Online Appendixes

D Sensitivity analysis

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

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

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

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

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

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

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

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

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

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

D.2 Sensitivity to sampling problems

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

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

D.2.1 Probability samples

First, we must define *imperfect sampling weights*.

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
\text{bias}[\widehat{y}'_{F,H}] &= \mathbb{E}[\widehat{y}'_{F,H}] - y_{F,H} && \text{(D.4)}\\
&= \sum_{i \in F} w_i' \mathbb{E}[I_i] y_{i,H} - \sum_{i \in F} y_{i,H} && \text{(D.5)}\\
&= \sum_{i \in F} \frac{\pi_i}{\pi_i'} y_{i,H} - \sum_{i \in F} y_{i,H} && \text{(D.6)}\\
&= \sum_{i \in F} y_{i,H}(\epsilon_i - 1). && \text{(D.7)}
\end{aligned}$$

Now, recall that, for any a_i and b_i,

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

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

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

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

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

■

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

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

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

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

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

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

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

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

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

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

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

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

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

where $\to$ *means 'is consistent and unbiased for,' and* $K_F = cor_F(y_{i,H}, \epsilon_i)cv_F(y_{i,H})cv_F(\epsilon_i)$.

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

D.2.2 Relative probability samples

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

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

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

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

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

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

Then

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

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

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

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

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

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

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

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

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

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

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

and also that

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

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

$$\begin{aligned}
\text{bias}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}) &= \mathbb{E}[\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}] - \bar{\tilde{v}}_{H,\mathcal{A}\cap F} && \text{(D.25)}\\
&= \frac{\sum_{i\in H} \epsilon_i \tilde{v}_{i,\mathcal{A}\cap F}}{\sum_{i\in H} \epsilon_i} - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}, \widehat{N}'_H)}{N'_H} - \frac{\sum_{i\in H} \tilde{v}_{i,\mathcal{A}\cap F}}{N_H} && \text{(D.26)}\\
&= \left(\frac{\sum_{i\in H} \epsilon_i \tilde{v}_{i,\mathcal{A}\cap F}}{\sum_{i\in H} \epsilon_i} - \frac{\sum_{i\in H} \tilde{v}_{i,\mathcal{A}\cap F}}{N_H}\right) - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}, \widehat{N}'_H)}{N'_H} && \text{(D.27)}\\
&= \left(\frac{\sum_{i\in H} \epsilon_i \tilde{v}_{i,\mathcal{A}\cap F} - \frac{1}{N_H}\sum_{i\in H} \tilde{v}_{i,\mathcal{A}\cap F}\sum_{i\in H}\epsilon_i}{\sum_{i\in H} \epsilon_i}\right) - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}, \widehat{N}'_H)}{N'_H} && \text{(D.28)}\\
&= \left(\frac{\text{cov}_H(\tilde{v}_{i,\mathcal{A}\cap F}, \epsilon_i)}{\bar{\epsilon}}\right) - \frac{\text{cov}(\widehat{\bar{\tilde{v}}}'_{H,\mathcal{A}\cap F}, \widehat{N}'_H)}{N'_H}, && \text{(D.29)}
\end{aligned}$$

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

■

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

D.2.3 Summary and results for all estimators

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

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

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

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

D.3 Combined sensitivity results

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

D.3.1 Generalized scale-up

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

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

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

Then

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

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

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

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

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

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

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

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

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

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

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

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

Suppose also researchers have also obtained a relative probability sample s_H to estimate $\bar{v}_{H,F}$ but that the researchers have imperfect relative sampling weights. Call the *imperfect relative sampling weights $w_i'^H = \frac{1}{c'\pi_i'^H}$, call the true probabilities of inclusion π_i, and define $\epsilon_i^H = \frac{\pi_i^H}{\pi_i'^H}$. Suppose also that the researcher's estimate of $N_{\mathcal{A}\cap F}$ is incorrect, so that $\widehat{N}_{\mathcal{A}\cap F} = c_1 \cdot N_{\mathcal{A}\cap F}$. Suppose that the reporting condition (Equation C.5) of Result C.2 is incorrect, so that $\tilde{v}_{H,\mathcal{A}\cap F} = c_2 \cdot v_{H,\mathcal{A}\cap F}$. Finally, suppose that the probe alter condition is incorrect, so that $\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = c_3 \cdot \frac{v_{H,F}}{N_F}$. Call the estimator for $\bar{v}_{H,F}$ under these imperfect conditions $\widehat{\bar{v}}'^{\star}_{H,F}$.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

tor,

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

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

D.3.2 Modified basic scale-up

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which is our result. ■

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

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

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

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

$$\widehat{N}_H'^{\star} = \frac{\widehat{y}'_{F,H}}{\widehat{\bar{d}}'^{\star}_{F,F}}. \tag{D.53}$$

Then

$$\widehat{N}_H'^{\star} \rightsquigarrow \frac{(1+K_{F_1})}{(1+K_{F_2})} \cdot \frac{c_1}{c_2\ c_3} \cdot \frac{\tau_F\ \delta_F}{\eta_F} \cdot N_H, \tag{D.54}$$

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

Proof:

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

For the numerator, Result D.9 shows that

$$\widehat{\bar{d}}'^{\star}_{F,F} \to \bar{\epsilon}^F(1+K_{F_2}) \cdot \frac{c_2\ c_3}{c_1} \cdot \bar{d}_{F,F}. \tag{D.55}$$

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

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

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

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

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

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

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

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

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

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

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

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

E Approximate unbiasedness of compound ratio estimators

E.1 Overview

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

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

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

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