Online Appendix

A Notation

We follow the notation used in Feehan and Salganik (2016a) and Feehan, Mahy, and Salganik (2017). For convenience, here is a table summarizing key features of the notation:

Table 2: Notation for network reporting quantities used in this paper.

Quantity	Explanation
U	the entire population
F	the frame population (typically adults over a certain age)
$\|U\| = N$	size of the entire population, $\|U\|$ (i.e., everyone who could ever be interviewed or reported about)
$\|F\| = N_F$	size of the frame population, $\|F\|$ (i.e., everyone who could ever be interviewed)
$y_{i,A} = y(i, A)$	out-reports from i about connections to A (e.g. i's reported number of siblings in group A)
$y_{F,A} = y(F, A)$	total out-reports from the frame population F about connections to A
$y^+_{F,A} = y^+(F, A)$	true positive out-reports from the frame population F about connections to A (i.e., the sum of reported connections that actually lead to people in A)
$d_{i,F} = d(i, F)$	number of network connections from i to F, i.e., number of i's siblings in F (which is not necessarily the same as the number of reported connections from i to F)
$d_{A,B} = d(A, B)$	number of network connections from group $A \subset U$ to group $B \subset U$ (note that A and B could be the same group, they could be entirely distinct groups, or they could overlap partially)
$\bar{d}_{A,B} = \bar{d}(A, B) = \frac{d_{A,B}}{N_A}$	average number of network connections from group A to group B, per member of group A (note that we always take averages with respect to the first subscript)
$\bar{v}_{A,F} = \bar{v}(A, F) = \frac{v_{A,F}}{\|A\|}$	average visibility (number of in-reports) per member of A
s	a probability sample of people from the frame population F
π_i	the probability that $i \in F$ is included in the sample, which comes from the sampling design

Table 2: Notation for network reporting quantities used in this paper.

Quantity	Explanation
τ_F	the true positive rate for out-reports from F
δ_F	the degree ratio of hidden population members relative to frame population members
η_F	the false positive rate for out-reports from F
α	a demographic group (e.g., women aged 45-54 in 2010)
F_α	frame population members who are in demographic group α
$\lvert F_\alpha \rvert$	the number of frame population members who are in demographic group α
$\lvert N_\alpha \rvert$	the number of people in the entire population U who are also in demographic group α
D_α	the number of deaths in demographic group α (e.g. number of deaths among women aged 45-54 in 2010)
$M_\alpha = \frac{D_\alpha}{N_\alpha}$	the death rate in demographic group α (e.g. the death rate for women aged 45-54 in 2010; exposure is approximated by the population size)
Σ	the set of all sibships in the population
σ	a specific sibship $\sigma \in \Sigma$
$\sigma[i]$	the specific sibship containing person i

Sibship structure

We define Σ to be the set of sibships in the population, and we use σ to index the sibships in Σ. Σ is a partition of the population, meaning that each population member i is in one and only one sibship $\sigma \in \Sigma$. We will sometimes denote the sibship containing i by $\sigma[i]$.

B Estimands

This appendix focuses on developing several important *estimands*: population-level relationships that form the basis for the different death rate estimators that are developed in subsequent appendixes.

We shall see that researchers have two important questions to answer when forming a death rate estimator from sibling histories: (1) should reports about siblings be adjusted for visibility at the individual or at the aggregate level?; and (2) should information about the survey respondent be included or excluded from the reports? Taken together, these two questions lead to four different ways that visible death rates can be estimated from sibling histories.

This appendix begins by developing general expressions for visibility in each of the four cases: individual visibility when respondents are included in reports; individual visibility when respondents are excluded from reports; aggregate visibility when respondents are included in reports; and aggregate visibility when respondents are excluded from reports.

Next, we use these expressions for visibility to develop population-level identities for (1) the number of deaths; (2) the amount of exposure; and (3) the visible death rate in each case. These identities hold in a census when reporting is perfectly accurate; later appendixes will show how estimates will be affected by sampling and by different types of reporting error.

B.1 Visibility and whether or not the respondent is included

As described in the main paper (Section 3), a critical step in making estimates using sibling histories is to adjust for the *visibility* of reported siblings–that is, to account for the fact that each sibling could be reported multiple times in a census of the frame population. In this section, we derive some important relationships that will be helpful in developing estimators that adjust for visibility at both the individual and at the aggregate level. We will see that the decision to include or exclude information about respondents from sibling reports affects how visibility is calculated. Therefore, we develop expressions for visibility both when the respondent is included in reports, and when the respondent is not included in reports.

Figure 6 illustrates the intuition behind the results we derive. The Figure shows sibships and reporting networks under the usual network reporting situation, where respondents are not included in their reports (panels a and b), and in an alternate situation, where respondents are included in their reports (panels c and d). Figure 6 illustrates the fact that whether or not the respondent includes herself in reports will only affect the visibility of siblings who are in the frame population: the visibility of node 11, which is dead (and thus not in the frame population) is unchanged under the two different scenarios. The other three nodes, on the other hand, have visibility 2 in panel b, when respondents are not included in their reports; and they have visibility 3 in panel d, when respondents are included in their reports.

B.2 Expressions for individual visibility

Individual visibility estimation is based on separately adjusting for the visibility of each person being reported about. Below, in Appendix F, we will use the expressions developed here to derive estimators for deaths and exposure using an individual visibility approach.

When the respondent is included in the reports

When the respondent is included in the reports, a reporting identity says that for person i in sibship $\sigma[i]$,

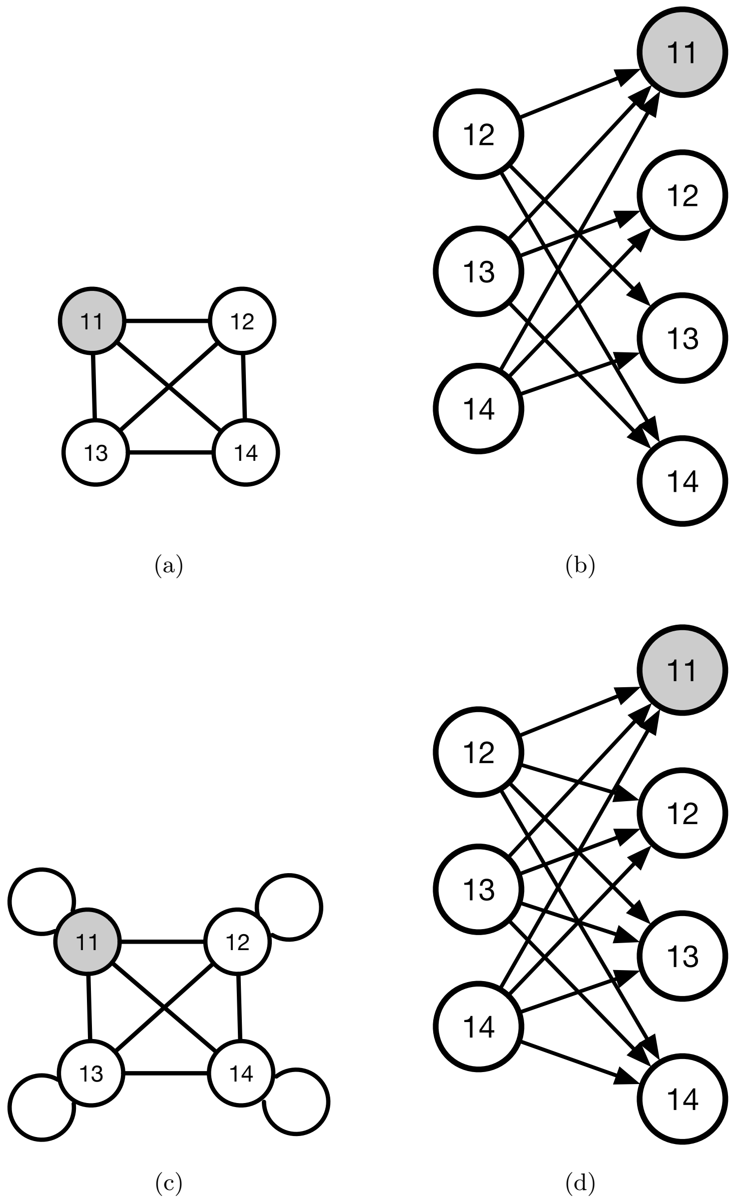

Figure 6: Network reporting when respondents are not included in their reports. (a) A single sibship. (b) The reporting network for the sibship in (a), when respondents are not included in reports. (c) The same sibship as (a) but with each sibling considered connected to herself. (d) The reporting network for the sibship in (c), when respondents are included in their reports.

$$v'(i, \sigma[i] \cap F) = y'^{+}(\sigma[i] \cap F, i), \tag{18}$$

where we use the notation v' for visibilities when respondents include themselves in reports and y' for reports when respondents include themselves.

When reports are perfect, each sibling will be reported once for each member of the sibship on the frame population, so Equation 18 can be written as

$$\begin{aligned} v'(i, \sigma[i] \cap F) &= y'^{+}(\sigma[i] \cap F, i) \\ &= d'(\sigma[i] \cap F, i) && \text{(if reporting is perfect)} \\ &= |\sigma \cap F| \\ &= \bar{y}(\sigma[i] \cap F, \sigma[i] \cap F) + 1 \\ &= y(j, F) + 1 && \text{for any } j \in \sigma[i] \cap F. \end{aligned} \tag{19}$$

Equation 19 shows that, when analyzing reports about individual i made by respondent j, i's visibility under perfect reporting is $y(j, F) + 1$.

When the respondent is not included in the reports

When the respondent does not include herself in her reports, a reporting identity says that, for population member i in sibship $\sigma[i]$,

$$v(i, \sigma[i] \cap F) = y^{+}(\sigma[i] \cap F, i). \tag{20}$$

When reporting is perfect, $y^{+}(\sigma \cap F, i)$ is the number of i's siblings who are on the frame population. Within a given sibship, this quantity will depend upon whether or not i is herself on the frame population: in a sibship σ with $|\sigma \cap F|$ members in F, there will be $|\sigma \cap F|$ siblings who can report a sibling i who is not in F; on the other hand, when i is in F, then i counts towards the size of $|\sigma \cap F|$ and so there are $|\sigma \cap F| - 1$ *other* siblings who can report i.

Mathematically, when reports are perfect, Equation 20 can be written as

$$
\begin{aligned}
v(i, \sigma \cap F) &= y^{+}(\sigma \cap F, i) \\
&= \begin{cases} |\sigma \cap F| & \text{if } i \notin F \\ |\sigma \cap F| - 1 & \text{if } \in F \end{cases} \qquad \text{(if reporting is perfect)} \\
&= \begin{cases} \bar{y}(\sigma \cap F, \sigma \cap F) + 1 & \text{if } i \notin F \\ \bar{y}(\sigma \cap F, \sigma \cap F) & \text{if } \in F \end{cases} \\
&= \begin{cases} y(j, F) + 1 & \text{if } i \notin F \\ y(j, F) & \text{if } \in F \end{cases} \qquad \text{for any } j \in \sigma[i] \cap F.
\end{aligned} \tag{21}
$$

Equation 21 shows that, when analyzing reports about individual i made by respondent j, i's visibility under perfect reporting is either $y(j, F)$ or $y(j, F) + 1$, depending on whether i is or is not in the frame population.

B.3 Individual visibility estimands for mortality

B.3.1 Individual visibility estimation including the respondent

When the respondent is included in the sibling reports, and when reports are perfect, Equation 19 shows that the visibility of each sibling is $\bar{y}(\sigma \cap F, \sigma \cap F) + 1$. Note that when reports are perfect, $\bar{y}(\sigma \cap F, \sigma \cap F) = y(i, \sigma \cap F)$ and $y'(i, \sigma \cap F) = y(i, \sigma \cap F) + 1$ for all $i \in \sigma \cap F$.

Estimand for deaths using reports that include the respondent:

$$
\begin{aligned}
D'^{V}_{\alpha,ind} &= \sum_{i \in F} \frac{y'(i, D_\alpha)}{y(i, F) + 1} \\
&= \sum_{i \in F} \frac{y'(i, D_\alpha)}{y'(i, F)}
\end{aligned} \tag{22}
$$

Estimand for exposure using reports that include the respondent:

$$
\begin{aligned}
N'^{V}_{\alpha,ind} &= \sum_{i \in F} \frac{y'(i, N_\alpha)}{y(i, F) + 1} \\
&= \sum_{i \in F} \frac{y'(i, N_\alpha)}{y'(i, F)}
\end{aligned} \tag{23}
$$

Estimand for the death rate using reports that include the respondent:

$$
\begin{aligned}
M'^{V}_{\alpha,ind} &= \frac{D'^{V}_{\alpha,ind}}{N'^{V}_{\alpha,ind}} \\
&= \frac{\sum_{i\in F} \frac{y'(i,D_\alpha)}{y'(i,F)}}{\sum_{i\in F} \frac{y'(i,N_\alpha)}{y'(i,F)}}
\end{aligned}
\tag{24}
$$

B.3.2 Individual visibility estimands not including the respondent

When the respondent does not include herself in reports about deaths and exposure, and when reports are perfect, Equation 21 shows that the visibility of a member i of a sibship σ reported by respondent $j \in \sigma[i] \cap F$ is given by:

$$
v(i, \sigma \cap F) = \begin{cases} |\sigma \cap F| = \bar{y}(\sigma \cap F, \sigma \cap F) + 1 & \text{if } i \notin F \\ |\sigma \cap F| - 1 = \bar{y}(\sigma \cap F, \sigma \cap F) & \text{if } \in F \end{cases} \quad \text{(if reporting is perfect).}
$$

Since $\bar{y}(\sigma \cap F, \sigma \cap F) = y(j, \sigma \cap F)$ for each $j \in \sigma \cap F$ when reports are perfect, we use the reported quantities $y(j, \sigma \cap F)$ and $1 + y(j, \sigma \cap F)$ to estimate visibilities for siblings in the frame and not in the frame, respectively.

Estimand for deaths using reports that don't include the respondent:

$$
\begin{aligned}
D^{V}_{\alpha,ind} &= \sum_{i \in F} \Bigg[\underbrace{\frac{y(i, D_\alpha \cap F)}{y(i,F)}}_{D_\alpha \cap F = \emptyset} + \underbrace{\frac{y(i, D_\alpha - F)}{y(i,F)+1}}_{\text{siblings not in } F} \Bigg] \quad \text{(when reports are perfect)} \\
&= \sum_{i \in F} \frac{y(i, D_\alpha)}{y(i,F)+1}.
\end{aligned}
\tag{25}
$$

In going from the first line to the second, we make use of the fact that deaths cannot be on the frame population, so that $y(i, D_\alpha \cap F) = 0$ for all i, and so $y(i, D_\alpha) = y(i, D_\alpha - F)$ for all i.

Estimand for exposure using reports that don't include the respondent:

$$
N^{V}_{\alpha,ind} = \sum_{i \in F} \Bigg[\underbrace{\frac{y(i, N_\alpha \cap F)}{y(i,F)}}_{\text{siblings in } F} + \underbrace{\frac{y(i, N_\alpha - F)}{y(i,F)+1}}_{\text{siblings not in } F} \Bigg] \quad \text{(when reports are perfect).}
\tag{26}
$$

Estimand for the death rate using reports that don't include the respondent:

The estimand for the death rate is the ratio of the estimand for the number of deaths and the estimand for the exposure:

$$M^V_{\alpha,ind} = \frac{\sum_{i \in F} \frac{y(i,D_\alpha)}{y(i,F)+1}.}{\sum_{i \in F} \left[\frac{y(i,N_\alpha \cap F)}{y(i,F)} + \frac{y(i,N_\alpha - F)}{y(i,F)+1} \right].} \tag{27}$$

B.4 Expressions for aggregate visibility

Aggregate visibility estimation is based on adjusting for the average visibility of the people being reported about. In this section, we develop expressions for the average visibility of siblings who are in some group $A \subset U$. By deriving results for a general group A, we will obtain expressions that can be readily used for both deaths and exposure (see Appendix E).

When the respondent is included in the reports

When reports include the respondent, the average visibility of siblings in a group A is:

$$\begin{aligned} \bar{v}'(A,F) &= |A|^{-1} \sum_{i \in A} v'(i,F) \\ &= |A|^{-1} \sum_{i \in A} y'^{+}(F,i). \end{aligned} \tag{28}$$

When reporting is perfect, this becomes

$$\begin{aligned} \bar{v}'(A,F) &= |A|^{-1} \sum_{i \in A} y'^{+}(F,i) \\ &= |A|^{-1} \sum_{i \in A} [\bar{y}(\sigma[i] \cap F, \sigma[i] \cap F) + 1] \quad \text{(if reporting is perfect)} \\ &= |A|^{-1} \sum_{\sigma \in \Sigma} \sum_{i \in \sigma \cap A} |\sigma \cap F| \\ &= |A|^{-1} \sum_{\sigma \in \Sigma} [|\sigma \cap A| \; |\sigma \cap F|] \\ &= \sum_{\sigma \in \Sigma} \underbrace{\frac{|\sigma \cap A|}{|A|}}_{\substack{\text{fraction of} \\ A \text{ in} \\ \text{in sibship } \sigma}} \underbrace{|\sigma \cap F|}_{\substack{\text{\# sibship} \\ \text{members} \\ \text{on frame}}}. \end{aligned} \tag{29}$$

Equation 29 shows that the visibility can be expressed as a weighted average of the number of siblings on the frame across all of the sibships in the population, where the weights are given by the proportion of A in each sibship.

Unlike the results for individual visibility estimation, we see no way to convert Equation 29 into a sample-based estimator for visibility. As we will see below, when estimating a death rate using an aggregate visibility estimator, we will take the ratio of two aggregate visibility estimators. Under the assumption that the visibility of the numerator and denominator are the same the visibilities will cancel, so that aggregate visibility does not need to be directly estimated.

When the respondent is not included in the reports

When reports do not include the respondent, the average visibility of siblings in a group A is:

$$\begin{aligned}\bar{v}(A,F) &= |A|^{-1}\sum_{i\in A} v(i,F)\\ &= |A|^{-1}\sum_{i\in A} y^{+}(F,i)\\ &= |A|^{-1}\sum_{\sigma\in\Sigma}\sum_{i\in\sigma\cap A} y^{+}(F,i).\end{aligned}$$

When reporting is perfect, this becomes

$$\begin{aligned}\bar{v}(A,F) &= |A|^{-1}\sum_{\sigma\in\Sigma}\sum_{i\in\sigma\cap A} y^{+}(F,i)\\ &= |A|^{-1}\sum_{\sigma\in\Sigma}\left[\underbrace{\sum_{i\in\sigma\cap A\cap F}\bar{y}(\sigma[i]\cap F,\sigma[i]\cap F)}_{\text{members of } A \text{ on frame}}+\underbrace{\sum_{i\in\sigma\cap A-F}(\bar{y}(\sigma[i]\cap F,\sigma[i]\cap F)+1)}_{\text{members of } A \text{ not on frame}}\right] \quad \text{(if reporting is perfect)}\\ &= |A|^{-1}\sum_{\sigma\in\Sigma}\left[\sum_{i\in\sigma\cap A}\bar{y}(\sigma[i]\cap F,\sigma[i]\cap F)+\sum_{i\in\sigma\cap A-F}1\right]\\ &= \sum_{\sigma\in\Sigma}\frac{|\sigma\cap A|}{|A|}\bar{y}(\sigma\cap F,\sigma\cap F)+\sum_{\sigma\in\Sigma}\frac{|\sigma\cap A-F|}{|A|}\\ &= \sum_{\sigma\in\Sigma}\frac{|\sigma\cap A|}{|A|}(|\sigma\cap F|-1)+\sum_{\sigma\in\Sigma}\frac{|\sigma\cap A-F|}{|A|}\\ &= \sum_{\sigma\in\Sigma}\left[\frac{|\sigma\cap A|}{|A|}|\sigma\cap F|-\frac{|\sigma\cap A|}{|A|}+\frac{|\sigma\cap A-F|}{|A|}\right]\\ &= \sum_{\sigma\in\Sigma}\frac{|\sigma\cap A|}{|A|}\left[|\sigma\cap F|-\frac{|\sigma\cap F\cap A|}{|\sigma\cap A|}\right].\end{aligned} \tag{30}$$

Comparing Equation 30 to the analogous expression we derived for the situation where respondents do include themselves in reports (Equation 29), we find that the two expressions differ according to the second factor in the sum: when respondents include themselves in reports, this second factor is always the number of siblings in the frame population, $|\sigma\cap F|$; when respondents do not include themselves, this second factor is the number of siblings in the frame population minus the proportion of siblings in A that is on the frame population, i.e., $|\sigma\cap F| - |\sigma\cap F\cap A|/|\sigma\cap A|$.

Note also that, by using the average visibility we derived for reports where respondents include themselves (Equation 29), we have an alternate way to write the result in Equation 30:

$$
\begin{aligned}
\bar{v}(A,F) &= \sum_{\sigma\in\Sigma} \frac{|\sigma\cap A|}{|A|}\left[|\sigma\cap F| - \frac{|\sigma\cap F\cap A|}{|\sigma\cap A|}\right] \\
&= \bar{v}'(A,F) - \frac{|F\cap A|}{|A|}.
\end{aligned} \tag{31}
$$

Equation 31 shows that the aggregate visibility when respondents are not included in reports is equal to the aggregate visibility when respondents are included in reports, minus the fraction of the set A that is in the frame population.

B.5 Aggregate visibility estimands for mortality

In Section B.4, we saw that the expressions for the average visibility did not readily lend themselves to forming estimators. In the situation where we wish to estimate death rates, however, the condition that visibility is the same for deaths and for exposure leads to an estimator where the visibilities cancel, meaning that they do not have to be directly estimated. We can then investigate the sensitivity of estimated death rates to different visibilities as part of the broader sensitivity framework (Section E).

For the estimands below, we write aggregate visibilities for a set A as $\bar{v}(A,F)$ and $\bar{v}'(A,F)$, bearing in mind that this cancellation will take place in the estimand for the death rate.

B.5.1 Aggregate visibility estimands not including the respondent

Estimand for deaths using reports that don't include the respondent

$$
D^{V}_{\alpha,agg} = \frac{y_{F,D_\alpha}}{\bar{v}_{D_\alpha,F}}. \tag{32}
$$

Estimand for exposure using reports that don't include the respondent

$$
N^{V}_{\alpha,agg} = \frac{y_{F,N_\alpha}}{\bar{v}_{N_\alpha,F}}. \tag{33}
$$

Estimand for the death rate using reports that don't include the respondent

$$
\begin{aligned}
M^{V}_{\alpha,agg} &= \frac{D^{V}_{\alpha,agg}}{N^{V}_{\alpha,agg}} \\
&= \frac{y_{F,D_\alpha}}{\bar{v}_{D_\alpha,F}} \frac{\bar{v}_{N_\alpha,F}}{y_{F,N_\alpha}} \\
&= \frac{y_{F,D_\alpha}}{y_{F,N_\alpha}} \frac{\bar{v}_{N_\alpha,F}}{\bar{v}_{D_\alpha,F}}.
\end{aligned} \tag{34}
$$

B.5.2 Aggregate visibility estimands including the respondent

Estimand for deaths using reports that include the respondent

$$D'^{V}_{\alpha,agg} = \frac{y'_{F,D_\alpha}}{\bar{v}'_{D_\alpha,F}}. \tag{35}$$

Estimand for exposure using reports that include the respondents

$$N'^{V}_{\alpha,agg} = \frac{y'_{F,N_\alpha}}{\bar{v}'_{N_\alpha,F}}. \tag{36}$$

Estimand for the death rate using reports that include the respondent

$$\begin{aligned} M'^{V}_{\alpha,agg} &= \frac{D'^{V}_{\alpha,agg}}{N'^{V}_{\alpha,agg}} \\ &= \frac{y'_{F,D_\alpha}}{\bar{v}'_{D_\alpha,F}} \frac{\bar{v}'_{N_\alpha,F}}{y'_{F,N_\alpha}} \\ &= \frac{y'_{F,D_\alpha}}{y'_{F,N_\alpha}} \frac{\bar{v}'_{N_\alpha,F}}{\bar{v}'_{D_\alpha,F}}. \end{aligned} \tag{37}$$

C Sensitivity to invisible deaths

Reports about siblings can only tell us about the *visible* population – i.e., the group of people who have siblings on the frame population who can provide information about their survival. In this Appendix, we develop expressions that relate the death rate in the visible population to the death rate in the entire population. This expression will help researchers understand how different death rates in the visible population can be expected to be from death rates in the entire population.

In order to analyze the sensitivity of sibling survival estimates to invisible deaths, we need to develop notation that can be used to distinguish between visible and invisible deaths. For a demographic group α (for example, women aged 15-25 in 2018), let

$$\begin{aligned} p^V_{D_\alpha} &= \frac{D^V_\alpha}{D^V_\alpha + D^I_\alpha}, \quad \text{be the fraction of deaths that is visible;} \\ p^V_{N_\alpha} &= \frac{N^V_\alpha}{N^V_\alpha + N^I_\alpha}, \quad \text{be the fraction of exposure that is visible.} \end{aligned} \tag{38}$$

We define analogous quantities for the fraction of deaths and exposure that is invisible, $p^I_{D_\alpha}$ and $p^I_{N_\alpha}$.

Note that

$$\begin{aligned} \frac{p^V_{D_\alpha}}{p^V_{N_\alpha}} &= \frac{D^V_\alpha}{D^V_\alpha + D^I_\alpha} \times \frac{N^V_\alpha + N^I_\alpha}{N^V_\alpha} \\ &= \frac{M^V_\alpha}{M_\alpha}. \end{aligned} \tag{39}$$

Thus, the ratio of the fraction of deaths that is visible to the fraction of exposure that is visible is equal to the ratio of the visible death rate to the total death rate.

Result H.2 shows that the total death rate M_α can be understood as a weighted harmonic mean of the invisible death rate M^I_α and the visible death rate M^V_α, where the weights are given by the number of visible and invisible deaths. We now use this insight to develop Result C.1, which helps us understand the formal relationship between the invisible death rate, the visible death rate, and the total death rate.

Result C.1. *Suppose that, for a demographic group α, the invisible death rate (M^I_α) and the visible death rate (M^V_α) differ by a factor of K, so that*

$$M^I_\alpha = K M^V_\alpha \tag{40}$$

for $K > 0$. Then

$$M_\alpha = M^V_\alpha \left[\frac{K}{p^I_{D_\alpha} + K(1 - p^I_{D_\alpha})} \right], \tag{41}$$

where $p^I_{D_\alpha}$ is the proportion of deaths that is invisible.

Proof. Using the fact that M_α is the weighted harmonic mean of M^I_α and M^V_α, we find

$$\begin{aligned} M_\alpha &= \left[\frac{p^I_{D_\alpha}}{M^I_\alpha} + \frac{p^V_{D_\alpha}}{M^V_\alpha} \right]^{-1} \\ &= \left[\frac{p^I_{D_\alpha}}{K M^V_\alpha} + \frac{p^V_{D_\alpha}}{M^V_\alpha} \right]^{-1} \\ &= \left[\frac{p^I_{D_\alpha} + K p^V_{D_\alpha}}{K M^V_\alpha} \right]^{-1} \\ &= M^V_\alpha \left[\frac{K}{p^I_{D_\alpha} + K p^V_{D_\alpha}} \right] \\ &= M^V_\alpha \left[\frac{K}{p^I_{D_\alpha} + K(1 - p^I_{D_\alpha})} \right]. \end{aligned}$$

□

Result C.1 reveals that there is a relationship between between K, the difference between the visible and invisible death rates, and $p^I_{D_\alpha}$, which is related to the number of invisible deaths relative to the number of visible deaths. Equation 41 shows that

- when $K = 1$, $M^V_\alpha = M_\alpha$
- when $p^I_{D_\alpha} = 0$, $p^V_{D_\alpha} = 1$ and so $M^V_\alpha = M_\alpha$

It can also be helpful to use Result C.1 to obtain an expression for the relative error that would follow from using the visible death rate M^V_α as an estimate of the total death rate M_α:

$$\begin{aligned}\frac{M^V_\alpha - M_\alpha}{M_\alpha} &= \frac{M^V_\alpha}{M_\alpha} - 1 \\ &= \frac{p^V_{D_\alpha} + K(1 - p^V_{D_\alpha})}{K} - 1 \\ &= p^I_{D_\alpha}\left(\frac{1-K}{K}\right).\end{aligned} \tag{42}$$

In order to further develop intuition about how large we might expect biases due to invisible deaths to be, we can investigate different scenarios. For example, suppose that 10% of deaths are invisible, and the death rate is 20% higher among the invisible population than among the visible population. Then $K = 1.2$, $p^I_{D_\alpha} = 0.1$, and the relative error calculated from Equation 42 is about -.017; in other words, in this scenario, death rate estimates based on the visible population alone will be too low by about 1.7 percent.

Figure 7 illustrates this relative error for a range of values of K and $p^I_{D_\alpha}$.

Next, Result C.2 provides a second expression that analyzes the formal relationship between the invisible death rate, the visible death rate, and the total death rate; this second result is parameterized in terms of $p^I_{N_\alpha}$, the proportion of exposure that is invisible. This is the relationship used in the main text.

Result C.2. *Suppose that, for a demographic group α, the invisible death rate (M^I_α) and the visible death rate (M^V_α) differ by a factor of K, so that*

$$M^I_\alpha = K M^V_\alpha \tag{43}$$

for $K > 0$. Then

$$M_\alpha = M^V_\alpha \left[1 + p^I_{N_\alpha}(K-1)\right], \tag{44}$$

where $p^I_{N_\alpha}$ is the proportion of exposure that is invisible.

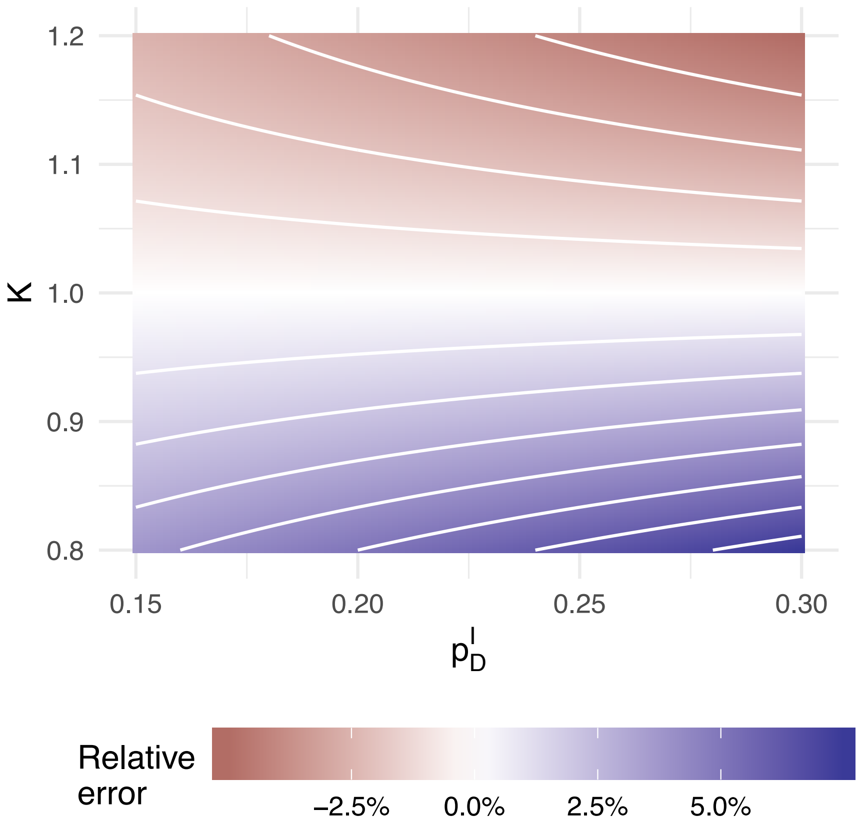

Figure 7: Illustration of the relative error in using the visible death rate M^V as an estimate for the total death rate M. The proportion of deaths that is invisible, $p^I_{D_\alpha}$, varies along the x axis; the relationship between the visible and invisible death rates, captured by the parameter K (Equation 40), varies along the y axis. The colors show the percentage relative error; so if 20% of deaths are invisible ($p^I_{D_\alpha} = 0.2$) and the invisible death rate is 10% higher than the visible death rate ($K = 1.1$), the relative error is about 2 percent. Relative error increases as K gets farther away from 1 and as $p^V_{D_\alpha}$ increases.

Proof. Using the fact that M_α is the weighted arithmetic mean of M_α^I and M_α^V, we find

$$\begin{aligned} M_\alpha &= p_{N_\alpha}^V M_\alpha^V + p_{N_\alpha}^I M_\alpha^I \\ &= (1 - p_{N_\alpha}^I) M_\alpha^V + p_{N_\alpha}^I K M_\alpha^V \\ &= M_\alpha^V - M_\alpha^V p_{N_\alpha}^I + K p_{N_\alpha}^I M_\alpha^V \\ &= M_\alpha^V \left[1 + p_{N_\alpha}^I (K - 1) \right]. \end{aligned}$$

□

Again, it can be helpful to use Result C.2 to obtain an expression for the relative error that would follow from using the visible death rate M_α^V as an estimate of the total death rate M_α:

$$\begin{aligned} \frac{M_\alpha^V - M_\alpha}{M_\alpha} &= \frac{M_\alpha^V}{M_\alpha} - 1 \\ &= \frac{1}{1 + p_{N_\alpha}^I (K - 1)} - 1 \\ &= \frac{p_{N_\alpha}^I (1 - K)}{1 - p_{N_\alpha}^I (1 - K)}. \end{aligned} \tag{45}$$

We can further develop intuition about how large we might expect biases due to invisible deaths to be, we can investigate different scenarios. For example, suppose that 10% of exposure is invisible, and the death rate is 20% higher among the invisible population than among the visible population. Then $K = 1.2$, $p_{N_\alpha}^I = 0.1$, and the relative error from Equation 45 is about -0.019; in other words, in this scenario, death rate estimates based on the visible population alone will be too low by about 1.9 percent.

Figure 4b illustrates this second expression for the relative error over a range of values of K and $p_{N_\alpha}^I$.

D Sampling

In Appendix B, we developed several estimands based on sibling reports. These estimands describe quantities that could be estimated from a census of the frame population. In practice, researchers do not have a census of the frame population, but rather a sample from the frame population. This section develops some results that will be helpful in understanding how to develop sample-based estimators for the estimands in Appendix B.

D.1 Sampling setup

We use the design-based sampling framework described in Sarndal, Swensson, and Wretman (2003), repeating a few key definitions here for convenience. We assume we have a probability sample s from a frame population F; common frame populations include all adults, all adults aged 15-59, and in many DHS surveys, all women aged 15-59. The random variable I_i takes the value 1 when $i \in F$ is included in the sample, and 0 otherwise. Each $i \in F$ has a nonzero *probability of inclusion* $\pi_i = \mathbb{E}[I_i]$ and the sampling weights are given by $w_i = \frac{1}{\pi_i}$.

Suppose some quantity y_i is defined for every $i \in F$. Then the *Horvitz Thompson estimator* for the population total $Y = \sum_{i \in F} y_i$ from a probability sample s is given by

$$\widehat{Y} = \sum_{i \in s} w_i y_i.$$

Sarndal, Swensson, and Wretman (2003) shows that Horvitz-Thompson estimators are consistent and unbiased[12], a fact that will be useful below.

Result D.1. *Suppose a Horvitz-Thompson estimator*

$$\widehat{Y}^{HT} = \sum_{i \in s} w_i y_i$$

is design-unbiased for a total Y. Then $\widehat{Y}^{HT}$ is also (design) consistent for Y.

Proof. This result follows from taking the sampling design to assign $\pi_i = 1$ for all $i \in F$. We then have $s = F$ and $w_i = 1$ for all i. Since the estimator is unbiased, design consistency follows.

□

Next, we state a Result that is helpful when devising estimators that are ratios of other estimators.

Result D.2. *Suppose that $\widehat{y}_1, \ldots, \widehat{y}_n$ are estimators that are consistent and unbiased for $Y_1, \ldots, Y_n$ respectively. Then the compound ratio estimator*

$$\widehat{R} = \frac{\widehat{y}_1 \cdots \widehat{y}_k}{\widehat{y}_{k+1} \cdots \widehat{y}_n}.$$

[12] In this paper, we use the framework of design-based sampling, so the properties of estimators – such as unbiasedness and consistency – are with respect to the probability sampling mechanism. There are many types of consistency; we refer in this work to design-consistency, also called Fisher consistency.

is consistent and essentially unbiased for $R = (Y_1 \dots Y_k)/(Y_{k+1} \dots Y_n)$.

Proof. See J. N. K. Rao and Pereira (1968), Wolter (2007) (pg. 233), and Feehan and Salganik (2016a) for more details.

□

The references that derive the compound ratio estimator in Result D.2 show that, technically, the estimator is biased. However, the bias has been shown to be of order $O(n^{-1})$; that is, the bias goes to 0 as the sample size $n \to \infty$. Further, Feehan and Salganik (2016a) investigated compound ratio bias in network reporting estimators and found suggestive evidence that it is likely to be very small. This finding is consistent with the literature on ratio estimators, which has revealed that bias is often very small. Thus, in these technical results, we refer to estimators whose only bias arises from being compound ratios – i.e., whose bias is of order $O(n^{-1})$ – as *essentially unbiased.*

D.2 Adjusting for visibility in samples

In this section, we briefly state key results that explain which conditions are required for individual and aggregate visibility estimators to be consistent and unbiased. More detailed derivations of these results can be found elsewhere.

Individual visibility estimation from a probability sample

Definition D.1. The **individual visibility** estimator for a total $\boldsymbol{Y} = \sum_{i \in U} y_i$ is defined to be

$$\widehat{Y}^{\text{ind}} = \sum_{i \in s} w_i \sum_{j \sim i} \frac{y_j}{v_{j,F}}, \tag{46}$$

where $j \sim i$ indexes the neighbors j of respondent i, w_i is the design weight for respondent i, and $v_{j,F}$ is the visibility of person j to the frame population F. In the situation where there are no errors in reporting, the visibility $v_{j,F}$ has also been called the multiplicity of person j (Sirken 1970).

Result D.3. *Suppose that there are no false positive reports. The individual visibility estimator is consistent and unbiased for the total* $\boldsymbol{Y}$.

Proof. First, we show that the estimator is unbiased. If π_i is i's probability of inclusion under the sampling design, then the design weights are $w_i = \frac{1}{\pi_i}$. Thus, we have

$$\begin{aligned}
\widehat{\boldsymbol{Y}}^{\text{ind}} &= \sum_{i \in s} \frac{1}{\pi_i} \sum_{j \sim i} \frac{y_j}{v_{j,F}} \\
\iff \mathbb{E}[\widehat{\boldsymbol{Y}}^{\text{ind}}] &= \sum_{i \in F} \mathbb{E}[I_i] \frac{1}{\pi_i} \sum_{j \sim i} \frac{y_j}{v_{j,F}} \\
&= \sum_{i \in F} \sum_{j \sim i} \frac{y_j}{v_{j,F}} \\
&= \sum_{l \in F} v_{l,F} \frac{y_l}{v_{l,F}} \\
&= \sum_{l \in F} y_l.
\end{aligned}$$

The last step follows because, as long as there are no false positive reports, in a census of F, each unit j appears once for each time it is visible to F; that is, $v_{j,F} = y_{F,j}$ (see Feehan and Salganik (2016a) for details).

Note that the derivation above reveals that the individual visibility estimator can be written as a Horvitz-Thompson estimator (Thompson 2002; Sirken 1970): to see how, define $z_i = \sum_{j \sim i} \frac{y_j}{v_{j,F}}$ for all $i \in F$. The individual visibility estimator in Equation 46 then becomes $\widehat{Y} = \sum_{i \in s} w_i z_i$. Since $\widehat{\boldsymbol{Y}}^{\text{ind}}$ can be written as a Horvitz-Thompson estimator, Result D.1 shows that unbiasedness implies consistency.

□

Good references for individual visibility estimators include Sirken (1970) and Lavallee (2007).

Aggregate visibility estimation from a probability sample

Definition D.2. The **aggregate visibility** estimator for a total $\boldsymbol{Y} = \sum_{i \in U} y_i$ is defined to be

$$\widehat{Y}^{\text{agg}} = \frac{\sum_{i \in s} w_i \sum_{j \sim i} y_j}{\widehat{\bar{v}}_{Y,F}},$$

where $\bar{v}_{Y,F} = N^{-1} \sum_{l \in U} v_{l,F}$ is the average of the individual visibilities of each person who could be reported about in a census of F.

The network scale-up estimator, the network survival estimator, and related approaches are examples of estimators based on the idea of aggregate visibility (Bernard et al. 1989; Feehan and Salganik 2016a; Feehan, Mahy, and Salganik 2017).

Result D.4. *Suppose that there are no false positive reports. Then the aggregate visibility estimator is consistent and essentially unbiased for the total* $\boldsymbol{Y}$*.*

Proof. See Feehan and Salganik (2016a).

□

E Estimating death rates using aggregate visibility

This section presents results for death rate estimators that adjust for visibility at the aggregate level; that is, the sum of reported connections to siblings across respondents is adjusted for using the average visibility of the reported siblings.

Section B.4 showed that there is no one obvious way to estimate the aggregate visibility of deaths or exposure using data collected from the sibling histories. However, if the visibility of deaths is the same as the visibility of exposure, then the death rate can be estimated without directly estimating the visibility of deaths or exposure, since these two quantities will cancel. This condition is analogous to the requirement, discussed at length in the sibling survival literature, that there be no correlation between sibship size and mortality. We suggest that researchers assess the sensitivity of aggregate visibility estimators to this condition using the sensitivity framework we develop below.

We will first provide an estimator that is based on excluding information about the survey respondents and only using reports about respondents' siblings. We will see that this approach produces an estimator that is essentially the one recommended in the DHS program's official documentation (Rutstein and Guillermo Rojas 2006, pg. 156) and by other researchers (e.g., Masquelier 2013). An additional estimator, based on including information about the survey respondents, turns out to be very similar, and follows as a Corollary of the derivation of the without-respondent estimator.

The relationship derived in Equation 34 suggests

$$\begin{aligned}\widehat{M}_{\alpha}^{V} &= \frac{\widehat{D}_{\alpha}^{V}}{\widehat{N}_{\alpha}^{V}} \\ &= \frac{\sum_{i\in s} w_i \; y(i, D_{\alpha})}{\sum_{i\in s} w_i \; y(i, N_{\alpha})}.\end{aligned} \tag{47}$$

Equation 47 is based on the idea that we can plug in sample-based aggregate visibility estimators for the number of siblings who died and the number of siblings who contribute exposure. Result E.1 formally states the conditions are required for the estimator to produce consistent and unbiased estimates for the visible death rate.

Result E.1. ***(Aggregate visibility estimation for death rates)*** *Suppose we have a probability sample $s \subset F$ and the associated weights w_i for all $i \in s$. Suppose that, in aggregate, there are no false positive reports about deaths or exposure, so that $y(F, D_\alpha) = y^+(F, D_\alpha)$ and $y(F, N_\alpha) = y^+(F, N_\alpha)$. Suppose also that the visibility of deaths and exposure to the frame population are the same, and nonzero, so that $\bar{v}(D_\alpha, F) = \bar{v}(N_\alpha, F) > 0$. Then*

$$\begin{aligned}\widehat{M}^V &= \frac{\widehat{y}(F, D_\alpha)}{\widehat{y}(F, N_\alpha)} \\ &= \frac{\sum_{i \in s} w_i \; y(i, D_\alpha)}{\sum_{i \in s} w_i \; y(i, N_\alpha)}\end{aligned} \tag{48}$$

is consistent and essentially unbiased for $M^V_\alpha = \frac{D^V_\alpha}{N^V_\alpha}$, where the exposure N^V_α is approximated by the number of visible people, living or dead, in group α.

Proof. The numerator and denominator are Horvitz-Thompson estimators, and are therefore consistent and unbiased for the population quantities $y(F, D_\alpha)$ and $y(F, N_\alpha)$. By Result D.2, $\widehat{M}^V$ is then consistent and essentially unbiased for the estimand $\frac{y(F,D_\alpha)}{y(F,N_\alpha)}$, so it remains to show that this estimand is equal to the visible death rate. Using the fact that $\bar{v}(D_\alpha, F) = \bar{v}(N_\alpha, F)$, we have

$$\frac{y(F, D_\alpha)}{y(F, N_\alpha)} = \frac{y(F, D_\alpha)}{y(F, N_\alpha)} \; \frac{\bar{v}(N_\alpha, F)}{\bar{v}(D_\alpha, F)}.$$

Since there are no false positive reports about death or exposure, we have

$$\frac{y(F, D_\alpha)}{y(F, N_\alpha)} \; \frac{\bar{v}(N_\alpha, F)}{\bar{v}(D_\alpha, F)} = \frac{y^+(F, D_\alpha)}{y^+(F, N_\alpha)} \; \frac{\bar{v}(N_\alpha, F)}{\bar{v}(D_\alpha, F)}.$$

By the aggregate reporting identity, $y^+(F, D_\alpha)/\bar{v}(D_\alpha, F) = D^V_\alpha$ and $y^+(F, N_\alpha)/\bar{v}(N_\alpha, F) = N^V_\alpha$. Therefore, we have

$$\frac{y^+(F, D_\alpha)}{y^+(F, N_\alpha)} \; \frac{\bar{v}(N_\alpha, F)}{\bar{v}(D_\alpha, F)} = \frac{D^V_\alpha}{N^V_\alpha} = M^V_\alpha.$$

□

To recap, Result E.1 shows that the estimator for visible death rates relies on the important condition that the visibility of deaths is the same as the visibility of exposure. This requirement does not mean that reports have to be perfectly accurate; in particular, omitting siblings will only cause a problem if it happens at a different rate for dead siblings than it does for living siblings. We discuss this issue in more detail when we develop the sensitivity analysis framework, below.

The Corollary below says that, when respondents are included in reports, then an analogous set of conditions to the ones required by Result E.1 will allow consistent and unbiased estimation of the death rate.

Corollary E.1. ***(Aggregate visibility estimation for death rates when respondents are included in reports).*** *Result E.1 holds under analogous conditions when reports include the respondents. Specifically, suppose we have a probability sample $s \subset F$ and the associated weights w_i for all $i \in s$. Suppose that there are no false positive reports about deaths or exposure, so that $y'(F, D_\alpha) = y'^{+}(F, D_\alpha)$ and $y'(F, N_\alpha) = y'^{+}(F, N_\alpha)$. Suppose also that the visibility of deaths and exposure to the frame population are the same, and nonzero, so that $\bar{v}'(D_\alpha, F) = \bar{v}'(N_\alpha, F) > 0$. Then*

$$\begin{aligned} \widehat{M}'^{V} &= \frac{\widehat{y}'(F, D_\alpha)}{\widehat{y}'(F, N_\alpha)} \\ &= \frac{\sum_{i \in s} w_i \; y'(i, D_\alpha)}{\sum_{i \in s} w_i \; y'(i, N_\alpha)} \end{aligned} \tag{49}$$

is consistent and essentially unbiased for $M'^{V}_\alpha = \frac{D'^{V}_\alpha}{N'^{V}_\alpha}$, where the exposure N^V_α is approximated by the number of visible people, living or dead, in group α.

Proof. The proof follows the same steps as the proof of Result E.1.

□

It is important to note that the aggregate visibility estimator based on not including information about respondents (Result E.1) is consistent and unbiased for a visible death rate M^V_α, and that the aggregate visibility estimator based on including information about respondents (Corollary E.1) is consistent and unbiased for a different visible death rate M'^{V}_α. In other words, the *definition* of the visible and invisible populations will be affected by whether or not respondents are included in the sibling reports. We analyze this issue in more detail in Appendix G.1.

Sensitivity of the aggregate visibility death rate estimator

We now turn to the sensitivity of the aggregate visibility death rate estimator to the various conditions it relies upon. Our approach is to introduce several quantities, called *adjustment factors*, that capture the degree to which the conditions the estimator relies upon are satisfied. These adjustment factors will be equal to 1 under ideal conditions, and will be different from 1 when a condition required by the death rate estimator is not satisfied. Our approach is closely related to other network reporting analyses, including Feehan and Salganik (2016a) and Feehan, Mahy, and Salganik (2017).

We consider here the case where respondents are not included in reports; however, an analogous sensitivity expression would result from extending our approach to the case where respondents are included in sibling reports.

The first adjustment factor is the *true positive rate* for reports about visible deaths, $\tau(F, D_\alpha)$; it is defined as

$$\tau(F, D_\alpha) = \frac{\text{average \# times a visible death in } \alpha \text{ would be reported by someone in } F}{\text{average number of connections a visible death in } \alpha \text{ has to } F} = \frac{\bar{v}(D^V_\alpha, F)}{\bar{d}(D^V_\alpha, F)}. \tag{50}$$

Note that we can also write $\tau(F, D_\alpha) = \frac{\bar{y}^+(F, D_\alpha)}{\bar{d}(F, D^V_\alpha)}$, since $v(D^V_\alpha, F) = y^+(F, D_\alpha)$.

The second adjustment factor is the *precision* for reports about visible deaths, $\eta(F, D_\alpha)$; it is defined as

$$\eta(F, D_\alpha) = \frac{\text{\# of reported connections from } F \text{ to } D_\alpha \text{ that actually lead to } D_\alpha}{\text{\# of reported connections from } F \text{ to } D_\alpha} = \frac{y^+(F, D_\alpha)}{y(F, D_\alpha)}. \tag{51}$$

We define analogous adjustment factors $\tau(F, N_\alpha)$ and $\eta(F, N_\alpha)$ for F's reports about siblings' exposure.

We can use the adjustment factors introduced in the previous section to decompose the aggregate visibility sibling survival estimator as

$$\begin{aligned} M_\alpha &= \frac{D_\alpha}{N_\alpha} \\ &= \frac{D^V_\alpha}{N^V_\alpha} \times \frac{p^V_{N_\alpha}}{p^V_{D_\alpha}} \\ &= \frac{D^V_\alpha}{N^V_\alpha} \times \frac{M_\alpha}{M^V_\alpha} \\ &= \frac{y_{F,D_\alpha}}{y_{F,N_\alpha}} \times \frac{\bar{d}^V(N_\alpha, F)}{\bar{d}^V(D_\alpha, F)} \times \frac{\eta(F, D_\alpha)}{\eta(F, N_\alpha)} \times \frac{\tau(F, N_\alpha)}{\tau(F, D_\alpha)} \times \frac{M_\alpha}{M^V_\alpha}. \end{aligned} \tag{52}$$

In order to simplify the eventual framework, let us introduce two quantities that capture net reporting about deaths and about exposure:

$$\gamma(F, D_\alpha) = \frac{\tau(F, D_\alpha)}{\eta(F, D_\alpha)},$$

and

$$\gamma(F, N_\alpha) = \frac{\tau(F, N_\alpha)}{\eta(F, N_\alpha)}.$$

The final step is to incorporate the expression for sensitivity to invisible siblings. Let the visible and invisible death rates differ by a factor K so that $M_\alpha^I = K M_\alpha^V$. Equation 41 then tells us that

$$\frac{M_\alpha^V}{M_\alpha} = \frac{p_{D_\alpha}^V + K(1 - p_{D_\alpha}^V)}{K}. \tag{53}$$

Combining Equation 41 and Equation 52, and substituting $\gamma(F, D_\alpha)$ and $\gamma(F, N_\alpha)$, we obtain an expression that relates the aggregate visibility estimand to the true death rate:

$$M_\alpha = \underbrace{\frac{y(F, D_\alpha)}{y(F, N_\alpha)}}_{\substack{\text{aggregate}\\ \text{mulitiplicity}\\ \text{estimand}}} \times \underbrace{\frac{\bar{d}^V(N_\alpha, F)}{\bar{d}^V(D_\alpha, F)}}_{\substack{\text{degree}\\ \text{ratio}}} \times \underbrace{\frac{\gamma(F, N_\alpha)}{\gamma(F, D_\alpha)}}_{\substack{\text{reporting}\\ \text{accuracy}}} \times \underbrace{\left[\frac{K}{p_{D_\alpha}^I + K(1 - p_{D_\alpha}^I)}\right]}_{\substack{\text{difference between}\\ \text{invisible and}\\ \text{visible}\\ \text{populations}}}. \tag{54}$$

This decomposition relates the quantities we can observe or estimate from a survey— i.e., $y(F, D_\alpha)$ and $y(F, N_\alpha)$–to the quantity that we actually wish to estimate, i.e., M_α.

The decomposition in Equation 54 produces two groups of factors that are the ratio of (i) an adjustment factor for deaths; and (ii) the same adjustment factor for exposure (for example, $\frac{\gamma(F,D_\alpha)}{\gamma(F,N_\alpha)}$). To the extent that reporting about deaths and reporting about exposure is similar, this is advantageous: these adjustment factors can cancel or counteract one another. Intuitively, Equation 52 reveals that the death rate estimator is quite robust to situations in which respondents' reports are imperfect, but imperfect in similar ways for deaths and for people who didn't die.

Equation 52 parameterizes the difference between the visible and visible populations in terms of $p_{D_\alpha}^D$, the proportion of deaths that is invisible. Researchers may prefer to parameterize this factor in terms of $p_{N_\alpha}^I$, the proportion of exposure that is invisible. In that case, Equation 52 becomes:

$$M_\alpha = \underbrace{\frac{y(F, D_\alpha)}{y(F, N_\alpha)}}_{\substack{\text{aggregate}\\ \text{mulitiplicity}\\ \text{estimand}}} \times \underbrace{\frac{\bar{d}^V(N_\alpha, F)}{\bar{d}^V(D_\alpha, F)}}_{\substack{\text{degree}\\ \text{ratio}}} \times \underbrace{\frac{\gamma(F, N_\alpha)}{\gamma(F, D_\alpha)}}_{\substack{\text{reporting}\\ \text{accuracy}}} \times \underbrace{\left[1 + p_{N_\alpha}^I (K - 1)\right]}_{\substack{\text{difference between}\\ \text{invisible and}\\ \text{visible}\\ \text{populations}}}. \tag{55}$$

Sensitivity for reports that include the respondent

The derivations that led to the sensitivity framework for the aggregate visibility estimand (Equation 54 and Equation 55) can also be used to develop a sensitivity framework for the aggregate visibility estimand with respondents included in reports: only slight modifications are necessary to account for the fact that respondents include themselves in reports. We do not repeat all of the derivations in detail but, for convenience, Equation 56 spells out the sensitivity framework for the aggregate visibility estimand with respondents included in reports.

$$M_\alpha = \underbrace{\frac{y'(F, D_\alpha)}{y'(F, N_\alpha)}}_{\substack{\text{aggregate}\\\text{mulitiplicity}\\\text{estimand}\\\text{including}\\\text{respondent}}} \times \underbrace{\frac{\bar{d}'(N_\alpha^V, F)}{\bar{d}'(D_\alpha^V, F)}}_{\substack{\text{degree}\\\text{ratio}}} \times \underbrace{\frac{\gamma'(F, N_\alpha)}{\gamma'(F, D_\alpha)}}_{\substack{\text{reporting}\\\text{accuracy}}} \times \underbrace{\left[1 + p'^{I}_{N_\alpha}(K' - 1)\right]}_{\substack{\text{difference between}\\\text{invisible and}\\\text{visible}\\\text{populations}}}, \tag{56}$$

where

- $\gamma'(F, N_\alpha) = \frac{\tau'(F,N_\alpha)}{\eta'(F,N_\alpha)}$
- $\gamma'(F, D_\alpha) = \frac{\tau'(F,D_\alpha)}{\eta'(F,D_\alpha)}$
- reporting parameters for with-respondent reports are defined analogous to the without-respondent parameters; for example, $\tau'(i, N_\alpha^V) = \frac{y'^{+}(i,D_\alpha^V \cap \sigma)}{d'^{+}(D_\alpha^V \cap \sigma, i)}$
- $p'^{I}_{N_\alpha}$ is the proportion of deaths that is invisible with respect to $M_\alpha'^{V}$ (ie, the visible death rate when respondents are included in reports, which is different from M_α^V, the visible death rate when respondents are excluded from reports)
- K' is a parameterization of the difference between the invisible and visible death rates when respondents are included in reports

F Estimating death rates using individual visibility

This section presents results for death rate estimators that adjust for visibility at the individual level; that is, each reported connection to a siblings is adjusted for using the visibility of that specific reported sibling.

First, we provide an estimator that is based on excluding information about the survey respondents and only using reports about respondents' siblings.

Result F.1. ***(Individual visibility estimation for death rates)*** *Suppose that reports are accurate at the individual level, so that $y(i, D_\alpha) = d(i, D_\alpha)$ and $y(i, N_\alpha) = d(i, N_\alpha)$. Then*

$$\begin{aligned} \widehat{M}^V &= \frac{\widehat{D_\alpha}}{\widehat{N_\alpha}} \\ &= \frac{\sum_{i \in s} w_i \frac{y(i,D_\alpha)}{y(i,F)+1}}{\sum_{i \in s} w_i \left[\frac{y(i,N_\alpha \cap F)}{y(i,F)} + \frac{y(i,N_\alpha - F)}{y(i,F)+1}\right]} \end{aligned} \tag{57}$$

is consistent and essentially unbiased for $M_\alpha^V = \frac{D_\alpha^V}{N_\alpha^V}$, where the exposure N_α^V is approximated by the number of visible people, living or dead, in group α.

Proof. The numerator and denominator are Horvitz-Thompson estimators, and therefore

$$\begin{aligned}
\sum_{i \in s} w_i \frac{y(i, D_\alpha)}{y(i, F) + 1} \longrightarrow \sum_{i \in F} \frac{y(i, D_\alpha)}{y(i, F) + 1} & \\
= \sum_{i \in F} \frac{d(i, D_\alpha)}{d(i, F) + 1} & \quad \text{by the perfect reporting condition} \\
= D_\alpha^V. & \quad \text{by the derivation in Equation 25}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{i \in s} w_i \left[\frac{y(i, N_\alpha \cap F)}{y(i, F)} + \frac{y(i, N_\alpha - F)}{y(i, F) + 1} \right] \longrightarrow \sum_{i \in F} \left[\frac{y(i, N_\alpha \cap F)}{y(i, F)} + \frac{y(i, N_\alpha - F)}{y(i, F) + 1} \right] & \\
= \sum_{i \in F} \left[\frac{d(i, N_\alpha \cap F)}{d(i, F)} + \frac{d(i, N_\alpha - F)}{d(i, F) + 1} \right] & \quad \text{by perfect reporting condition} \\
= N_\alpha & \quad \text{by relationship in Equation 26}
\end{aligned}$$

Therefore, the numerator of Equation 58 is consistent and unbiased for the visible deaths, D_α^V, and the denominator of Equation 58 is consistent and unbiased for the visible exposure N_α^V. Result D.2 then shows that the ratio of these two estimators, $\widehat{M}^V = \frac{\widehat{D}_\alpha}{\widehat{N}_\alpha}$ will be consistent and essentially unbiased for, $\frac{D_\alpha^V}{N_\alpha^V}$, the visible death rate.

□

Result F.1 applies in the situation where respondents are not included in the sibship reports. We now turn to a Corollary that addresses the situation where information about respondents is included in the sibship reports.

Corollary F.1. ***(Individual visibility estimation for death rates when respondents are included in reports)*** *Result F.1 holds under analogous conditions when reports include the respondents. Specifically, suppose we have a probability sample $s \subset F$ and the associated weights w_i for all $i \in s$. Suppose that reports are accurate at the individual level, so that $y'(i, D_\alpha) = d'(i, D_\alpha)$ and $y'(i, N_\alpha) = d'(i, N_\alpha)$. Then*

$$\begin{aligned}
\widehat{M}'^V &= \frac{\widehat{D'_\alpha}}{\widehat{N'_\alpha}} \\
&= \frac{\sum_{i \in s} w_i \frac{y'(i, D_\alpha)}{y'(i, F)}}{\sum_{i \in s} w_i \frac{y'(i, N_\alpha)}{y'(i, F)}}
\end{aligned} \tag{58}$$

is consistent and essentially unbiased for $M_\alpha'^V = \frac{D_\alpha'^V}{N_\alpha'^V}$, where the exposure N_α^V is approximated by the number of visible people, living or dead, in group α.

Proof. The proof is essentially the same as the proof of Result F.1; we repeat it here for clarity. The numerator and denominator are Horvitz-Thompson estimators, and therefore

$$\begin{aligned}
\sum_{i \in s} w_i \frac{y'(i, D_\alpha)}{y'(i, F)} &\longrightarrow \sum_{i \in F} \frac{y'(i, D_\alpha)}{y'(i, F)} \\
&= \sum_{i \in F} \frac{d'(i, D_\alpha)}{d'(i, F)} && \text{by the perfect reporting condition} \\
&= D'^{V}_{\alpha}. && \text{by the derivation in Equation 22}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{i \in s} w_i \frac{y'(i, N_\alpha}{y'(i, F)} &\longrightarrow \sum_{i \in F} \frac{y'(i, N_\alpha)}{y'(i, F)} \\
&= \sum_{i \in F} \frac{d'(i, N_\alpha)}{d'(i, F)} && \text{by perfect reporting condition} \\
&= N'_{\alpha}. && \text{by derivation in Equation 23}
\end{aligned}$$

Therefore, the numerator of the estimator (Equation 58) is consistent and unbiased for the visible deaths, D'^{V}_{α}, and the denominator of Equation 58 is consistent and unbiased for the visible exposure N'^{V}_{α}. Result D.2 then shows that the ratio of these two estimators, $\widehat{M}'^{V} = \frac{\widehat{D}'_\alpha}{\widehat{N}'_\alpha}$ will be consistent and essentially unbiased for, $\frac{D'^{V}_{\alpha}}{N'^{V}_{\alpha}}$, the visible death rate when respondents are included in reports.

□

F.1 Sensitivity of individual death rate estimator

Now we turn to an analysis of the sensitivity of the individual visibility estimator. We follow the same approach that we did for the aggregate visibility estimator: we introduce a series of adjustment factors that relate reports to the underlying sibship network. Using these adjustment factors, we develop expressions that show how reporting errors and other factors will affect estimated death rates.

Adjustment factors for deaths

We start by defining individual-level adjustment factors that will be useful in understanding how reporting errors can affect the individual estimator. For a particular respondent $i \in F$ in sibship σ, let

$$\tau(i, D_\alpha \cap \sigma) = \frac{\#\text{ sibs } i \text{ reports having died in group } \alpha \text{ that actually died in group } \alpha}{\#\text{ siblings of } i \text{ that actually died in group } \alpha} = \frac{y^+(i, D^V_\alpha \cap \sigma)}{d^V(D_\alpha \cap \sigma, i)},$$

and let

$$\eta(i, D_\alpha \cap \sigma) = \frac{\text{\# siblings } i \text{ reports having died in group } \alpha \text{ that actually died in group } \alpha}{\text{\# of siblings } i \text{ reports having died in group } \alpha} = \frac{y^+(i, D_\alpha \cap \sigma)}{y(i, D_\alpha \cap \sigma)}.$$

$\tau(i, D_\alpha \cap \sigma)$ and $\eta(i, D_\alpha \cap \sigma)$ are the individual-level analogues of the quantities we introduced in the previous section[13]. Analogous reporting quantities can be defined for reports about exposure among siblings, $\tau(i, N_\alpha \cap \sigma)$ and $\eta(i, N_\alpha \cap \sigma)$, and for reports about siblings' membership in the frame population, $\tau(i, \sigma \cap F)$ and $\eta(i, \sigma \cap F)$.

We also define a combined adjustment factor for reports about deaths made by each individual i in sibship σ[14]:

$$\gamma(i, D_\alpha \cap \sigma) = \frac{\tau(i, D_\alpha \cap \sigma)}{\eta(i, D_\alpha \cap \sigma)} \frac{\eta'(i, \sigma \cap F)}{\tau'(i, \sigma \cap F)}.$$

(Following our notational convention, τ' and η' refer to individual-level true positive rates and precision for reports that include the respondent.) $\gamma_{i, D_\alpha \cap \sigma}$ can be considered the net reporting parameter for respondent i. Using $\gamma(i, D_\alpha)$ instead of $\tau(i, D_\alpha)$ and $\eta(i, D_\alpha)$ will help simplify the expressions we derive below.

Note that, by definition,

$$\gamma(i, D_\alpha \cap \sigma) = \frac{y(i, D_\alpha \cap \sigma)}{d(i, D_\alpha \cap \sigma)} \frac{[d(i, F \cap \sigma) + 1]}{[y(i, F \cap \sigma) + 1]}.$$

This means that

$$\frac{y(i, D_\alpha \cap \sigma)}{y(i, F \cap \sigma) + 1} = \frac{d(i, D_\alpha \cap \sigma)}{d(i, F \cap \sigma) + 1} \gamma(i, D_\alpha \cap \sigma), \tag{59}$$

a relationship that will prove useful below.

Adjustment factors for exposure

The exposure in the denominator of the estimator in Result F.1 is the sum of two terms: one term related to reports about siblings on the frame population and one term related to reports about siblings not on the frame population. Therefore, we require two types of reporting terms: one

[13] In the special case where $d^V(i, D_\alpha \cap \sigma) = 0$, we define $\tau(i, D_\alpha \cap \sigma) = 0$; similarly, we define all τ quantities to be zero when their denominators are zero. (Additional τ quantities are introduced below.)

[14] In the special case when $\eta(i, D_\alpha \cap \sigma) = 0$, we define $\gamma(i, D_\alpha \cap \sigma) = 0$. Similarly, we define all γ quantities to be zero when their denominators are zero. (Additional γ quantities are introduced below.)

appropriate for reports about siblings on the frame population, and one appropriate for reports about siblings who are not on the frame population. This distinction is necessary because the individual-level visibility adjustment is different for siblings who are on and off the frame population.

For reported exposure among siblings on the frame population, we define net reporting factors:

$$\gamma(i, N_\alpha \cap F) = \frac{\tau(i, N_\alpha \cap F)}{\eta(i, N_\alpha \cap F)} \frac{\eta(i, \sigma \cap F)}{\tau(i, \sigma \cap F)}.$$

Similarly, for reported exposure among siblings not on the frame population, we define net reporting factors:

$$\gamma(i, N_\alpha - F) = \frac{\tau(i, N_\alpha - F)}{\eta(i, N_\alpha - F)} \frac{\eta'(i, \sigma \cap F)}{\tau'(i, \sigma \cap F)}.$$

(Again, following our notational convention, τ' and η' refer to individual-level true positive rates and precision for reports that include the respondent.)

Sensitivity of deaths

Using these adjustment factors, we can now derive an expression that relates reported quantities to the actual visible death rate. We'll do this separately for the numerator and the denominator; starting with the numerator, deaths. It will make sense to begin by considering reports about deaths in a particular sibship σ:

$$\begin{aligned}
\sum_{i \in \sigma \cap F} \frac{y(i, D_\alpha)}{y(i, F) + 1} &= \sum_{i \in \sigma \cap F} \frac{d(i, D_\alpha)}{d(i, F) + 1} \gamma(i, D_\alpha) && \text{from Equation 59} \\
&= \bar{d}(\sigma \cap F, D_\alpha) \sum_{i \in \sigma \cap F} \frac{1}{d(i, F) + 1} \gamma(i, D_\alpha) && \text{since } d(i, D_\alpha) = \bar{d}(\sigma \cap F, D_\alpha) \ \forall i \in \sigma \cap F \\
&= \frac{\bar{d}(\sigma \cap F, D_\alpha)}{|\sigma \cap F|} \sum_{i \in \sigma \cap F} \gamma(i, D_\alpha) && \text{since } d(i, F) + 1 = |\sigma \cap F| \\
&= D_\alpha^V \ \bar{\gamma}(\sigma \cap F, D_\alpha),
\end{aligned} \tag{60}$$

where we have defined $\bar{\gamma}(\sigma \cap F, D_\alpha) = |\sigma \cap F|^{-1} \sum_{i \in \sigma \cap F} \gamma(i, D_\alpha)$ to be the average net reporting factor for deaths in sibship σ.

The derivation in Equation 60 is based on the idea that the net adjustment factor $\gamma(i, D_\alpha \cap \sigma)$ relates the individual reports to the actual underlying sibship network; by applying the net adjustment factor, we develop an understanding of how the actual number of visible deaths in the sibship, $D_{\alpha \cap \sigma}^V$ is related to the individual visibility estimate.

Equation 60 shows how the reported connections to deaths for one sibship can be expressed as the product of the actual number of deaths in the sibship and the average net reporting factor for the sibship, $\bar{\gamma}(\sigma \cap F, D_\alpha)$. Summing across sibships produces an expression for the total visible deaths in terms of all of the reports:

$$
\begin{aligned}
\sum_{i \in F} \frac{y(i, D_\alpha \cap \sigma)}{y(i, \sigma \cap F) + 1} &= \sum_{\sigma \in \Sigma} \sum_{i \in \sigma} \frac{y(i, D_\alpha \cap \sigma)}{y(i, \sigma \cap F) + 1} \\
&= \sum_{\sigma \in \Sigma} D^V_{\alpha \cap \sigma} \; \bar{\gamma}(\sigma \cap F, D_\alpha \cap \sigma)
\end{aligned}
$$

To keep notation as minimal as possible, we introduce the abbreviations $\bar{\gamma}_{\sigma,D} = \bar{\gamma}(\sigma \cap F, D_\alpha \cap \sigma)$, the average γ for deaths in sibship σ; $\bar{\gamma}_D = |\Sigma|^{-1} \sum_{\sigma \in \Sigma} \bar{\gamma}_{\sigma,D}$, the average γ across sibships; and $D_\sigma = D^V_{\sigma \cap \alpha}$, sibship σ's visible deaths in group α. To continue simplifying the expression:

$$
\begin{aligned}
&= D^V_\alpha \bar{\gamma}_D + |\Sigma| \mathrm{cov}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \\
&= D^V_\alpha \left[\bar{\gamma}_D + \frac{|\Sigma| \mathrm{cov}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D})}{D^V_\alpha} \right] \\
&= D^V_\alpha \left[\bar{\gamma}_D + \frac{|\Sigma| \mathrm{cor}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \mathrm{sd}_\Sigma(D_\sigma) \mathrm{sd}_\Sigma(\bar{\gamma}_{\sigma,D})}{D^V_\alpha} \right] \\
&= D^V_\alpha \left[\bar{\gamma}_D + \mathrm{cor}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \mathrm{cv}_\Sigma(D_\sigma) \mathrm{sd}_\Sigma(\bar{\gamma}_{\sigma,D}) \right] \\
&= D^V_\alpha \left[\bar{\gamma}_D + \mathrm{cor}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \mathrm{cv}_\Sigma(D_\sigma) \mathrm{cv}_\Sigma(\bar{\gamma}_{\sigma,D}) \bar{\gamma}_D \right] \\
&= D^V_\alpha \; \bar{\gamma}_D \; \left[1 + \mathrm{cor}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \mathrm{cv}_\Sigma(D_\sigma) \mathrm{cv}_\Sigma(\bar{\gamma}_{\sigma,D}) \right].
\end{aligned} \tag{61}
$$

Sensitivity of exposure

Equation 61 is an expression for the sensitivity of the numerator of the individual visibility estimator. We now wish to derive a sensitivity expression for the denominator of the individual visibility estimator. However, the denominator of the individual visibility estimator is more complex than the numerator because the denominator involves two terms: one for reports about exposure to frame population members and one for reports about exposure to non frame population members. Considering each of these two terms separately, an argument parallel to the one for the sensitivity of deaths (Equation 61) shows that

$$
\sum_{i \in \sigma \cap F} \frac{y(i, N_\alpha \cap F)}{y(i, F)} = N^V_{\alpha \cap F} \; \bar{\gamma}(\sigma \cap F, N_\alpha \cap F), \tag{62}
$$

and that

$$
\sum_{i \in \sigma \cap F} \frac{y(i, N_\alpha - F)}{y(i, F) + 1} = N^V_{\alpha - F} \; \bar{\gamma}(\sigma \cap F, N_\alpha - F). \tag{63}
$$

For a particular sibship σ, $N^V_\alpha = N^V_{\alpha \cap F} + N^V_{\alpha - F}$, so we have

$$\begin{aligned}\sum_{i \in \sigma \cap F} \frac{y(i, N_\alpha \cap F)}{y(i,F)} + \sum_{i \in \sigma \cap F} \frac{y(i, N_\alpha - F)}{y(i,F)+1} &= N^V_{\alpha \cap F}\ \bar{\gamma}(\sigma \cap F, N_\alpha \cap F) + N^V_{\alpha - F}\ \bar{\gamma}(\sigma \cap F, N_\alpha - F) \\ &= N^V_{\alpha \cap \sigma} \Big[p_{F|N_\alpha \cap \sigma} \bar{\gamma}(\sigma \cap F, N_\alpha \cap F) \\ &\quad + (1 - p_{F|N_\alpha \cap \sigma}) \bar{\gamma}(\sigma \cap F, N_\alpha - F) \Big], \end{aligned} \tag{64}$$

where we have defined $p_{F|N_\alpha \cap \sigma} = |\sigma \cap N_\alpha \cap F| / |\sigma \cap N_\alpha|$ to be the proportion of siblings with expsoure that is on the frame population.

Equation 64 shows that, for a single sibship, we can write the individual visibility estimand as the visible exposure in the sibship, $N^V_{\alpha \cap \sigma}$, times a weighted average of the net reporting factor for exposure on the frame and exposure not on the frame. In order to simplify this expression, for a sibship σ, we define

$$\bar{\gamma}^*(\sigma \cap F, N_\alpha) = p_{F|N_\alpha \cap \sigma} \bar{\gamma}(\sigma \cap F, N_\alpha \cap F) + (1 - p_{F|N_\alpha \cap \sigma}) \bar{\gamma}(\sigma \cap F, N_\alpha - F).$$

Having defined γ^*, we can rewrite Equation 64 as

$$\sum_{i \in \sigma \cap F} \frac{y(i, N_\alpha \cap F)}{y(i,F)} + \sum_{i \in \sigma \cap F} \frac{y(i, N_\alpha - F)}{y(i,F)+1} = N^V_{\alpha \cap \sigma}\ \gamma^*(\sigma \cap F, N_\alpha). \tag{65}$$

This need to define γ^* is awkward, but necessary because of the way that the denominator of the individual visibility estimator mixes together siblings with different visibilities.

Equation 65 shows the relationship between reports and the exposure for one sibship. Following a derivation parallel to the one for deaths above (producing Equation 61), we can add up over all sibships to obtain an expression for the population-level visible exposure:

$$\sum_{i \in F} \left[\frac{y(i, N_\alpha \cap F)}{y(i,F)} + \frac{y(i, N_\alpha - F)}{y(i,F)+1} \right] = N^V_\alpha\ \bar{\gamma}^\star_N\ \left[1 + \text{cor}_\Sigma(N_\sigma, \bar{\gamma}^\star_{\sigma,N}) \text{cv}_\Sigma(N_\sigma) \text{cv}_\Sigma(\bar{\gamma}^\star_{\sigma,D}) \right]. \tag{66}$$

Sensitivity of death rates

Combining the expression for sensitivity of the reported deaths (the numerator, Equation 61) and the expression for the sensitivity of reported exposure (the denominator, Equation 66), we have

$$
\begin{aligned}
\text{individual estimand} &= \frac{\sum_{i \in F} \frac{y(i, D_\alpha \cap \sigma)}{y_(i, \sigma \cap F)+1}}{\sum_{i \in F} \left[\frac{y(i, N_\alpha \cap F)}{y(i,F)} + \frac{y(i, N_\alpha - F)}{y(i,F)+1} \right]} \\
&= \frac{D_\alpha^V \;\; \bar{\gamma}_D \;\; [1 + \text{cor}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \text{cv}(D_\sigma) \text{cv}(\bar{\gamma}_{\sigma,D})]}{N_\alpha^V \;\; \bar{\gamma}_N^\star \;\; \left[1 + \text{cor}_\Sigma(N_\sigma, \bar{\gamma}_{\sigma,N}^\star) \text{cv}_\Sigma(N_\sigma) \text{cv}_\Sigma(\bar{\gamma}_{\sigma,D}^\star)\right]} \\
&= M_\alpha^V \times \frac{\bar{\gamma}_D}{\bar{\gamma}_N^\star} \times \frac{1 + K_D}{1 + K_N},
\end{aligned}
\tag{67}
$$

where we have simplified the expression by introducing two aggregate factors

- $K_D = \text{cor}_\Sigma(D_\sigma, \bar{\gamma}_{\sigma,D}) \text{cv}(D_\sigma) \text{cv}(\bar{\gamma}_{\sigma,D})$
- $K_N = \text{cor}_\Sigma(N_\sigma, \bar{\gamma}_{\sigma,N}^\star) \text{cv}_\Sigma(N_\sigma) \text{cv}_\Sigma(\bar{\gamma}_{\sigma,N}^\star)$

K_D and K_N are complex, but the intuition is that they capture the relationship between sibship-level reporting factors and deaths (for K_D) or exposure (for K_N). Taking deaths as an example, when reporting is perfect, there is no relationship between sibship deaths and reporting, so that $K_D = 0$. When reporting is not perfect, the sign of K_D is determined by the correlation factor (since the other two factors are non-negative). When reporting tends to omit deaths in sibships with more deaths, then $K_D > 1$; conversely, when reporting tends to omit deaths in sibships with fewer deaths, then $K_D < 1$.

The final step is to incorporate the expression for sensitivity to invisible siblings. Let the visible and invisible death rates differ by a factor K so that $M_\alpha^I = K M_\alpha^V$. Equation 41 then tells us that

$$
\frac{M_\alpha^V}{M_\alpha} = \frac{p_{D_\alpha}^V + K(1 - p_{D_\alpha}^V)}{K}.
\tag{68}
$$

Combining Equation 41 and Equation 67, we obtain an expression that relates the individual visibility estimand to the true death rate:

$$
\begin{aligned}
M_\alpha &= \frac{\sum_{i \in F} \frac{y(i, D_\alpha \cap \sigma)}{y(i, \sigma \cap F)+1}}{\sum_{i \in F} \left[\frac{y(i, N_\alpha \cap F)}{y(i,F)} + \frac{y(i, N_\alpha - F)}{y(i,F)+1} \right]} \times \frac{\bar{\gamma}_N^\star}{\bar{\gamma}_D} \times \frac{1 + K_N}{1 + K_D} \times \left[\frac{K}{p_{D_\alpha}^I + K(1 - p_{D_\alpha}^I)} \right] \\
&= \underbrace{\frac{\widehat{D}_\alpha^V}{\widehat{N}_\alpha^V}}_{\substack{\text{individual} \\ \text{multiplicity} \\ \text{estimand}}} \times \underbrace{\frac{\bar{\gamma}_N^\star}{\bar{\gamma}_D}}_{\substack{\text{average} \\ \text{reporting} \\ \text{adj. factors}}} \times \underbrace{\frac{1 + K_N}{1 + K_D}}_{\substack{\text{correlation} \\ \text{between} \\ \text{adj. factors} \\ \text{and qoi}}} \times \underbrace{\left[\frac{K}{p_{D_\alpha}^I + K(1 - p_{D_\alpha}^I)} \right]}_{\substack{\text{difference} \\ \text{between} \\ \text{visible and} \\ \text{invisible}}}.
\end{aligned}
\tag{69}
$$

There are a few important things to note about Equation 69. First, there is no structural term analogous to the factor $\frac{\bar{d}_{N_\alpha,F}^V}{\bar{d}_{D_\alpha,F}^V}$ from the aggregate sensitivity framework; adjusting at the individual level eliminates the need for this condition. Second, the factors related to reporting errors are more complex than the analogous group of factors in the aggregate estimator. It is no longer the case that reporting errors can cancel each other out if they are the same for the numerator and denominator on average; the condition in Equation 69 requires that the relationship between adjustment factors and sibship characteristics, captured by K_D and K_N, also agree in order for cancellation to go through. Third, note that the averages in Equation 69 are taken across sibships, and not individuals. This means that, in order to collect data on the individual-level adjustment factors in Equation 69, we would need detailed information about sibships, which seems likely to pose a challenge to data collection efforts.

Finally, Equation 69 parameterizes the difference between the visible and visible populations in terms of $p^D_{D_\alpha}$, the proportion of deaths that is invisible. Researchers may prefer to parameterize this factor in terms of $p^I_{N_\alpha}$, the proportion of exposure that is invisible. In that case, Equation 69 becomes:

$$M_\alpha = \underbrace{\frac{\widehat{D}^V_\alpha}{\widehat{N}^V_\alpha}}_{\substack{\text{individual}\\\text{multiplicity}\\\text{estimand}}} \times \underbrace{\frac{\bar{\gamma}^\star_N}{\bar{\gamma}_D}}_{\substack{\text{average}\\\text{reporting}\\\text{adj. factors}}} \times \underbrace{\frac{1+K_N}{1+K_D}}_{\substack{\text{correlation}\\\text{between}\\\text{adj. factors}\\\text{and qoi}}} \times \underbrace{\left[1+p^I_{N_\alpha}(K-1)\right]}_{\substack{\text{difference}\\\text{between}\\\text{visible and}\\\text{invisible}}}. \tag{70}$$

Sensitivity for reports that include the respondent

The derivations that led to the sensitivity framework for the individual visibility estimand (Equation 69 and Equation 70) can also be used to develop a sensitivity framework for the individual visibility estimand with respondents included in reports: only slight modifications are necessary to account for the fact that respondents include themselves in reports. We do not repeat all of the derivations in detail but, for convenience, Equation 71 spells out the sensitivity framework for the individual visibility estimand with respondents included in reports.

$$M_\alpha = \underbrace{\frac{\widehat{D}'^V_\alpha}{\widehat{N}'^V_\alpha}}_{\substack{\text{individual}\\\text{multiplicity}\\\text{estimand}\\\text{including}\\\text{respondents}}} \times \underbrace{\frac{\bar{\gamma}'_N}{\bar{\gamma}'_D}}_{\substack{\text{average}\\\text{reporting}\\\text{adj. factors}}} \times \underbrace{\frac{1+K'_N}{1+K'_D}}_{\substack{\text{correlation}\\\text{between}\\\text{adj. factors}\\\text{and qoi}}} \times \underbrace{\left[1+p'^I_{N_\alpha}(K'-1)\right]}_{\substack{\text{difference}\\\text{between}\\\text{visible and}\\\text{invisible}}}, \tag{71}$$

where:

- $\bar{\gamma}'_D = \frac{\tau'(i,D_\alpha)\eta'(i,F)}{\eta'(i,D^V_\alpha)\tau'(i,F)}$
- $\bar{\gamma}'_N = \frac{\tau'(i,N_\alpha)\eta'(i,F)}{\eta'(i,N^V_\alpha)\tau'(i,F)}$
- reporting parameters for with-respondent reports are defined analogous to the without-respondent parameters; for example, $\tau'(i, N^V_\alpha) = \frac{y'^+(i,D^V_\alpha \cap \sigma)}{d'^+(D^V_\alpha \cap \sigma, i)}$

- $K'_D = \mathrm{cor}_\Sigma(D^V_\sigma, \bar{\gamma}'_{\sigma,D})\mathrm{cv}(D^V_\sigma)\mathrm{cv}(\bar{\gamma}'_{\sigma,D})$
- $K'_N = \mathrm{cor}_\Sigma(N'^V_\sigma, \bar{\gamma}'_{\sigma,N'^V})\mathrm{cv}_\Sigma(N'^V_\sigma)\mathrm{cv}_\Sigma(\bar{\gamma}'_{\sigma,N})$
- $p'^I_{N_\alpha}$ is the proportion of deaths that is invisible with respect to M'^V_α (ie, the visible death rate when respondents are included in reports, which is different from M^V_α, the visible death rate when respondents are excluded from reports)
- K' is a parameterization of the difference between the invisible and visible death rates when respondents are included in reports

G Comparing the four estimators

Our results suggest that there are four possible approaches to analyzing sibling histories: there is the decision to include or exclude respondents from sibling reports; and there is the choice between the aggregate and individual visibility estimators. In this Appendix, we discuss the differences between these four approaches in greater depth. We first investigate the impact of deciding to include or exclude respondents from reports, and we argue that it is preferable to exclude respondents. Then we turn to a discussion of the individual versus aggregate visibility estimator; our analysis leads us to suggest that, in the absence of additional information about adjustment factors, the individual visibility estimator is preferable.

G.1 The difference between M'^V_α and M^V_α

Above, we mentioned that the *definition* of the visible population is affected by whether or not we include respondents themselves in sibling reports. In this section, we explain how M'^V_α – i.e., the death rate when respondents are included in reports – differs from M^V_α – i.e., the death rate in the visible population when respondents are not included in reports.

We will focus on visibility at the individual level (but note that the definition of the visible population is not affected by the decision to adjust for visibility at the individual or at the aggregate level). Recall from Section B.2 that, under perfect reporting,

$$v'(i, F) = |\sigma \cap F|,$$

and

$$v(i, F) = \begin{cases} |\sigma \cap F| - 1 & \text{when } i \in F \\ |\sigma \cap F| & \text{when } i \notin F. \end{cases}$$

Thus, for a particular sibling i, we can write the relationship between $v'(i, F)$ and $v(i, F)$ as

$$v'(i,F) = \begin{cases} v(i,F) + 1 & \text{when } i \in F \\ v(i,F) & \text{when } i \notin F. \end{cases}$$

When switching from including respondents to not including respondents, the only people whose visibility is affected are in the frame population F. In particular, this means that deaths – who are never on the frame population – will not have their visibility affected by whether or not the respondent is included in reports.

For exposure, on the other hand, individuals $i \in N_\alpha \cap F$ will have their visibility affected by whether or not respondents are included in reports. We have two cases:

Case 1: $N_\alpha \cap F = \phi$

Example: men in any age group when only women are interviewed.

In this case, visibility does not change whether or not respondents are included. So $v(i,F) = v'(i,F) = d(j,F) + 1 \ \forall j \in \sigma[i] \cap F$.

Case 2: $N_\alpha \cap F \neq \phi$

Example: women aged 30-35 in a typical DHS survey.

In this case, the visibility of deaths does not change based on whether or not respondents are included in reports. However, the visibility of exposure *does* change. The key question is: whose visibility switches from $v'(i,F) > 0$ to $v(i,F) = 0$? This is the group of people who become invisible when respondents are excluded from reports.

Since

$$v(i,F) = v'(i,F) - 1 \ \forall i \in (N_\alpha \cap F) - D_\alpha,$$

this can only happen when $v'(i,F) = 1$. So, for each $i \in N_\alpha$, we have

$$v'(i,F) - v(i,F) = \begin{cases} 1 & \text{if } i \in \alpha \cap F \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the set of people who become invisible when respondents are excluded from reports is precisely the set of siblings i such that $v'(i,F) = 1$ and $i \in N_\alpha \cap F$. The number of people whose exposure becomes invisible when switching from including respondents to not including respondents is

$$N_\alpha^{V\prime} - N_\alpha^V = \sum_{i \in N_\alpha \cap F} \mathbb{1}_{[v'(i,F)=1]}.$$

Upshot

The visible death rate including respondents in reports is

$$M_\alpha'^V = \frac{D_\alpha^{V\prime}}{N_\alpha^{V\prime}}.$$

The visible death rate not including respondents in reports is

$$\begin{aligned} M_\alpha^V &= \frac{D_\alpha^V}{N_\alpha^V} \\ &= \frac{D_\alpha'^V}{N_\alpha'^V - C} \end{aligned}$$

where $C = \sum_{i \in N_\alpha \cap F} \mathbb{1}_{[v'(i,F)=1]}$.

C is a factor that captures the difference in visibility due to including or not including respondents. C will tend to be bigger, inducing a bigger difference between M'^V and M^V, when

- $|\alpha \cap F|$ is bigger
- the distribution of $v'(i,F)$ in $|\alpha \cap F|$ is smaller, meaning that more values of $v'(i,F) = 1$.

G.2 A simple model to study including and excluding respondents from reports

In this section, we develop a simple model that can be used to illustrate how including respondents in reports changes the definition of the visible and invisible populations. The results of our model agree with previous models in suggesting that researchers exclude respondents from the denominator of sibling history estimates (Trussell and Rodriguez 1990), even though our model makes no assumptions about the parametric form of the distribution of sibship sizes.

For the purposes of this model, we will assume that we have a homogenous population whose members all face the same probability of death q. (So, we disregard differences in age and sex.) We introduce some notation for the model, which will be used in this section alone. Let N be the set of everyone in the population, which is also the set of people who are exposed to the possibility of death. People are organized into sibships, and $s_i \geq 0$ describes the number of siblings i has, living or dead; thus, i's sibship has $s_i + 1$ members. We make no additional assumptions about the distribution of sibship sizes.

Let Q_i be a random variable whose outcome determines whether or not person $i \in N$ dies. We take $\mathbb{E}[Q_i] = q$ for all i and $\text{cov}[Q_i, Q_j] = 0$ for all $i, j \in N$, meaning that everyone in the population faces the same probability of death q and deaths are not correlated with one another. (In particular, deaths are not correlated within sibships.) The expected proportion of people who die is thus q.

Our model describes a stochastic process; one realization of this process produces a finite population. We are interested in the number of reported deaths and the amount of reported exposure when (i) respondents do not include themselves in reports; and (ii) respondents do include themselves in reports. In both cases, we assume reporting is perfect, and we assume that everyone who is alive is in the frame population. We investigate the population-level reports, and the definition of the invisible and visible populations in both cases.

Case 1: Respondents are not included in reports

Visible and invisible death rates

When respondents are not included in reports, the total number of visible deaths will be

$$D^V = \sum_{i \in N} \underbrace{Q_i}_{\substack{\text{prob } i \\ \text{dies}}} \times \underbrace{[1 - \Pi_{j \sim i} Q_j]}_{\substack{\text{prob} \\ \text{at least one of } i\text{'s} \\ \text{sibs survives}}}, \tag{72}$$

where $j \sim i$ indexes the siblings j of person i. In expectation, we have

$$\mathbb{E}[D^V] = \sum_{i \in N} q \times (1 - q^{s_i}) = q \sum_{i \in N} (1 - q^{s_i}). \tag{73}$$

Since $\mathbb{E}[D] = |N|q$ and $D^I = D - D^V$, $\mathbb{E}[D^I] = \mathbb{E}[D] - \mathbb{E}[D^V]$. So we also have

$$\mathbb{E}[D^I] = |N|q - \sum_{i \in N} q \times (1 - q^{s_i}) = q \sum_{i \in N} q^{s_i}. \tag{74}$$

The total amount of visible exposure will be

$$N^V = \sum_{i \in N} 1 \times \underbrace{[1 - \Pi_{j \sim i} Q_j]}_{\substack{\text{prob} \\ \text{at least one of } i\text{'s} \\ \text{sibs survives}}}.$$

Taking expectations, we have

$$\mathbb{E}[N^V] = \sum_{i \in N} (1 - q^{s_i}). \tag{75}$$

Since $N^I = N - N^V$, we have $\mathbb{E}[N^I] = |N| - \mathbb{E}[N^V]$. So

$$\mathbb{E}[N^I] = \sum_{i \in N} q^{s_i}. \tag{76}$$

At the population level, the visible death rate is

$$\begin{aligned} M^V &\approx \frac{\mathbb{E}[D^V]}{\mathbb{E}[N^V]} \\ &= \frac{q \sum_{i \in N}(1 - q^{s_i})}{\sum_{i \in N}(1 - q^{s_i})} = q, \end{aligned}$$

and the invisible death rate is

$$\begin{aligned} M^I &\approx \frac{\mathbb{E}[D^I]}{\mathbb{E}[N^I]} \\ &= \frac{q \sum_{i \in N} q^{s_i}}{\sum_{i \in N} q^{s_i}} = q. \end{aligned}$$

(In both cases, the approximation is to the first order, and is a consequence of the fact that the death rate is a ratio of random variables; in most cases, we expect this approximation to be highly accurate.) Thus, under this model, when respondents are not counted in the reports, the invisible and visible death rates are both approximately q, the model's underlying probability of death.

Reporting quantities

The finite population total number of deaths reported will be

$$y(F, D^V) = \sum_{i \in F} y(i, D) = \sum_{i \in F} \sum_{j \sim i} Q_j.$$

Since only people who are alive will be in the frame population, this becomes

$$y(F, D^V) = \sum_{i \in N} (1 - Q_i)\ y(i, D) = \sum_{i \in F} (1 - Q_i) \sum_{j \sim i} Q_j.$$

In expectation, we have

$$
\begin{aligned}
\mathbb{E}[y(F, D^V)] &= \mathbb{E}\left[\sum_{i \in N} (1 - Q_i) \sum_{j \sim i} Q_j\right] \\
&= \sum_{i \in N} \mathbb{E}\left[(1 - Q_i)\right] \; \mathbb{E}\left[\sum_{i \sim j} Q_i\right] \\
&= \sum_{i \in N} \mathbb{E}\left[(1 - Q_i)\right] \; s_i q \\
&= |N| \; (1 - q) \; s_i q.
\end{aligned} \tag{77}
$$

The finite population total amount of exposure reported will be

$$
y(F, N^V) = \sum_{i \in F} y(i, N) = \sum_{i \in F} s_i.
$$

In expectation, we have

$$
\begin{aligned}
\mathbb{E}[y(F, N^V)] &= \mathbb{E}\left[\sum_{i \in N} (1 - Q_i) \; y(i, N)\right] \\
&= |N| \; (1 - q) \; s_i.
\end{aligned} \tag{78}
$$

Thus, in this case, the finite population ratio of expected reports about deaths (Equation 77) and expected reports about exposure (Equation 78) is approximately the probability of death q. In other words, under this model, excluding respondents from reports (i) induces the visible and invisible populations to have the same death rate; and (ii) means that the aggregate visibility estimator produces essentially unbiased estimates for the visible death rate.

Case 2: Respondents are included in reports

Visible and invisible death rates

When respondents are included in reports, the total number of visible deaths will be

$$
D'^V = \sum_{i \in N} \underbrace{Q_i}_{\substack{\text{prob } i \\ \text{dies}}} \times \underbrace{\left[1 - \Pi_{j \sim i} Q_j\right]}_{\substack{\text{prob} \\ \text{at least one of } i\text{'s} \\ \text{sibs survives}}},
$$

which is the same expression as Equation 72, when respondents are not included in reports. Thus, the expected values for all of the deaths are the same, i.e., $\mathbb{E}[D^V] = \mathbb{E}[D'^V]$ and $\mathbb{E}[D^I] = \mathbb{E}[D'^I]$.

The total amount of visible exposure will be

$$N'^V = \sum_{i \in N} \left[\underbrace{Q_i}_{\substack{\text{prob} \\ i \text{ dies}}} \times \underbrace{[1 - \Pi_{j \sim i} Q_j]}_{\substack{\text{prob} \\ \text{at least one of } i\text{'s} \\ \text{sibs survives}}} + \underbrace{(1 - Q_i)}_{\substack{\text{prob} \\ i \text{ survives}}} \times 1 \right]$$

Taking expectations, we have

$$\begin{aligned} \mathbb{E}[N'^V] &= \sum_{i \in N} [q(1 - q^{s_i}) + (1 - q)] \\ &= q \sum_{i \in N} (1 - q^{s_i}) + |N|(1 - q) \\ &= |N|q - q \sum_{i \in N} q^{s_i} + |N| - |N|q \\ &= |N| - q \sum_{i \in N} q^{s_i}. \end{aligned} \tag{79}$$

Since $N'^I = N - N'^V$, we have $\mathbb{E}[N'^I] = |N| - \mathbb{E}[N'^V]$. So

$$\mathbb{E}[N'^I] = q \sum_{i \in N} q^{s_i}. \tag{80}$$

At the population level, the visible death rate is thus

$$\begin{aligned} M'^V &\approx \frac{\mathbb{E}[D'^V]}{\mathbb{E}[N'^V]} \\ &= \frac{q \sum_{i \in N} (1 - q^{s_i})}{|N| - q \sum_{i \in N} q^{s_i}} \\ &= \frac{q|N| - q \sum_{i \in N} q^{s_i}}{|N| - q \sum_{i \in N} q^{s_i}}. \end{aligned} \tag{81}$$

In general, Equation 81 is not equal to q. The invisible death rate is

$$\begin{aligned} M'^I &\approx \frac{\mathbb{E}[D'^I]}{\mathbb{E}[N'^I]} \\ &= \frac{q \sum_{i \in N} q^{s_i}}{q \sum_{i \in N} q^{s_i}} = 1. \end{aligned} \tag{82}$$

Reporting quantities

The finite population total number of deaths reported will be

$$y'(F, D^V) = \sum_{i \in F} y(i, D) = \sum_{i \in F} \sum_{j \sim i} Q_j.$$

This is the same as case 1; thus, in expectation,

$$\mathbb{E}[y(F, D^V)] = |N| \ (1-q) \ s_i q. \tag{83}$$

The finite population total amount of exposure reported will be

$$y'(F, N^V) = \sum_{i \in F} (y(i, N) + 1) = \sum_{i \in F} (s_i + 1),$$

where the plus one adds the exposure of the respondent who is, by definition, alive. In expectation, we have

$$\begin{aligned} \mathbb{E}[y'(F, N^V)] &= \mathbb{E}\left[\sum_{i \in N} (1 - Q_i) \ [y(i, N) + 1]\right] \\ &= |N| \ (1-q) \ (1 + s_i) \\ &= |N| \ (1-q) + |N| \ (1-q) s_i. \end{aligned} \tag{84}$$

Thus, in this case, the finite population ratio of expected reports about deaths (Equation 83) and expected reports about exposure (Equation 84) does not equal the probability of death q; there is an extra term in the denominator. In fact, this term is precisely the C factor discussed in Appendix G.1; that is, under this model, $C = |N|(1 - q)$.

In this case, people can only be invisible if they die, making the invisible death rate equal to 1. The visible death rate, on the other hand, will in general be different from q. In other words, under this model, including respondents in reports (i) induces a difference in the death rates of the visible and invisible populations, even though everyone in the population has the same probability of death q; and (ii) means that the aggregate visibility estimator does not necessarily produce essentially unbiased estimates for the visible death rate.

Summary

To recap, we introduced a model in which all members of a population have the same probability of dying. Under this model, we saw that it was appealing to exclude respondents from sibling reports; when respondents are excluded, the visible and invisible populations have the same death rate, and that death rate is equal to the probability of an individual dying. On the other hand, including respondents in sibling reports induced a difference in death rates between the visible and invisible populations. Thus, this model agrees with Trussell and Rodriguez (1990) in suggesting that it is most reasonable to exclude respondents from sibling reports.

G.3 Differences between aggregate and individual visibility

The aggregate and individual visibility estimators can produce different results; for example female death rates estimates in Figure 3 are lower for the individual visibility estimator than for the aggregate visibility estimator. On the other hand, Figure 3 also shows that results for males are highly consistent with one another. What explains when and how aggregate and individual visibility estimates differ? In this appendix, we address this question in two stages: first we derive a relationship between the aggregate visibility of deaths and the aggregate visibility of exposure; and, second, we considering two heuristic approximations for the aggregate visibility of exposure that empirically account for most of the difference between the aggregate and individual visibility death rate estimates in Malawi.

We start by deriving a relationship between the aggregate visibility of deaths and the aggregate visibility of exposure. The basis for this relationship is the adjustment factors called the *visibility ratio* in the aggregate sensitivity framework (Equation 14). The visibility ratio captures the condition, required by the aggregate visibility estimator, that the average visibility of exposure be equal to the average visibility of deaths. This condition is not required by the individual visibility estimator; thus, it is a possible source of differences between the two estimators.

We will start from results about aggregate visibility derived in Appendix B.4. Equation 31 showed that, for a group $A \subset U$:

$$\bar{v}(A, F) = \bar{v}'(A, F) - \frac{|F \cap A|}{|A|}. \tag{85}$$

Equation 85 is stated in terms of a generic group A. We now investigate what Equation 85 implies for deaths and for exposure. Deaths will never be on the sampling frame; thus, for deaths, $\frac{|F \cap D_\alpha^V|}{|D_\alpha^V|} = 0$. For exposure, on the other hand, $|F \cap N_\alpha^V|/|N_\alpha^V|$ is the proportion of people who contribute exposure that is also on the sampling frame. This quantity will depend on whether or not survivors in group α would be expected to be on the sampling frame. In a typical Demographic and Health Survey, the sampling frame for sibling histories will be women of reproductive age. Thus, we expect that:

$$|F \cap N_\alpha^V|/|N_\alpha^V| \approx \begin{cases} 1 & \text{if } \alpha \text{ is women in a reproductive age group} \\ 0 & \text{otherwise.} \end{cases} \tag{86}$$

Equation 86 is an approximation because in the first case, when people in α are on the frame population, some of those who contribute exposure will also die; thus, the true value will be somewhat less than 1. So Equation 86 is an approximation based on the idea that the number of people who die will usually be small relative to the amount of people who contribute exposure.

In the paper, we focus on reports that exclude respondents from the denominator; thus, we are most interested in $\bar{v}(N_\alpha^V, F)$ and $\bar{v}(D_\alpha^V, F)$. It is important to distinguish between two cases, based on which group α's death rate is being estimated.

Case 1: $N_\alpha \cap F = \phi$

Example: men in any age group when only women are interviewed.

We have $|D_\alpha^V \cap F|/|D_\alpha^V| = |N_\alpha^V \cap F|/|N_\alpha^V| = 0$. Thus, in this case, it seems reasonable to assume that $\bar{v}(N_\alpha^V, F) = \bar{v}(D_\alpha^V, F)$, as the estimates in DHS reports do.

Case 2: $N_\alpha \cap F \neq \phi$

Example: women aged 30-35 in a typical DHS survey.

We have $|D_\alpha^V \cap F|/|D_\alpha^V| = 0$, but $|N_\alpha^V \cap F|/|N_\alpha^V| \approx 1$. Thus, our analysis suggests that, in the absence of additional information about the adjustment factors, it is most natural to expect that $\bar{v}(N_\alpha^V, F) \approx \bar{v}(D_\alpha^V, F) - 1$.

The aggregate visibility estimator in Equation 7 (and used in DHS reports) assumes that $\bar{v}(N_\alpha^V, F) = \bar{v}(D_\alpha^V, F)$. So, our analysis suggests that this condition is reasonable when the group α is not on the frame population (like men in most DHS surveys); however, when most members of α will be on the frame population (like women aged 15-49 in most DHS surveys), our analysis suggests that it is more natural to assume that $\bar{v}(N_\alpha^V, F) = \bar{v}(D_\alpha^V, F) - 1$.

In order to illustrate this analysis, we propose two heuristic ways to approximate $\bar{v}(N_\alpha^V, F)$. Then we use these approximations to adjust aggregate visibility estimates in Malawi. We shall see that these approximations account for most of the difference between the individual and aggregate visibility estimates.

The idea behind the two approximations is to use survey respondents' visibilities to approximate the visibility of exposure. Of course, deaths (who we do not interview) also contribute exposure, but between ages of 15 and 49, we expect survivors to outnumber deaths by a considerable margin, even when death rates are high. So we expect the average visibility of survey respondents to be similar to the average visibility of exposure.

The first approximation uses

$$\bar{v}(N_\alpha^V, F) \approx \bar{y}(F, F). \quad \text{(All ages approximation)} \tag{87}$$

Equation 87 approximates the visibility of exposure in group α by using the average visibility of survey respondents. Under this approximation, estimates from the aggregate visibility estimator in Equation 7 should be adjusted by a factor of $\frac{\widehat{\bar{y}}(F,F)}{\widehat{\bar{y}}(F,F)+1}$ when group α overlaps with the frame population.

We call Equation 87 the *all-ages* approximation, to distinguish it from the second approximation:

$$\bar{v}(N_\alpha^V, F) \approx \bar{y}(F_\alpha, F). \quad \text{(Age-specific approximation)} \tag{88}$$

Equation 88 takes the idea from Equation 87 and applies it to the specific group α. This approximation is motivated by the form of individual visibility estimator. Under this approximation, estimates from the aggregate visibility estimator in Equation 7 should be adjusted by a factor of $\frac{\widehat{\bar{y}}(F_\alpha,F)}{\widehat{\bar{y}}(F_\alpha,F)+1}$ when group α overlaps with the frame population.

Figure 8 illustrates using the Malawi example from the main paper. The figure shows the individual and aggregate visibility estimates, along with aggregate visibility estimates that have been adjusted using (i) the approximation based on all ages; and (ii) the age-specific approximations. Note that the adjustments only affect groups α whose members could potentially also be members of the frame population; in this case, that is women aged 15-49. Figure 8 shows that the two heuristic approximations do an excellent job of approximating the individual visibility results.

To recap, our derivations suggest that when the group α is not on the frame population, it may be reasonable to assume that $\bar{v}(N_\alpha^V, F) = \bar{v}(D_\alpha^V, F)$, like the aggregate visibility estimator does. However, when the group α overlaps with the frame population – for example, when it includes women aged 15-49 on a DHS survey – then it seems more reasonable to assume that $\bar{v}(N_\alpha^V, F) = \bar{v}(D_\alpha^V, F) - 1$. To illustrate, we approximated $\bar{v}(N_\alpha^V, F)$ in two different ways. Using the adjustment factors suggested by our approximations, Figure 8 adjusted aggregate visibility estimates, and the resulting adjusted estimates were very close to the individual visibility estimates. Thus, for the 2000 Malawi DHS, the difference between the aggregate and individual visibility estimates appears to be explained by the implicit assumption about the visibility adjustment factor made by the aggregate visibility estimator.

H Useful facts

Covariances

The following fact will be useful in some of our analysis below; see, for example, Feehan and Salganik (2016a) for a derivation.

Fact H.1. *Suppose we have a finite population U of size N and that $a_i, b_i \in \mathbb{R}$ are defined for all*

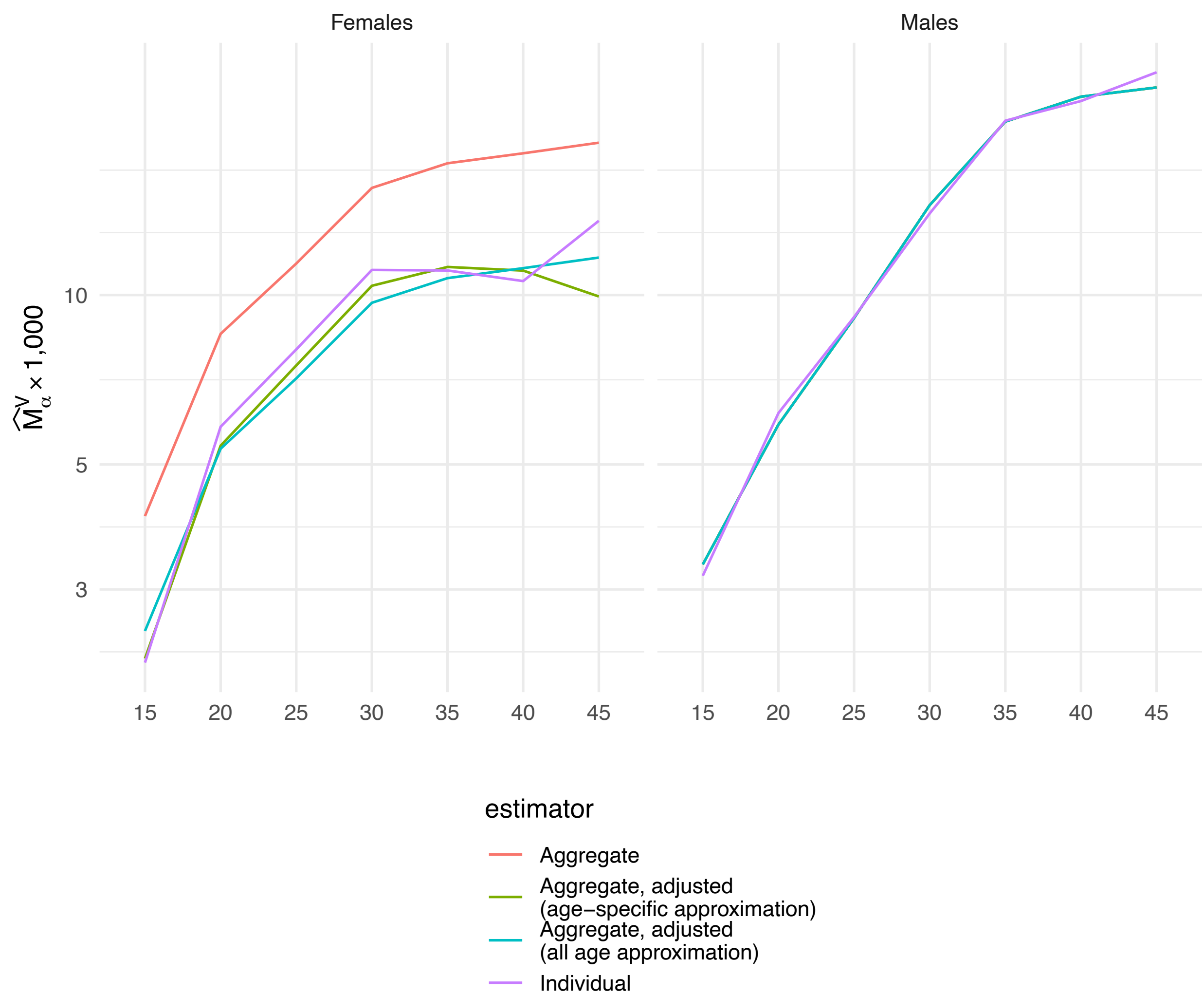

Figure 8: Comparing three variants of adjusted aggregate visibility estimates to the unadjusted aggregate visibility estimates and individual visibility estimates from the 2000 Malawi DHS.

$i \in N$. Then

$$\sum_{i \in U} a_i b_i = N \left[\bar{a}\bar{b} + cov_U(a_i, b_i) \right],$$

where $\bar{a} = N^{-1} \sum_{i \in U} a_i$, $\bar{b} = N^{-1} \sum_{i \in U} b_i$, and $cov_U(a_i, b_i)$ is the finite population covariance of the a_i and b_i values.

Aggregating death rates across groups

In order to develop our sensitivity frameworks, we need a couple of technical results. These results will help us understand how death rate estimates can be affected by invisible deaths and invisible exposure. (See Feehan and Wrigley-Field (2019) for a more in-depth discussion of this topic.)

Demographers frequently use the weighted arithmetic mean. However, many other means exist; in understanding how the visible and invisible death rates aggregate, the harmonic mean will play an important role. So, for convenience, we now review the definition of the weighted harmonic mean:

Definition H.1. Let $\boldsymbol{x}, \boldsymbol{w} \in \mathbb{R}^n$ and let $x_i > 0$ and $w_i > 0$ for all i. Then the **Weighted Harmonic Mean** of the $\boldsymbol{x}$ values, with weights given by the $\boldsymbol{w}$ values, is

$$H[\boldsymbol{x}; \boldsymbol{w}] = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n \frac{w_i}{x_i}}.$$

For comparison, the usual **weighted arithmetic mean** is given by

$$A[\boldsymbol{x}; \boldsymbol{w}] = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}. \tag{89}$$

The next derivation shows that the weights can be rescaled without affecting the weighted harmonic mean.

Result H.1. *Let $\boldsymbol{x}, \boldsymbol{w} \in \mathbb{R}^n$ and let $x_i > 0$ and $w_i > 0$ for all i. Let $\boldsymbol{w'}$ be defined so that $w'_i = K w_i$ for all i and for $K > 0$. Then $H[\boldsymbol{x}; \boldsymbol{w}] = H[\boldsymbol{x}; \boldsymbol{w'}]$.*

Proof. This property follows directly from the definition of the weighted harmonic mean (Definition H.1):

$$H[\boldsymbol{x}; \boldsymbol{w'}] = \frac{\sum_{i=1}^n w'_i}{\sum_{i=1}^n \frac{w'_i}{x_i}} = \frac{\sum_{i=1}^n K\ w_i}{\sum_{i=1}^n \frac{K\ w_i}{x_i}} = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n \frac{w_i}{x_i}} = H[\boldsymbol{x}; \boldsymbol{w}].$$

□

We can connect these insights about harmonic means to deepen our understanding of how death rates aggregate across groups, as Result H.2 shows:

Result H.2. *Suppose two demographic groups have death rates* $M_1 = \frac{D_1}{N_1}$ *and* $M_2 = \frac{D_2}{N_2}$*, where* D_1 *is the number of deaths in the first group,* N_1 *is the person-years of exposure in the first group,* D_2 *is the number of deaths in the second group, and* N_2 *is the person-years of exposure in the second group. Now suppose we combine the two groups and treat them as one aggregate group. Then the death rate for the aggregate group is the weighted harmonic mean of the death rates in the subgroups, with weights given by the number of deaths in each subgroup:*

$$M_{agg} = H[(M_1, M_2); (D_1, D_2)] = \frac{D_1 + D_2}{N_1 + N_2}.$$

Proof. See Feehan and Wrigley-Field (2021).

□

Next, we will see that when groups are aggregated with a focus on the amount of exposure, the weighted arithmetic mean describes the resulting aggregate death rate.

Result H.3. *Suppose two demographic groups have death rates* $M_1 = \frac{D_1}{N_1}$ *and* $M_2 = \frac{D_2}{N_2}$*, where* D_1 *is the number of deaths in the first group,* N_1 *is the person-years of exposure in the first group,* D_2 *is the number of deaths in the second group, and* N_2 *is the person-years of exposure in the second group. Now suppose we combine the two groups and treat them as one aggregate group. Then the death rate for the aggregate group is the weighted arithmetic mean of the death rates in the subgroups, with weights given by the amount of exposure in each subgroup:*

$$M_{agg} = A[(M_1, M_2); (N_1, N_2)] = \frac{D_1 + D_2}{N_1 + N_2}.$$

Proof. By the definition of the weighted arithmetic mean (Equation 89), we have

$$\begin{aligned} A[(M_1, M_2); (N_1, N_2)] &= \frac{N_1 M_1 + N_2 M_2}{N_1 + N_2} \\ &= \frac{D_1 + D_2}{N_1 + N_2} = M_{\text{agg}}. \end{aligned}$$

□

Taken together, Results H.2 and H.3 will be useful in constructing sensitivity frameworks because they allow us to parameterize the difference between the visible and invisible populations in terms of either the fraction of deaths that is invisible (leading to the harmonic mean relationship) or the fraction of exposure that is invisible (leading to the arithmetic mean relationship). Which result to use depends on whether researchers want to parameterize sensitivity using the proportion of deaths that is invisible (Result H.2) or the proportion of exposure that is invisible (Result H.3).

Note that Results H.2 and H.3 can be extended to more than two groups; in general, the death rate for an aggregation of several groups will be given by the weighted harmonic (arithmetic) mean of the component death rates, with the weights given by the number of deaths (amount of exposure) in each component group.

I Variance estimation

Variance estimators can be useful in at least two ways. First, variance estimators enable researchers to understand and communicate the sampling variance associated with any point estimate. Second, before a survey design is decided upon, variance estimators can be used to understand how big a sample is needed to estimate a quantity at a given level of precision. Somewhat surprisingly, little work has formally analyzed the variance of sibling history estimates[15].

All of the sibling estimators discussed in this study are variants of a ratio or compound ratio estimator. A standard result in the survey sampling literature shows that the variance of such an estimator can be estimated using a Taylor approximation. Here, we state this result and explain how it relates to the sibling survival estimators.

Sarndal, Swensson, and Wretman (2003, sec 5.6) shows that the relative variance of any ratio estimator of the form $\widehat{M} = \frac{\widehat{D}}{\widehat{N}}$ can be approximated by

$$\widehat{\text{Var}}[\widehat{M}] \approx \frac{1}{\widehat{N}^2} \left[\widehat{\text{Var}}[\widehat{D}] + \widehat{M}^2 \ \widehat{\text{Var}}[\widehat{N}] - 2\widehat{M}\widehat{\text{Cov}}[\widehat{D}, \widehat{N}] \right]. \tag{90}$$

Multiplying Equation 90 through by $\frac{1}{\widehat{M}^2}$, we obtain an expression for the approximate relative variance

$$\frac{\widehat{\text{Var}}[\widehat{M}]}{\widehat{M}^2} = \widehat{\text{Rel-Var}}[\widehat{M}^2] \approx \widehat{\text{Rel-Var}}[\widehat{D}] + \widehat{\text{Rel-Var}}[\widehat{N}] - 2 \ \widehat{\text{Rel-Cov}}[\widehat{D}, \widehat{N}], \tag{91}$$

where $\widehat{\text{Rel-Var}}[\widehat{X}] = \widehat{\text{Var}}[\widehat{X}]/\widehat{X}^2$ is the relative sampling variance and $\widehat{\text{Rel-Cov}}[\widehat{X}, \widehat{Y}] = \widehat{\text{Cov}}[\widehat{X}, \widehat{Y}]/\widehat{X}\widehat{Y}$ is the relative sampling covariance.

[15] Hanley, Hagen, and Shiferaw (1996) studied the sisterhood method, which is closely related to the sibling survival method. By approximating the proportion of sisters reported dead as a binomial variable, Hanley, Hagen, and Shiferaw (1996) derives an expression for the variance of a sisterhood estimate. Although this approach has been very useful, future work could likely improve upon Hanley, Hagen, and Shiferaw (1996)'s results; their expression is based on several simplifications and does not appear to account for the complex design used in almost all surveys that collect sibling history data.

In the context of estimating death rates, the $\widehat{M}$ in Equation 90 is the estimated death rate, the $\widehat{D}$ is the reports about deaths, and the $\widehat{N}$ is the reports about exposure. Equation 90 and Equation 91 show for a given level of mortality, the estimated sampling uncertainty will be lower when

- the reports exposure about deaths and exposure are estimated from the sample with a high degree of precision, making $\widehat{\text{Rel-Var}}[\widehat{D}]$ and $\widehat{\text{Rel-Var}}[\widehat{N}]$ small
- the sampling design produces a high, positive covariance between the estimated numerators and denominators, making $\widehat{\text{Rel-Cov}}[\widehat{D}, \widehat{N}]$ large

J Simulation study

We conducted a simulation study with two goals: (1) we wanted to confirm the correctness of our analytical results, including the sensitivity frameworks; and, (2) we wanted to empirically compare the performance of the four possible estimators. We based our simulation on the sibships reported about in the 2000 Malawi DHS study. Our aim was not to re-construct a perfectly authentic population-level sibship structure; instead, we wanted a reasonably realistic population that could help us achieve the goals of confirming the analytical results and understanding the differences between the four possible estimators.

J.1 Constructing the universe

We start with the observed set of sibships that are reported about in the sibling history module; there is one sibship reported about for each survey respondent. We assume that these sibships are distinct, ignoring the fact that more than one member of a sibship may have been interviewed in the study. We use these sibships as the basis for constructing a pseudo-population of sibships as follows (Figure 9):

1. Starting from the 13,161 sibships in the dataset, sample M_{sibships} to form the pseudo-population of sibships. We sample with replacement, so some sibships are sampled multiple times, and each sibship is sampled with probability proportional to its visibility to the frame population, i.e. the number of sibship members, including the respondent, who were eligible to respond to the survey.
2. For each of the M_{sibships} resampled sibships, with probability $\frac{1}{2}$, we flip the sexes of the reported siblings. (This accounts for the fact that only females were interviewed in Malawi; without this step, we would end up with an unrealistic gender distribution.)
3. We then form a universe of siblings from the individual siblings corresponding to the (possibly gender-flipped) resampled sibships.

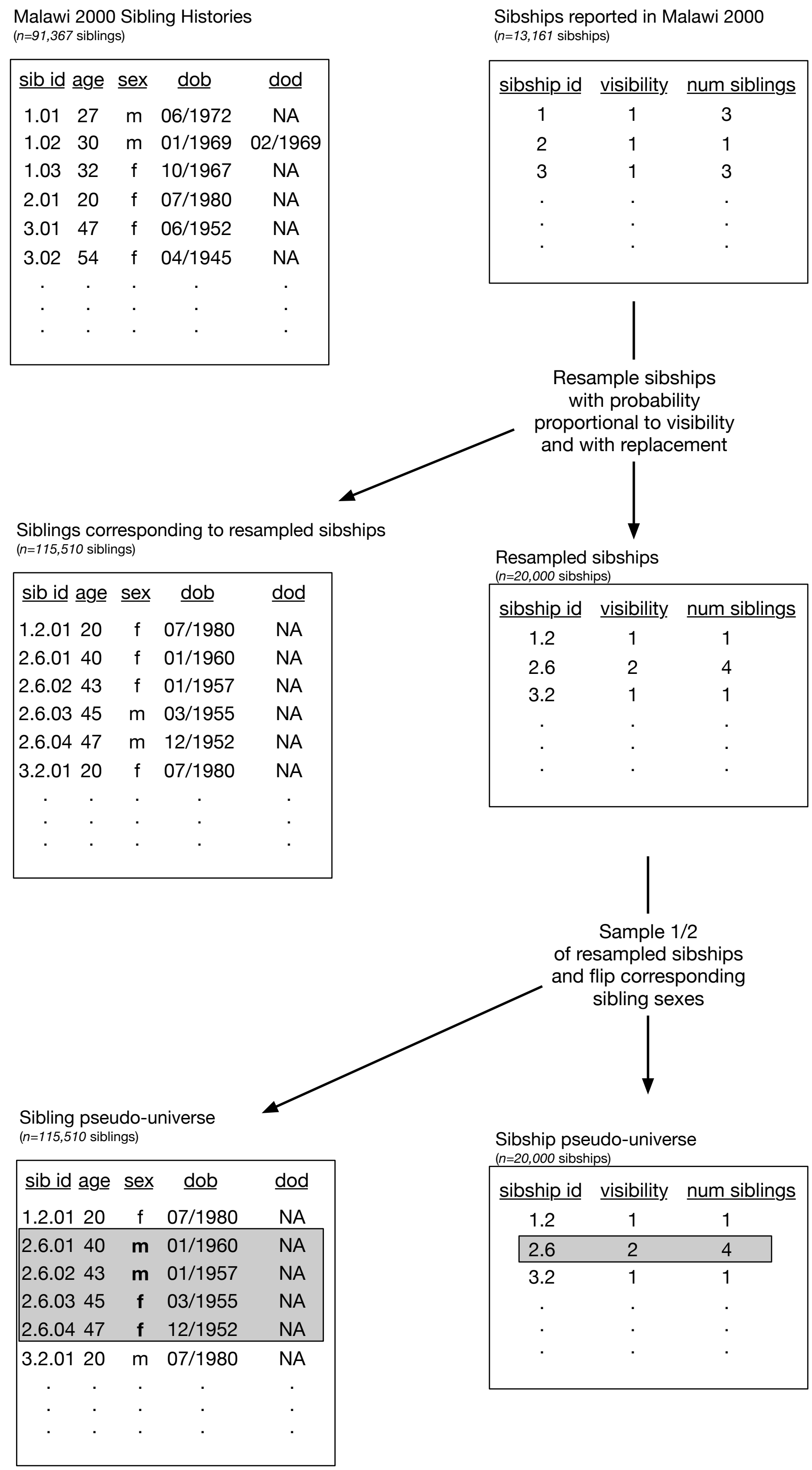

Figure 9: Overview of the method used to construct a pseudo-universe that forms the basis for the simulation study.

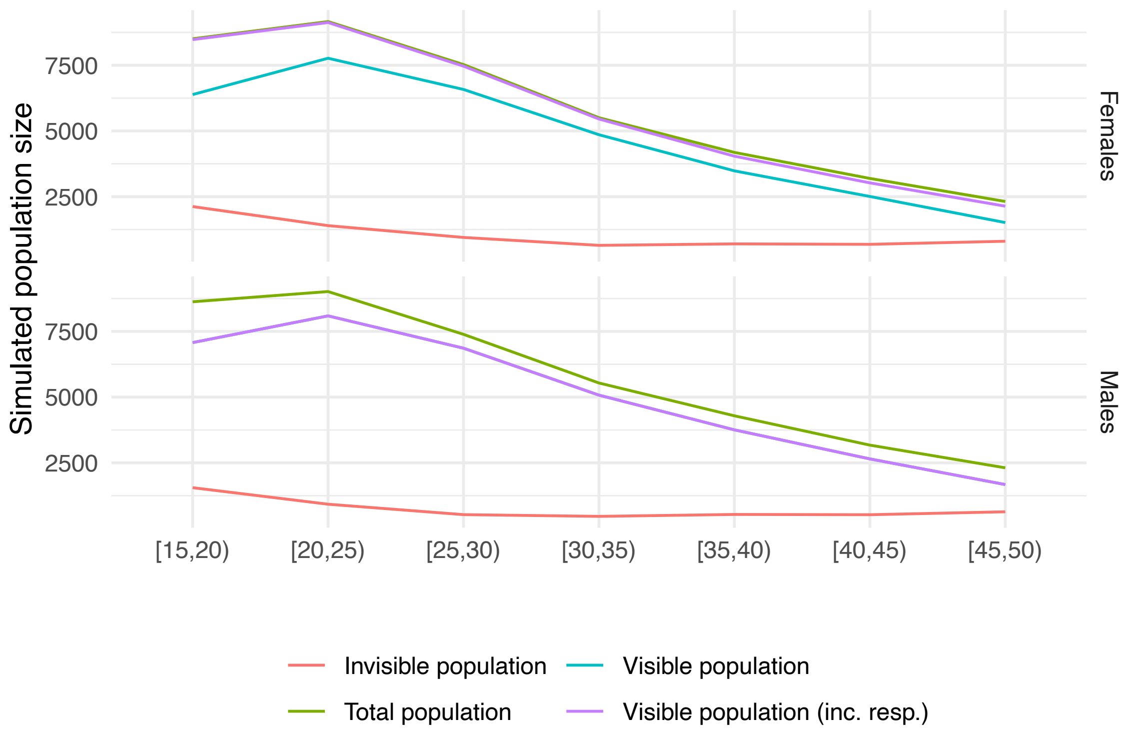

Figure 10: Age-sex distribution of the simulated population.

The result is a universe of siblings who are assigned to sibships that is approximately representative of the 2000 Malawi sibship population. Figure 10 and Figure 11 show the age-sex distribution and death rates in the simulated universe. From the universe of siblings and sibships, we create a sibship network `igraph` object. We also create a census dataset, and use it as the basis for calculating true death rates by age and sex.

J.2 Simulating reporting error and sampling

Having created a pseudo-universe of siblings linked together in sibships, the goal is to use this population as the basis for simulated sibling history surveys under different scenarios. By *scenario*, we mean a set of aggregate reporting parameters together with a sampling fraction. We now describe how we simulate sibling history surveys under several different scenarios.

Using the sibling universe, we create several population-level reporting networks, one for each combination of aggregate reporting parameters $\tau_D = \{0.8, 1\}$ and $\tau_N = \{0.8, 1\}$. We assume there are no false positives throughout (i.e., $\eta_D = 1$ and $\eta_N = 1$).

For each population reporting network, we calculate all of the aggregate and individual-level adjustment factors.

For each population reporting network, we simulate $M = 1000$ sample surveys for each sampling

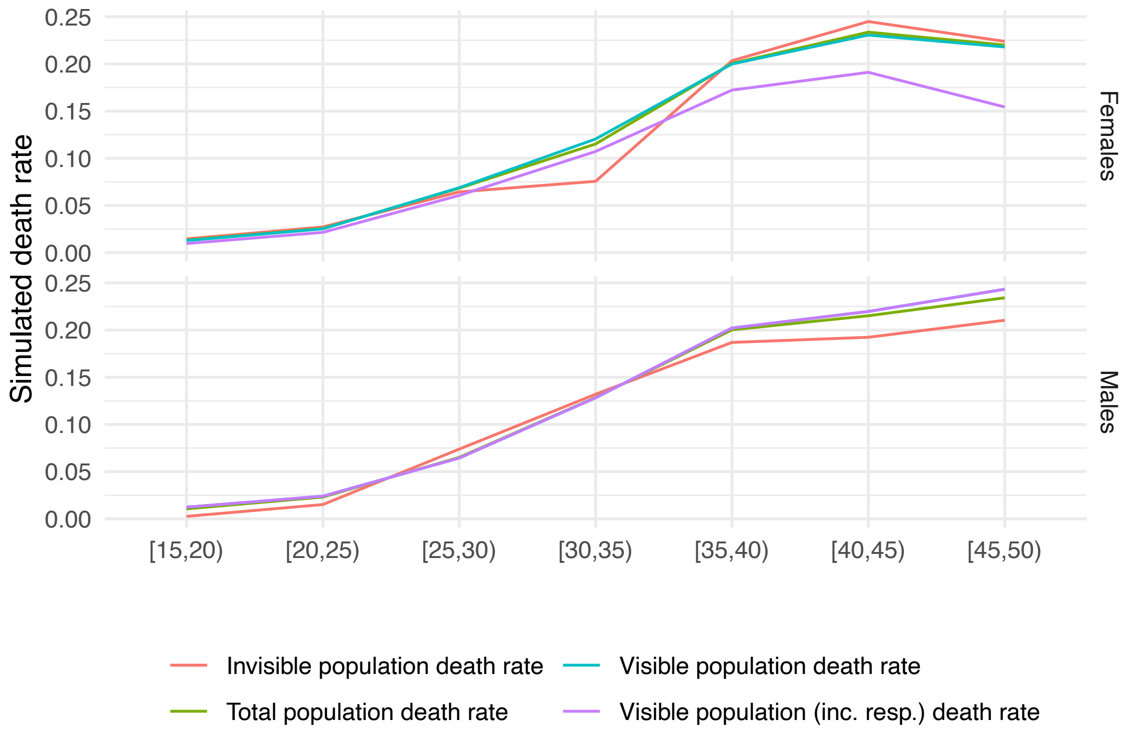

Figure 11: Age- and sex-specific death rates among members of the simulated population.

fraction in $f = \{0.05, 0.1, 0.15, 0.3, 0.6\}$.

Finally we calculate the four estimates for each simulated survey: aggregate visibility including/excluding the respondent and individual visibility including/excluding the respondent. Within each scenario, these M estimates are the simulated sampling distribution for the given estimator.

To recap, we generate a pseudo-universe of people linked together in sibships based on the actual sibships reported in the 2000 Malawi DHS survey. From this pseudo-universe, we generate many simulated surveys of different sample sizes, and different assumptions about reporting errors. By calculating death rate estimates using the four estimators for each of these simulated surveys and comparing the estimates to the known underlying truth, we can assess how well the sampling distributions of the four estimators recover true death rates under different reporting conditions and sample sizes.

J.3 Results

J.3.1 Confirming the accuracy of the sensitivity framework

First, we examine plots that compare the estimands for the aggregate and individual visibility estimators to the true underlying age-specific visible death rates. By examining the estimands, we remove the complication of sampling and focus on the quantity that each estimator would estimate if the entire frame population were interviewed. These plots will confirm the correctness of our analytical frameworks and provide some intuition about how estimators are affected by reporting errors.

In the main text, Figure 2 compares the true visible death rates (x axis) to the adjusted and unadjusted aggregate death rate estimands (y axis). We repeat that figure below, as Figure 12, for convenience. As we discuss in the main text, two important features emerge from Figure 12: first, the adjusted estimands all lie on the diagonal $y = x$ line, confirming the correctness of the sensitivity framework for the aggregate estimator (Equation 54). Second, by comparing the unadjusted estimates across the four reporting scenarios, it is clear that the unadjusted estimands for the scenario in which $\tau_D = 0.8$ and $\tau_N = 0.8$ (top-left panel) are nearly as accurate as the scenario in which $\tau_D = 1$ and $\tau_N = 1$ (bottom-right panel). Intuitively, this suggests that in some cases imperfect reporting may not be very problematic for the aggregate sibling survival estimator, as long as the imperfect reporting is similar for deaths and for exposure; in that case, Figure 12 shows that the reporting errors can cancel out (confirming the intuition from Equation 54).

Figure 13 is analogous to Figure 12, but for the aggregate visibility estimator with reports including the respondent; similarly, Figure 14 and Figure 15 are analogous to Figure 12, but for the individual visibility estimator (Figure 14) and for the individual visibility estimator with reports including the respondent (Figure 15). In all three plots, the adjusted estimands lie on the diagonal $y = x$ line, confirming the correctness of the sensitivity framework. The intuition about reporting errors also carries over to Figure 14 and Figure 15: in each case, the estimands for the scenario in which $\tau_D = 0.8$ and $\tau_N = 0.8$ are nearly as accurate as the scenario in which $\tau_D = 1$ and $\tau_N = 1$, suggesting that reporting errors about deaths and exposure may not be problematic, if the imperfect reporting is similar for deaths and for exposure. However, Figure 13 reveals that reporting errors for the aggregate visibility estimator including the respondent do not follow the same pattern as the other three estimators. Instead, Figure 13 shows that for both the scenario in which $\tau_D = 0.8$ and $\tau_N = 0.8$, and also the scenario in which $\tau_D = 1$ and $\tau_N = 1$, unadjusted estimates of female death rates are inaccurate. We suspect that this can be explained by the fact that, in our simulation, females are members of the frame population; in this case, the discussion in Appendix G.3 suggests that a better aggregate visibility estimator would subtract one from the aggregate exposure in the denominator. We do not pursue this question further here, since we do not recommend using the aggregate visibility estimator with reports including the respondent.

J.3.2 Comparing the estimators

Next, we investigate the performance of the individual and aggregate visibility estimators. In order to evaluate each estimator, we calculated the relative mean square error across the $K = 1000$ simulated surveys for each scenario. The relative mean square error of a set of estimates $\widehat{\vec{M}}^V$ is

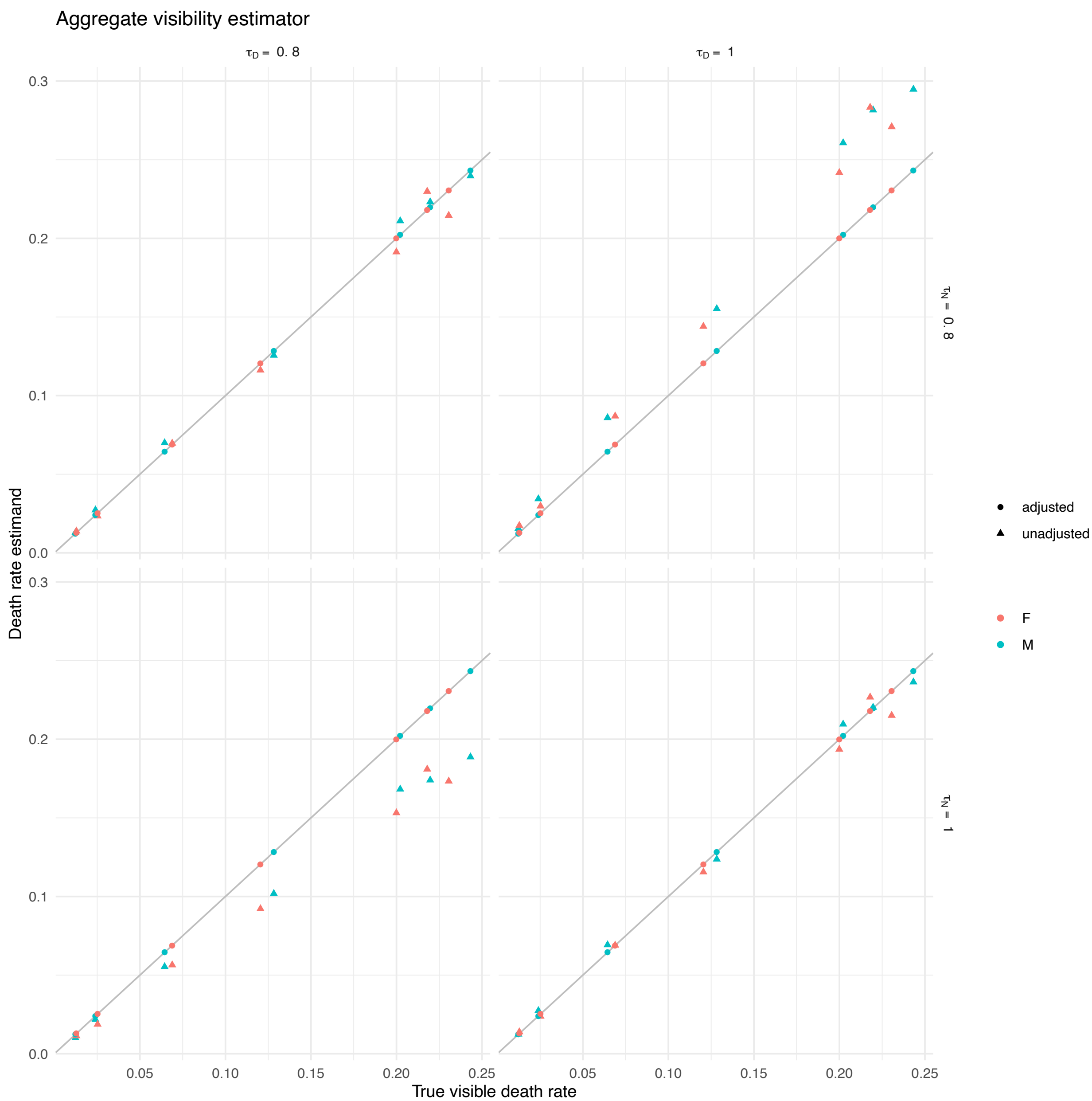

Figure 12: True age-specific death rates (x axis) against adjusted and unadjusted death rate aggregate visibility estimands (y axis). Male death rates are in blue and females are in red. The four panels show four different reporting scenarios. The adjusted estimands all agree with the true underlying death rates, illustrating the correctness of the aggregate sensitivity framework in Equation 54. These results are for the aggregate estimator that does not include the respondent.

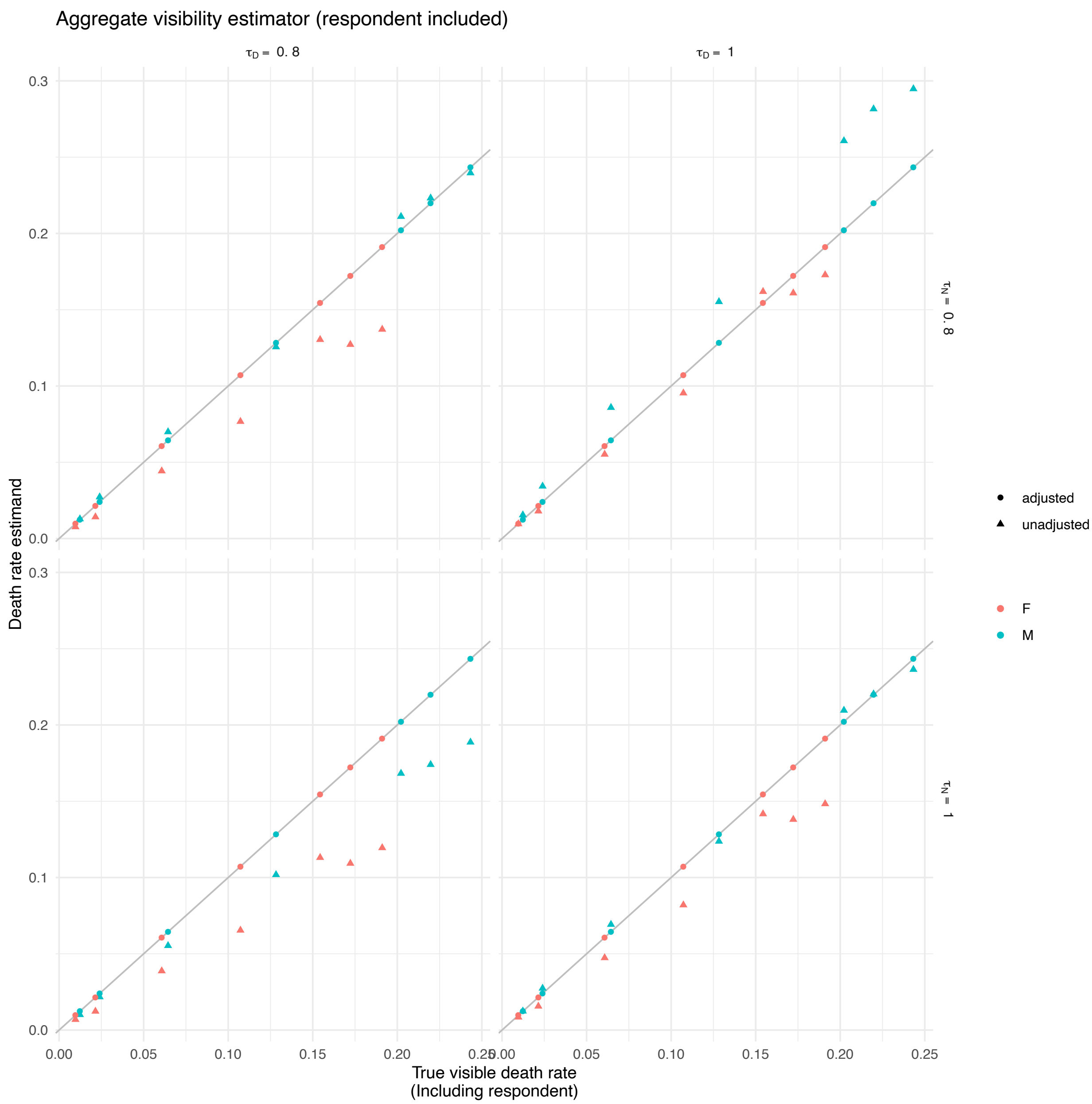

Figure 13: True age-specific death rates (x axis) against adjusted and unadjusted death rate aggregate visibility estimands with respondents included in reports (y axis). Male death rates are in blue and females are in red. The four panels show four different reporting scenarios. The adjusted estimands all agree with the true underlying death rates, illustrating the correctness of the aggregate sensitivity framework in Equation 56.

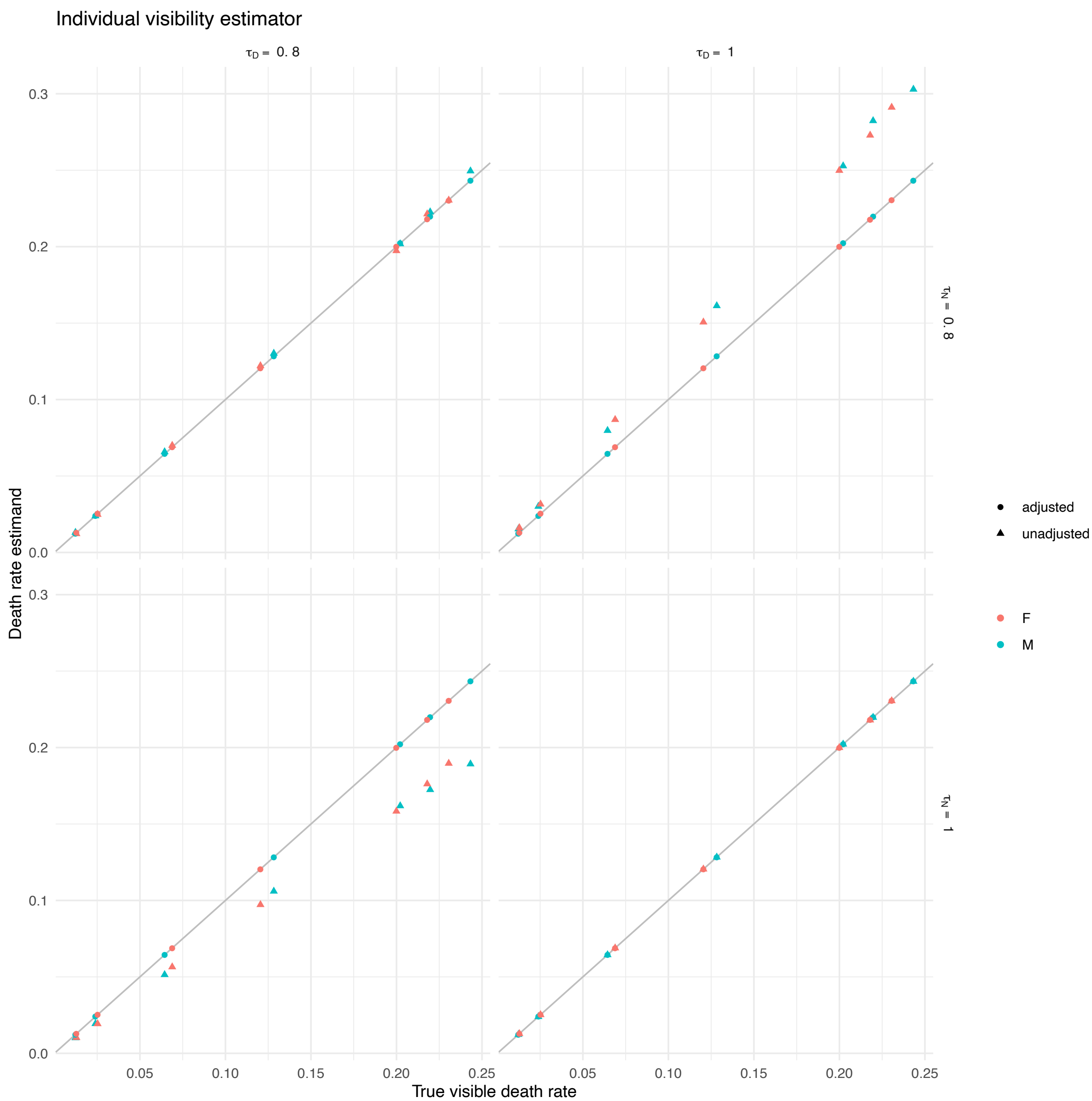

Figure 14: True age-specific death rates (x axis) against adjusted and unadjusted death rate individual visibility estimands (y axis). Male death rates are in blue and females are in red. The four panels show four different reporting scenarios. The adjusted estimands all agree with the true underlying death rates, illustrating the correctness of the individual sensitivity framework in Equation 69 These results are for the individual estimator that does not include the respondent.

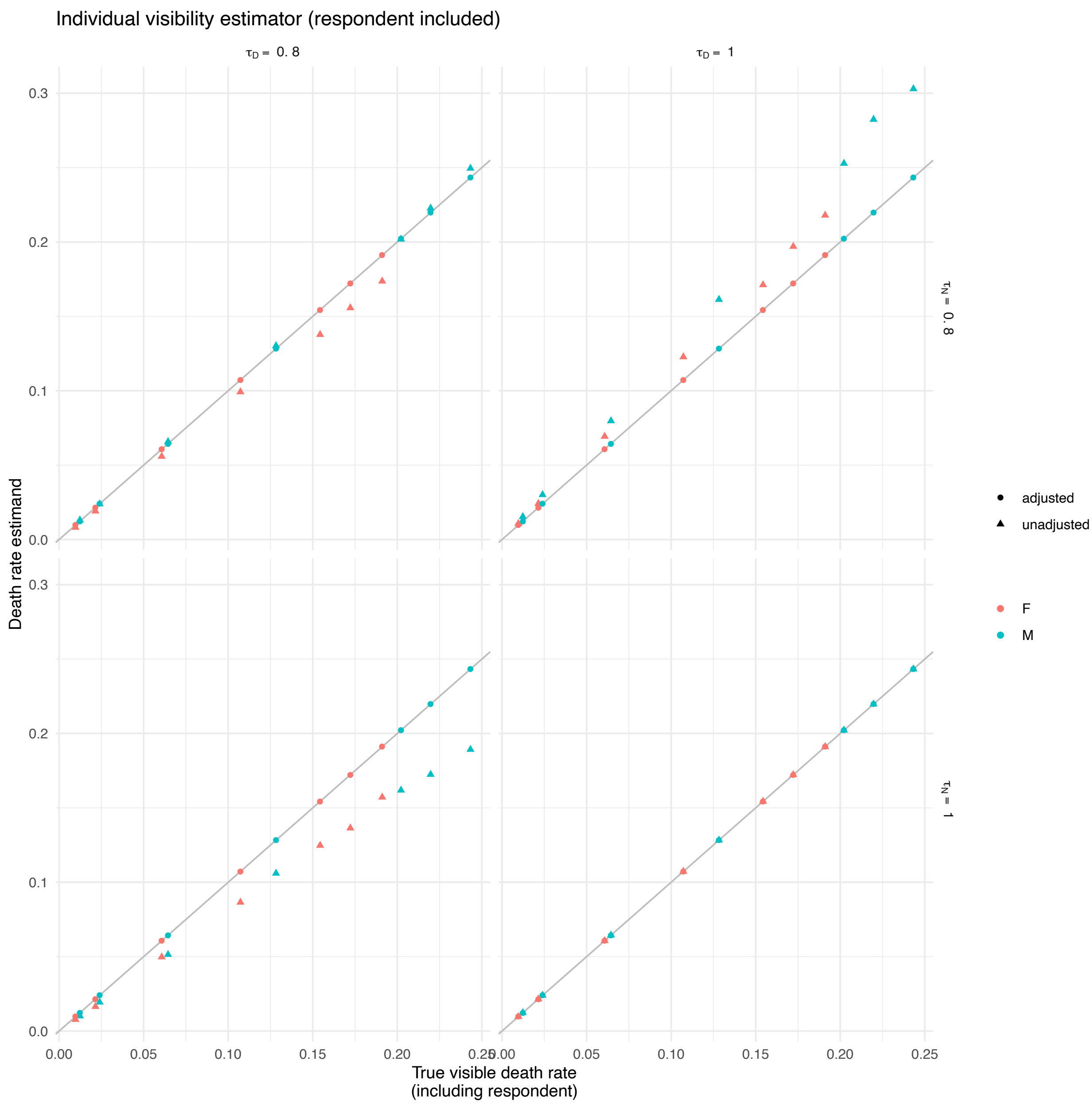

Figure 15: True age-specific death rates (x axis) against adjusted and unadjusted death rate individual visibility estimands with respondents included in reports (y axis). Male death rates are in blue and females are in red. The four panels show four different reporting scenarios. The adjusted estimands all agree with the true underlying death rates, illustrating the correctness of the individual sensitivity framework in Equation 71.

$$\text{rel-MSE}(\widehat{\vec{M}}^V) = \frac{\sum_{i=1}^K (\widehat{M}_i^V - M^V)^2}{K\ (M^V)^2}.$$

The relative mean square error can be decomposed into the sum of the squared relative bias:

$$\text{rel-bias}^2(\widehat{\vec{M}}^V) = \left(\frac{\sum_{i=1}^K \widehat{M}_i^V - M^V}{K\ M^V} \right)^2,$$

and the relative variance:

$$\text{rel-var}(\widehat{\vec{M}}^V) = \frac{\sum_{i=1}^K (\widehat{M}_i^V - \bar{M})^2}{(M^V)^2},$$

where $\bar{M} = K^{-1} \sum_i \widehat{M}_i^V$ is the average of the K estimates. Thus, we have

$$\text{rel-MSE}(\widehat{\vec{M}}^V) = \text{rel-bias}^2(\widehat{\vec{M}}^V) + \text{rel-var}(\widehat{\vec{M}}^V).$$

Figure 16 compares the two approaches when the sampling fraction (i.e., the sample size relative to the population size) is 0.05 and reporting is perfect. The figure shows the relative MSE as well as its decomposition into relative squared bias and relative variance. Several observations can be made about Figure 16. First, the magnitude of the relative MSE is comparable for the individual and aggregate visibility estimators, though it is slightly higher using the individual visibility estimator for the oldest age females. Second, the decomposition of the MSE into squared bias and variance reveals that the individual visibility estimator is essentially unbiased – all of its MSE comes from variance. The aggregate visibility estimator, on the other hand, is slightly biased, even in this favorable simulation setup.

Figure 17 compares the two estimators when the sampling fraction is 0.05 but reporting is imperfect; in this scenario, the true positive rate for reports about deaths is 0.8, meaning that some deaths are omitted from reports. Several observations can be made about Figure 17. First, compared to the perfect reporting results shown in Figure 16, the level of relative MSE is higher in Figure 17, presumably due to the reporting errors. Second, again due to the reporting errors, we see now that both the aggregate and individual visibility estimators are biased. Generally, the MSE are again similar between the two approaches, now with some cells producing slightly lower MSE for the individual visibility estimator.

Importantly, both Figure 16 and Figure 17 show the results of scenarios in which there is, by design, no relationship between visibility and mortality. In situations where such a relationship exists, we would expect the individual visibility estimator to tend to perform better than the aggregate visibility estimator, since the individual visibility estimator avoids having to make the assumption that the visibility of deaths is equal to the visibility of exposure.

Future research could extend this simulation analysis to better understand the tradeoffs between these two approaches; our results here (i) confirm that our analytical results are correct; and (ii) illustrate that, in relatively favorable situations, the error from the aggregate and individual visibility estimators is roughly comparable. A deeper understanding of the difference in the two estimators is an important topic for future research.

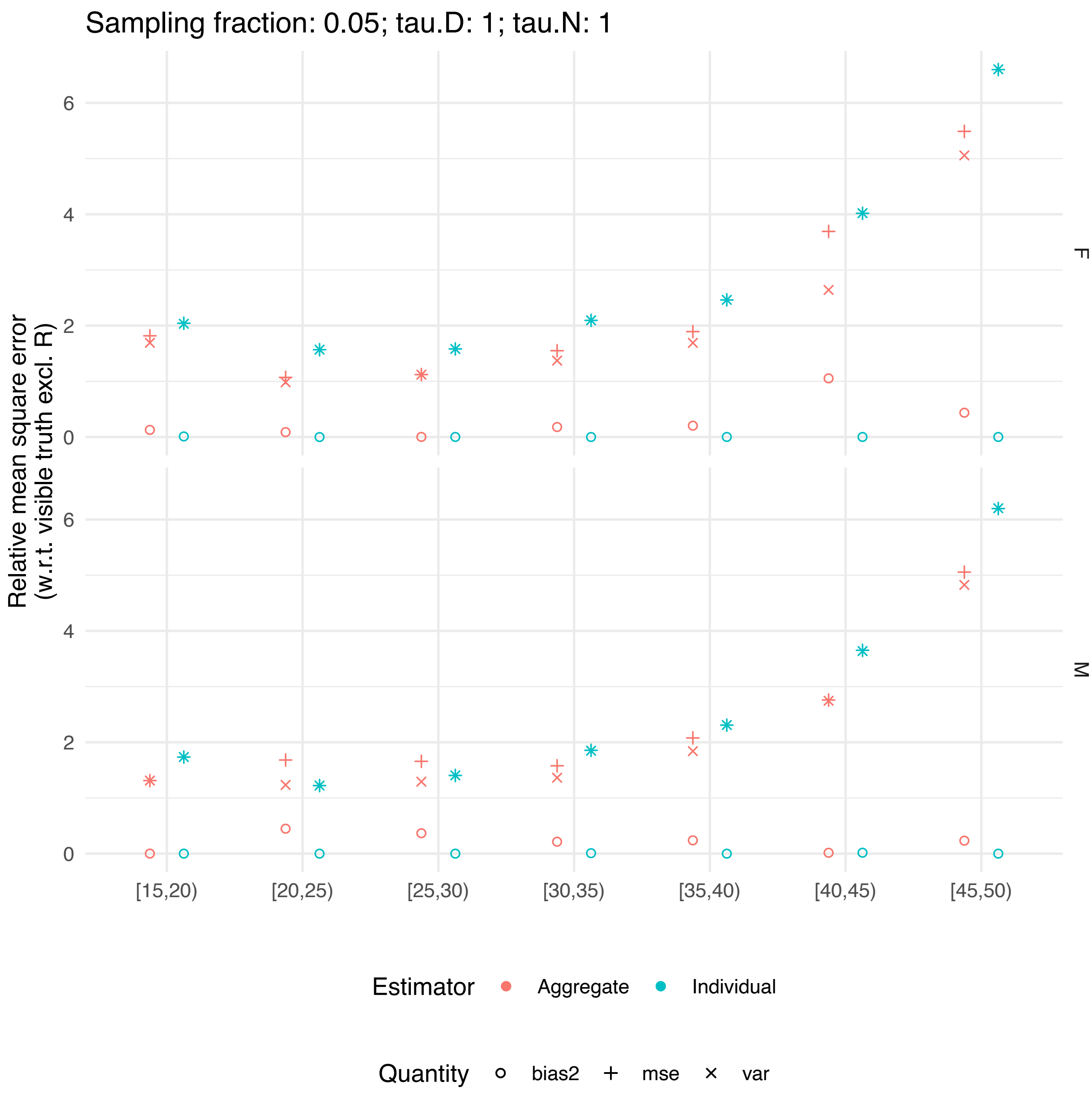

Figure 16: Mean square error, squared bias, and variance for the individual and aggregate visibility estimators when the sampling fraction is 0.05.

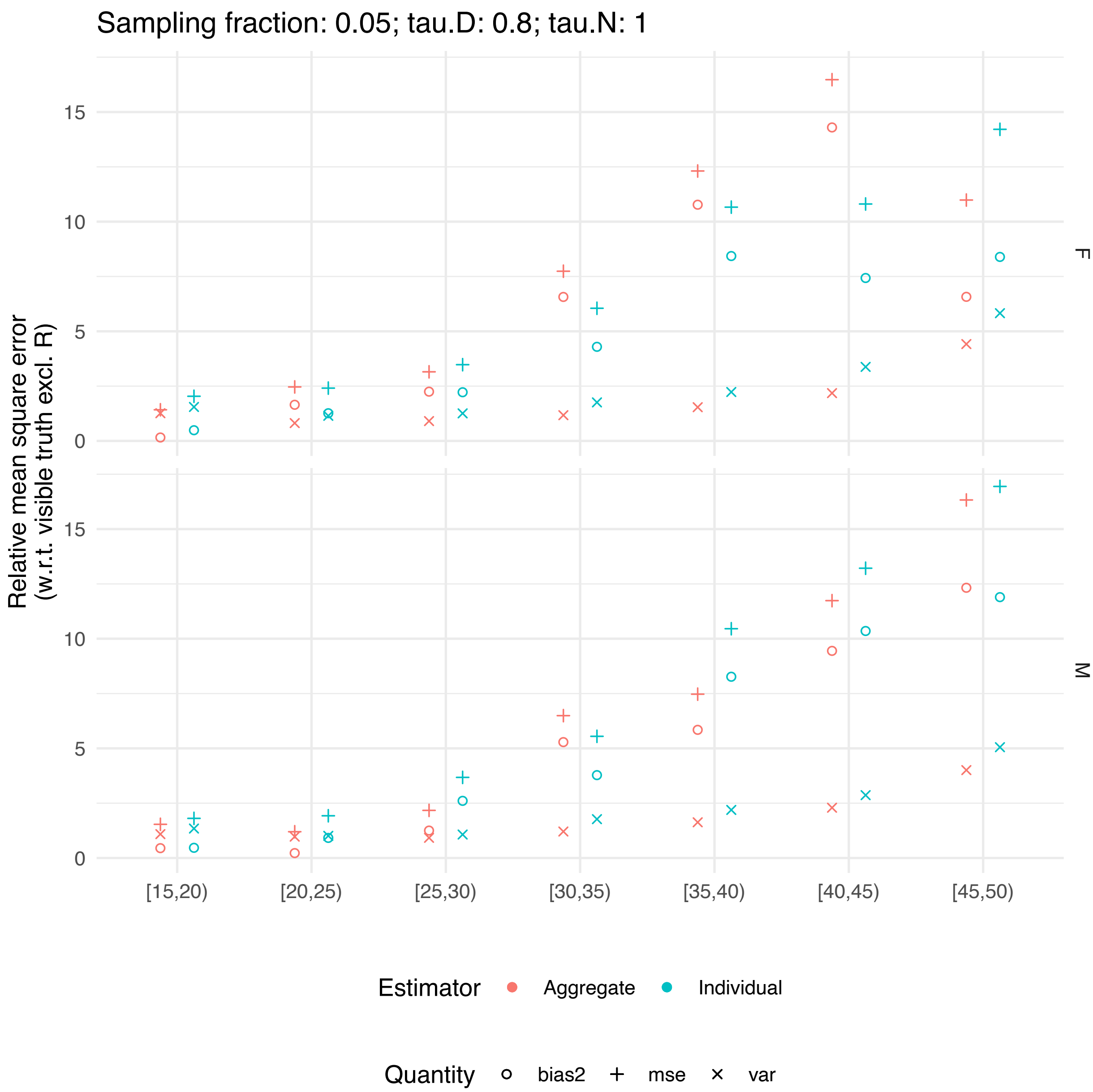

Figure 17: Mean square error, squared bias, and variance for the individual and aggregate visibility estimators when the sampling fraction is 0.05.

References

Bernard, H. R., Johnsen, E. C., Killworth, P. D., & Robinson, S. (1989). Estimating the size of an average personal network and of an event subpopulation. In M. Kochen (Ed.), *The small world* (pp. 159–175). Norwood, NJ: Ablex Publishing.

Feehan, D. M., Mahy, M., & Salganik, M. J. (2017). The network survival method for estimating adult mortality: Evidence from a survey experiment in Rwanda. *Demography, 54,* 1503–1528.

Feehan, D. M., & Salganik, M. J. (2016a). Generalizing the network scale-up method: A new estimator for the size of hidden populations. *Sociological Methodology, 46,* 153–186.

Feehan, D. M., & Wrigley-Field, E. (2021). How do populations aggregate? *Demographic Research, 44,* 363–378. https://doi.org/10.4054/DemRes.2021.44.15

Hanley, J. A., Hagen, C. A., & Shiferaw, T. (1996). Confidence intervals and sample-size calculations for the sisterhood method of estimating maternal mortality. *Studies in Family Planning, 27*, 220–227.

Lavallee, P. (2007). *Indirect sampling*. New York, NY: Springer-Verlag.

Masquelier, B. (2013). Adult mortality from sibling survival data: A reappraisal of selection biases. *Demography, 50,* 207–28.

Rao, J. N. K., & Pereira, N. P. (1968). On double ratio estimators. *Sankhyā?: The Indian Journal of Statistics, Series A (1961–2002), 30,* 83–90.

Rutstein, S. O., & Guillermo Rojas, M. C. S. (2006). *Guide to DHS statistics*. Calverton, MD: ORC Macro.

Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. New York, NY: Springer Verlag.

Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the American Statistical Association, 65,* 257–266.

Thompson, S. K. (2002). *Sampling* (2nd ed.). Hoboken, NJ: Wiley.

Trussell, J., & Rodriguez, G. (1990). A note on the sisterhood estimator of maternal mortality. *Studies in Family Planning, 21,* 344–346.

Wolter, K. (2007). *Introduction to variance estimation* (2nd ed.). New York, NY: Springer.