = c_1 \cdot N_{\mathcal{A}}$. Suppose also that the reporting condition (Equation B.6) of Result B.3 is incorrect, so that $y_{F,\mathcal{A}} = c_2 \cdot d_{F,\mathcal{A}}$. Suppose also that the probe alter condition (Equation B.7) of Result B.3 is incorrect, so that $\bar{d}_{\mathcal{A},F} = c_3 \cdot \bar{d}_{F,F}$. Call the estimator for $\bar{d}_{F,F}$ under these imperfect conditions $\widetilde{\widehat{d}}{}^{\star}_{F,F}$.

Finally, suppose that the basic scale-up conditions do not hold; that is, suppose that there are false positive reports, so that $y^+_{F,H} = \eta_F y_{F,H}$; suppose that there are false negative reports, so that $\bar{v}_{H,F} = \tau_F \bar{d}_{H,F}$; and suppose that the average personal network size of hidden population members is not equal to the average personal network size of frame population members, so that $\bar{d}_{H,F} = \delta_F \bar{d}_{F,F}$.

Let the modified basic scale-up estimator for N_H in this situation be

$$\widehat{N}_H^{\prime\star} = \frac{\widehat{y}_{F,H}^{\prime}}{\widetilde{d}_{F,F}^{\prime\star}}.$$

(D.53)

Then

$$\widehat{N}_H^{\prime\star} \rightsquigarrow \frac{(1 + K_{F_1})}{(1 + K_{F_2})} \cdot \frac{c_1}{c_2\,c_3} \cdot \frac{\tau_F\,\delta_F}{\eta_F} \cdot N_H,$$

(D.54)

where $\rightsquigarrow$ means 'is consistent and essentially unbiased for'; $K_{F_1} = cor_F(y_{i,H}, \epsilon_i^F)\,cv_F(y_{i,H})\,cv_F(\epsilon_i^F)$; and $K_{F_2} = cor_F(y_{i,\mathcal{A}}, \epsilon_i^F)\,cv_F(y_{i,\mathcal{A}})\,cv_F(\epsilon_i^F)$.

Proof:

The modified basic scale-up estimator is formed from a ratio of estimators for the numerator $(y_{F,H})$ and denominator $(\bar{d}_{F,F})$. We have already derived results for each of the numerator and the denominator separately; our approach will therefore be to combine them. We must account for the fact that, in addition to the assumptions required for the estimator of the numerator and the denominator, the modified basic scale-up estimator also requires the additional conditions that there are no false positive reports, that there are no false negative reports, and that the degree ratio is one.

For the numerator, Result D.9 shows that

$$\widetilde{d}_{F,F}^{\star} \to \bar{\epsilon}^F (1 + K_{F_2}) \cdot \frac{c_2\,c_3}{c_1} \cdot \bar{d}_{F,F}.$$

(D.55)

Thus, Expression D.55 shows sensitivity of the denominator of the modified basic scale-up estimator to violations of all of the conditions it relies upon.

A54

Turning now to the numerator of the modified basic scale-up estimator, Corollary D.3 shows that

$$\widehat{y}'_{F,H} \to y_{F,H} \cdot \bar{\epsilon} \cdot (1 + K_{F_1}), \tag{D.56}$$

where $K_{F_1} = \mathrm{cor}_F(y_{i,H}, \epsilon_i^F)\mathrm{cv}_F(y_{i,H})\mathrm{cv}_F(\epsilon_i^F)$. Thus, Expression D.56 shows sensitivity of the numerator of the modified basic scale-up estimator to violations of all of the conditions it relies upon.

Using the fact that a ratio estimator is consistent and essentially unbiased for the ratio of the estimand of its numerator and denominator (see Online Appendix E and Sarndal et al. (1992, chap. 5)), we therefore have

$$\widehat{N}'^{\star}_H \rightsquigarrow \frac{(1 + K_{F_1})}{(1 + K_{F_2})} \cdot \frac{c_1}{c_2\, c_3} \cdot \frac{y_{F,H}}{\bar{d}_{F,F}}. \tag{D.57}$$

Finally, by assumption, we have $y_{F,H} = y_{F,H}^+/\eta_F$, and $\bar{v}_{H,F} = \bar{d}_{F,F}/(\tau_F\, \delta_F)$. Substituting these assumptions into Expression D.58 produces

$$\widehat{N}'^{\star}_H \rightsquigarrow \frac{(1 + K_{F_1})}{(1 + K_{F_2})} \cdot \frac{c_1}{c_2\, c_3} \cdot \frac{\tau_F\, \delta_F}{\eta_F} \cdot \frac{y_{F,H}^+}{\bar{v}_{F,F}} \tag{D.58}$$

By the argument in Section 2 and Appendix A, $N_H = y_{F,H}^+/\bar{v}_{H,F}$. Substituting N_H for $y_{F,H}^+/\bar{v}_{H,F}$ in the expression above completes the proof. ∎

Corollary D.11 *From Result D.10, it follows that, for the modified basic scale-up estimator,*

$$\widehat{N}'^{\star}_H \cdot \underbrace{\frac{(1 + K_{F_2})}{(1 + K_{F_1})}}_{\substack{sampling \\ conditions}} \cdot \underbrace{\frac{c_2\, c_3}{c_1}}_{\substack{known \\ population \\ conditions}} \cdot \underbrace{\frac{\eta_F}{\tau_F\, \delta_F}}_{\substack{basic \\ scale\text{-}up \\ conditions}} \rightsquigarrow N_H. \tag{D.59}$$

Researchers who wish to conduct a sensitivity analysis for estimates made using the generalized scale-up method can therefore (1) assume values or ranges of values for K_{F_1}, K_{F_2}, c_1, c_2, c_3, δ_F, τ_F, and η_F; and (2) use Corollary D.11 to determine the resulting values of N_H. Thus, researchers can use this approach to explore the sensitivity of their estimates to all of the assumptions they had to make, individually and jointly.

E   Approximate unbiasedness of compound ratio estimators

E.1   Overview

Several of the estimators we propose are nonlinear, which means that they are not design-unbiased (Sarndal et al., 1992). While ratio estimators are common in survey sampling and the bias of these estimators is commonly regarded as insignificant (Sarndal et al., 1992), several of the estimators we propose are somewhat more complex than standard ratio estimators. In fact, all of our nonlinear estimators turn out to all be special cases of a ratio of ratios (Table E.1), which is also known as a double ratio estimator (Rao and Pereira, 1968). Any double ratio can be written

$$R_d = \frac{R_1}{R_0} = \frac{\frac{\bar{y}_1}{\bar{x}_1}}{\frac{\bar{y}_0}{\bar{x}_0}} = \frac{\bar{y}_1 \bar{x}_0}{\bar{x}_1 \bar{y}_0}. \tag{E.1}$$

If we have unbiased estimators for each of the four terms, we can estimate R_d by

$$\widehat{r}_d = \frac{\widehat{\bar{y}}_1 \widehat{\bar{x}}_0}{\widehat{\bar{x}}_1 \widehat{\bar{y}}_0}. \tag{E.2}$$

In this appendix we investigate when we can expect the biases in our estimators to be small enough to be negligible; we conclude that, in practice, the bias is typically negligible when compared to sampling and non-sampling error.

E.2   The general case

We will focus on the relative bias in our estimator, $\widehat{r}_d$. The relative bias is given by

$$B_d = \frac{\mathbb{E}[\widehat{r}_d] - R_d}{R_d}. \tag{E.3}$$

B_d expresses the bias in our estimator $\widehat{r}_d$ in terms of the true value; a relative bias of 0.5, for example, means that our estimator is typically 0.5 times bigger than the true value. This is a natural quantity to consider because estimators that have small relative bias have small bias in substantive terms.

Our approach will be to follow Rao and Pereira (1968) in using a Taylor series to form an approximation to the relative bias. This is accomplished in Result E.1.

Result E.1 *(Rao and Pereira, 1968) If $\widehat{\bar{x}}_0$, $\widehat{\bar{x}}_1$, $\widehat{\bar{y}}_0$, and $\widehat{\bar{y}}_1$ are unbiased estimators, and $|(\widehat{\bar{x}}_1 - \bar{x}_1)/\bar{x}_1| < 1$ and $|(\widehat{\bar{y}}_0 - \bar{y}_0)/\bar{y}_0| < 1$, then the relative bias of the double ratio estimator, B_d, is approximated by*

$$B_d = \frac{\mathbb{E}[\widehat{r}_d] - R}{R} \approx B'_d = C_{\widehat{\bar{x}}_1,\widehat{\bar{y}}_0} - C_{\widehat{\bar{x}}_1,\widehat{\bar{y}}_1} - C_{\widehat{\bar{y}}_0,\widehat{\bar{y}}_1} - C_{\widehat{\bar{x}}_0,\widehat{\bar{x}}_1} - C_{\widehat{\bar{x}}_0,\widehat{\bar{y}}_0} + C_{\widehat{\bar{y}}_1,\widehat{\bar{x}}_0} + C^2_{\widehat{\bar{y}}_0} + C^2_{\widehat{\bar{x}}_1},$$
$$\tag{E.4}$$

where $C_{\widehat{\bar{x}},\widehat{\bar{y}}} = \frac{cov(\widehat{\bar{x}},\widehat{\bar{y}})}{\bar{x}\bar{y}}$ is the relative covariance between $\widehat{\bar{x}}$ and $\widehat{\bar{y}}$, and $C^2_{\widehat{\bar{y}}} = \frac{var(\widehat{\bar{y}})}{\widehat{\bar{y}}^2}$.

Proof: Define

$$\delta_{\widehat{\bar{x}}_0} = \frac{\widehat{\bar{x}}_0 - \bar{x}_0}{\bar{x}_0}, \tag{E.5}$$

Estimator	Reference	Form	$\hat{\bar{x}}_0$	$\hat{\bar{y}}_1$	$\hat{\bar{x}}_1$	$\hat{\bar{y}}_0$	Approx. rel. bias
$\hat{\phi}_F$	Res. B.6	$K\hat{\bar{x}}_0/\hat{\bar{y}}_0$	$\sum_{i\in s_F} y_{i,A_{F_1}}/\pi_i$	-	-	$\sum_{i\in s_F} y_{i,A_{F_2}}/\pi_i$	$C^2_{\hat{\bar{y}}_0} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$
$\hat{\tilde{v}}_{H,F}$	Res. C.2	$K\hat{\bar{x}}_0/\hat{\bar{y}}_0$	$\sum_{i\in s_H} \tilde{v}_{i,A_H\cap F}/c\pi_i$	-	-	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\hat{\bar{y}}_0} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$
$\hat{\tilde{d}}_{H,F}$	Res. C.5	$K\hat{\bar{x}}_0/\hat{\bar{y}}_0$	$\sum_{i\in s_H} y_{i,A_H\cap F}/c\pi_i$	-	-	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\hat{\bar{y}}_0} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$
$\hat{\delta}_F$	Res. C.6	$K\hat{\bar{x}}_0/(\hat{\bar{y}}_0\,\hat{\bar{x}}_1)$	$\sum_{i\in s_H} y_{i,A_H\cap F}/c\pi_i$	-	$\sum_{i\in s_F} y_{i,A_F}/\pi_i$	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\hat{\bar{y}}_0} + C^2_{\hat{\bar{x}}_1} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$
$\hat{\tau}_F$	Res. C.7	$K\hat{\bar{x}}_0/(\hat{\bar{y}}_0\,\hat{\bar{x}}_1)$	$\sum_{i\in s_H} \tilde{v}_{i,A_H\cap F}/c\pi_i$	-	$\sum_{i\in s_H} y_{i,A_H\cap F}/\pi_i$	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\hat{\bar{y}}_0} + C^2_{\hat{\bar{x}}_1} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$
$\hat{N}_H$	Res. C.8	$K\hat{\bar{y}}_1\hat{\bar{x}}_0/\hat{\bar{y}}_0$	$\sum_{i\in s_H} 1/c\pi_i$	$\sum_{i\in s_F} y_{i,H}/\pi_i$	-	$\sum_{i\in s_H} \tilde{v}_{i,A_H\cap F}/c\pi_i$	$C^2_{\hat{\bar{y}}_0} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$
$\hat{N}_H$	Res. C.10	$K\hat{\bar{x}}_0/\hat{\bar{y}}_0$	$\sum_{i\in s_F} y_{i,H}/\pi_i$	-	-	$\sum_{i\in s_F}\sum_j y_{i,A_j}/\pi_i$	$C^2_{\hat{\bar{y}}_0} - C_{\hat{\bar{y}}_0,\hat{\bar{x}}_0}$

Table E.1: Description of the general form of the nonlinear estimators we propose. K is a constant, $\hat{\bar{y}}_1$ and $\hat{\bar{x}}_1$ are taken from s_F, while $\hat{\bar{x}}_0$ and $\hat{\bar{y}}_0$ are taken from s_H. Our nonlinear estimators are all special cases of the double ratio estimator, which we define and discuss below. Note that the estimator for $\hat{N}_H$ that involves adjusting a basic scale-up estimate (Result C.10) would, in practice, take these adjustment factors from other studies; we therefore assume that these adjustment factors are independent of the quantities that go into the scale-up estimate, and treat them as constants.

with analogous definitions for $\delta_{\widehat{x}_1}$, $\delta_{\widehat{y}_1}$, and $\delta_{\widehat{y}_0}$. We can express r_d as

$$\widehat{r}_d = R\frac{(1 + \delta_{\widehat{y}_1})(1 + \delta_{\widehat{x}_0})}{(1 + \delta_{\widehat{y}_0})(1 + \delta_{\widehat{x}_1})}. \qquad (\text{E.6})$$

The relative bias then becomes

$$B_d = \frac{\mathbb{E}[\widehat{r}_d] - R}{R} = \mathbb{E}\left[\frac{(1 + \delta_{\widehat{y}_1})(1 + \delta_{\widehat{x}_0})}{(1 + \delta_{\widehat{y}_0})(1 + \delta_{\widehat{x}_1})}\right] - 1. \qquad (\text{E.7})$$

The strategy is now to expand the two factors in the denominator and to then discard high-order terms. What remains will be an approximation to the true relative bias.

Recall that if $|x| < 1$ then $\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$ and, in particular, $\frac{1}{1+x} = 1 - x^2 + x^3 - \cdots$. We'll make use of this expansion for the two factors in the denominator of Equation E.7; that is, we assume that $|\delta_{\bar{y}_0}| < 1$ and $|\delta_{\bar{x}_1}| < 1$. Then we have

$$B_d = \mathbb{E}\left[(1 + \delta_{\widehat{y}_1})(1 + \delta_{\widehat{x}_0})(1 - \delta_{\widehat{y}_0} + \delta_{\widehat{y}_0}^2 - \cdots)(1 - \delta_{\widehat{x}_1} + \delta_{\widehat{x}_1}^2 - \cdots)\right] - 1 \qquad (\text{E.8})$$

If we multiply this out and retain only terms up to order 2, we obtain the following approximation:

$$B_d \approx \mathbb{E}\left[\delta_{\widehat{x}_1}\delta_{\widehat{y}_0} + \delta_{\widehat{x}_0}\delta_{\widehat{y}_1} - \delta_{\widehat{x}_0}\delta_{\widehat{y}_0} - \delta_{\widehat{x}_0}\delta_{\widehat{x}_1} - \delta_{\widehat{x}_1}\delta_{\widehat{y}_1} - \delta_{\widehat{y}_0}\delta_{\widehat{y}_1} + \delta_{\widehat{x}_0} + \delta_{\widehat{y}_1} - \delta_{\widehat{x}_1} - \delta_{\widehat{y}_0} - \delta_{\widehat{y}_0}^2 - \delta_{\widehat{x}_1}^2\right].$$
$$(\text{E.9})$$

Since we assumed that the estimators for the individual components of r_d are unbiased, we know that

$$\mathbb{E}[\delta_{\widehat{x}_1}] = 0, \qquad (\text{E.10})$$

A59

We can also determine that

$$\mathbb{E}[\delta_{\widehat{\bar{x}}_1} \delta_{\widehat{\bar{y}}_1}] = \frac{\mathrm{cov}(\widehat{\bar{x}}_1, \widehat{\bar{y}}_1)}{\bar{x}_1 \bar{y}_1}, \qquad (E.11)$$

and, that

$$\mathbb{E}[\delta_{\widehat{\bar{x}}_1}^2] = \frac{\mathrm{var}(\widehat{\bar{x}}_1)}{\bar{x}_1^2}. \qquad (E.12)$$

Applying these relationships to Equation E.9, we find

$$B_d \approx C_{\widehat{\bar{x}}_0,\widehat{\bar{y}}_1} + C_{\widehat{\bar{x}}_1,\widehat{\bar{y}}_0} - C_{\widehat{\bar{x}}_0,\widehat{\bar{x}}_1} - C_{\widehat{\bar{x}}_0,\widehat{\bar{y}}_0} - C_{\widehat{\bar{x}}_1,\widehat{\bar{y}}_1} - C_{\widehat{\bar{y}}_0,\widehat{\bar{y}}_1} + C_{\bar{x}_1}^2 + C_{\bar{y}_0}^2, \qquad (E.13)$$

which is our result. ∎

Result E.1 is useful because it reveals the behavior of double ratio estimators in quite general contexts. To understand what it says a bit more intuitively, note that Result E.1 is framed in terms of the relative covariances and variances of the *estimators* $\widehat{\bar{x}}_0$, $\widehat{\bar{x}}_1$, $\widehat{\bar{y}}_0$, and $\widehat{\bar{y}}_1$. In the special case of simple random sampling with replacement, we can re-write the approximation in terms of the finite population variances and covariances and a constant, κ:

$$B_d' = \kappa \left[C_{x_1,y_0} - C_{x_1,y_1} - C_{y_0,y_1} - C_{x_0,x_1} - C_{x_0,y_0} + C_{y_1,x_0} + C_{y_0}^2 + C_{x_1}^2 \right], \qquad (E.14)$$

where $\kappa = \left(\frac{1}{n} - \frac{1}{N} \right)$, n is our sample size, and N is the size of the population. In the case of simple random sampling, the relative bias depends upon the finite population variances of the underlying population values and the size of our sample.

For designs other than simple random sampling, there is no analogous expression as simple as Equation E.14. However, speaking roughly, if we have an idea that our

sampling plan has a typical design effect (deff) for the quantities inside the square brackets in Equation E.14, then we can see that we would simply replace the κ in Equation E.14 by $(\kappa \cdot \text{deff})$ in order to get a sense of the approximate relative bias.

Notice, also, that Result E.1 is framed largely in terms of relative covariances. When we apply Result E.1, we will often make use of the fact that the relative covariances can be expressed in terms of correlations and coefficients of variation as follows:

$$C_{\widehat{\bar{x}},\widehat{\bar{y}}} = \frac{\text{cov}(\widehat{\bar{x}},\widehat{\bar{y}})}{\bar{x}\bar{y}} = \frac{\rho_{\widehat{\bar{x}},\widehat{\bar{y}}}\sqrt{\text{var}(\widehat{\bar{x}})}\sqrt{\text{var}(\widehat{\bar{y}})}}{\bar{x}\bar{y}} \tag{E.15}$$

$$= \rho_{\widehat{\bar{x}},\widehat{\bar{y}}}\,\text{cv}(\widehat{\bar{x}})\,\text{cv}(\widehat{\bar{y}}), \tag{E.16}$$

where $\rho_{\widehat{\bar{x}},\widehat{\bar{y}}}$ is the correlation between the estimators $\widehat{\bar{x}}$ and $\widehat{\bar{y}}$, and $\text{cv}(\widehat{\bar{x}}) = \frac{\sqrt{\text{var}(\widehat{\bar{x}})}}{\bar{x}}$ is the coefficient of variation of the estimator $\widehat{\bar{x}}$. We will also make use of the fact that $C_{\widehat{\bar{x}}}^2 = \text{cv}(\widehat{\bar{x}})^2$.

E.3   Applying Result E.1 to scale-up

We now apply Result E.1 to understand the biases in the nonlinear estimators we propose for realistic situations. For each particular estimator, we can simplify the expression in Result E.1. In order to do so, we first remove terms that do not appear in the estimator itself (for example, in $\widehat{\delta}_F$, there is no $\widehat{\bar{y}}_1$). Additionally, we assume that the estimates produced from a sample from the frame population and a sample from the hidden population will be independent of one another, meaning that their correlation will be 0. Table E.1 summarizes the nonlinear estimators we propose, along with the specific version of the approximate relative bias from Result E.1 that

applies.

Finally, in order to give a sense of the magnitude of the coefficients of variation and correlations found in real studies, we estimated the quantities that go into the approximate relative bias from the studies available to us. Table E.2 shows the coefficients of variation for the estimated degree (the values of $\widehat{\bar{x}}_1$ for $\widehat{\delta}_F$) in surveys from Rwanda, the United States, and Curitiba, Brazil. Further, Tables E.3 and E.4 show the relevant coefficients of variation and pairwise correlations for all remaining quantities using data from Curitiba, Brazil (currently, the only setting where we have data from a sample of the hidden population). For all values in these tables, the estimated variance of the estimators is calculated using the bootstrap methods presented in Section F.1.

Since we have both a sample from the frame population and a sample from the hidden population in Curitiba, we can compute numerical estimates of the bias of each nonlinear estimator in the context of that study. We can see that in this study bias caused by the nonlinearity of the estimator was not a big problem: in each case, the estimated approximate bias was less than one percent of the estimate (Table E.5).

To conclude, we derived an expression for the approximate relative bias in double ratio estimators in general. We then simplified the approximation for each specific nonlinear estimator that we propose. Finally, we used data from a real scale-up study in Curitiba, Brazil to estimate magnitude of the biases caused by the non-linearity of the estimators in a specific scale-up study. From these results, we conclude that theses estimators are essentially unbiased, and that sampling error and non-sampling error will dominate any bias introduced by the nonlinear form of the estimators.

$\widehat{\mathrm{cv}}(\widehat{\bar{d}})$	source
0.05	Rwanda
0.10	Curitiba
0.02	US

Table E.2: Estimated coefficients of variation for the average degree from 3 different scale-up surveys. These play a role in the approximate relative bias for the estimate of $\widehat{\delta}_F$. Our approximation tells us that the larger these values are, the worse the relative bias will be. The estimates were computed using the rescaled bootstrap procedure.

	estimated coef. of variation
$\sum_{i \in s_H} y_{i,\mathcal{A} \cap F}/c\pi_i$	0.08
$\sum_{i \in s_H} \tilde{v}_{i,\mathcal{A} \cap F}/c\pi_i$	0.08
$\sum_{i \in s_H} 1/c\pi_i$	0.06

Table E.3: Estimated coefficients of variation for quantities derived from a sample from the hidden population. These quantities play a role in the approximate relative bias for the estimate of all of the nonlinear estimators we propose. The estimates were computed using the respondent-driven sampling bootstrap procedure (Salganik, 2006).

F   Variance estimation and confidence intervals

In addition to producing point estimates, researchers must also produce confidence intervals around their estimates. The procedure currently used by scale-up researchers begins with the variance estimator proposed in Killworth et al. (1998b):

$$\widehat{se}(\widehat{N}_H) = \sqrt{\frac{N \cdot \widehat{N}_H}{\sum_{i \in s_F} \widehat{d}_{i,U}}}, \tag{F.1}$$

and then produces a confidence interval:

$$\widehat{N}_H \pm z_{1-\alpha/2}\widehat{se}(\widehat{N}_H), \tag{F.2}$$

where $1-\alpha$ is the desired confidence level (typically 0.95), and $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard Normal distribution.

	estimated correlation
$\widehat{\text{cor}}(\sum_{i \in s_H} y_{i,\mathcal{A} \cap F}/c\pi_i, \sum_{i \in s_H} \tilde{v}_{i,\mathcal{A} \cap F}/c\pi_i)$	0.92
$\widehat{\text{cor}}(\sum_{i \in s_H} y_{i,\mathcal{A} \cap F}/c\pi_i, \sum_{i \in s_H} 1/c\pi_i)$	0.71
$\widehat{\text{cor}}(\sum_{i \in s_H} \tilde{v}_{i,\mathcal{A} \cap F}/c\pi_i, \sum_{i \in s_H} 1/c\pi_i)$	0.68

Table E.4: Estimated pairwise correlations for quantities derived from a sample from the hidden population. These quantities play a role in the approximate relative bias for the estimate of all of the nonlinear estimators we propose.

	approx. rel. bias, B_d	estimate	estimated absolute bias
$\widehat{\tau}_F$	0.0005	0.77	0.0004
$\widehat{\delta}_F$	0.0105	0.69	0.0073
$\widehat{N}_H$	0.0026	114498.00	298.0000

Table E.5: Approximate relative bias in the estimates of the nonlinear quantities using data taken from the Curitiba study, the point estimates produced by the Curitiba study, and the estimated implied absolute bias. For each quantity, the bias is very small.

Unfortunately, the variance estimator (Equation F.1) was derived from the basic scale-up model (Equation 11), and so it suffers from the limitations of that model. In particular, it has three main problems, none of which seem to have been appreciated in the scale-up literature and all of which lead it to underestimate the variance in most situations. First, the variance estimator in Equation F.1 does not include any information about the procedure used to sample respondents, which can lead to problems when complex sampling designs, such as stratified, multi-stage designs, are used. Second, it implicitly assumes that the researchers have learned about $\sum_{i \in s_F} d_{i,U}$ independent alters, which is not true if there are barrier effects (i.e., non-random social mixing). Finally, like virtually all variance estimators, it only provides a measure of uncertainty introduced by sampling but not other possible sources of error.

To address the first two problems but not the third, we propose that researchers used the rescaled bootstrap variance estimation procedure (Rao and Wu, 1988; Rao et al., 1992; Rust and Rao, 1996) with the percentile method; a combination that, for

convenience, we will refer to as the rescaled bootstrap. This procedure, described in more detail below, has strong theoretical foundations; does not depend on the basic scale-up model; can handle both simple and complex sample designs; and can be used for both the basic scale-up estimator and the generalized scale-up estimator.

In addition to the theoretical reasons to prefer the rescaled bootstrap, empirically, we find that the rescaled bootstrap produces intervals with slightly better coverage properties in three real scale-up studies. In particular, using the internal consistency check procedure proposed in Killworth et al. (1998a) for all groups of known size in three real scale-up datasets—one collected via simple random sampling (McCarty et al., 2001) and two collected via complex sample designs (Salganik et al., 2011a; Rwanda Biomedical Center, 2012)—we produced a size estimate using the basic scale-up estimator (Equation 12), and we produced confidence intervals using (1) the current procedure (Equation F.1); (2) the simple bootstrap (which does not account for complex sample designs) with the percentile method; and (3) the rescaled bootstrap (which does account for complex sample designs) with the percentile method.

This empirical evaluation (Figure F.1) produced three main results. First, as expected, we found that the current confidence interval procedure produces intervals with bad coverage properties: purported 95% confidence intervals had empirical coverage rates of about 5%. This poor performance does not seem to have been widely appreciated in the scale-up literature. Second, also consistent with expectation, we found that the rescaled bootstrap produced wider intervals than both the current procedure and the simple bootstrap, especially in the case of complex sample designs. Third, and somewhat surprisingly, the rescaled bootstrap did not work well in an absolute sense: purported 95% confidence intervals had empirical coverage rates of about 10%, only slightly better than the current procedure.

We speculate that there are two possible reasons for the surprisingly poor cover-

A65

age rates of the rescaled bootstrap. The first is bias in the basic scale-up estimator. As described in detail in Sarndal et al. (1992, Sec 5.2), bias in an estimator can degrade the coverage rates for confidence intervals. For example, if Native Americans (one of the groups in the study of McCarty et al. (2001)) have smaller personal networks than other Americans, then there will be a downward bias in the estimated number of Native Americans (Equation 20). This bias will necessarily degrade the coverage properties of any confidence interval procedure, especially if the bias ratio $\left(bias(\hat{N}_H)/se(\hat{N}_H)\right)$ is large (see Sarndal et al. (1992, Sec 5.2)). The second possible reason for the surprisingly poor coverage rates could also be some unknown problem with the rescaled bootstrap or the percentile method. Because (i) the rescaled bootstrap and percentile method have strong theoretical foundations (Rao and Wu, 1988; Rao et al., 1992; Rust and Rao, 1996; Efron and Tibshirani, 1993) and (ii) we expect that the basic scale-up estimates are biased in most situations (see Equation 20), we believe that the main reason for the poor coverage is the bias. However, we also believe that future research should explore the properties of the rescaled bootstrap and percentile method in greater detail.

An additional concern about these empirical results is that they only apply to the basic scale-up estimator and not the generalized scale-up estimator. Unfortunately, we cannot assess the performance of the rescaled bootstrap procedure when used with the generalized scale-up estimator because the generalized scale-up estimator has not yet been used for populations of known size.

These empirical results, and the theoretical arguments that follow, lead us to three conclusions. First, confidence intervals from the rescaled bootstrap are preferable to intervals from the current procedure. Second, researchers should expect that the confidence intervals from the rescaled bootstrap procedure will be anti-conservative (i.e., they will be too small). Third, creating confidence intervals around scale-up

A66

estimates is an important area for further research.

Next in Section F.1 we review the standard bootstrap and rescaled bootstrap; describe how we applied these methods to three real scale-up datasets; and describe the results in Figure F.1 in greater detail. Finally, in Section F.2 we describe how researchers can use the rescaled bootstrap with the generalized scale-up estimator.

F.1 Variance estimation with a sample from F

The goal of a bootstrap variance estimation procedure is to put a confidence interval around an estimate $\widehat{N}_H$ that is derived from a sample s_F. The most standard bootstrap procedure has three steps. First, researchers generate B replicate samples, $s_F^{(1)}, s_F^{(2)}, \ldots, s_F^{(B)}$ by randomly sampling with replacement from s_F. Second, these replicate samples are then used to produce a set of replicate estimates, $\widehat{N}_H^{(1)}, \widehat{N}_H^{(2)}, \ldots, \widehat{N}_H^{(B)}$. Finally, the replicate estimates are combined to produce a confidence interval; for example, by the percentile method which chooses the 2.5th and 97.5th percentiles of the B estimates (Fig. F.2) (Efron and Tibshirani, 1993).

When the original sample can be modeled as a simple random sample, this standard bootstrap procedure is appropriate. For example, consider the scale-up study of McCarty et al. (2001) that was based on telephone survey of 1,261 Americans selected via random digit dialing.[12] We can approximate the sampling design as simple random sampling, and draw $B = 10,000$ replicate samples of size 1,261. In this case the bootstrap confidence intervals are, as expected, larger than the confidence intervals from Equation F.1, since they account for the clustering of responses with respondent; on average, they are 2.05 times wider.

---

[12]The original data file includes 1,375 respondents. From these cases, 113 respondents who had missing data for some of the aggregated relational data questions and 1 respondent who answered 7 for all questions (see Zheng et al. (2006)). Further, consistent with common practice (e.g., Zheng et al. (2006)), we top coded all responses at 30, affecting 0.26% of responses.
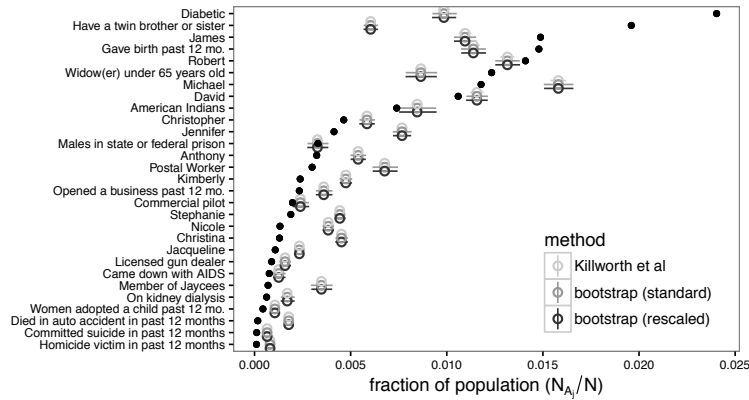
(a) United States (simple random sample)
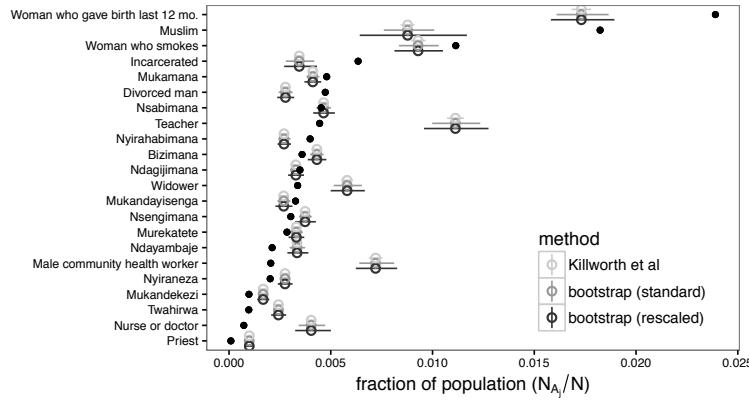
(b) Rwanda (stratified, multi-stage)
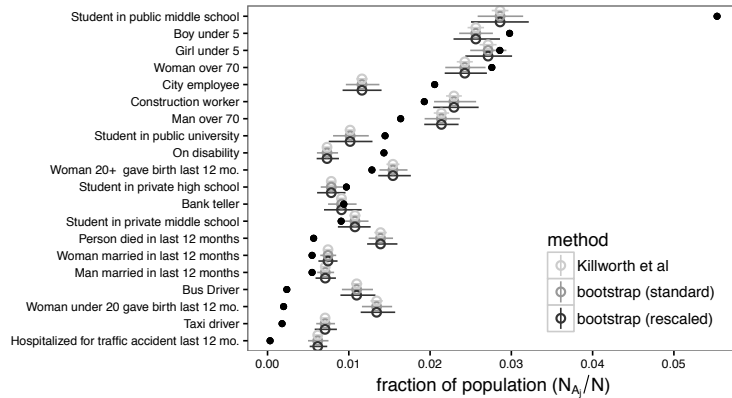
(c) Curitiba, Brazil (multi-stage)

Figure F.1: Assessing confidence interval procedures using scale-up studies in the United States (McCarty et al., 2001), Rwanda (Rwanda Biomedical Center, 2012), and Curitiba, Brazil (Salganik et al., 2011a). The true size of each group is shown with a black dot. Estimates made use the basic scale-up estimator are shown with circles. The rescaled bootstrap confidence intervals include the true group size for 3.4%, 9.1%, and 15.0% of the groups in the US, Rwanda, and Curitiba, respectively. The standard bootstrap confidence intervals include the true group size for 3.4%, 9.1%, and 10.0% of the groups. The currently used procedure (Equation F.1), contains the true group size for 3.4%, 9.1%, and 5.0% of the groups.
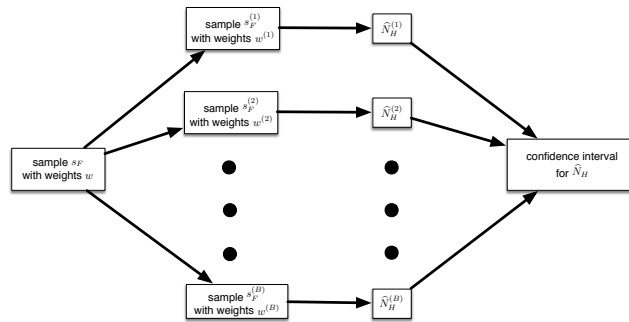
A68

Figure F.2: Schematic of the bootstrap procedure to put a confidence interval around $\widehat{N}_H$ when there is a sample from the frame s_F.

This standard bootstrap procedure, however, can perform poorly when the original data are collected with a complex sample design (Shao, 2003). To deal with this problem, Rust and Rao (1996) proposed the rescaled bootstrap procedure that works well when the data are collected with a general multistage sampling design, a class of designs that includes most designs that would be used for face-to-face scale-up surveys. For example, it includes stratified two-stage cluster sampling with oversampling (as was used in a recent scale-up study in Rwanda (Rwanda Biomedical Center, 2012)) and three-stage element sampling (as was used in a recent scale-up study in Curitiba, Brazil (Salganik et al., 2011a)); a full description of the designs included in this class is presented in Rust and Rao (1996).

The rescaled bootstrap includes two conceptual changes from the standard bootstrap. First, it approximates the actual sampling design by a closely related one that is much easier to work with. In particular, if we assume that primary sampling units (PSUs) are selected with replacement and that all subsequent stages of sampling are conducted independently each time a given PSU is selected, then we can use the

A69

with-replacement sampling framework in which variance estimation is much easier; see Sarndal et al. (1992) Result 4.5.1 for a more formal version of this claim. It is important to note that this approximation is generally conservative because with-replacement sampling usually results in higher variance than without-replacement sampling. Therefore, we will be estimating the variance for a design that has higher variance than the actual design. In practice, this difference is usually small because the sampling fraction in each stratum is usually small (Rao et al., 1992; Rust and Rao, 1996); see Sarndal et al. (1992) Section 4.6 for a more formal treatment. To estimate the variance in this idealized with-replacement design, resampling should be done independently in each stratum and the units that are resampled with replacement should be entire PSUs, not respondents.

This change—resampling PSUs, not respondents—introduces the need for a second change in the resampling procedure. It is known that the standard bootstrap procedure is off by a factor of $(n-1)/n$ where n is the sample size (Rao and Wu, 1988). Thus, when the sample size is very small, the bootstrap will tend to underestimate the variance. While this issue is typically ignored, it can become important when we resample PSUs rather than respondents. In particular, the number of sampled PSUs in stratum h, n_h, can be small in complex sample designs. At the extreme, in a design with two sampled PSUs per stratum, which is not uncommon, the standard bootstrap would be expected to produce a 50% underestimate of the variance. Therefore, Rao et al. (1992) developed the rescaled bootstrap, whereby the bootstrap sample size is slightly smaller than the original sample size and the sample weights are rescaled to account for this difference. Rust and Rao (1996) recommend that if the original sample includes n_h PSUs in strata h, then researchers should resample $n_h - 1$ PSUs and rescale the respondent weights by $n_h/(n_h - 1)$. That is, the weight for the jth

person in PSU i in the bth replicate sample is

$$w_{ij}^{(b)} = w_{ij} \times \frac{n_h}{(n_h - 1)} \times r_i^{(b)} \tag{F.3}$$

where w_{ij} is the original weight for the jth unit in the ith PSU, n_h is the number of PSUs in strata h, and $r_i^{(b)}$ is the number of times the ith PSU was selected in replicate sample b.

In Figure 1, we compared the three different procedures for putting confidence intervals around the basic scale-up estimator: the current procedure (Killworth et al., 1998b), the standard bootstrap with the percentile method, and the rescaled bootstrap with the percentile method. We made this comparison using data from scale-up studies in the United States, Rwanda,[13] and Curitiba, Brazil.[14] As expected, the rescaled bootstrap produced confidence intervals that are larger than those from the standard bootstrap, which in turn are larger than those from the current scale-up

---

[13]The scale-up study in Rwanda used stratified two-stage cluster sampling with unequal probability of selection across strata in order to oversample urban areas. Briefly, the sample design divided Rwanda into five strata: Kigali City, North, East, South, and West. At the first stage, PSUs—in this case villages—were selected with probability proportional to size and without replacement within each stratum with oversampling in the Kigali City stratum. This approach resulted in a sample of 130 PSUs: 35 from Kigali City, 24 from East, 19 from North, 26 from South, and 26 from West. At the second stage, 20 households were selected via simple random sampling without replacement from each PSU in Kigali City and 15 households from each PSU in other strata. Finally, all members of the sampled household over the age of 15 were interviewed. The study included a survey experiment which randomized respondents to report about one of two different personal networks; to keep things simple, we use responses about only one personal network here. For full details see Rwanda Biomedical Center (2012). The original data file includes 2,406 respondents. From these cases, we removed 2 respondents who had missing data for some of the aggregated relational data questions. Further, consistent with common practice (e.g., Zheng et al. (2006)), we top coded all responses at 30, affecting 0.12% of responses.

[14]The scale-up study in Curitiba, Brazil used two-stage element sampling where 54 primary sampling units (PSUs)—in this case census tracks—were selected with probability proportional to their estimated number of housing units and without replacement. Then, within each cluster, eight secondary sampling units (SSUs)—in this case people—were selected with equal probability without replacement. For full details see Salganik et al. (2011a). The original data file includes 500 respondents. From these cases, we removed no respondents who had missing data for some of the aggregated relational data questions. Further, consistent with common practice (e.g., Zheng et al. (2006)), we top coded all responses at 30, affecting 0.58% of responses.

variance estimation procedure. In the study from Curitiba, the rescaled bootstrap procedure produced confidence intervals 1.17 times larger than the standard bootstrap and 2.84 times larger than the current procedure. In the Rwanda case, the rescaled bootstrap procedure produced confidence intervals 1.35 times larger than the standard bootstrap and 2.65 times larger than the current procedure.

Finally, Figure F.1 shows the estimated confidence intervals for the groups of known size in the three studies described above. The coverage rates for the bootstrap confidence intervals for the US, Rwanda, and Curitiba, are 3.4%, 9.1%, 15.0%. While this is far from ideal, we note that it is slightly better than the currently used procedure (Equation F.2), which produced coverage rates of 3.4%, 9.1%, 5.0%, and it is also slightly better than the standard bootstrap, which produced coverage rates of 3.4%, 9.1%, and 10.0%.

F.2 Variance estimation with sample from F and H

Producing confidence intervals around the generalized scale-up estimator is more difficult than the basic scale-up estimator because the generalized estimator has uncertainty from two different samples: the sample from the hidden population and the sample from the frame population. To capture all of this uncertainty, we propose combining replicate samples from the frame population with independent replicate samples from the hidden population in order to produce a set of replicate estimates. More formally, given s_F, a sample from the frame population, and an independent sample s_H from the hidden population, we seek to produce a set of B bootstrap replicate samples for s_F and s_H, $s_F^{(1)}, s_F^{(2)}, \ldots, s_F^{(B)}$ and $s_H^{(1)}, s_H^{(2)}, \ldots, s_H^{(B)}$, which are then combined to produce a set of B bootstrap estimates: $\widehat{N}_H^{(1)} = f(s_F^{(1)}, s_H^{(1)})$, $\widehat{N}_H^{(2)} = f(s_F^{(2)}, s_H^{(2)}), \ldots \widehat{N}_H^{(B)} = f(s_F^{(B)}, s_H^{(B)})$. Finally, these B replicate estimates are
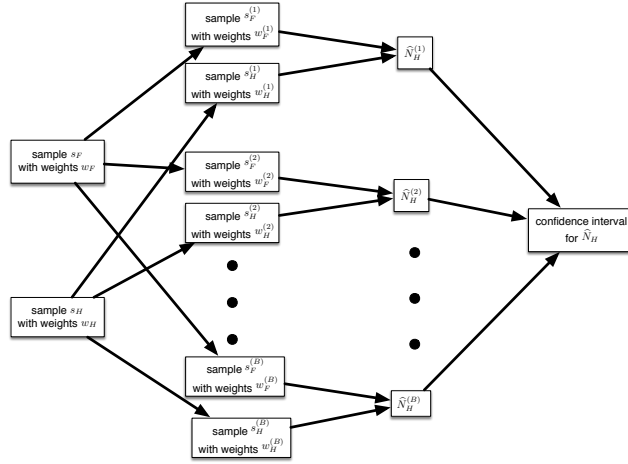
Figure F.3: Schematic of the bootstrap procedure to put a confidence interval around $\widehat{N}_H$ when there is a sample from the frame s_F and a sample from the hidden population s_H.

converted into a confidence interval using the percentile method (Fig. F.3).

Because of the challenges involved in sampling hard-to-reach populations, the two most likely sampling designs for s_H will be time-location sampling and respondent-driven sampling. If s_H was selected with time-location sampling, we recommend treating the design as a two-stage element sample (see Karon and Wejnert (2012)) and using the procedure of Rust and Rao (1996). If s_H was selected with respondent-driven sampling, as was done in a recent study of heavy drug users in Curitiba, Brazil (Salganik et al., 2011b), we recommend using the best available bootstrap method for respondent-driven sampling data, which at the present time is the procedure introduced in Salganik (2006). One implementation detail of that particular bootstrap procedure is that it requires researchers to divide the sample of the hidden population into two mutually exclusive groups. In this case, we recommend dividing the hidden population into those who are above and below the median of their estimated visibility $\widehat{v}_{i,F}$ in order to capture some of the extra uncertainty introduced if there are strong tendencies for more hidden members of the hidden population to

recruit each other.

Because the generalized scale-up estimator has never been used for groups of known size, we cannot explore the coverage rate of the proposed procedure. However, based on experience with respondent-driven sampling, we suspect that variance estimation procedures for hidden populations will underestimate the actual uncertainty in the estimates (Goel and Salganik, 2009, 2010; Yamanis et al., 2013; Verdery et al., 2013; Rohe, 2015). If this is the case, then the intervals around the generalized scale-up estimates will be anti-conservative.

In conclusion, Sec. F.1 presents a bootstrap procedure for simple and complex sample designs from the sampling frame, and Sec. F.2 extends these results to account for the sampling variability introduced by having a sample from the hidden population. We have shown that the performance of these procedures on three real scale-up datasets is consistent with theoretical expectations. Additional research in this area, which is beyond the scope of this paper, could adopt a total survey error approach and attempt to quantify all sources of uncertainty in the estimates, not just sampling uncertainty. Additional research could also explore the properties and sensitivity of these confidence interval procedures though simulation.

G   Simulation study

In this appendix, we describe a simulation study comparing the performance of the generalized and basic network scale-up estimators. The results of these simulations confirm and illustrate several of the analytical results in Section 3 of the paper. Most importantly, the simulations show that the generalized network scale-up estimator is unbiased for all of the situations explored by the simulation, while the basic network scale-up estimator is biased for all but a few special cases. Moreover, our analytical

results correctly predict the bias of the basic network scale-up estimator in each case.

Our simulation study is intentionally simple in order to clearly illustrate our analytical results; it is not designed to be a realistic model of any scale-up study. Concretely, our simulations compare the performance of generalized and basic scale-up estimators as three important quantities vary: (1) the size of the frame population F, relative to the size of the entire population, U; (2) the extent to which people's network connections are not formed completely at random, also called the amount of inhomogenous mixing; and (3) the accuracy of reporting, as captured by the true positive rate τ_F (see Equation 18).

We simulate populations consisting of $N = 5,000$ people, using a stochastic blockmodel (White et al., 1976; Wasserman and Faust, 1994) to randomly generate networks with different amounts of inhomogenous mixing. Stochastic block models assume population members can be grouped into different *blocks*. For any pair of people, i and j, the probability that there is an edge between i and j is completely determined by the block memberships of i and j.

In our simulation model, each person can be either in or out of the frame population F and each person can also be either in or out of the hidden population H, producing four possible blocks: FH, $F\neg H$, $\neg F\neg H$, and $\neg FH$. (Here, we use the logical negation symbol, $\neg$, to denote not being in a group.) The probability of an edge between any two people i and j is then governed by a Bernoulli distribution whose mean is a function of the two block memberships:

$$\Pr(i \leftrightarrow j) \sim \text{Bernoulli}(\mu_{g(i),g(j)}), \qquad (\text{G.1})$$

where $g(i)$ is the block containing i, $g(j)$ is the block containing j, $i \leftrightarrow j$ denotes an undirected edge between i and j, and $\mu_{g(i),g(j)}$ is the probability of an edge be-

A75

tween a member of group $g(i)$ and a member of group $g(j)$. In a network with a no inhomogenous mixing (equivalent to an Erdos-Renyi random graph), $\mu_{g(i),g(j)}$ will be the same for all i and j. On the other hand, in a network with a high level of inhomogenous mixing, $\mu_{g(i),g(j)}$ will be relatively small when $g(i) \neq g(j)$ and $\mu_{g(i),g(j)}$ will be relatively large when $g(i) = g(j)^{15}$.

Each random network drawn under our simulation model depends on seven parameters. The first four parameters describe population size and group memberships; they are:

- N, the size of the population

- p_F, the fraction of people in the frame population

- p_H, the fraction of people in the hidden population

- $p_{F|H}$, the fraction of hidden population members also in the frame population

The next three parameters govern the amount of inhomogenous mixing in the network that connects people to each other:

- ζ, the probability of an edge between two people who are both in the same block.

- ξ, the relative probability of an edge between two vertices that differ in frame population membership. For example, a value of 0.6 would mean that the chances of having a connection between a particular person in F and a particular person not in F is 60% of the chance of a connection between two members of F or two members of $\neg F$.

---

[15]Computer code to perform the simulations was written in R (R Core Team, 2014) and used the following packages: devtools (Wickham and Chang, 2013); functional (Danenberg, 2013); ggplot2 (Wickham, 2009); igraph (Csardi and Nepusz, 2006); networkreporting (Feehan and Salganik, 2014); plyr (Wickham, 2011); sampling (Tillé and Matei, 2015); and stringr (Wickham, 2012).

$$\mathbf{M} = \begin{array}{c} \\ F\ H \\ F\neg H \\ \neg F\ H \\ \neg F\neg H \end{array} \begin{array}{cccc} F\ H & F\neg H & \neg F\ H & \neg F\neg H \\ \left(\begin{array}{cccc} \zeta & \rho\cdot\zeta & \xi\cdot\zeta & \xi\cdot\rho\cdot\zeta \\ \rho\cdot\zeta & \zeta & \xi\cdot\rho\cdot\zeta & \xi\cdot\zeta \\ \xi\cdot\zeta & \xi\cdot\rho\cdot\zeta & \zeta & \rho\cdot\zeta \\ \xi\cdot\rho\cdot\zeta & \xi\cdot\zeta & \rho\cdot\zeta & \zeta \end{array} \right) \end{array} \qquad (\text{G.2})$$

Figure G.1: The mixing matrix used to generate a random network using the stochastic block model. Entry (i, j) in the matrix describes the probability of an edge between two people, one of whom is in group i and one in group j. The probabilities are governed by ζ, ξ, and ρ. In our simulations, we generate networks with different amounts of inhomogenous mixing between hidden population members and non-hidden population members by fixing $\zeta = 0.05$ and $\xi = 0.4$, and then varying ρ from 0.1 (extreme inhomogenous mixing between hidden and non-hidden population members) to 1 (perfectly random mixing between hidden and non-hidden population members).

- ρ, the relative probability of an edge between two vertices that differ in hidden population membership. For example, a value of 0.8 would mean that the chances of having a connection between a particular person in H and a particular person not in H is 80% of the chance of a connection between two members of H or two members of $\neg H$.

Together, the parameters ζ, ξ, and ρ are used to construct the mixing matrix M (Figure G). Note that varying the parameter ρ will change several structural features of the network in addition to the amount of inhomogenous mixing; for example, changing ρ will alter the degree distribution. Our analytical results show that the generalized network scale-up estimator is robust to changes in these structural features.

The final parameter, τ_F, is used to control the amount of imperfect reporting. After randomly drawing a network using the stochastic block model, we generate a reporting network as follows:

1. convert all undirected edges $i \leftrightarrow j$ in the social network into two directed reporting edges in the reporting network: one $i \to j$ and one $j \to i$

2. select a fraction, $1 - \tau_F$, of the edges that lead from members of the frame pop-

ulation to members of the hidden population uniformly at random and remove them from the reporting graph.

Given a simple random sample of 500 members of the frame population and a relative probability sample of 30 members of the hidden population, the reporting graph is then used to compute the basic and generalized scale-up estimates for the size of the hidden population.

Across our simulations, we fix five of the parameters at constant values ($N = 5,000$; $p_F = 0.03$; $p_{F|H} = 1$; $\zeta = 0.05$; $\xi = 0.4$). We systematically explore varying the remaining parameters: we investigate ρ for values from 0.1 to 1 in increments of 0.1; we investigate p_F for values 0.1, 0.5, and 1; and we investigate τ_F for 0.1, 0.5, and 1. For each combination of the parameter values, we generate 10 random networks. Within each random network, we simulate 500 surveys. Each survey consists of two samples: a probability sample from the frame population, with sample size of 500; and a relative probability from the hidden population of size 30, with inclusion proportional to each hidden population member's personal network size. For each unique combination of parameters, we averaged the results across the surveys and across the randomly generated networks.