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

Dennis M. Feehan† and Matthew J. Salganik‡§

April 21, 2015

Abstract

The network scale-up method enables researchers to estimate the size 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. We propose a new generalized scale-up estimator that can be used in settings with non-random social mixing and imperfect awareness about membership in the hidden population. Further, 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, we develop interpretable adjustment factors that can be applied to the basic scale-up estimator. We conclude with practical recommendations for the design and analysis of future studies.

*The authors thank Alexandre Abdo, Francisco Bastos, Russ Bernard, Neilane Bertoni, Dimitri Fazito, Sharad Goel, Wolfgang Hladik, Jake Hofman, Mike Hout, Karen Levy, Rob Lyerla, Mary Mahy, Chris McCarty, Maeve Mello, Tyler McCormick, Damon Phillips, Justin Rao, Adam Slez, and Tian Zheng for helpful discussions. This research was supported by The Joint United Nations Programme on HIV/AIDS (UNAIDS), NSF (CNS-0905086), and NIH/NICHD (R01-HD062366, R01-HD075666, & R24-HD047879). Some of this research was conducted while MJS was an employee Microsoft Research. The opinions expressed here represent the views of the authors and not the funding agencies. Upon publication, we will make replication materials available through the archive of the Office of Population Research at Princeton University. Many of the methods described in this paper can be implemented in R package: networkreporting.

†Office of Population Research, Princeton University, Princeton, NJ, USA
‡Department of Sociology, Princeton University, Princeton, NJ, USA
1 Introduction

Many important problems in social science, public health, and public policy require estimates of the size 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 policy makers are unsatisfied with the ability of current statistical methods to provide these estimates (UNAIDS, 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 size 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: (i) social ties are formed completely at random (i.e., random mixing), (ii) respondents are perfectly aware of the characteristics of their alters, and (iii) respondents are able to provide accurate answers to survey questions about their personal networks. From the basic scale-up model Killworth et al. (1998b) 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% (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.

Unfortunately, the three assumptions underlying the basic scale-up model have all been shown to be
<table>
<thead>
<tr>
<th>Hidden population(s)</th>
<th>Location</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality in earthquake</td>
<td>Mexico City, Mexico</td>
<td>(Bernard et al., 1989)</td>
</tr>
<tr>
<td>Rape victims</td>
<td>Mexico City, Mexico</td>
<td>(Bernard et al., 1991)</td>
</tr>
<tr>
<td>HIV prevalence, rape, and homelessness</td>
<td>U.S.</td>
<td>(Killworth et al., 1998b)</td>
</tr>
<tr>
<td>Heroin use</td>
<td>14 U.S. cities</td>
<td>(Kadushin et al., 2006)</td>
</tr>
<tr>
<td>Choking incidents in children</td>
<td>Italy</td>
<td>(Snidero et al., 2007, 2009, 2012)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Ukraine</td>
<td>(Paniotto et al., 2009)</td>
</tr>
<tr>
<td>Heavy drug users</td>
<td>Curitiba, Brazil</td>
<td>(Salganik et al., 2011a)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Kerman, Iran</td>
<td>(Shokoohi et al., 2012)</td>
</tr>
<tr>
<td>Men who have sex with men</td>
<td>Japan</td>
<td>(Ezoe et al., 2012)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Almaty, Kazakhstan</td>
<td>(Scutelniciuc, 2012a)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Moldova</td>
<td>(Scutelniciuc, 2012b)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Thailand</td>
<td>(Aramrattan and Kanato, 2012)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Rwanda</td>
<td>(Rwanda Biomedical Center, 2012)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Chongqing, China</td>
<td>(Guo et al., 2013)</td>
</tr>
<tr>
<td>Groups most at-risk for HIV/AIDS</td>
<td>Tabriz, Iran</td>
<td>(Khounigh et al., 2014)</td>
</tr>
<tr>
<td>Men who have sex with men</td>
<td>Taiyuan, China</td>
<td>(Jing et al., 2014)</td>
</tr>
<tr>
<td>Drug and alcohol users</td>
<td>Kerman, Iran</td>
<td>(Sheikhzadeh et al., 2014)</td>
</tr>
</tbody>
</table>

Table 1: Network scale-up studies that have been completed.

problematic. Scale-up researchers call violations of the random mixing assumption barrier effects (Killworth et al., 2006; Zheng et al., 2006; Maltiel et al., 2013); they call violations of the perfect awareness assumption transmission error (Shelley et al., 1995, 2006; Killworth et al., 2006; Salganik et al., 2011b; Maltiel et al., 2013); and they call violations of the respondent accuracy assumption recall error (Killworth et al., 2003, 2006; McCormick and Zheng, 2007; Maltiel et al., 2013). In order to address some of these three problems, Maltiel et al. (2013) developed a more complex, flexible version of the basic scale-up model, but the quality of estimates from that model is still unknown.

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), 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 it permits researchers to make estimates in a much wider range of situations.

In the next section, 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 estimators. This comparison highlights the fact that the generalized scale-up estimator is consistent in a much wider range of situations than the basic scale-up estimator. We then propose a decomposition that separates the difference between them into three measurable and substantively meaningful factors (Equation 15). In Section 4, we
propose a new variance estimation procedure for both basic and generalized scale-up estimators, and we
demonstrate that it outperforms the existing variance estimator. In Section 5 we make practical recommend-
dations for the design and analysis of future scale-up studies, and in Section 6, we conclude with a summary
and an outline of next steps. Online Appendices A - G 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 as-
sumptions 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 sam-
ping (Lavallée, 2007). Consider a population of 7 people, 2 of whom are drug injectors (Figure 1(a)). 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 $i$.

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 1(b) 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:

$$\text{total out-reports} = \text{total in-reports}.$$  \hspace{1cm} (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. Further, let $y_{i,H}$ be the total number of out-reports
from person $i$ (i.e., person $i$’s answer to the question “How many drug injectors do you know?”). For example,
Figure 1(b) shows that person 5 would report knowing 1 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 1(b) shows person 5 would be reported as a drug injector by 3 people so $v_{5,U} = 3$. Since total out-reports must equal total in-reports, it must be the case that

$$y_{U,H} = v_{U,U},$$  \hspace{1cm} (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

$$N_H = \frac{y_{U,H}}{v_{U,U}/N_H}. \quad (3)$$

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. In standard survey research, restricting interviews to adults does not cause problems, but that is not the case with the scale-up method. As we show in Section 5.2, failure to consider the differences between the frame population and entire population requires the introduction of an awkward adjustment factor. Rather than attempting...
to make an adjustment after the fact, we derive 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 1(c)). 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 re-write Equation 4 as

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

(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). Since hidden populations like drug injectors are often stigmatized, it is reasonable to suspect that false negatives 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 introduce a positive bias into estimates from the generalized scale-up estimator. Therefore, in Online Appendix A we (i) formally define an interpretable measure of false positive reports, the precision of out-reports; (ii) analytically show the bias in size estimates as a function of the precisions of out-reports; and (iii) 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 population ($N_H$), 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 sample from the frame population and $\bar{v}_{H,F}$ with a sample from the hidden population.

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

$$\hat{y}_{F,H} = \sum_{i \in s} \frac{y_{i,H}}{\pi_i},$$  

where $s$ denotes the sample, $y_{i,H}$ denotes the reported number of connections between $i$ and $H$, and $\pi_i$ is $i$’s probability of inclusion from a conventional probabilistic sampling design from the frame population. Because $\hat{y}_{F,H}$ is a standard Horvitz-Thompson estimator, it is consistent and unbiased (Sarndal et al., 1992). This result is stated formally in Online Appendix B (Result B.1).

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. Fortunately, as we show below, estimating $\bar{v}_{H,F}$ only requires a relative probability sampling design in which we know respondents’ probabilities of inclusion up to a constant of proportionality, $c\pi_i$ (see Online Appendix C.1). Two widely used sampling methods designed for hard-to-reach populations offer this possibility: respondent-driven sampling (Heckathorn, 1997) and time-location sampling (Karon and Wejnert, 2012). Of course, any problems with these sampling methods could introduce problems into estimates.

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, etc.). 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 et al., 2011b; 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, etc). 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 Online Appendix C.4 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 \( \bar{v}_{i,A_j} \) 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:

\[
\hat{\bar{v}}_{H,F} = \frac{N_F}{N_A} \sum_{i \in s_H} \frac{\sum_{j} \bar{v}_{i,A_j}/(c_{\pi_i})}{\sum_{i \in s_H} 1/(c_{\pi_i})}.
\] (7)

Equation 7 is a standard weighted sample mean (Sarndal et al., 1992, Sec. 5.7) multiplied by a constant, \( N_F/N_A \). Result C.2 shows that, in order for this estimator to be consistent and essentially unbiased,\(^1\) two main conditions must be satisfied.

The first condition required for the estimator in Equation 7 is the *probe alter condition*, which 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:

\[
\frac{v_{H,A}}{N_A} = \frac{v_{H,F}}{N_F},
\] (8)

where \( v_{H,A} \) 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, \( N_A \) 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

\(^1\)We use the term essentially unbiased because Equation 7 is not, strictly speaking, unbiased; however, the large literature on ratio estimation confirms that these estimators are, for all practical purposes, unbiased (e.g. Sarndal et al., 1992, chap. 5). For a more precise discussion of this issue, see Online Appendix E.
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 Online Appendix C: (i) Result C.3 presents three other algebraically equivalent formulations of probe alter condition, some of which offer additional intuition; (ii) Result C.4 provides a method to empirically test the probe alter condition; and (iii) Result C.5 quantifies the bias in estimates when the probe alter condition is not satisfied.

The second condition required for the estimator in Equation 7 is **accurate aggregate reports about visibility**: 

$$\tilde{v}_{H,A} = v_{H,A},$$  \hspace{1cm} (9) 

where $\tilde{v}_{H,A}$ is the total reported visibility of members of the hidden population to the probe alters ($\sum_{i \in H} \sum_{j \in J} \tilde{v}_{i,Aj}$) and $v_{H,A}$ is the total actual visibility of members of the hidden population to the probe alters ($\sum_{i \in H} \sum_{j \in J} v_{i,Aj}$). 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, 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 over-estimate how much others know about them (Gilovich et al., 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 where 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 ($\tilde{v}_{i,A}$) and the information that we want ($v_{i,A}$). Unfortunately, there is currently no empirical evidence about the possible magnitude of these two problems in the context of scale-up studies. However, Result C.4 quantifies the bias introduced into estimates if the assumption of accurate aggregate reports about visibility is not correct.

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. Further, we can combine these components to form the generalized scale-up estimator:

\[
\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{v}_{H,F}}.
\] (10)

Result C.9 demonstrates that $\hat{N}_H$ will be consistent and essentially unbiased if the estimator for the numerator is consistent and unbiased, if the estimator for the denominator is consistent and essentially unbiased, and if there are no false positive reports.

One attractive feature of the generalized scale-up estimator (Equation 10) is that it is a simple combination of estimators that are themselves relatively simple. This structure enables us to derive very general robustness results about the impact of violations of various assumptions, either individually or jointly, on the estimated size of the hidden population. These robustness results are summarized in Online Appendix D.

### 3 Comparison between the generalized and basic scale-up estimators

Having introduced the generalized scale-up estimator, we now compare it to the basic scale-up estimator, which was used in most of the studies listed in Table 1. The analytic comparison below shows that whenever the basic scale-up estimator is consistent, the generalized scale-up estimator will also be consistent. Moreover, the results show that the generalized scale-up estimator is consistent in many situations where the basic scale-up estimator is inconsistent. These analytic results are illustrated with simulation results in Online Appendix G.

While the generalized scale-up estimator is derived from an identity relating in-reports and out-reports, the basic scale-up estimator was derived from a social network model. In order 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 et al., 1998b). Together, these three assumptions lead to the probabilistic model:

\[
y_{i,A_j} = d_{i,A_j} \sim \text{Binomial} \left( d_{i,U} \frac{N_{A_j}}{N} \right),
\] (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 a basic scale-up estimator:

\[
\hat{N}_H = \frac{\sum_{i \in s} y_{i,H}}{\sum_{i \in s} d_{i,U}} \times N. \tag{12}
\]

Killworth et al. (1998b) showed that Equation 12 is the maximum-likelihood estimator for \(N_H\) under the basic scale-up model, conditional on the assumption that \(d_{i,U}\) is known for each \(i \in s\).

In order to understand when the basic and generalized scale-up estimators might be the same, we will assume that we have a census of the entire population. In this case, we can write the basic scale-up estimator in Equation 12 as

\[
\hat{N}_H = \frac{y_{F,H}}{d_{F,U}} \times N = \frac{y_{F,H}}{\bar{d}_{U,F}}, \tag{13}
\]

where \(d_{F,U} = \sum_{i \in F} d_{i,U}\) and \(\bar{d}_{U,F} = d_{U,F}/N = d_{F,U}/N\).

Further, in the case of a census, we can rewrite the generalized scale-up estimator (Equation 10) as

\[
\hat{N}_H = \frac{y_{F,H}}{\bar{v}_{H,F}}. \tag{14}
\]

By comparing Equations 13 and 14, we conclude that the two estimators will be the same when \(\bar{v}_{H,F} = \bar{d}_{U,F}\).

This comparison reveals that the basic scale-up estimator is consistent only in the special case when \(\bar{v}_{H,F} = \bar{d}_{U,F}\), a result that follows from the fact that the generalized scale-up estimator is consistent (Result C.9). If researchers know that \(\bar{v}_{H,F} = \bar{d}_{U,F}\), then the basic scale-up estimator can be used to produce consistent estimates without the need for a sample from the hidden population. On the other hand, if it is possible that \(\bar{v}_{H,F} \neq \bar{d}_{U,F}\), then we recommend collecting additional data from a sample of the hidden population in order to use the generalized scale-up estimator.

In order to add intuition about situations where the basic and generalized scale-up estimators are different, we propose a decomposition that separates the difference between the two estimators into three measurable and substantively meaningful adjustment factors:

\[
N_H = \left(\frac{y_{F,H}}{\bar{d}_{U,F}}\right) \times \left(\frac{1}{d_{F,U}/d_{U,F}}\right) \times \left(\frac{1}{d_{H,F}/d_{F,F}}\right) \times \left(\frac{1}{\bar{v}_{H,F}/\bar{d}_{H,F}}\right) = \left(\frac{y_{F,H}}{\bar{v}_{H,F}}\right). \tag{15}
\]
We now describe each of the three adjustment factors in turn. First, we define the frame ratio, $\phi_F$, to be

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

$\phi_F$ can range from zero to infinity, and in most practical situations we expect $\phi_F$ will be greater than one. Result B.6 shows that we can make consistent and essentially unbiased estimates of $\phi_F$ from a sample of $F$.\(^2\)

Next, we define the degree ratio $\delta_F$ to be

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

$\delta_F$ ranges from zero to infinity, and it is less than one when the hidden population members have, on average, fewer connections to the frame population than frame population members. Result C.7 shows that we can to make consistent and essentially unbiased estimates of $\delta_F$ from samples of $F$ and $H$.

Finally, we define the true positive rate, $\tau_F$, to be

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

$\tau_F$ relates network degree to network reports.\(^3\) $\tau_F$ 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 $\tau_F$ to be to 0. Result C.8 shows that we can to make consistent and essentially unbiased estimates of $\tau_F$ from a sample of $H$.

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:

$$\text{bias}(\hat{N}_H^{\text{basic}}) \equiv \hat{N}_H^{\text{basic}} - N_H = \hat{N}_H^{\text{basic}} \left[1 - (\phi_F \delta_F \tau_F)\right].$$  \hfill (19)

Equation 20 reveals that if the product of the three adjustment factors ($\phi_F \delta_F \tau_F$) is 1, then the basic

\(^2\)Note that, since $\bar{d}_{U,F} = (N_F/N) \bar{d}_{F,U}$, an equivalent expression for the frame ratio is

$$\phi_F = \frac{\bar{d}_{F,F}}{\bar{d}_{F,U} (N_F/N)} = \frac{\bar{d}_{F,F}}{\bar{d}_{F,U} N_F} N.$$  

\(^3\)Note that the fact that in-reports must equal out-reports means that $\tau_F$ can also be defined

$$\tau_F = \frac{\text{# reported edges from } F \text{ actually connected to } H}{\text{# edges connecting } F \text{ and } H} = \frac{\bar{y}_{F,H}^\uparrow}{\bar{d}_{F,H}}.$$  

Here we have written $y_{F,H}^\uparrow$ to mean the true positive reports among the $y_{F,H}$; see Online Appendix A for a detailed explanation.
The comparison between the basic and generalized scale-up estimators leads to two main conclusions. First, our comparison shows that the generalized scale-up estimator is consistent in a much wider range of situations than the basic scale-up estimator, a result that we further confirm and illustrate with a simulation study in Online Appendix G. Second, as Equation 15 shows, if the adjustment factors $\phi_F$, $\delta_F$, and $\tau_F$ are known, then they can be used to improve basic scale-up estimates.

4 Variance estimation

In addition to point estimates, the focus of the paper thus far, researchers must also put accurate confidence intervals around their estimates. Unfortunately, the variance estimator currently used with the basic scale-up estimator has several problems and cannot be easily extended to accommodate the new generalized scale-up estimator (Equation 10). Therefore, we propose a new variance estimation approach based on the rescaled bootstrap (Rao et al., 1992; Rust and Rao, 1996). This approach both improves variance estimation for the basic scale-up estimator and enables variance estimation for the generalized scale-up estimator.

The variance estimator currently used in scale-up studies was proposed in Killworth et al. (1998b):

$$\hat{s}_e(\bar{N}_H) = \sqrt{\frac{N \cdot \bar{N}_H}{\sum_{i \in s} d_{i,U}}}.$$  \hspace{1cm} (21)

This estimator 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, all of which lead it to produce intervals that tend to be too small to include the true value as often as they should. First, the variance estimator in Equation 21 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 $\sum_{i \in s} d_{i,U}$ 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 measurement error. To address the first two problems (but not the third), in Online Appendix F we propose a rescaled bootstrap variance estimation procedure (Rao et al., 1992; Rust and Rao, 1996) that 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.

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. However, we can compare the coverage properties of several variance estimators when used in conjunction with the basic scale-up estimator using the interval validation procedure proposed in Killworth et al. (1998a). In particular, for 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)—for each group of known size, we produced a size estimate using the basic scale-up estimator (Equation 12), and we produced confidence intervals using (i) the current variance estimation procedure (Equation 21); (ii) the simple bootstrap, which does not account for complex sample designs; and (iii) the rescaled bootstrap, which does account for complex sample designs. This empirical evaluation (Figure 2) produced three main results. First, as expected, we found that the current variance estimation procedure performed poorly: purported 95% confidence intervals had empirical coverage rates of about 5%. Second, also consistent with expectation, we found that the rescaled bootstrap procedure 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 procedure 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 do not know the exact reason for the poor coverage rate of the rescaled bootstrap, but we suspect that it is caused by biases in the basic scale-up estimator (see Equation 15) and non-sampling errors not captured in the resampling procedure.

These empirical results, and the theoretical arguments in Online Appendix F, lead us to three conclusions. First, confidence intervals from the rescaled bootstrap procedure are preferable to intervals from the current scale-up variance estimation procedure. Second, researchers should expect that the confidence intervals from the rescaled bootstrap procedure will understate the true uncertainty in scale-up estimates until there is a substantially better understanding of non-sampling errors. Third, variance estimation for the scale-up method is an important area for further research.

5 Recommendations for practice

Our analytic results in Sections 2 and 3 lead to 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 also collect a sample from the hidden population so that they can use the generalized scale-up estimator. In Section 5.1 we present recommendations for researchers who can collect a sample from both the frame population, $F$, and the hidden population, $H$. However, we also acknowledge that it might not be possible to

---

4 Computer code to perform these calculations was written in R (R Core Team, 2014) and used the following packages: networkreporting (Feehan and Salganik, 2014); ggplot2 (Wickham, 2009); devtools (Wickham and Chang, 2013); plyr (Wickham, 2011); functional (Danenberg, 2013); and stringr (Wickham, 2012).
Figure 2: Assessing variance estimation 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 21), contains the true group size for 0.0%, 9.1%, and 5.0% of the groups.
collect a sample from the hidden population in all situations, so in Section 5.2 we present recommendations for researchers who can only select a sample from the frame population.

## 5.1 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):

\[
\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{\theta}_{H,F}}.
\]  

(22)

For researchers using the generalized scale-up estimator we have three specific recommendations. First, 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 C.5 for more advice about choosing probe alters). Second, because the generalized scale-up estimator depends on assumptions (e.g., assumptions about the probe alters and reporting behavior), we recommend that researchers assess the robustness of their estimates to any assumptions that may be particularly problematic in their setting. Fortunately, the generalized scale-up estimator is a simple combination of relatively simple estimators so it is possible to assess robustness analytically; see Online Appendix A and Online Appendix D for more information. Finally, we recommend that researchers produce confidence intervals around their estimate using the rescaled bootstrap procedure presented in Section 4, keeping in mind that this will likely produce intervals that are too small.

We also have two 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 this data so that the basic and generalized estimators can be compared. Also, we recommend that researchers publish estimates of \( \delta_F \) and \( \tau_F \), although these quantities play no role in the generalized scale-up estimator (Fig. 3). As a body of evidence about these adjustment factors accumulates (e.g., Salganik et al. (2011a); 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.

## 5.2 Estimation with only a sample from \( F \)

If researchers cannot collect a sample from the hidden population, we have four recommendations. First, we recommend two simple changes to the basic scale-up estimator that remove the need to adjust for the frame
Figure 3: 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 $\delta_F$ and $\tau_F$ in order to aid other researchers.

ratio, $\phi_F$. Recall, that the basic scale-up estimator that has been used in previous studies is (see Section 3):

$$\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{d}_{F,U}} \times N = \frac{\hat{y}_{F,H}}{\hat{d}_{F,U}/N}. \quad (23)$$

Instead of Equation 23, we suggest a new 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):

$$\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{d}_{F,F}} \times N = \frac{\hat{y}_{F,H}}{\hat{d}_{F,F}/N_F}. \quad (24)$$

There are two differences between our proposed basic scale-up estimator (Equation 24) and the one used in previous studies (Equation 23). First, we recommend that researchers estimate $\hat{d}_{F,F}$ (i.e., the total number of connections between adults and adults) rather than $\hat{d}_{F,U}$ (i.e., the total number of connections between adults and everyone). In order 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 Online Appendix B.4 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 $N$. These two simple changes remove the need to adjust for the frame ratio $\phi_F$, and thereby eliminate an assumption about an unmeasured quantity. An adjusted
version of this proposed basic scale-up estimator would then be:

\[
\hat{N}_H = \frac{\hat{y}_{F,H}}{(d_F/N_F)} \times \frac{1}{\delta_F} \times \frac{1}{\tau_F}
\]

Our second recommendation for researchers using a basic scale-up estimator is to be explicit about the values that they assume for \(\delta_F\) and \(\tau_F\). Further, like researchers using the generalized scale-up estimator, we recommend that researchers using a basic scale-up estimator assess the robustness of their estimates to any assumptions that may be particularly problematic in their setting. These robustness assessments can be done analytically using the ideas developed in Online Appendix A and Online Appendix D. Finally, we recommend that researchers produce confidence intervals around their estimate using the rescaled bootstrap procedure presented in Section 4, while explicitly accounting for the fact that there is uncertainty around assumed adjustment factors, and bearing in mind that this will likely produce intervals that are too small.

6 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 non-random 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, revealing that the generalized scale-up estimator will be consistent in a much wider range of situations.

The approach that 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 where 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 where these studies are not already planned, researchers can collect their own data from the hidden population, or they can borrow estimated adjustment factors from other published studies.

The generalized scale-up estimator, like all estimators, depends on a number of assumptions. We think two 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 (Online Appendix A), 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. The results in Table D.1 show how violations of these assumptions—no false positive reports and accurate aggregate awareness about visibility—impact the resulting size estimates.

Our results and their limitations highlight several directions for further work, and we think that two are most urgent. First, researchers need practical techniques for estimating the rate of false positive reporting. These estimates, combined with the estimator in Online Appendix A, would permit the relaxation of one of the most important remaining assumptions made by all scale-up studies to date. Second, the proposed bootstrap variance estimator improves on the current variance estimation procedure, but still performs poorly in an absolute sense. We believe that more research on non-sampling error could lead to more realistic confidence intervals. We hope that the framework developed in this paper will provide a basis for this and other future research.
References


Wickham, H. (2012). *stringr: Make it easier to work with strings*. R package version 0.6.2.


Online Appendices
A Estimation with 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,

\[
y_{i,H} = y_{i,H}^+ + y_{i,H}^-.
\]

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

\[
y_{F,H} = y_{F,H}^+ + y_{F,H}^-.
\]

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

\[
y_{F,H}^+ = v_{H,F} \quad \text{(A.3)}
\]

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

\[
N_H = \frac{y_{F,H}^+}{\bar{v}_{H,F}}. \quad \text{(A.4)}
\]

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, $\eta_F$: 

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

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. Even if false-positive reports are possible, the precision allows us to derive an identity that relates out-reports to $N_H$:

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

Equation A.6 then suggests the estimator:

$$\hat{N}_H = \frac{\hat{\eta}_F \hat{y}_{F,H}}{\hat{\bar{v}}_{H,F}}. \quad (A.7)$$

If we could find a consistent and essentially unbiased estimator for $\eta_F$, 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 a study such as this.

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, $\phi_F$, 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 $\{s_1, \ldots, s_j, \ldots, s_{\text{max}}\}$, 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, $y_{F,H}$

Estimating the total number of out-reports is a straightforward application of a standard survey estimator, as long as we have a probability sample from the frame.

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 $\pi_i$ (Sec. B.1). Then the estimator given by

$$\hat{y}_{F,H} = \sum_{i \in s_F} y_{i,H} / \pi_i$$  \hspace{1cm} (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 \textit{probe alters}, $A$, is a multiset. The size of $A$ is $N_A = \sum_j N_{A_j}$.

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

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 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 make it clear 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.

\textbf{Property B.2} Suppose we have a sample $s_F$ from $F$ taken using a probability sampling design with probabilities of inclusion $\pi_i$ (Section B.1). Then

$$\hat{y}_{F,A} = \sum_{i \in s_F} y_{i,A}/\pi_i$$

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

\textbf{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

$$\hat{y}_{F,A} = \sum_{i \in s_F} a_i/\pi_i.$$  \hfill (B.3)

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

A4
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, $d_{i,U}$ must be conceptually well-defined, but it may not be very easy to directly observe. 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).\footnote{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).} 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 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 5.2 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 $F$ ($\bar{d}_{F,F}$).\footnote{Although we have framed our discussion here in terms of $\bar{d}_{F,F}$, the same ideas apply to $\bar{d}_{U,F}$ and $\bar{d}_{F,U}$.} 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, $A$

Result B.3, below, shows that the known population estimator will produce consistent and unbiased estimates of average network degree if (i) $y_{F,A} = d_{F,A}$ (reporting condition); and (ii) $\bar{d}_{A,F} = \bar{d}_{F,F}$ (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 \( A \) will enable consistent and unbiased estimates.

First, the reporting condition \( (y_{F,A} = d_{F,A}) \) 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 et al., 2006; Salganik et al., 2011b; Maltiel et al., 2013). 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., 2013). 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., 2013).

Second, the probe alter condition \( (\bar{d}_{A,F} = \bar{d}_{F,F}) \) 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 \( (\bar{d}_{A,F}) \). Basic algebraic manipulation shows that the probe alter condition can be written as:

\[
\frac{\sum_j \bar{d}_{A_j,F} N_{A_j}}{\sum_j N_{A_j}} = \bar{d}_{F,F}.
\]

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 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 robustness 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 robustness 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 $\pi_i$ (see Section B.1). Suppose also that we have a set of known populations, $A$. Then the known population estimator given by

$$\hat{\bar{d}}_{F,F} = \frac{\sum_{i \in s_F} \sum_j y_{i,A_j} / \pi_i}{N_A} \quad (B.5)$$

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

$$y_{F,A} = d_{F,A}, \quad \text{(reporting condition)} \quad (B.6)$$

and if

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

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

**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 $\pi_i$ (see Section C.1). Suppose also that we have a set of known populations, $A$. Then the known population estimator given by

$$\hat{\bar{d}}_{U,F} = \frac{\sum_{i \in s_F} \sum_j y_{i,A_j} / \pi_i}{N_A} \quad (B.8)$$

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

$$y_{F,A} = d_{F,A}, \quad \text{(reporting condition)} \quad (B.9)$$
and if

\[ \bar{d}_{A,F} = \bar{d}_{U,F}. \]  
(probe alter condition) \hspace{1cm} (B.10)

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

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,

\[ \hat{\bar{d}}_{F,U} = \frac{N}{N_F} \]  
(B.11)

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

**B.5 Estimating the frame ratio, \( \phi_F \)**

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, \( \phi_F \).

**Result B.6** The estimator

\[ \hat{\phi}_F = \frac{\hat{\bar{d}}_{F,F}}{\hat{\bar{d}}_{U,F}} \]  
(B.12)

is consistent and essentially unbiased for \( \phi_F \) if \( \hat{\bar{d}}_{F,F} \) is consistent and essentially unbiased for \( \bar{d}_{F,F} \) and \( \hat{\bar{d}}_{U,F} \) 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 \( A_{F_1} \) for \( \bar{d}_{F,F} \), and \( A_{F_2} \) for \( \bar{d}_{U,F} \), we obtain

\[ \hat{\phi}_F = \frac{N_{A_{F_2}}}{N_{A_{F_1}}} \sum_{i \in s_F} \frac{\sum_{A_j \in A_{F_1}} y_{i,A_j}/\pi_i}{\sum_{i \in s_F} \sum_{A_k \in A_{F_2}} y_{i,A_k}/\pi_i}. \]  
(B.13)

In our discussion of \( \bar{d}_{F,F} \) (Result B.3) and \( \bar{d}_{U,F} \) (Result B.4), we concluded that we want the known populations \( A_{F_1} \) used for \( \bar{d}_{F,F} \) to be typical of members of \( F \) in their connections to \( F \). An analogous argument
shows that we want the known populations $A_{F_2}$ used for $\hat{d}_{U,F}$ 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 $\hat{d}_{F,F}$ and $\hat{d}_{U,F}$. If researchers wish to estimate $\phi_F$ directly, one approach would be to choose $A_{F_2}$ 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 $A_{F_1}$. In this case, researchers would also want $\frac{N_{A_{F_1}}}{N_{A_{F_2}}} \approx \frac{N_F}{N}$. This complication is one of the reasons we recommend in Section 5 that future scale-up studies estimate $\bar{d}_{F,F}$ directly, thus avoiding the need to estimate $\phi_F$ 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. First, we define the general requirements that our sampling design for $H$ must satisfy (Section C.1). Then we describe a flexible data collection procedure called the game of contacts (Section C.2). Next, we introduce some background material on estimation using questions about multisets (Section C.3) and present an estimator for $\bar{v}_{H,F}$, the average number of in-reports among the members of the hidden population (Section C.4). Then, we present estimators for the two adjustment factors introduced in Section 3: the degree ratio, $\delta_F$, and the true positive rate, $\tau_F$ (Section C.6). Finally, we present formal results for four different estimators for $N_H$ (Section C.7).

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 Hajek 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 $A \cap F$ 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 $A \cap F$ 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).

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...
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).

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

$$\hat{y}_{H,A} = \frac{\sum_{i \in s_H} y_{i,A} / (cπ_i)}{\sum_{i \in s_H} 1 / (cπ_i)}$$

(C.1)

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

**Proof:** Note that the $c$ in the relative probabilities of inclusion $cπ_i$ cancel, so that

$$\hat{y}_{H,A} = \frac{\sum_{i \in s_H} y_{i,A} / (π_i)}{\sum_{i \in s_H} 1 / (π_i)}$$

(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

$$\hat{y}_{H,A} = \frac{\sum_{i \in s_H} a_i / π_i}{\sum_{i \in s_H} 1 / π_i}.$$  

(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 $\frac{1}{N_H} \sum_{i \in H} a_i = y_{H,A}/N_H$. 

**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 two conditions are satisfied: one about reporting and one about the visibility of the hidden population to the probe alters.

**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 $cπ_i$ for all sampled elements (Sec. C.1). Then

$$\hat{v}_{H,F} = \frac{N_F}{N_{A\cap F}} \frac{\sum_{i \in s_H} \sum_j \bar{v}_{i,A_j \cap F} / (cπ_i)}{\sum_{i \in s_H} 1 / (cπ_i)}$$

(C.4)
is consistent and essentially unbiased for $\bar{v}_{H,F}$ if

$$\bar{v}_{H,A \cap F} = v_{H,A \cap F}, \quad \text{(reporting condition)}$$

and

$$\frac{v_{H,A \cap F}}{N_{A \cap F}} = \frac{v_{H,F}}{N_F}. \quad \text{(probe alter condition)}$$

**Proof:** Property C.1 holds for estimating $\bar{v}_{F,A \cap F}$ from $\bar{v}_{i,A \cap F}$, just as it holds for estimating $\bar{y}_{H,A \cap F}$ from $\bar{y}_{i,A \cap F}$. Applying Property C.1 here, we conclude that the estimator is consistent and essentially unbiased for

$$\frac{N_F}{N_{A \cap F}} \bar{v}_{H,A \cap F} = \frac{N_F}{N_{A \cap F}} \bar{v}_{H,F}. \quad \text{(C.7)}$$

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

$$\frac{N_F}{N_{A \cap F}} \bar{v}_{H,A \cap F} = \frac{N_F}{N_{A \cap F}} v_{H,A \cap F}. \quad \text{(C.8)}$$

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

$$\frac{N_F}{N_{A \cap F}} v_{H,A \cap F} = \frac{N_F}{N_{A \cap F}} \frac{v_{H,F}}{N_H} \frac{N_H}{N_F} \quad \text{(C.9)}$$

$$= \bar{v}_{H,F} \quad \text{(C.10)}$$

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

$$N_{A \cap F} = \frac{N_F}{N} N_A. \quad \text{(C.11)}$$

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

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

\[
\tau_i = \frac{v_{H,i}}{d_{i,H}},
\]

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

\[
\tau_F = \frac{1}{N_F} \sum_{i \in F} \tau_i.
\]

In general, \( \tau_F \neq \tau_F \). To see this, note that while \( \tau_F \) is the average of the individual-level true positive rates with each individual weighted equally, \( \tau_F \) 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 \( \tau_F \) in terms of the \( \tau_i \):

\[
\tau_F = \frac{\sum_{i \in F} \tau_i d_{i,H}}{\sum_{i \in F} d_{i,H}},
\]

since multiplying each \( \tau_i \) 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) \( \frac{v_{H,A \cap F}}{N_{A \cap F}} = \frac{v_{H,F}}{N_F} \)

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

(iii) \( \tau_{A \cap F} \bar{d}_{A \cap F,H} + \text{cov}_{A \cap F}(\tau_i, d_{i,H}) = \tau_F \bar{d}_{F,H} + \text{cov}_F(\tau_i, d_{i,H}) \)

(iv) \( \bar{y}^+_{F,H} = \frac{\sum_{i \in F} \bar{y}_{\lambda_i \cap A \cap F} N_{\lambda_i \cap F}}{\sum_{i \in F} N_{\lambda_i \cap F}} \)

where \( \text{cov}_F \) is the finite-population covariance taken over the set \( F \).\(^7\)

\(^7\)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.
**Proof:** First, we show that

\[
\tau_{A \cap F} \bar{d}_{A \cap F, H} = \tau_F \bar{d}_{F, H} \iff \frac{v_{H, A \cap F}}{N_{A \cap F}} = \frac{v_{H, F}}{N_F}.
\]

(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 \(\tau_{A \cap F} \bar{d}_{A \cap F, H} = v_{H, A \cap F}/N_A\). We conclude that \((i) \iff (ii)\).

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

\[
\tau_F \bar{d}_{F, H} = \sum_{i \in F} \tau_i d_{i, H} = N_F [\tau_F \bar{d}_{F, H} + \text{cov}_F(\tau_i, d_{i, H})].
\]

(C.16)

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

\[
\tau_F \bar{d}_{F, H} = \tau_F \bar{d}_{F, H} + \text{cov}_F(\tau_i, d_{i, H}).
\]

(C.17)

The same argument shows that

\[
\bar{d}_{A \cap F, H} \tau_{A \cap F} = \tau_{A \cap F} \bar{d}_{A \cap F, H} + \text{cov}_{A \cap F}(\tau_i, d_{i, H}).
\]

(C.18)

So we conclude that \((ii) \iff (iii)\).

Finally, we show that \((iv)\) is equivalent to \((i)\). In Appendix A, showed that \(\bar{y}_{F, H}^+ = v_{H, F}\) (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

\[
\frac{v_{H, A \cap F}}{N_{A \cap F}} = \frac{\sum_j v_{H, A_j \cap F}}{\sum_j N_{A_j \cap F}} = \frac{\sum_j \bar{y}_{A_j \cap F, H}^+}{\sum_j N_{A_j \cap F}} = \frac{\sum_j \bar{y}_{A_j \cap F, H} N_{A_j \cap F}}{\sum_j N_{A_j \cap F}}.
\]

(C.19)

So we conclude that \((i) \iff (iv)\).

Since \((i) \iff (ii)\) and \((ii) \iff (iii)\), it follows that \((i) \iff (iii)\). Furthermore, since \((i) \iff (iv)\), 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 \( A \cap F \):

\[
\frac{\bar{y}_{F,H}^+}{\bar{y}_{F,H}} = \frac{\bar{y}_{A \cap F,H}^+}{\bar{y}_{A \cap F,H}}
\]  

(C.20)

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

\[
\bar{y}_{F,H} = \bar{y}_{A \cap F,H}.
\]  

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

\[
\bar{y}_{F,H}^+ = \bar{y}_{A \cap F,H}^+.
\]  

(C.23)

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

\[
\frac{\bar{y}_{F,H}^+}{\bar{y}_{F,H}} = \frac{\bar{y}_{A \cap F,H}^+}{\bar{y}_{A \cap F,H}},
\]  

(C.24)

by multiplying the left-hand side by \( \frac{N_{F}}{N_{F}} \) and the right-hand side by \( \frac{N_{A \cap F}}{N_{A \cap F}} \). 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 \( \bar{y}_{A \cap F,H} \); and (ii) researchers assume that the precision of out-reports from the frame population is the same as the precision of out-reports from \( A \cap F \), then they can evaluate how well the probe alter condition is satisfied empirically by comparing \( \hat{y}_{F,H} \) and \( \hat{y}_{A \cap F,H} \).

Finally, we can foresee three 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. Finally, a third problem might arise due to errors in administrative records which would cause researchers to have incorrect information about the size of the multiset of probe alters on the frame \( (N_{\mathcal{A} \cap F}) \). Fortunately, as we show in Result C.5, 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.

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

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

\[
\hat{v}^*_{\mathcal{F}, \mathcal{H}} = \frac{1}{c_1} \frac{N_F}{N_{\mathcal{A} \cap F}} \frac{\sum_{i \in \mathcal{H}} \sum_j \hat{v}_{i, \mathcal{A} \cap F} / (c \pi_i)}{\sum_{i \in \mathcal{H}} 1 / (c \pi_i)}. \tag{C.25}
\]

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 \( \hat{v}^*_{\mathcal{F}, \mathcal{H}} \) is consistent and essentially unbiased for

\[
\frac{1}{c_1} \frac{N_F}{N_{\mathcal{A} \cap F}} \frac{\hat{v}_{\mathcal{H}, \mathcal{A} \cap F}}{N_H} = \frac{c_2}{c_1} \frac{N_F}{N_{\mathcal{A} \cap F}} \frac{v_{\mathcal{H}, \mathcal{A} \cap F}}{N_H} = \frac{c_3 c_2}{c_1} \frac{N_F}{N_{\mathcal{A} \cap F}} \frac{v_{\mathcal{H}, \mathcal{F}}}{N_H}. \tag{C.26}
\]

In words, the estimand is now incorrect by \( \frac{c_3 c_2}{c_1} \). Since \( \hat{v}^*_{\mathcal{F}, \mathcal{H}} \) is consistent and essentially unbiased for \( \hat{v}_{\mathcal{F}, \mathcal{H}} \), we conclude that \( \hat{v}^*_{\mathcal{F}, \mathcal{H}} \) is consistent and essentially unbiased for \( \frac{c_3 c_2}{c_1} \hat{v}_{\mathcal{F}, \mathcal{H}} \). Note that if the assumptions needed for Result C.2 hold, then \( c_1 = 1, c_2 = 1, \) and \( c_3 = 1 \), giving us our original result.

We present similar robustness results for all of our estimators in Appendix D.

### 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 robustness 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 \( \frac{v_{H,A}}{N_A} = \frac{v_{H,F}}{N_F} \). 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 \( \bar{v}_{H,A} = v_{H,A} \). 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., 2013).

In practice it might be difficult to meet each of these three conditions exactly, therefore we recommend a robustness 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, \( \delta_F \), then they should design the probe alters \( A \) so that they can be asked of both members of the hidden population sample and members of the frame population sample (see Result C.7). 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: $\delta_F$ and $\tau_F$

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

$$\delta_F = \frac{\bar{d}_{H,F}}{d_{F,F}} \quad (C.27)$$

and the true positive rate,

$$\tau_F = \frac{\bar{v}_{H,F}}{d_{H,F}}. \quad (C.28)$$

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.27 and C.28 make clear, both adjustment factors involve $\bar{d}_{H,F}$ so we first present an estimator for that quantity.

Result C.6 Suppose we have a sample $s_H$ taken from the hidden population using a relative probability sampling design with relative probabilities of inclusion denoted $c\pi_i$ (Sec C.1). Then the estimator given by

$$\hat{d}_{H,F} = \frac{N_F}{N_{A\cap F}} \frac{\sum_{i \in s_H} \sum_j y_{i,(A\cap F)}/(c\pi_i)}{\sum_{i \in s_H} 1/(c\pi_i)} \quad (C.29)$$

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

$$y_{H,A\cap F} = d_{H,A\cap F}, \quad \text{(reporting condition)} \quad (C.30)$$
\[ d_{A \cap F, H} = d_{F, H}. \quad \text{(probe alter condition)} \]  

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

\[
\frac{N_F}{N_{A \cap F}} \frac{y_{H, A \cap F}}{N_H} = \frac{N_F}{N_{A \cap F}} \frac{y_{H, A \cap F}}{N_H}.
\]  

(C.32)

Under the reporting condition (Equation C.30) this becomes

\[
\frac{N_F y_{H, A \cap F}}{N_H N_{A \cap F}} = \frac{N_F d_{H, A \cap F}}{N_H N_{A \cap F}}.
\]  

(C.33)

Finally, applying the probe alter condition in Equation C.31, we have

\[
\frac{N_F d_{H, A \cap F}}{N_H N_{A \cap F}} = \frac{N_F d_{F, H}}{N_H N_F}.
\]  

(C.34)

\[ = d_{H, F}. \]  

(C.35)

\[ \blacksquare \]

Result C.6 requires that reports are, in total, correct (Equation C.30). Like Result C.2, Result C.6 also requires us to know the size of the probe alters on the frame, \( N_{A \cap F} \). In some cases, this may not be readily available, but it may be reasonable to assume that

\[
N_{A \cap F} = \frac{N_F}{N} N_A.
\]  

(C.36)

Furthermore, if \( A \) is chosen so that all of its members are in \( F \), then \( N_{A \cap F} = N_A \) 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 \( A \cap F \); we can just ask about connections to \( A \). Result C.6 also requires a specific rate of connectivity between the probe alters and the hidden population (Equation C.31). We discussed some of the consequences of these assumptions in the main text, where we reviewed the implications of our results for practice (Section 5).

**C.6.1 Estimating the degree ratio, \( \delta_F \)**

We can combine our estimator for \( d_{H, F} \) (Result C.6) and our estimator for \( d_{F, F} \) (Result B.3), to estimate the degree ratio, \( \delta_F \).
**Result C.7** The estimator

\[
\hat{\delta}_F = \frac{\hat{d}_{H,F}}{\hat{d}_{F,F}}
\]  
(C.37)

is consistent and essentially unbiased for \(\delta_F\) if \(\hat{d}_{H,F}\) is consistent and essentially unbiased for \(d_{H,F}\) and \(\hat{d}_{F,F}\) is consistent and essentially unbiased for \(d_{F,F}\).

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

More concretely, combing the estimators in Result C.6 and Result B.3, results in an estimator for \(\hat{\delta}_F\) with the following form:

\[
\hat{\delta}_F = \frac{\sum_{i \in s_H} \sum_{A_j \in A_H} y_{i,(A_j \cap F)}/(c_i^H)}{\sum_{i \in s_H} 1/(c_i^H)} \frac{1}{\sum_{i \in s_F} \sum_{A_k \in A_F} y_{i,A_k}/\pi_i^F}
\]  
(C.38)

If the probe alters for the frame population and the hidden population are the same, so that \(A_H = A_F = A\), and if the probe alters are randomly distributed in the frame population in the sense that

\[
N_{A \cap F} = N_A \frac{N_F}{N},
\]  
(C.39)

then we can reduce the constants in front of Equation C.38 to

\[
\frac{N_F}{N_{A \cap F}} \frac{1}{N_A} = \frac{N}{N_A} = N.
\]  
(C.40)

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 \(A\) drop out. This fact allows researchers to use groups defined by first names (e.g., people named Michael) in the probe alters \(A\), even if the size of these groups is not known, as long as it is reasonable to assume that \(A\) satisfies Equation C.39 (c.f., Salganik et al. (2011a)).

**C.6.2 Estimating the true positive rate, \(\tau_F\)**

We can combine our estimator for \(\hat{v}_{H,F}\) (Result C.2) and our estimator for \(\hat{d}_{H,F}\) (Result C.6) to estimate the true positive rate \(\tau_F\).

**Result C.8** The estimator

\[
\hat{\tau}_F = \frac{\hat{v}_{H,F}}{\hat{d}_{H,F}}
\]  
(C.41)
is consistent and essentially unbiased for $\tau_F$ if $\hat{\bar{v}}_{H,F}$ is a consistent and essentially unbiased estimator of $\bar{v}_{H,F}$ and if $\hat{d}_{H,F}$ is a consistent and essentially unbiased estimator of $d_{H,F}$.

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

More concretely, combing the estimator in Result C.2 and Result C.6 yields an estimator for $\hat{\tau}_F$ with the following form:

$$
\hat{\tau}_F = \frac{\sum_{i \in s_H} \tilde{v}_i A_H / (c \pi_i)}{\sum_{i \in s_H} y_i A_H / (c \pi_i)}.  \tag{C.42}
$$

All of the factors involving the size of $A$ drop out of Equation C.42. This fact allows researchers to use groups defined by first names (e.g., people named Michael) in the probe alters $A$, 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.9** The generalized scale-up estimator given by

$$
\tilde{N}_H = \frac{\hat{y}_{F,H}}{\bar{v}_{H,F}} \tag{C.43}
$$

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

**Proof:** From the properties of a compound ratio estimator, we know that our estimator is consistent and essentially unbiased for $y_{F,H}/\bar{v}_{H,F}$ (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.10** The adjusted basic scale-up estimator given by

$$
\tilde{N}_H = \frac{\hat{y}_{F,H}}{\hat{d}_{U,F} \phi_F} \frac{1}{\hat{\delta}_F} \frac{1}{\hat{\tau}_F} \tag{C.44}
$$

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 Appendix E, we know that this compound ratio estimator will be consistent and essentially unbiased for $y_{F,H} / (\hat{d}_{U,F} \phi_F \hat{\delta}_F \hat{\tau}_F)$. 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 $N_H$. \hfill \blacksquare

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

\[
\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{d}_{F,F}} \frac{1}{\delta_F} \frac{1}{\tau_F} \tag{C.45}
\]

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 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 $N_H$. \hfill \blacksquare

**Result C.12** The adjusted scale-up estimator

\[
\hat{N}_H = \frac{\hat{y}_{F,H}}{\hat{d}_{F,F}} \frac{1}{\delta_F} \frac{1}{\tau_F} \hat{\eta}_F \tag{C.46}
\]

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

**Proof:** From the results in 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 $\hat{y}_{F,H}$ by construction and the product of the denominators is $\bar{v}_{H,F}$ by construction, leaving us with $\hat{y}_{F,H}/\bar{v}_{H,F}$. By the argument in Appendix A this quantity is equal to $N_H$. \hfill \blacksquare

**D Robustness**

All of the estimators that we propose require that specific conditions hold in order to produce consistent and essentially unbiased estimates. Result C.5 shows how we can quantify the extent to which one of these estimators is robust to violations of the conditions it depends upon.

Table D.1 shows results analogous to Result C.5 for all of the estimators we propose. We do not prove each one individually, since the derivations all follow the pattern of Result C.5 very closely. Researchers who wish to understand the how their estimates are affected by the assumptions they make can use Table D.1 to produce sensitivity analysis.
<table>
<thead>
<tr>
<th>Estimator</th>
<th>Imperfect assumptions</th>
<th>Effective estimand</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{d}_{F,F} ) (Result B.3)</td>
<td>(i) ( \hat{N}_A = c_1 N_A )</td>
<td>( \frac{c_2}{c_1} \hat{d}_{F,F} )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{d}<em>{A,F} = c_2 \hat{d}</em>{F,F} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iii) ( y_{F,A} = c_3 d_{F,A} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{d}_{U,F} ) (Result B.4)</td>
<td>(i) ( \hat{N}_A = c_1 N_A )</td>
<td>( \frac{c_2}{c_1} \hat{d}_{U,F} )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{d}<em>{A,F} = c_2 \hat{d}</em>{U,F} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iii) ( y_{F,A} = c_3 d_{F,A} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{\phi}_F ) (Result B.6)</td>
<td>(i) ( \hat{d}<em>{F,F} \rightsquigarrow c_1 \hat{d}</em>{F,F} )</td>
<td>( \frac{c_1}{c_2} \phi_F )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{d}<em>{U,F} \rightsquigarrow c_2 \hat{d}</em>{U,F} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{v}_{H,F} ) (Result C.2)</td>
<td>(i) ( \hat{N}<em>{A \cap F} = c_1 N</em>{A \cap F} )</td>
<td>( \frac{c_3}{c_1} \hat{v}_{H,F} )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{v}<em>{H,A \cap F} = c_2 v</em>{H,A \cap F} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iii) ( \frac{v_{H,A \cap F}}{N_{A \cap F}} = c_3 \frac{v_{H,F}}{N_F} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{\delta}_F ) (Result C.7)</td>
<td>(i) ( \hat{d}<em>{H,F} \rightsquigarrow c_1 \hat{d}</em>{H,F} )</td>
<td>( \frac{c_1}{c_2} \delta_F )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{d}<em>{F,F} \rightsquigarrow c_2 \hat{d}</em>{F,F} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{\tau}_F ) (Result C.8)</td>
<td>(i) ( \hat{v}<em>{H,F} \rightsquigarrow c_1 \hat{v}</em>{H,F} )</td>
<td>( \frac{c_1}{c_2} \tau_F )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{d}<em>{H,F} \rightsquigarrow c_2 \hat{d}</em>{H,F} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{N}_H ) (Result C.9)</td>
<td>(i) ( \hat{v}<em>{H,F} \rightsquigarrow c_1 \hat{v}</em>{H,F} )</td>
<td>( \frac{1}{c_1} N_H )</td>
</tr>
<tr>
<td>( \hat{N}_H ) (Result C.11)</td>
<td>(i) ( \hat{d}<em>{F,F} \rightsquigarrow c_1 \hat{d}</em>{F,F} )</td>
<td>( \frac{1}{c_1 c_2 c_3} N_H )</td>
</tr>
<tr>
<td></td>
<td>(ii) ( \hat{\delta}_F \rightsquigarrow c_2 \delta_F )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iii) ( \hat{\tau}_F \rightsquigarrow c_3 \tau_F )</td>
<td></td>
</tr>
</tbody>
</table>

Table D.1: Summary of estimators’ robustness to imperfect assumptions. The first column lists several estimators we discuss in the main body and appendixes. The consistency and approximate unbiasedness of each estimator relies upon particular 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 \( \hat{d}_{F,F} \) 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 \( A \) have been chosen so that \( \hat{d}_{A,F} = 1.1 \hat{d}_{F,F} \). Then \( c_2 = 1.1 \). Looking at the third column, we can see that our estimator will then be consistent and essentially unbiased for \( 1.1 \times \hat{d}_{F,F} \) instead of \( \hat{d}_{F,F} \).
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

\[ R_d = \frac{R_1}{R_0} = \frac{\bar{y}_1}{\bar{x}_1} = \frac{\bar{y}_1\bar{x}_0}{\bar{x}_1\bar{y}_0}. \]  

(E.1)

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

\[ \hat{r}_d = \frac{\hat{y}_1\hat{x}_0}{\hat{x}_1\hat{y}_0}. \]  

(E.2)

In this appendix we investigate when we can expect the biases in our estimators to be small enough to be negligible.

E.2  The general case

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

\[ B_d = \frac{\mathbb{E}[\hat{r}_d] - R_d}{R_d}. \]  

(E.3)

\( B_d \) expresses the bias in our estimator \( \hat{r}_d \) 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 \( \tilde{x}_0, \tilde{x}_1, \tilde{y}_0, \text{ and } \tilde{y}_1 \) are unbiased estimators, and \( |(\tilde{x}_1 - \bar{x}_1)/\bar{x}_1| < 1 \)
Table E.1: Description of the general form of the nonlinear estimators we propose. $K$ is a constant, $\hat{y}_j$ and $\hat{x}_1$ are taken from $s_F$, while $\hat{x}_0$ and $\hat{y}_0$ 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.11) 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.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Reference</th>
<th>Form</th>
<th>$\hat{x}_0$</th>
<th>$\hat{y}_1$</th>
<th>$\hat{x}_1$</th>
<th>$\hat{y}_0$</th>
<th>Approx. rel. bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_F$</td>
<td>Res. B.6</td>
<td>$K\hat{x}<em>0/\hat{y}<em>0$ $\sum</em>{i \in s_F} y_i,</em>{A_F}/\pi_i$</td>
<td>-</td>
<td>-</td>
<td>$\sum_{i \in s_F} y_i,_{A_F}/\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
<tr>
<td>$\hat{v}_{H,F}$</td>
<td>Res. C.2</td>
<td>$K\hat{x}<em>0/\hat{y}<em>0$ $\sum</em>{i \in s_H} v_i,</em>{A_H \cap F}/c\pi_i$</td>
<td>-</td>
<td>-</td>
<td>$\sum_{i \in s_H} v_i,_{A_H \cap F}/c\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
<tr>
<td>$\hat{d}_{H,F}$</td>
<td>Res. C.6</td>
<td>$K\hat{x}<em>0/\hat{y}<em>0$ $\sum</em>{i \in s_H} y_i,</em>{A_H \cap F}/c\pi_i$</td>
<td>-</td>
<td>-</td>
<td>$\sum_{i \in s_H} y_i,_{A_H \cap F}/c\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
<tr>
<td>$\hat{\delta}_F$</td>
<td>Res. C.7</td>
<td>$K\hat{x}_0/\hat{y}<em>0 \hat{x}<em>1$ $\sum</em>{i \in s_H} y_i,</em>{A_H \cap F}/c\pi_i$</td>
<td>-</td>
<td>$\sum_{i \in s_H} y_i,_{A_H}/\pi_i$</td>
<td>$\sum_{i \in s_H} y_i,_{A_H}/c\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} + C^2</em>{\hat{x}<em>1} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
<tr>
<td>$\hat{r}_F$</td>
<td>Res. C.8</td>
<td>$K\hat{x}_0/\hat{y}<em>0 \hat{x}<em>1$ $\sum</em>{i \in s_H} v_i,</em>{A_H \cap F}/c\pi_i$</td>
<td>-</td>
<td>$\sum_{i \in s_H} v_i,_{A_H \cap F}/c\pi_i$</td>
<td>$\sum_{i \in s_H} 1/c\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} + C^2</em>{\hat{x}<em>1} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
<tr>
<td>$\hat{N}_H$</td>
<td>Res. C.9</td>
<td>$K\hat{y}_1\hat{x}_0/\hat{y}<em>0$ $\sum</em>{i \in s_H} 1/c\pi_i$</td>
<td>$\sum_{i \in s_F} y_i,_{H}/\pi_i$</td>
<td>-</td>
<td>$\sum_{i \in s_H} v_i,_{A_H \cap F}/c\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
<tr>
<td>$\tilde{N}_H$</td>
<td>Res. C.11</td>
<td>$K\hat{x}<em>0/\hat{y}<em>0$ $\sum</em>{i \in s_F} y_i,</em>{H}/\pi_i$</td>
<td>-</td>
<td>-</td>
<td>$\sum_{i \in s_F} \sum_{j} y_i,_{A_j}/\pi_i$</td>
<td>$C^2_{\hat{y}<em>0} - C</em>{\hat{y}_0}\hat{x}_0$</td>
<td></td>
</tr>
</tbody>
</table>
and \(|(\hat{y}_0 - \bar{y}_0)/\bar{y}_0| < 1\), then the relative bias of the double ratio estimator, \(B_d\), is approximated by

\[
B_d = \frac{\mathbb{E}[\hat{r}_d] - R}{R} \approx B'_d = C_{\hat{x}_1, \bar{y}_0} - C_{\hat{x}_1, \hat{y}_1} - C_{\hat{y}_0, \hat{y}_1} - C_{\hat{x}_0, \hat{y}_1} + C_{\hat{y}_0, \hat{y}_0} + C_{\hat{y}_1, \bar{x}_0} + C_{\hat{x}_0, \bar{x}_0} + C_{\hat{x}_1, \bar{x}_1},
\]

(E.4)

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

**Proof:** Define

\[
\hat{\delta}_{x_0} = \frac{\hat{x}_0 - \bar{x}_0}{x_0},
\]

(E.5)

with analogous definitions for \(\hat{\delta}_{x_1}, \hat{\delta}_{y_1}\), and \(\hat{\delta}_{y_0}\). We can express \(r_d\) as

\[
\hat{r}_d = \frac{(1 + \hat{\delta}_{y_1})(1 + \hat{\delta}_{x_0})}{1 + \hat{\delta}_{y_0}} (1 + \hat{\delta}_{x_1}).
\]

(E.6)

The relative bias then becomes

\[
B_d = \frac{\mathbb{E}[\hat{r}_d] - R}{R} = \mathbb{E} \left[ \frac{(1 + \hat{\delta}_{y_1})(1 + \hat{\delta}_{x_0})}{1 + \hat{\delta}_{y_0}} (1 + \hat{\delta}_{x_1}) \right] - 1.
\]

(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 \(|\delta_{y_0}| < 1\) and \(|\delta_{x_1}| < 1\). Then we have

\[
B_d = \mathbb{E} \left[ (1 + \hat{\delta}_{y_1})(1 + \hat{\delta}_{x_0}) (1 - \delta_{y_0} + \delta_{y_0}^2 - \cdots)(1 - \delta_{x_1} + \delta_{x_1}^2 - \cdots) \right] - 1
\]

(E.8)

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

\[
B_d \approx \mathbb{E} \left[ \delta_{x_1} \delta_{y_0} + \delta_{x_2} \delta_{y_1} - \delta_{x_0} \delta_{y_0} - \delta_{x_0} \delta_{x_1} - \delta_{x_0} \delta_{y_1} - \delta_{x_1} \delta_{y_1} + \delta_{x_0} \delta_{y_0} + \delta_{y_1} - \delta_{y_0} - \delta_{x_0}^2 - \delta_{x_1}^2 \right].
\]

(E.9)

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

\[
\mathbb{E}[\delta_{x_1}] = 0,
\]

(E.10)
We can also determine that

\[ E[\delta_{x_1}, \delta_{y_1}] = \frac{\text{cov}(\hat{x}_{1}, \hat{y}_{1})}{\bar{x}_1 \bar{y}_1}, \quad (E.11) \]

and, that

\[ E[\delta^2_{x_1}] = \frac{\text{var}(\hat{x}_{1})}{\bar{x}_1^2}. \quad (E.12) \]

Applying these relationships to Equation E.9, we find

\[ B_d \approx \kappa \left[ C_{x_0,y_0} - C_{x_1,y_0} - C_{y_0,x_0} + C_{y_0,x_1} - C_{x_0,y_0} + C_{y_1,x_0} + C_{y_1,x_1} \right], \quad (E.13) \]

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 \( \hat{x}_0, \hat{x}_1, \hat{y}_0, \) and \( \hat{y}_1 \). 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, \( \kappa \):

\[ B'_d = \kappa \left[ C_{x_1,y_0} - C_{x_1,y_1} - C_{y_0,x_1} - C_{y_0,x_0} + C_{y_1,x_0} + C_{x_0,x_1} + C_{y_0,y_1} + C_{y_1,y_0} \right], \quad (E.14) \]

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 \( \kappa \) 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:
\[ C_{\hat{x}, \hat{y}} = \frac{\text{cov}(\hat{x}, \hat{y})}{\hat{x} \hat{y}} = \frac{\rho_{\hat{x}, \hat{y}} \sqrt{\text{var}(\hat{x})} \sqrt{\text{var}(\hat{y})}}{\hat{x} \hat{y}} = \rho_{\hat{x}, \hat{y}} \text{cv}(\hat{x}) \text{cv}(\hat{y}), \]  

(E.15)

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

### 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 \( \hat{\delta}_F \), there is no \( \hat{y}_1 \)). 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 \( \hat{x}_1 \) for \( \hat{\delta}_F \)) 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 began by deriving 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,
conclude that these estimators are essentially unbiased, and that sampling error and non-sampling error will dominate any bias introduced by the nonlinear form of the estimators.

<table>
<thead>
<tr>
<th>$\hat{cv}(d)$</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>Rwanda</td>
</tr>
<tr>
<td>0.09</td>
<td>Curitiba</td>
</tr>
<tr>
<td>0.02</td>
<td>US</td>
</tr>
</tbody>
</table>

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 $\hat{d}_F$. 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.

<table>
<thead>
<tr>
<th>estimated coef. of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_{i \in s_H} y_i, A \cap F / c \pi_i$</td>
</tr>
<tr>
<td>$\sum_{i \in s_H} \tilde{v}_i, A \cap F / c \pi_i$</td>
</tr>
<tr>
<td>$\sum_{i \in s_H} 1 / c \pi_i$</td>
</tr>
</tbody>
</table>

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).

<table>
<thead>
<tr>
<th>estimated correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{cor}(\sum_{i \in s_H} y_i, A \cap F / c \pi_i, \sum_{i \in s_H} \tilde{v}_i, A \cap F / c \pi_i)$</td>
</tr>
<tr>
<td>$\text{cor}(\sum_{i \in s_H} y_i, A \cap F / c \pi_i, \sum_{i \in s_H} 1 / c \pi_i)$</td>
</tr>
<tr>
<td>$\text{cor}(\sum_{i \in s_H} \tilde{v}<em>i, A \cap F / c \pi_i, \sum</em>{i \in s_H} 1 / c \pi_i)$</td>
</tr>
</tbody>
</table>

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.

F Variance estimation

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 $\hat{N}_H$ 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, $\hat{N}_H^{(1)}, \hat{N}_H^{(2)}, \ldots, \hat{N}_H^{(B)}$. 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.1) (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
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.

<table>
<thead>
<tr>
<th></th>
<th>approx. rel. bias, $B_d$</th>
<th>estimate</th>
<th>estimated absolute bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\tau}_F$</td>
<td>0.0005</td>
<td>0.77</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\hat{\delta}_F$</td>
<td>0.0086</td>
<td>0.69</td>
<td>0.0059</td>
</tr>
<tr>
<td>$\hat{N}_H$</td>
<td>0.0027</td>
<td>114498</td>
<td>304</td>
</tr>
</tbody>
</table>

Figure F.1: Schematic of the bootstrap procedure to put a confidence interval around $\hat{N}_H$ when there is a sample from the frame $s_F$.

survey of 1,261 Americans selected via random digit dialing.\(^8\) 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 21, since they account for the clustering of responses with respondent; on average, they are 2.05 times wider.

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), propose a modified 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).

\(^8\)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.
This bootstrap procedure proposed by Rust and Rao (1996) includes two conceptual changes from the standard bootstrap. First, we approximate 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 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 uncorrected 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 \(j^{th}\) person in PSU \(i\) in the \(b^{th}\) replicate sample is

\[
\hat{w}_{ij}^{(b)} = w_{ij} \times \frac{n_h}{(n_h - 1)} \times r_i^{(b)}
\]

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

In Figure 1 in the main paper, we compared the three different procedures for putting confidence intervals around the basic scale-up estimator: the current scale-up variance estimator (Killworth et al., 1998b), the standard bootstrap, and the rescaled bootstrap. We made this comparison using data from recent scale-up

A32
studies in the United States, Rwanda,9 and Curitiba, Brazil.10 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 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 variance estimator. In the Rwanda case, the rescaled bootstrap procedure produced confidence intervals 1.58 times larger than the standard bootstrap and 2.95 times larger than the current variance estimator.

The standard bootstrap procedure for simple random samples and the rescaled bootstrap procedure for complex sample designs are both well-grounded statistically, but they are not perfect. Most importantly, they only capture sampling error and do not capture other forms of survey error such as response error, sampling frame error, and interviewer effects. Therefore, these procedures should be expected to produce 95% intervals that do not have the desired coverage probabilities. In fact, this is what we find in practice: Figure 2 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%, 13.6%, 15.0%. While this is far from ideal, we note that it is slightly better than the currently used procedure (Equation 21), which produced coverage rates of 0.0%, 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%. We believe that these theoretical and empirical arguments demonstrate three conclusions. First, bootstrap confidence intervals, which do not assume the basic scale-up model, are preferable to the current variance estimation procedure; second, until there is a substantially better understanding of non-sampling errors, researchers should expect that the bootstrap confidence intervals will understate true uncertainty; and, third, variance estimation for the scale-up method is an important topic for future research.

9The 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. For full details see Rwanda Biomedical Center (2012). The original data file includes 4,669 respondents. From these cases, we removed 6 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 0.23% of responses.

10The 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.
F.2 Variance estimation with sample from $F$ and $H$

In this paper we showed that the scale-up estimates could be improved if information from people on the frame was combined with information from people in the hidden population. Producing confidence intervals around the generalized scale-up estimator is more difficult because the generalized estimator has uncertainty from two different samples. 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: $\hat{N}_H^{(1)} = f(s_F^{(1)}, s_H^{(1)}), \hat{N}_H^{(2)} = f(s_F^{(2)}, s_H^{(2)}), \ldots, \hat{N}_H^{(B)} = f(s_F^{(B)}, s_H^{(B)})$. Finally, these $B$ replicate estimates are converted into a confidence interval using the percentile method (Fig. F.2).

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 $s_H$ is 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 this bootstrap procedure is that it requires researchers to divide the sample of the hidden population into two mutually exclusive groups. In

Figure F.2: Schematic of the bootstrap procedure to put a confidence interval around $\hat{N}_H$ when there is a sample from the frame $s_F$ and a sample from the hidden population $s_H$. 
this case, we recommend dividing the hidden population into those who are above and below the median of their estimated visibility $\hat{v}_{i,F}$ 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 method 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). If this is the case, then the intervals around the generalized scale-up estimates will be too small.

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 robustness of these variance estimation procedures through simulation.

G Simulation study

G.1 Overview

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.

G.2 Simulation design

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 $\tau_F$ (see Equation 18).

We simulate populations consisting of $N = 10,000$ people, using a stochastic block-model (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:

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

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 between 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 $g(i) = g(j)$. We use the igraph library to conduct our simulations (Csárdi and Nepusz, 2006).

Each random network drawn under our simulation model depends on six parameters. The first three 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

For simplicity, we assume that each person’s membership in the frame population ($F$) and the hidden population ($H$), is independent. This means that, for example, the number of people who are in both $F$ and $H$ will be $N \times p_F \times p_H$.

The next two parameters govern the amount of inhomogenous mixing in the network that connects people to each other; they are:

- $\zeta$, the probability of an edge between two people who are both in the same block.
\[
M = \begin{pmatrix}
  F & H & F & H & \neg F & H & \neg F & \neg H \\
  F & H & \zeta & \rho \cdot \zeta & \rho \cdot \zeta & \rho^2 \cdot \zeta & \rho \cdot \zeta & \rho^2 \cdot \zeta \\
  \neg F & H & \rho \cdot \zeta & \zeta & \rho^2 \cdot \zeta & \rho \cdot \zeta & \zeta & \rho \cdot \zeta \\
  \neg F & \neg H & \rho^2 \cdot \zeta & \rho \cdot \zeta & \rho \cdot \zeta & \zeta & \rho \cdot \zeta & \zeta \\
\end{pmatrix}
\]  

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 \(\zeta\) and \(\rho\). In our simulations, we generate networks with different amounts of inhomogenous mixing between the four groups by fixing \(\zeta\) and varying \(\rho\) from 0.1 (extreme inhomogenous mixing) to 1 (perfectly random mixing).

- \(\rho\), the relative probability of an edge between two vertices that are not in the same group. For example, a value of 0.8 would mean that the chances of having a connection between a particular person in \(F\) and a particular person not in \(F\) is 80% of the chance of a connection between two members of \(F\) or two members of \(\neg F\).

Together, the parameters \(\zeta\) and \(\rho\) are used to construct the mixing matrix \(M\) (Figure G.2), which is used as the basis for the probability of randomly creating an edges between each pair of people in the population. Note that varying the parameter \(\rho\) will change several structural features of the network in addition to the amount of inhomogenous mixing; for example, changing \(\rho\) 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, \(\tau_F\), 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 population to members of the hidden population uniformly at random and remove them from the reporting graph.

Given a census of the relevant populations, 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 three of the parameters at constant values \((N = 10,000; p_F = 0.03; \zeta = 0.05)\). We systematically explore varying the remaining three parameters, \(p_F\), \(\rho\), and \(\tau_F\), from 0.1 to 1 in increments of 0.1. For each combination of the parameter values, we repeated our simulation 10 times, for a total of \(10 \times 10 \times 10 \times 10 = 10,000\) iterations. For each unique combination of parameters, we averaged the results across the replications.
G.3 Results

In this section, we focus on a representative sample of the simulation results in order to illustrate the main conclusions while keeping the plots relatively simple. The complete simulation results are consistent with the analysis presented here.

Figure G.2 shows the estimated size of the hidden population for the generalized and basic scale-up estimators. The generalized scale-up estimator is exactly correct for all parameter combinations. The basic scale-up estimator works well when the assumptions it relies upon hold. In our simulations, these assumptions hold in special cases, such as when $p_F = 1$, $\rho = 1$, and $\tau_F = 1$. With these parameter values, (1) the social network is an Erdos-Renyi random graph (since $\rho = 1$), meaning that in expectation $\bar{d}_{H,F}/\bar{d}_{F,F} = \delta_F = 1$; (2), $U = F$ (since $p_F = 1$), meaning $\bar{d}_{U,F} = \bar{d}_{F,F}$; and, (3), there is perfect reporting (since $\tau_F = 1$), meaning $\bar{v}_{H,F} = \bar{d}_{H,F}$. Together, (1), (2), and (3) mean that the basic scale-up estimator’s assumptions hold for this particular combination of parameters (Equation 15). In general, however, Figure G.2 shows that the basic scale-up estimator can be biased (as described in Section 3).

Further, our analytical results can be used to predict the bias in the basic scale-up estimator. In Section 3, we propose a decomposition of the difference between the basic and generalized scale-up estimators, and derive an expression for the bias in the basic scale-up estimator (Equation 20). Figure G.3 illustrates the empirical bias in the basic scale-up estimator from our simulations for the same combination of parameter values shown in Figure G.2. Our analytical expression (Equation 20) predicts the empirical bias (Figure G.3) exactly in all cases.
Figure G.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 inhomogenous mixing is varied from high ($\rho = 0.1$) to low ($\rho = 1$). The columns show results for different sizes of the frame population, from small (left column, $p_F = 0.1$), to medium (middle column, $p_F = 0.5$), and 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 a moderate amount of true positives (middle row, $\tau_F = 0.5$), and 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 basic scale-up estimate is about 100, while the generalized scale-up estimate is 300. The generalized scale-up estimator is exactly correct for all parameter combinations, while the basic scale-up estimator is only correct for certain special cases (e.g., when $\rho = 1$, and $\tau_F = 1$). These results confirm and illustrate the discussion in Section 3.
Figure G.3: Bias (solid line) and predicted bias (open circles) in the basic scale-up estimates for the same parameter configurations depicted in Figure G.2. The predicted bias from our analytical results (Equation 20) matches the bias observed in our simulation study exactly.