Estimating Adult Death Rates From Sibling Histories: A Network Approach

Dennis M. Feehan,
Department of Demography, University of California, Berkeley, CA, USA;

Gabriel M. Borges
Brazilian Institute of Geography and Statistics (IBGE), Rio de Janeiro, Brazil;

Abstract

Hundreds of millions of people live in countries that do not have complete death registration systems, meaning that most deaths are not recorded and that critical quantities, such as life expectancy, cannot be directly measured. The sibling survival method is a leading approach to estimating adult mortality in the absence of death registration. The idea is to ask survey respondents to enumerate their siblings and to report about their survival status. In many countries and periods, sibling survival data are the only nationally representative source of information about adult mortality. Although a vast amount of sibling survival data has been collected, important methodological questions about the method remain unresolved. To help make progress on this issue, we propose reframing the sibling survival method as a network sampling problem. This approach enables a formal derivation of statistical estimators for sibling survival data. Our derivation clarifies the precise conditions that sibling history estimates rely on, leads to internal consistency checks that can help assess data and reporting quality, and reveals important quantities that could potentially be measured to relax assumptions in the future. We introduce the R package siblingsurvival, which implements the methods we describe.

Keywords

Networks; Mortality; Demographic and health surveys; Sampling

Introduction

Death rates at adult ages are a core component of population health and a central topic of study for demography. Unfortunately, most of the world’s poorest countries do not have complete death registration systems, meaning that most people die without ever having...
their existence officially recorded (AbouZahr et al. 2015; Setel et al. 2007). This lack of complete death registration means that critical quantities, such as life expectancy, cannot be directly measured. Improving death registration systems is the long-term solution to this scandal of invisibility, but progress has been very slow (Mikkelsen et al. 2015). Until complete death registration systems are available everywhere, sample-based approaches to adult mortality estimation will continue to play a critical role in understanding population health and well-being.

The leading approach to collecting information about adult mortality in the absence of death registration is the sibling survival method (Brass 1975; Rutenberg and Sullivan 1991). The idea is to ask survey respondents to report the number of siblings they have and to then ask for each sibling’s gender, date of birth, and (where appropriate) date of death. This data collection strategy produces sibling histories that contain information about the survival status of all of the members of the respondent’s sibship.

Because high-quality household surveys are routinely conducted in most countries—including countries that lack death registration systems—the sibling survival method offers the opportunity to try to estimate adult death rates in many places that have no other nationally representative adult mortality data. Over the past two decades, a vast amount of sibling history data has been collected; for example, as a part of the Demographic and Health Surveys (DHS) program alone, sibling histories have been collected in more than 150 surveys from dozens of countries around the world (Corsi et al. 2012; Fabic et al. 2012).

However, understanding how to analyze sibling histories has proven to be very challenging. Researchers have long been aware that the method suffers from many possible sources of bias (Gakidou and King 2006; Graham et al. 1989; Masquelier 2013; Reniers et al. 2011; Trussell and Rodriguez 1990). Previous studies have concluded that sibling history estimates can be problematic if (1) there are sibships with no surviving members who could potentially be sampled and interviewed in the survey; (2) more generally, there is a relationship between sibship size and mortality (e.g., larger sibships face higher death rates); and (3) respondents’ reports about their siblings are inaccurate (e.g., respondents omit siblings or misreport a sibling’s survival status). There has also been confusion about whether the survival status of the respondent should be included in the calculations, given that respondents are always alive (Masquelier 2013; Reniers et al. 2011).

Researchers have worked on addressing these concerns about the sibling survival method in three main ways: they have collected empirical information about possible sources of bias in sibling reports (e.g., Helleringer, Pison, Kanté et al. 2014), they have used microsimulation to illustrate how large certain sources of bias can be under different scenarios (Masquelier 2013), and they have used regression models to pool information from different countries and periods (Gakidou and King 2006; Obermeyer et al. 2010; Timaeus and Jasseh 2004). Together, these studies have produced many important insights about the sibling survival method. However, these insights have not yet brought about a consensus on how sibling histories should be analyzed. Currently, there is partial evidence about many individual sources of possible bias, but there is no way to integrate all of this evidence. Thus, even if we knew the exact size and direction of all the different sources of possible error, we still
would not understand how the errors would combine to affect estimated death rates. More generally, little has been proven about the precise conditions under which sibling survival estimates can be expected to have attractive statistical properties, such as consistency or unbiasedness.

In this study, our goal is to help resolve some of the methodological uncertainty about sibling survival. Our analysis is based on the insight that the sibling relation induces a particular type of social network among the members of a population. In this network, two people are connected if they are siblings; thus, estimating death rates from sibling histories can be understood as a problem in network sampling. Starting from the principles of network reporting, we describe how a sibling survival estimator can be mathematically derived. Deriving an estimator from first principles in this way enables us to (1) clarify the precise assumptions that the estimator requires in order to be consistent, unbiased, and efficient; (2) describe how violations of any and all assumptions can combine to affect estimated death rates; (3) identify quantities that could potentially be measured in the future to relax assumptions; and (4) develop internal consistency checks that can be used to assess data and reporting quality in a given sample.

Setup

Figure 1 illustrates how we understand sibling histories as a network reporting problem. Panel a shows a small population whose members are connected if they are siblings; that is, two nodes are connected if they have the same mother. Clear nodes are alive, and gray nodes are dead at the time of the survey. Because the sibling relation is transitive, the network is entirely composed of fully connected components, or cliques; each of these cliques is one sibship. Panel b shows one specific sibship, and panel c illustrates the bipartite reporting network that is generated when all the surviving members of that sibship are asked to report about their siblings. In the bipartite reporting network, each directed edge represents one sibling reporting about another. For example, the edge 13 → 12 indicates that node 13 reports about node 12. See Feehan and Salganik (2016a) for a more detailed description of bipartite reporting networks.

Our quantity of interest is $M_\alpha$, the death rate for a specific group $\alpha$ (for example, $\alpha$ might be all women aged 30–34 in 2018). $M_\alpha$ is defined as

$$M_\alpha = \frac{D_\alpha}{N_\alpha},$$

where $D_\alpha$ is the number of deaths in group $\alpha$, and $N_\alpha$ is the person-years of exposure among members of group $\alpha$. We can develop an estimator for $M_\alpha$ by separately estimating the numerator and the denominator of Eq. (1); thus, the challenge is to derive sibling history-based estimators for $D_\alpha$ and $N_\alpha$.

---

1 Respondents are typically asked to consider siblings to be all children born to their mother.
2 The dead person (in gray) cannot be interviewed and thus is not shown on the left side of the bipartite reporting network.
Panel c of Figure 1 illustrates that each sibling can potentially be reported as many times as she has living sibship members who are eligible to respond to the survey (Gakidou and King 2006; Masquelier 2013; Sirken 1970). Inferences from sibling reports must somehow account for this fact. Our approach is to distinguish between two groups of people: the first group is people who have no siblings who are eligible to respond to the survey. These people will never appear in the sibling history data: they are invisible to the sibling histories. The second group is visible people who do have siblings eligible to respond to the survey.

Formally, let $U$ be the population being studied, and let $F \subset U$ be the frame population, which is the set of all people who are eligible to respond to the survey. We define person $j$’s visibility, $v(j, F)$, to be the number of living siblings who would report person $j$ in a census of $F$. Everyone in the population is either visible ($v(j, F) > 0$) or invisible ($v(j, F) = 0$). Thus, we can write the number of deaths in group $\alpha$ as follows:

$$D_\alpha = \sum_{j \in D_\alpha} 1 + \sum_{j \in D_\alpha} 1 = D^V_\alpha + D^I_\alpha,$$

(2)

where $D^V_\alpha$ is the number of visible deaths that could be learned about using sibling reports, and $D^I_\alpha$ is the number of invisible deaths that cannot be learned about using sibling reports. We can define analogous quantities for the denominator $N_\alpha = N^V_\alpha + N^I_\alpha$, where the $N^V_\alpha$ is the visible exposure, and $N^I_\alpha$ is the invisible exposure. Finally, we define $M^I_\alpha = \frac{D^I_\alpha}{N^I_\alpha}$ to be the invisible death rate, $M^V_\alpha = \frac{D^V_\alpha}{N^V_\alpha}$ to be the visible death rate, and $M_\alpha = \frac{D^I_\alpha + D^V_\alpha}{N^I_\alpha + N^V_\alpha}$ to be the total death rate.

### Adjusting for Visibility in Sibling Reports

The visible death rate can be estimated from sibling histories using an expression of the form

$$\hat{M}^V_\alpha = \frac{\hat{D}^V_\alpha}{\hat{N}^V_\alpha},$$

(3)

where $\hat{D}^V_\alpha$ is an estimator for the number of visible deaths in group $\alpha$, and $\hat{N}^V_\alpha$ is an estimator for the amount of visible exposure in group $\alpha$. In estimating these two quantities, we face the challenge that even people who are visible to the sibling histories may still differ in the extent to which they are visible; for example, visible people from larger sibships may tend to have different death rates than visible people from smaller sibships. We address this

---

4The idea behind the notation $\gamma(\cdot, \cdot)$ is that the first argument is the set of people reporting, and the second argument is the set of people who are being reported about; thus, $\gamma(F, D^I_\alpha)$ is the total number of deaths in $D^I_\alpha$ reported by people in a census of the frame population $F$. 

*Demography. Author manuscript; available in PMC 2024 March 09.*
challenge by introducing statistical estimators that adjust for how visible reported siblings are.

We consider two different approaches to adjusting for differential visibility: aggregate visibility estimation and individual visibility estimation. These two approaches lead to two different estimators for the visible death rate. In both cases, we start by describing how to derive population-level relationships—that is, relationships that would be observed in a census. Of course, sibling histories are not typically collected in censuses, but starting from this perspective will enable us to derive relationships that can then form the basis for sample-based estimators.

**The Aggregate Visibility Approach**

The aggregate visibility approach is based on the idea that reports about siblings can first be aggregated, and then the aggregated reports can be adjusted to account for visibility (Bernard et al. 1989; Feehan and Salganik 2016a; Rutstein and Guillermo Rojas 2006). To illustrate this approach, we first focus on reports about visible deaths among siblings, $D^\alpha_V$.

Throughout the paper, we assume that there are no *false positive* reports—that is, we assume that respondents’ reports may omit siblings but that they never mistakenly include someone who is not truly a sibling. Of course, this could in fact happen, but our assumption makes the exposition much cleaner, and the results derived in the online appendix consider reporting with false positives.

Let $y(F,D^\alpha_H)$ be the total number of deaths that would be reported among respondents’ siblings in a census of the frame population. Section B. 4 of the online appendix shows that if there are no false positive reports, then the total number of reports about sibling deaths divided by the average visibility of visible deaths will be equal to the number of visible deaths: \[^5\]

$$D^\alpha_V = \frac{y(F,D^\alpha_H)}{\bar{v}(D^\alpha_H,F)}. \tag{4}$$

The idea is that in a census, the average death will be reported $\bar{v}(D^\alpha_H,F)$ times; thus, the total number of reports about deaths, $y(F,D^\alpha_H)$, can be divided by $\bar{v}(D^\alpha_H,F)$ to recover the total number of visible deaths. Section B. 4 of the online appendix shows that a similar analysis can be applied to the denominator of the visible death rate, yielding

$$N^\alpha_V = \frac{y(F,N^\alpha_H)}{\bar{v}(N^\alpha_H,F)}. \tag{4}$$

Thus, in a census,

---

\[^5\]To avoid overcomplicating notation, we use $D^\alpha_V$ to indicate both the number of visible deaths in group $\alpha$ and the set of visible deaths in group $\alpha$.
\[ M'_{\alpha} = \frac{D'_{\alpha}}{N_{\alpha}} = \frac{\tilde{y}(F, D'_{\alpha})}{\tilde{y}(F, N'_{\alpha})} \times \frac{\tilde{v}(N'_{\alpha}, F)}{\tilde{v}(D'_{\alpha}, F)} \]

(5)

Under the assumption that the visibility of deaths is the same as the visibility of exposure, \( \tilde{v}(D'_{\alpha}, F) = \tilde{v}(N'_{\alpha}, F) \), and Eq. (5) simplifies to the aggregate visibility estimand:

\[ M'_{\alpha, \text{agg}} = \frac{\tilde{y}(F, D'_{\alpha})}{\tilde{y}(F, N'_{\alpha})} \]

(6)

Finally, the population-level relationship in Eq. (6) motivates the sample-based estimator (section E of the online appendix):

\[ M'_{s, \text{agg}} = \frac{\sum_{i \in s} w_{i} \tilde{y}(i, D'_{\alpha})}{\sum_{i \in s} w_{i} \tilde{y}(i, N'_{\alpha})} \]

(7)

Because this approach is based on combining reports about all the siblings and then adjusting for the visibility of these aggregate reports, we call \( M'_{s, \text{agg}} \) an aggregate visibility estimator (Bernard et al. 1989; Feehan 2015; Feehan et al. 2017). In the online appendix, section E (Result E.1) formally derives the estimator; the derivation reveals that this approach can be expected to produce essentially unbiased estimates as long as reports about siblings are accurate and as long as there is no relationship between sibship visibility and mortality (i.e., as long as \( \tilde{v}(D'_{\alpha}, F) = \tilde{v}(N'_{\alpha}, F) \)).

The Individual Visibility Approach

The individual visibility approach is based on the idea that reports about siblings can be first adjusted for visibility and then aggregated (Feehan 2015; Gakidou and King 2006; Lavallee 2007; Sirken 1970). To illustrate this approach, consider reports about a specific deceased sibling \( j \in D_{\alpha} \) that are made in a census of the frame population. Let \( y(F, j) \) be the number of times that people in the frame population \( F \) report the deceased sibling \( j \). This quantity, \( y(F, j) \), will be equal to the visibility of \( j \) to \( F \), \( v(j, F) \), as long as there are no false positive reports. Thus, for every visible sibling \( j \), we have

\[ \frac{y(F, j)}{v(j, F)} = 1. \]

(8)

Summing over all visible deaths, we obtain

\[ D'_{\alpha} = \sum_{i \in \nu'_{\alpha}} 1 = \sum_{i \in \nu'_{\alpha}} \frac{y(F, j)}{v(j, F)} = \sum_{i \in F} \sum_{j \in \nu'_{\alpha}} \frac{y(i, j)}{v(j, F)}. \]
where the last step follows because $y(F, j) = \sum_{i \in F} y(i, j)$. Equation (9) expresses the number of visible deaths in terms of each survey respondent’s reported connections to each death ($y(i, j)$) and the visibility of each death ($v(j, F)$). In section B.2 of the online appendix, we show that when reporting is accurate, the visibility of any dead sibling, $j \in D^c_{\alpha}$, can be written as $v(j, F) = y(i, F) + 1$ for any survey respondent $i$ who is in the same sibship as $j$. Using this relationship, Eq. (9) can be rewritten as

$$D^c_{\alpha} = \sum_{i \in F} \sum_{j \in D^c_{\alpha}} \frac{y(i, j)}{y(i, F)} = \sum_{i \in F} \sum_{j \in D^c_{\alpha}} \frac{y(i, j)}{y(i, F) + 1} = \sum_{i \in F} \frac{y(i, D)}{y(i, F) + 1},$$  

(10)

where $y(i, F)$ is $i$’s reported number of siblings on the sampling frame. Equation (10) relates the population-level number of visible deaths $D^c_{\alpha}$ to survey respondents’ reports about deaths in their sibships, $y(i, D)$, and survey respondents’ reports about the number of frame population members in their sibships, $y(i, F)$.

A parallel argument found in section B.2 of the online appendix reveals that when reporting is accurate, $N^c_{\alpha}$ can be written as

$$N^c_{\alpha} = \sum_{i \in F} \left[ \frac{y(i, N_{\alpha} \cap F)}{y(i, F)} + \frac{y(i, N_{\alpha} \cap F)}{y(i, F) + 1} \right] \frac{y(i, D)}{y(i, F) + 1},$$  

(11)

where $y(i, N_{\alpha} \cap F)$ is $i$’s reported number of siblings on the sampling frame who contributed exposure; and $y(i, N_{\alpha} - F)$ is $i$’s reported number of siblings not on the sampling frame who contributed exposure. Combining Eqs. (10) and (11), we have the population-level individual visibility estimand:

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

(12)

Finally, the population-level relationship in Eq. (12) motivates a sample-based estimator for $M^c_{\alpha}$ (online appendix, section F):

$$\hat{M}^c_{\alpha, ind} = \frac{\sum_{i \in s} \left[ \frac{y(i, D)}{y(i, F) + 1} \right]}{\sum_{i \in s} \left[ \frac{y(i, N_{\alpha} \cap F)}{y(i, F)} + \frac{y(i, N_{\alpha} - F)}{y(i, F) + 1} \right]}.$$  

(13)
where \( i \) indexes survey respondents in the probability sample \( s \), and \( w_i \) is \( i \)'s sampling weight. Because this approach is based on adjusting for the visibility of each individual reported sibling, we call it individual visibility estimation. Section F of the online appendix (Result F.1) formally derives the estimator in Eq. (13), including the precise conditions required for it to provide consistent and essentially unbiased estimates of the visible death rate.

**Relationship to Previous Work**

To the best of our knowledge, our study is the first to derive the aggregate visibility estimator from first principles. However, the estimator itself is not new: the aggregate visibility estimator is probably the most common approach to estimating death rates from sibling history data. For example, Eq. (7) is the estimator used to produce age-specific adult death rate estimates in all DHS reports (Rutstein and Guillermo Rojas 2006). The estimator appears to have been first proposed in Rutenberg and Sullivan (1991), and it has since been the subject of several methodological analyses, including Masquelier (2013), Gakidou and King (2006), Hill et al. (2006), Timaeus and Jasseh (2004), Stanton et al. (2000), and Garenne et al. (1997). By focusing on the networked structure of sibling relations, our derivation reveals that the aggregate visibility estimator is related to other network estimation approaches, including the network scale-up method (Bernard et al. 1989), and the network survival estimator (Feehan et al. 2017).

The individual visibility estimator has its origins in multiplicity sampling (Feehan 2015; Lavallee 2007; Sirken 1970). In the context of sibling survival, an estimator similar to the one derived here was introduced by Gakidou and King (2006) and then further discussed by Obermeyer et al. (2010) and Masquelier (2013). The actual individual estimator in Eq. (13) is somewhat different from the one Gakidou and King (2006) proposed, but both are motivated by the idea that observed information can be used to adjust for visibility at the level of individual reports. We prove that the individual estimator introduced here is correct, and we provide code that other researchers can use to implement it.

**Framework for Sensitivity Analysis**

The individual and aggregate visibility estimators rely on several conditions to guarantee that they will produce consistent and essentially unbiased estimates of the death rate. These conditions make precise long-standing concerns researchers have had about sibling survival estimates. For example, researchers have often worried that inaccurate reports about siblings may lead to biased death rate estimates; our results reveal exactly how reports about siblings must be accurate in order to produce consistent and essentially unbiased estimates of death rates. They also reveal precise quantities that could potentially be measured to adjust sibling reports to account for reporting errors.

Sections E and F of the online appendix contain detailed derivations of the sensitivity frameworks for both the individual and aggregate visibility estimators; here, we present and discuss the results of that analysis. The simulation study in section J of the online appendix empirically illustrates the sensitivity frameworks and confirms their correctness.
Sensitivity of the Aggregate Visibility

The relationship between the population death rate in group $\alpha$ and the aggregate visibility estimand can be written as follows:

$$M_\alpha = \frac{M^V_{\text{agg}}}{\text{aggregate visibility estimand}} \times \frac{\bar{d}^V (N_\alpha, F)}{\text{visibility ratio}} \times \frac{\gamma (F, N_\alpha)}{\gamma (F, D_\alpha)} \times \left[ 1 + \rho_{\alpha} (K - 1) \right] \times \text{difference between invisible and visible populations}$$

(14)

Equation (14) shows that the visible death rate can be decomposed into the product of the aggregate visibility estimand and several adjustment factors. When all these adjustment factors are equal to 1, the aggregate visibility estimand is equal to the total death rate.\(^6\)

The first group of adjustment factors—called the visibility ratio—describes how a relationship between visibility and mortality would affect estimated death rates. It is the ratio of the average visibility of all siblings who contribute to exposure $\bar{d}^V (N_\alpha, F)$ to the average visibility of siblings who die $\bar{d}^V (D_\alpha, F)$.\(^7\) When there is no relationship between these two quantities, the visibility ratio will be 1. When, say, siblings who have died tend to be in less visible sibships than siblings overall, then this factor will tend to be greater than 1 (meaning that the death rate will be underestimated).

The next group of adjustment factors captures the extent to which reports about siblings are accurate. It is the ratio of a quantity that captures the net accuracy of reports about exposure ($\gamma (F, N_\alpha)$) to a quantity that captures the net accuracy of reports about deaths ($\gamma (F, D_\alpha)$).

Two particularly salient findings emerge from the derivation of this group of adjustment factors (online appendix, section E): first, the estimator requires that reporting be accurate in aggregate but not necessarily at the individual level; as long as reporting errors across individuals cancel out, estimates will not be affected. Second, this group of adjustment factors shows that aggregate visibility estimates will not be affected if reporting errors about deaths and reporting errors about exposure balance out. In other words, if respondents tend to, say, omit older siblings at a constant rate, independent of the survival status of older siblings ($\gamma (F, D_\alpha) = \gamma (F, N_\alpha) < 1$), then Eq. (14) reveals that death rates can still be accurate because reporting errors about deaths and about exposure will cancel out. Thus, Eq. (14) shows that the death rate estimator can be robust to situations in which respondents’ reports are imperfect, but imperfect in similar ways for siblings who die and siblings who survive.

\(^6\)More generally, if the product of these adjustment factors is 1, the aggregate visibility estimand will be the total death rate. This means that the conditions on which the estimator relies are sufficient but not necessary.

\(^7\)In the visibility ratio, $\bar{d} (A, B)$ refers to the true average number of sibship connections between the average member of group $A$ and group $B$; $\bar{d} (\cdot, \cdot)$ can differ from $\bar{v} (\cdot, \cdot)$ because $\bar{v} (\cdot, \cdot)$ could be affected by reporting errors. These reporting errors are accounted for in the reporting accuracy factor of the visibility framework. See the online appendix, section E, for details.
Finally, the last group captures the conditions needed to be able to use only information about visible deaths to estimate the total death rate. This group depends on two quantities: $p^{\alpha}_N$, the amount of exposure that is invisible; and $K = M^{\prime}_\alpha / M^{\prime}_V$, an index for how different the invisible and visible death rates are. Later in the paper, we will use an empirical example to illustrate this group of adjustment factors in more depth.

**Sensitivity of the Individual Visibility Estimator**

The derivation in section F of the online appendix reveals that the relationship between the population death rate in group $\alpha$ and the individual visibility estimand can be written as follows:

$$M^{\prime}_\alpha = \frac{M^{\prime\prime}_{\alpha, ind}}{\text{individual visibility estimand}} \times \frac{\gamma^{\prime}_\nu (1 + K_{\alpha})}{\gamma_\nu (1 + K_{\alpha})} \times \frac{[1 + p^{\alpha}_N (K - 1)]}{\text{difference between visible and indivivible populations}}$$

(15)

Equation (15) again decomposes the population death rate into the product of the individual visibility estimand and several groups of adjustment factors. The main insights from Eq. (14) also apply to the individual visibility expression in Eq. (15). However, a few differences between the two frameworks are noteworthy. First, Eq. (15) does not include any adjustment factors related to a visibility ratio. This is an advantage of the individual visibility estimator: it does not need to make any assumptions about the absence of a relationship between visibility and mortality (within the visible population). Second, the individual visibility expression in Eq. (15) has a more complex set of adjustment factors that capture reporting accuracy. This more complex expression describes the extent to which reporting errors are correlated with reports about deaths and reports about exposure; for example, if reporting tends to omit deaths in sibships that have more deaths, then $K_{\alpha} > 1$. More generally, even if deaths and exposure are underreported at the same average rate, such that the average reporting adjustment factor is equal to 1, problems can still arise if the reporting errors differ in their correlation with the number of sibship deaths and exposure. In general, reporting errors appear to be more complex under the individual visibility estimator; we discuss the implications of this later.

**Simulation Study**

To confirm the correctness of the sensitivity results, we conduct a simulation study. We first simulate a large population whose members are linked in sibships. We base the structure of these sibships and the age and sex composition of the population on sibling reports from the 2000 Malawi DHS. Next, we simulate deaths by age and sex in the population, using model-based estimates for adult death rates in Malawi in 2000–2005 as the “true” underlying mortality (United Nations Population Division 2020). We simulate sibling history interviews with each surviving member of the population under four sets of assumptions about reporting errors: perfect reporting, imperfect reporting about deaths (only 80% of deaths are reported, $\tau_\nu = .8$), imperfect reporting about exposure (only 80% of exposure is reported, $\tau_N = .8$), and imperfect reporting about deaths and exposure (\tau_{\nu, N} = .8)
and $\tau_s = .8$). These four sets of assumptions lead to four “censuses,” one for each set of reporting assumptions. Because our population is simulated, the exact death rate and visible death rate can be calculated by age and sex. For each set of reporting assumptions, we can also calculate the exact estimand for each of the four estimators we consider (aggregate and individual visibility, including and not including the respondent). Finally, we can calculate all the adjustment factors that appear in our sensitivity framework. If our sensitivity framework is correct, then the estimands that have been adjusted using the sensitivity framework should be equal to the underlying death rate.

For brevity, we focus here on the results for the aggregate visibility estimator not including the respondent; results for all four estimators can be found in the online appendix, section J. Figure 2 compares the true visible death rates (x-axis) with the adjusted and unadjusted aggregate death rate estimands (y-axis). Two important features emerge from Figure 2: first, all the adjusted estimands lie on the diagonal $y = x$ line, confirming the correctness of the sensitivity framework for the aggregate estimator (Eq. (14)). Second, by comparing the unadjusted estimates across the four reporting scenarios, it is clear that the unadjusted estimands for the scenario in which $\tau_D = .8$ and $\tau_N = .8$ (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 2 shows that the reporting errors can cancel out (confirming the intuition from Eq. (14)).

The online appendix (section J) describes our simulation setup in greater detail and provides results confirming that the sensitivity framework is correct for all four estimators.

**Empirical Illustration**

We motivate our technical results with an empirical example: the sibling history data from the 2000 Malawi DHS (Malawi National Statistical Office and ORC Macro 2001). We use this example for two reasons. First, we wanted the empirical example to be a DHS because these surveys are the largest available source of sibling history data; as of this writing, more than 150 DHS surveys in 60 countries have collected sibling histories over a period of about 30 years. Second, among DHSs, the 2000 Malawi DHS is preferable because it has low missingness in sibling reports and because its sample size of 13,220 women is close to average.\(^8\) Our analysis of the 2000 Malawi DHS sibling histories uses the *siblingsurvival* R package, which we created as a companion to this paper.\(^9\)

Figure 3 shows estimated death rates and confidence intervals for male and female death rates in Malawi over the seven-year period before interviews were conducted. For males, the individual and aggregate visibility estimates are qualitatively quite similar. For females, however, aggregate visibility estimates are systematically higher than individual visibility

\(^8\)The average DHS survey that included the sibling history module interviewed 14,224 women. (In DHS surveys, the sibling history module is typically asked only of women.)

\(^9\)The *siblingsurvival* package is open source and freely available: [https://www.github.com/dfeehan/siblingsurvival](https://www.github.com/dfeehan/siblingsurvival)
estimates, and these differences are larger than the estimated sampling variation. We discuss a possible explanation for this difference in more detail in the online appendix, section G.3.

**Variance Estimation**

Researchers need to be able to estimate the sampling uncertainty of death rate estimates, which can be calculated using Eqs. (7) and (13). We recommend that researchers use a resampling approach called the *rescaled bootstrap* to do so (Rao and Wu 1988; Rao et al. 1992). The rescaled bootstrap is appealing because (1) it accounts for the complex sampling design that is typical of surveys such as the DHS, (2) it enables researchers to use a single approach to estimating the sampling uncertainty for death rates and for other quantities (such as the internal consistency checks we introduce later), and (3) it has been successfully applied to other network reporting studies (e.g., Feehan et al. 2017). The *siblingsurvival* package uses the rescaled bootstrap to provide estimated sampling uncertainty for death rate estimates.

An alternate approach to estimating sampling uncertainty is to derive a mathematical expression that relates sampling variation to a function of study design parameters and population characteristics that are known or that can be approximated; this approach is discussed in section I of the online appendix, where we illustrate how linearization can be used to derive an approximate variance estimator for estimated death rates. There are also alternate resampling approaches to estimating sampling variance from survey data. For example, the Malawi DHS report uses a jackknife method to estimate sampling uncertainty for some quantities (but not adult death rates)\(^{10}\) (Malawi National Statistical Office and ORC Macro 2001). In practice, we expect the bootstrap approach discussed here to be most useful in empirical analyses, but future work could explore this topic in greater depth.

Figure 3 shows confidence intervals for estimated male and female death rates in Malawi over the seven-year period before interviews were conducted. To compare the amount of estimated sampling uncertainty for the aggregate and individual visibility estimates in Figure 3, we define the relative standard error of an estimate of \(\overline{M}_v\) to be \(\frac{SE(\overline{M}_v)}{\overline{M}_v}\), where \(SE(\overline{M}_v)\) is the rescaled bootstrap-estimated standard error. We then calculate the average of these relative standard errors across all age and sex groups for each estimator. The results suggest that the individual visibility estimator has slightly larger sampling variance than the aggregate visibility estimator (Table 1). This empirical finding is consistent with simulation results discussed in the online appendix, section J.

**Applying the Sensitivity Framework**

**Sensitivity to Invisible Deaths**—Equation (14) expresses the difference between the visible and total death rate in terms of two parameters: \(K\), an index for how different the visible and invisible death rates are; and \(p_{\alpha I}'\), the proportion of exposure that is invisible. Sibling history data cannot be used to determine the fraction of exposure that is invisible, \(p_{\alpha I}'\). However, we can try to approximate this quantity by taking advantage of the fact that we

---

\(^{10}\)The Malawi DHS report does not present estimated sampling errors for adult death rate estimates.
have a random sample of the frame population that is currently alive. We can use the sample
to estimate what fraction of respondents would be invisible to sibling histories at the time of
the survey—that is, to estimate the fraction of survey respondents who have no siblings in
the sampling frame.

Panel a of Figure 4 shows the estimated fraction of respondents to the 2000 Malawi DHS
who would not be visible to sibling histories. As our derivations reveal, it is this visibility
—and not sibship size per se—that matters for estimating death rates. Panel a gives an
approximate sense for the type of values that we might expect to see for \( \rho_{N(I)} \) by age. The
share of respondents who are invisible has a U-shaped relationship with age, reaching its
highest levels among the youngest and oldest survey respondents. This relationship is likely
due to the definition of the frame population, which includes women aged 15–49; age
groups close to the boundaries of the frame population will tend to have siblings who are
too old or too young to be included as respondents, reducing the visibility of these ages. At
worst, about 40% of women at ages 45–49 would be invisible to sibling histories; at best,
only about 15% of women ages 30–34 would be invisible to sibling histories.

What difference would this range of invisibility make to death rate estimates? The sensitivity
relationship in Eq. (14) reveals that the answer relies on understanding how different death
rates are in the invisible and visible populations. Panel b of Figure 4 illustrates by showing
the relative error that would result from (1) a range of differences between invisible and
visible death rates, from 20% higher death rates in the invisible population to 20% lower
death rates in the invisible population (\( K \) parameter, shown on the \( y \)-axis); and (2) a range
of different proportions of exposure that is invisible, from 15% of exposure invisible to 30%
of exposure invisible (\( \rho_{N(I)} \), shown on the \( x \)-axis). Even relatively large values for the two
parameters appear to result in modest relative errors.

**Reporting Errors**—Researchers have long been aware that reporting errors may affect the
quality of sibling survival estimates. We will illustrate two ways to assess the sensitivity of
aggregate visibility estimates to reporting errors. First, we will use Eq. (14) to assess how
much different levels of reporting error affect death rate estimates. Second, we will illustrate
how our network reporting framework leads to data quality checks that can be performed on
sibling history data.

**Analyzing the Impact of Reporting Errors Using the Sensitivity Framework**—
We will illustrate the sensitivity framework by focusing on aggregate visibility estimates
for brevity. The sensitivity framework in Eq. (14) shows that reporting errors will affect
the accuracy of aggregate visibility estimates through the ratio of two parameters: \( \gamma(F, N_{c}) \)
and \( \gamma(F, D_{c}) \). Importantly, in principle, it is possible to design a study that could measure
\( \gamma(F, N_{c}) \) and \( \gamma(F, D_{c}) \). In fact, some promising research on the sibling method to date
has compared sibling reports to ground-truth information at a demographic surveillance
site in southeastern Senegal (e.g., Helleringer, Pison, Kanté et al. 2014). Studies such
as that by Helleringer, Pison, Kanté et al. (2014) were designed to estimate somewhat
different reporting parameters from \( \gamma(F, D_{c}) \) and \( \gamma(F, N_{c}) \); thus, currently no direct evidence
about these parameters is available. However, to illustrate our sensitivity framework, we
can base some back-of-the-envelope calculations on the data reported by Helleringer and colleagues. Suppose that respondents never mistakenly count nonsiblings as siblings but that, on average, about 4% of living siblings in group $\alpha$ are omitted from respondents’ reports and about 9% of dead siblings in group $\alpha$ are omitted from respondents’ reports. Then $\gamma(F, N_\alpha) \approx 0.96$ and $\gamma(F, D_\alpha) \approx 0.91$. If these approximations were correct, Eq. (14) shows that the estimated death rates should be multiplied by about $0.95/0.91 = 1.05$ to adjust for reporting errors. In other words, under this scenario, the unadjusted aggregate visibility estimator produces estimates that have a relative error of about 5% because of imperfect reporting.\footnote{We do not recommend that this value derived from a back-of-the-envelope calculation be used in practice. We offer it as an illustration how our framework might be used when measurements of $\gamma(F, D_\alpha)$ and $\gamma(F, N_\alpha)$ are available.}

Internal Consistency Checks—A second strategy for assessing sensitivity to reporting errors is to perform data quality checks on sibling history data (e.g., Garenne and Friedberg 1997; Helleringer, Pison, Masquelier et al. 2014; Masquelier and Dutreuilh 2014; Rutenberg and Sullivan 1991; Stanton et al. 2000). We now show how our reporting framework enables us to introduce several new internal consistency checks that can be used to assess the accuracy of sibling reports. The idea is to use the network reporting framework to identify several quantities that can be estimated in two different ways using independent subsets of the sibling history data (e.g., Feehan and Cobb 2019). If reporting is highly accurate, then we expect these independent estimates to agree; when these independent estimates are very different, that suggests that there may be considerable amounts of reporting error.

We base our internal consistency checks on reports about siblings’ ages. For a particular age—say, 30—the symmetry of the sibling relation guarantees that in a census of the frame population with perfect reporting,

\[
\text{Number of connections people in } F \text{ aged 30 report to siblings in } F \text{ not aged 30} = \text{Number of connections people in } F \text{ not aged 30 report to siblings in } F \text{ aged 30}.
\]

(16)

The quantity on the left side of Eq. (16) can be estimated from the survey respondents aged 30, and the quantity on the right side of Eq. (16) can be estimated from all of the survey respondents who are not aged 30. These two estimates can then be compared; if reporting is accurate, then we expect the estimates to agree.

Formally, for age $\alpha$, we write $\gamma(F_{\alpha}, F_{\alpha})$ for the total reported connections from respondents aged $\alpha$ to siblings who are in $F$ but not aged $\alpha$; similarly, we write $\gamma(F_{\alpha}, F_{\alpha})$, for the total reported connections from respondents not aged $\alpha$ to siblings who are in $F$ and who are aged $\alpha$. In theory, these are the same quantity. From sibling history data, we can independently estimate $\gamma(F_{\alpha}, F_{\alpha})$ and $\gamma(F_{\alpha}, F_{\alpha})$ and then determine the similarity of these two independent estimates of the same quantity by calculating $\Delta$:

\[
\Delta_{\alpha} = \gamma(F_{\alpha}, F_{\alpha}) - \gamma(F_{\alpha}, F_{\alpha})
\]
When the two estimates agree, $\Delta \alpha$ is close to 0. If there is considerable reporting error, then $\Delta \alpha$ can be very different from 0.

Figure 5 illustrates this idea by showing internal consistency checks for each age from 15 to 49 from the 2000 Malawi DHS sibling histories (Malawi National Statistical Office and ORC Macro 2001). Each point shows the difference between two independent estimates for the same quantity. If these independent estimates agreed perfectly, they would all lie on the horizontal $y = 0$ line. The confidence intervals capture estimated sampling variation. Most of the confidence intervals include 0, suggesting that reports are internally consistent. However, the figure also suggests some misreporting, particularly between ages 20 and 25.

In general, we expect that such plots or quantitative summaries of internal consistency checks will be a useful way for researchers to assess the face validity of sibling history data. As we describe later, these internal consistency relationships could also form the basis for developing model-based approaches to analyzing sibling histories.

**Recommendations for Practice**

**Should the Respondent Be Included in the Denominator?**

The sibling survival literature has debated whether or not respondents should be included in the sibling reports. The concern is that respondents are, by definition, alive; thus, it seems possible that including respondents will bias estimated death rates downward (Masquelier 2013; Trussell and Rodriguez 1990). The estimates published in DHS reports, which use the aggregate visibility estimator, do not include the respondent. Gakidou and King (2006) argued that respondents should be included in sibling reports, but Masquelier (2013) disputed this notion.

In our framework, deciding whether to include respondents in the denominator of the estimator amounts to defining the visible and invisible populations. Because sibling history methods estimate a visible death rate, the appropriate question is, Which visible population’s death rate is more likely to be a good estimate for the total population death rate, $M_\alpha$?

To address this question, we analyze the difference between the two visible populations in section G.1 of the online appendix. Recall that $M_\alpha V = \frac{D_\alpha V}{N_\alpha V}$ is the visible death rate when respondents are excluded from sibling histories, and let $M_\alpha V' = \frac{D_\alpha V'}{N_\alpha V'}$ be the visible population when respondents are included in the reports. In section G.1 of the online appendix, we show that the relationship between $M_\alpha V$ and $M_\alpha V'$ can be written as

$$M_\alpha V' = \frac{D_\alpha V' - C}{N_\alpha V' - C},$$

where $C$ is the number of people who contribute exposure and who have visibility of exactly 1 when respondents are included in the denominator; that is, $C$ is the number of people in the population who are in group $\alpha$, are eligible to respond to the survey, and who have no siblings who would be eligible to respond to the survey. $C$ will tend to be bigger, inducing a
bigger difference between $M'_{\alpha V}$ and $M_{\alpha V}$, when (1) there is more overlap between the group $\alpha$ and the frame population; and (2) visibilities tend to be smaller, meaning that more people have a visibility exactly equal to 1 when respondents are included in the denominator.

Equation (17) reveals that the decision to include respondents in reports will move some people into the denominator, but it can never move anyone into the numerator.

A model developed in section G.2 of the online appendix shows that, in a simple situation in which everyone has the same probability of dying, it is most natural to exclude respondents from sibling reports. Under the model, when respondents are excluded, the invisible and visible populations have the same death rate, and that death rate can be estimated from sibling reports. Including respondents, on the other hand, induces a difference between the death rate in the visible and invisible populations (even though every individual faces the same probability of death). Thus, our model suggests that excluding respondents from reports is preferable, at least in the simple world it describes. This conclusion agrees with the earlier modeling work of Trussell and Rodriguez (1990), which also argued that respondents should be excluded from reports.

However, these are suggestive results: without additional information, there is no way to be certain that $M'_{\alpha V}$ or $M_{\alpha V}$ will produce a better approximation to the population death rate, $M_\alpha$, in a given population. The estimates presented here exclude respondents from reports. The full derivations of all of our estimators in sections F and E of the online appendix cover both including and not including the respondent in the denominator. Researchers who wish to include respondents in the denominator of the death rate can find the appropriate estimators there.

**Aggregate Versus Individual Visibility Estimator**

Our analysis suggests that both the individual and aggregate visibility estimators have strengths and limitations. Aggregate visibility estimates are typically based on the assumption that the visibility of deaths and the visibility of exposure are equal. The individual visibility estimator avoids this assumption altogether; thus, when no information about adjustment factors is available, we recommend using the individual visibility estimator. In section G.3 of the online appendix, we analyze the difference in aggregate and individual visibility estimates for Malawi, and we show that it is likely that this visibility assumption explains most of the difference between death rate estimates for females in Figure 3.

Table 1 and the simulation study in section J of the online appendix, however, suggest that the individual visibility estimator has slightly higher sampling variance than the aggregate visibility estimator. We hope that future research will continue to systematically compare the aggregate and individual visibility estimators; in the meantime, our view is that the individual visibility estimator’s relatively small loss in precision is a price worth paying to avoid having to make assumptions about the visibility of deaths and exposure.

Another disadvantage of the individual visibility estimator comes from comparing the aggregate sensitivity framework (Eq. (14)) with the individual sensitivity framework (Eq.
This comparison reveals that the quantities that would be needed to adjust the individual visibility estimator for reporting error are much more complex than the analogous quantities needed to adjust the aggregate visibility estimator. Thus, if data that can be used to estimate adjustment factors become more widely available, then we expect the relative appeal of the aggregate visibility estimator to increase.

To recap, we recommend excluding respondents from reports. We also recommend using the individual visibility estimator in the absence of any empirical estimates for adjustment factors. However, we expect empirical information about adjustment factors to be much easier to collect for the aggregate visibility estimator. Thus, as empirical information about adjustment factors becomes available, we expect the aggregate visibility estimator to be more attractive. In all cases, we recommend that researchers who produce sibling history-based estimates use the sensitivity frameworks to assess how estimated death rates are affected by the assumptions used to produce them.

**Discussion and Conclusion**

We showed how sibling history data can be understood as a type of network reporting. We explained how to derive network-based estimators for adult death rates, how to devise internal consistency checks, and how to understand how sensitive death rate estimates can be to the different conditions on which the estimators rely. We illustrated with an empirical example, based on our freely available R package *siblingsurvival*, and we outlined several recommendations for practitioners who wish to estimate death rates from sibling histories.

We see several important avenues for future research. Methodologically, a deeper comparison between the individual and aggregate visibility estimators would be useful. In particular, analytic results could help better explain our empirical finding that the individual visibility estimator has somewhat higher sampling variance than the aggregate visibility estimator. This analysis could also produce insights that might be useful for designing future data collection.

Our results also suggest next steps for developing models for death rate estimates based on sibling histories. This paper focused on design-based estimators for death rates using sibling histories. Future research can use these design-based estimators as the starting point for developing model-based estimators. For example, the internal consistency checks that we discussed could form the basis for model-based adjustments of sibling reports (see McCormick et al. (2010) for a similar approach that has been developed in the context of aggregate relational data). Our framework also offers a natural way to think about how to pool information across countries and periods.

Collecting more information on sibling reports from settings where gold-standard adult death rates are available is crucial; the Helleringer, Pison, Kanté et al. (2014) study offers a useful template for the type of study design that could help produce more information. This type of study investigates the properties of sibling history reports in small areas where a gold-standard underlying truth about adult death rates is available. Data collected in this way can produce the information needed to estimate the adjustment factors in the individual and
aggregate visibility sensitivity relationships. Combined with the framework introduced here, estimates for these adjustment factors could provide a principled way to adjust national-level sibling survival estimates from surveys such as the DHS, relaxing the conditions required for the estimates to be accurate.

Our analysis focused on death rate estimates for the period immediately preceding the survey. In principle, the estimators discussed here could also be used for more distant periods. However, the assumptions—although mathematically the same—presumably get stronger further into the past. Future work could investigate this topic in more depth.

Our framework can also be applied to other demographic estimation techniques related to sibling survival, such as methods in which people report about their parents, spouses, or children (Moultrie et al. 2014; United Nations Population Division 1983). For example, in settings with complex kinship structures that arise from social phenomena, such as polygamy or remarriage, our framework would enable exploring definitions of sibship that go beyond those people born to the same mother as the survey respondent. More generally, ideas from the sibling survival literature can be used to develop new methods for collecting data that have the potential to overcome some of the limitations of sibling histories. Feehan et al. (2017) explored how reports about two social network relationships could form the basis for death rate estimates at adult ages (see also Feehan et al. 2016). Future research could continue to explore how to collect reports about more general types of relations, such as broader kin relationships or other types of social networks, with the goal of producing information that is timely and accurate enough to estimate adult death rates. Future research could also explore how sibling reports might be used to estimate quantities other than adult mortality.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

For helpful feedback on earlier versions of the manuscript, the authors thank the participants in the 2018 Formal Demography Workshop at UC Berkeley, participants in the 2018 annual meeting of the Population Association of America session “Social capital and older adults in developing countries,” and Stephane Helleringer. The authors also thank the Berkeley Population Center, which is funded by the Eunice Kennedy Shriver National Institute of Child Health and Development (P2C HD 073964) and the Berkeley Center for the Economics and Demography of Aging, which is funded by the National Institute on Aging (5P30AG012839).

**References**

AbouZahr C, de Savigny D, Mikkelsen L, Setel PW, Lozano R, & Lopez AD (2015). Towards universal civil registration and vital statistics systems: The time is now. Lancet, 386, 1407–1418. [PubMed: 25971217]

Bernard HR, Johnsen EC, Killworth PD, & Robinson S (1989). Estimating the size of an average personal network and of an event subpopulation. In Kochen M (Ed.), The small world (pp. 159–175). Norwood, NJ: Ablex Publishing.

Brass W (1975). Methods for estimating fertility and mortality from limited and defective data. Chapel Hill, NC: Laboratories for Population Statistics.
Corsi DJ, Neuman M, Finlay JE, & Subramanian SV (2012). Demographic and Health Surveys: A Profile. International Journal of Epidemiology, 41, 1602–1613. [PubMed: 23148108]

Fabic MS, YoonJoung Choi Y, & Bird S (2012). A systematic review of Demographic and Health Surveys: Data availability and utilization for research. Bulletin of the World Health Organization, 90, 604–612. [PubMed: 22893744]

Feehan DM (2015). Network reporting methods (Doctoral dissertation). Program in Population Studies, Princeton University, Princeton, NJ. Retrieved from http://arks.princeton.edu/ark:/88435/dsp014m90dx89q

Feehan DM, & Cobb C (2019). Using an online sample to learn about an offline population. Demography, 56, 2377–2392. [PubMed: 3179232]

Feehan DM, Mahy M, & Salganik MJ (2017). The network survival method for estimating adult mortality: Evidence from a survey experiment in Rwanda. Demography, 54, 1503–1528. [PubMed: 28741073]

Feehan DM, & Salganik MJ (2016a). Generalizing the network scale-up method: A new estimator for the size of hidden populations. Sociological Methodology, 46, 153–186. [PubMed: 29375167]

Feehan DM, & Salganik MJ (2016b). Surveybootstrap: Tools for the bootstrap with survey data. Available from https://CRAN.R-project.org/package=surveybootstrap

Feehan DM, Umubyeyi A, Mahy M, Hladik W, & Salganik MJ (2016). Quantity versus quality: A survey experiment to improve the network scale-up method. American Journal of Epidemiology, 183, 747–757. [PubMed: 27015875]

Gakidou E, & King G (2006). Death by survey: Estimating adult mortality without selection bias from sibling survival data. Demography, 43, 569–585. [PubMed: 17051828]

Garenne M, & Friedberg F (1997). Accuracy of indirect estimates of maternal mortality: A simulation model. Studies in Family Planning, 28, 132–142. [PubMed: 9216033]

Garenne M, Sauerborn R, Nougatara A, Borchert M, Benzler J, & Diesfeld J (1997). Direct and indirect estimates of maternal mortality in rural Burkina Faso. Studies in Family Planning, 28, 54–61. [PubMed: 9097386]

Graham W, Brass W, & Snow RW (1989). Estimating maternal mortality: The sisterhood method. Studies in Family Planning, 20, 125–135. [PubMed: 2734809]

Helleringer S, Pison G, Kanté AM, Duthé G, & Andro A (2014). Reporting errors in siblings’ survival histories and their impact on adult mortality estimates: Results from a record linkage study in Senegal. Demography, 51, 387–411. [PubMed: 24493063]

Helleringer S, Pison G, Masquelier B, Kanté AM, Douillot L, Duthé G,… Delaunay V (2014). Improving the quality of adult mortality data collected in demographic surveys: Validation study of a new siblings’ survival questionnaire in Niakhar, Senegal. PLoS Med, 11(5), e1001652. 10.1371/journal.pmed.1001652 [PubMed: 24866715]

Hill K, El Arifeen S, Koening M, Al-Sabir A, Jamil K, & Ragers H (2006). How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches. Bulletin of the World Health Organization, 84, 173–180. 10.2471/blt.05.027714 [PubMed: 16583075]

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

Malawi National Statistical Office, & ORC Macro. (2001). Malawi Demographic and Health Survey 2000. Zomba, Malawi: National Statistical Office.

Masquelier B (2013). Adult mortality from sibling survival data: A reappraisal of selection biases. Demography, 50, 207–228. [PubMed: 23055235]

Masquelier B, & Dutreuilh C (2014). Sibship sizes and family sizes in survey data used to estimate mortality. Population (English Edition), 69, 221–238.

McCormick TH, Salganik MJ, & Zheng T (2010). How many people do you know? Efficiently estimating personal network size. Journal of the American Statistical Association, 105, 59–70. [PubMed: 23729943]

Mikkelsen L, Phillips DE, AbouZahr C, Setel PW, de Savigny D, Lozano R, & Lopez AD (2015). A global assessment of civil registration and vital statistics systems: Monitoring data quality and progress. Lancet, 386, 1395–1406. [PubMed: 25971218]
Moultrie TA, Dorrington RE, Hill AG, Hill K, Timaeus IM, & Zaba B (2014). Tools for demographic estimation. Paris, France: International Union for the Scientific Study of Population. Retrieved from http://demographicestimation.iussp.org/

Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD, & Murray CJL (2010). Measuring adult mortality using sibling survival: A new analytical method and new results for 44 countries, 1974–2006. PLoS Medicine, 7(4), e1000260. 10.1371/journal.pmed.1000260 [PubMed: 20405004]

Rao JNK, & Wu CFJ (1988). Resampling inference with complex survey data. Journal of the American Statistical Association, 83, 231–241.

Rao JNK, Wu CFJ, & Yue K (1992). Some recent work on resampling methods for complex surveys. Survey Methodology, 18, 209–217.

Reniers G, Masquelier B, & Gerland P (2011). Adult mortality in Africa. In Rogers RG & Crimmins EM (Eds.), International handbook of adult mortality (pp. 151–170). Dordrecht, the Netherlands: Springer.

Rutenberg N, & Sullivan JM (1991). Direct and indirect estimates of maternal mortality from the sisterhood method. In Proceedings of the Demographic and Health Surveys World Conference (Vol. 3, 1669–1696). Columbia, MD: IRD/Macro International.

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

Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S, & AbouZahr C (2007). A scandal of invisibility: Making everyone count by counting everyone. Lancet, 370, 1569–1577. [PubMed: 17992727]

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

Stanton C, Abderrahim N, & Hill K (2000). An assessment of DHS maternal mortality indicators. Studies in Family Planning, 31, 111–123. http://onlinelibrary.wiley.com/doi/10.1111/j.1728-4465.2000.00111.x/abstract [PubMed: 10907277]

Timaeus IM, & Jasseh M (2004). Adult mortality in sub-Saharan Africa: Evidence from Demographic and Health Surveys. Demography, 41, 757–772. http://www.springerlink.com/index/A2023R3756356V92.pdf [PubMed: 15622953]

Trussell J, & Rodriguez G (1990). A note on the sisterhood estimator of maternal mortality. Studies in Family Planning, 21, 344–346. [PubMed: 2075625]

United Nations Population Division. (1983). Manual X. Indirect techniques for demographic estimation (Report). New York, NY: United Nations.

United Nations Population Division. (2020). WPP2017: World Population Prospects 2017. New York, NY: United Nations. Retrieved from https://CRAN.R-project.org/package=wpp2017
Fig. 1.
Framing sibling survival as a network reporting estimator. Panel a shows a population connected through a sibship network. Because the sibling relation is transitive, the network is composed entirely of cliques. Panel b shows the network for a single sibship. Panel c displays a bipartite reporting graph for a single sibship.
Fig. 2.
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 female death rates 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 Eq. (14). These results are for the aggregate estimator that does not include the respondent.
Fig. 3.
Aggregate and individual visibility estimates for log-adult death rates from the Malawi 2000 DHS sibling histories. Estimates are for the seven years before the interview. Confidence bands show sampling uncertainty, accounting for the complex sample design; they were produced using the rescaled bootstrap method (Feehan and Salganik 2016b; Rao and Wu 1988). The aggregate visibility point estimates are the same as the adult death rate estimates presented in the Malawi DHS report (Malawi National Statistical Office and ORC Macro 2001).
Fig. 4.
Sensitivity to invisible deaths. Panel a shows the estimated fraction of respondents to the Malawi 2000 DHS who would not be visible in sibling history data. Panel b presents an illustration of the relative error in using the visible death rate $M^V$ as an estimate for the total death rate $M^T$. The proportion of exposure that is invisible, $p_{N\alpha}^I$, varies along the x-axis; the relationship between the visible and invisible death rates, captured by the parameter $K$ (Eq. (40)), varies along the y-axis. The colors show the percentage relative error; so if 20% of the population’s exposure is invisible ($p_{N\alpha}^I = .2$) and the invisible death rate is 10% higher than the visible death rate ($K = 1.1$), the relative error is about 2%. Relative error increases as $K$ gets further away from 1 and as $p_{N\alpha}^I$ increases.
Fig. 5.
Internal consistency checks using sibling reports from the 2000 Malawi DHS sibling histories. Each point shows the difference between two independent estimates for the same quantity. If these independent estimates agreed perfectly, they would all lie on the horizontal $y = 0$ line. Most of the checks’ confidence intervals include 0, suggesting that reports are internally consistent. However, the figure also suggests that there is some misreporting, particularly between ages 20 and 25. Confidence intervals show sampling uncertainty, accounting for the complex sample design; they were produced using the rescaled bootstrap method (Feehan and Salganik 2016b; Rao and Wu 1988).
Table 1

Comparison in average relative standard error for aggregate and individual visibility estimates of death rates across all ages for females and males

| Estimator          | Females | Males  |
|--------------------|---------|--------|
| Aggregate Visibility | 0.09    | 0.093  |
| Individual Visibility | 0.11    | 0.108  |