



# HHS Public Access

Author manuscript

*Demography*. Author manuscript; available in PMC 2015 July 01.

Published in final edited form as:

*Demography*. 2014 April ; 51(2): 387–411. doi:10.1007/s13524-013-0268-3.

## Reporting Errors in Siblings' Survival Histories and Their Impact on Adult Mortality Estimates: Results From a Record Linkage Study in Senegal

**Stéphane Helleringer,**

Columbia University, Mailman School of Public Health, 60 Haven Avenue, New York, NY 10032 USA, sh2813@mail.cumc.columbia.edu

**Gilles Pison,**

Institut National d'Études Démographiques, 133 Boulevard Davout, 75980 Paris Cedex 20, France

**Almamy M. Kanté,**

Columbia University, Mailman School of Public Health, 60 Haven Avenue, New York, NY 10032, USA

Ifakara Health Institute, PO Box 78373, Dar es Salaam, Tanzania

**Géraldine Duthé, and**

Columbia University, Mailman School of Public Health, 60 Haven Avenue, New York, NY 10032, USA

Institut National d'Études Démographiques, 133 Boulevard Davout, 75980 Paris Cedex 20, France

**Armelle Andro**

Institut National d'Études Démographiques, 133 Boulevard Davout, 75980 Paris Cedex 20, France

### Abstract

Estimates of adult mortality in countries with limited vital registration (e.g., sub-Saharan Africa) are often derived from information about the survival of a respondent's siblings. We evaluated the completeness and accuracy of such data through a record linkage study conducted in Bandafassi, located in southeastern Senegal. We linked at the individual level retrospective siblings' survival histories (SSH) reported by female respondents ( $n = 268$ ) to prospective mortality data and genealogies collected through a health and demographic surveillance system (HDSS).

Respondents often reported inaccurate lists of siblings. Additions to these lists were uncommon, but omissions were frequent: respondents omitted 3.8 % of their live sisters, 9.1 % of their deceased sisters, and 16.6 % of their sisters who had migrated out of the DSS area. Respondents underestimated the age at death of the siblings they reported during the interview, particularly among siblings who had died at older ages ( $> 45$  years). Restricting SSH data to person-years and events having occurred during a recent reference period reduced list errors but not age and date

errors. Overall, SSH data led to a 20 % underestimate of  $_{45q15}$  relative to HDSS data. Our study suggests new quality improvement strategies for SSH data and demonstrates the potential use of HDSS data for the validation of “unconventional” demographic techniques.

## Keywords

Adult mortality; Data quality; Reporting errors; Siblings’ survival histories; Senegal

---

## Introduction

With child mortality declining (Hill et al. 2012; Lozano et al. 2011), an increasing proportion of deaths in low- and middle-income countries (LMICs) will occur at adult ages in the next decades (Mathers and Loncar 2006). In LMICs, adult mortality rates may be high because of HIV/AIDS (Feeney 2001; Stover 2004), maternal mortality (Hogan et al. 2010), injuries (Garrib et al. 2011; Lagarde 2007), noncommunicable diseases (Duthé and Pison 2008; Mathers et al. 2009), or war (Obermeyer et al. 2008). Adult deaths threaten the livelihood of entire families (Yamano and Jayne 2004), prevent the schooling and development of children (Case and Ardington 2006; Evans and Miguel 2007), limit support to the elderly (Ainsworth and Dayton 2003; Kautz et al. 2010), and precipitate the closure of businesses (Chao et al. 2007). Each year, a significant proportion of all development assistance for health is thus allocated to interventions aimed at preventing premature deaths in adulthood (e.g., Goosby 2012). The ultimate impact of these interventions remains largely unknown, however, because few of the target countries have vital registration (VR) systems that permit measuring adult mortality accurately (Setel et al. 2007).

In countries with limited VR, estimates of adult mortality produced by international organizations are often extrapolated from child mortality rates using model life tables (e.g., Murray et al. 2003; United Nations 2010). Other approaches to obtaining national estimates of adult mortality consist of “unconventional” techniques (Hill et al. 2005), which use information on a respondent’s close relatives collected during censuses and surveys (Graham et al. 1989; Timaeus 1991b). Such data include questions on the number of deaths in a respondent’s household over the previous 12 months (Hill et al. 2006) and the survival of a respondent’s spouse, parents, or siblings (Timaeus 1991a, 1991b). Collecting siblings’ survival histories (SSH), in particular, is becoming increasingly popular. Demographic and Health Surveys (DHS) conducted in sub-Saharan and Asian countries now frequently include SSH modules, as did the WHO World Health Survey (Reniers et al. 2011).

In SSH, respondents are asked to list all siblings born to their biological mother by birth order and then to provide information about each sibling’s sex, survival status, and current age or age at death, as well as time elapsed since death if the sibling is deceased. These questions produce a “siblings’ data set,” in which adult mortality rates can be estimated directly because dates of birth and (possibly) death can be calculated for each reported sibling.<sup>1</sup>

---

<sup>1</sup>In such analyses, “adulthood” is most commonly defined as the age range 15–49 years or 15–59 years. The most commonly used indicators of adult mortality are thus  $_{35q15}$  or  $_{45q15}$ .

## Sample Selection Bias and Reporting Errors in Siblings' Survival Histories

SSH are relatively inexpensive to collect, but they suffer from several potential drawbacks. First, because adult deaths are rare events compared with births or child deaths,<sup>2</sup> SSH may still yield sample sizes that are too small to produce reliable estimates (Hill et al. 2006; Timaeus and Jasseh 2004).

Second, SSH are prone to sample selection bias (Trussell and Rodriguez 1990) because SSH surveys are based on random samples of individuals who have survived until the survey date rather than on random samples of individuals who were at risk of dying during the  $n$  years prior to the survey (Gakidou and King 2006). If adult survival is positively correlated with the size of one's sibship, then members of high-mortality sibships (e.g., those with zero survivors) will be underrepresented in SSH and adult mortality rates will be underestimated (Trussell and Rodriguez 1990). Analytical corrections for sample selection bias have been proposed (Gakidou and King 2006), but they may artificially inflate estimates of  ${}_{45}q_{15}$  (Masquelier 2013).

Third, SSH estimates may be affected by reporting errors. These occur when SSH reported by survey respondents differ from the "true" survival experience of their maternal siblings. Demographers have generally equated reporting errors in SSH with omissions of deceased siblings by survey respondents (e.g., Obermeyer et al. 2010), but reporting errors may be more complex. We distinguish four main types of reporting errors.

*List errors* occur when the reported list of maternal siblings differs from the true sibship of a respondent. Respondents may not list all their true maternal siblings (omission) or may include others who are not their true maternal siblings (addition). Omissions may happen, for example, when a respondent has not known one of her siblings, when s/he has not seen one of her siblings for an extended period of time (possibly because of migration), or when a sibling died long ago. Omissions may also happen because of interviewer behavior. SSH are highly repetitive questionnaires, with the same set of 5–10 questions being asked about each of sometimes more than 10 siblings. Thus, to save time, interviewers may be inclined to omit a subset of the respondent's siblings. Additions to siblings' lists may occur in areas where extended families coreside or where child fostering and orphanhood are widespread. In such settings, respondents are likely to have paternal half-siblings or cousins, which they may erroneously include in SSH.

*Vital status errors* occur when a respondent reports that one of her live siblings is dead at the time of the SSH survey, or when she reports that a deceased sibling is alive at the time of the survey. This may happen, for example, if a respondent has not been in touch with one of her sibling(s) for some time or may not be aware of health conditions or accidents having affected that sibling.

---

<sup>2</sup>The primary objectives of a Demographic and Health Survey often focus on estimating the total fertility rate and infant/child mortality rates in a given country. The sample size calculations for each DHS thus aim to achieve a given level of precision for these indicators. Births and infant/child deaths are significantly more common than adult deaths in sub-Saharan populations, however. As a result, even if a respondent reports about several adult siblings during SSH, SSH may still yield an insufficient number of reports of deaths to enable specific analyses.

*Age errors* refer to respondents' inaccurate reporting of the current age or age at death of one of their true maternal siblings.

*Date errors* refer to respondents' inaccurate reporting of the year in which their sibling(s) died. They are relevant in SSH analysis because analysts often restrict the siblings' data set to reports of deaths and person-years occurring during a reference period of a few years before the survey (e.g., 6–8 years). This practice has been adopted because SSH data are believed to be more complete in that reference period (Timaeus and Jasseh 2004). Date errors are important because data on dates of siblings' death have been used to investigate time trends in adult mortality (de Walque and Filmer 2013), estimate the effect of short-term mortality crises (e.g., de Walque and Verwimp 2010), and evaluate the benefits of public programs (Bendavid et al. 2012).

### Potential Effects of Reporting Errors on Adult Mortality Estimates

Reporting errors have potentially complex effects on SSH estimates of adult mortality. List errors affect estimates when the likelihood of omission/addition depends on the vital status of the sibling (Timaeus and Jasseh 2004). For example, if respondents are more likely to omit their siblings who died at adult ages than their live siblings, adult mortality rates will be underestimated. The direction of bias resulting from vital status errors depends on whether respondents are more likely to report deceased siblings as alive, or vice versa. Compared with list errors, however, vital status errors likely have a disproportionate impact on mortality estimates because they affect only the numerator of mortality rates.<sup>3</sup>

Errors about the current age of live siblings may affect the denominator of mortality rates. For example, if respondents underestimate the current ages of their siblings, then fewer person-years lived among these siblings will be counted, and (all else being equal) adult mortality rates will be overestimated.

The effects of date and age errors about a sibling's death depend on (1) the magnitude of these errors, and (2) the width of the age interval and the length of the reference period used in calculating adult mortality rates. We illustrate this point for calculations of  ${}_{45}M_{15}$  over the eight years prior to the survey with the Lexis diagram in Fig 1. Sibling deaths can be misplaced along the vertical axis (age error; e.g.,  $D_1$  and  $D_2$ ), along the horizontal axis (date error; e.g.,  $D_3$  and  $D_4$ ), or along both axes (e.g.,  $D_5$ ) of the Lexis diagram. When misreported deaths remain within the reference area of the diagram (e.g.,  $D_1^{SSH,1}$  and  $D_3^{SSH,1}$ ), age and date errors affect the count of person-years lived by deceased siblings (i.e., the denominator of  ${}_{45}M_{15}$ ) but not the number of events (i.e., the numerator of  ${}_{45}M_{15}$ ). In Fig. 1, for example, SSH data about  $D_1$  and  $D_3$  yield fewer person-years than expected but not fewer deaths. On the other hand, when misreported deaths are shifted outside the reference area (e.g.,  $D_1^{SSH,2}$  and  $D_3^{SSH,2}$ ), age and date errors also affect the numerator of adult mortality rates. Deaths, which are shifted outside the reference area by age and/or date

<sup>3</sup>For example, in a cohort of 20 women in which 10 deaths occur, all at age 35, the true death rate at adult ages  ${}_{45}M_{15}$  is  $10 / [(10 \times 20) + (10 \times 45)] = 15.4$  per 1,000. A list error in which a respondent omits one deceased sister would yield  ${}_{45}M_{15} = 9 / [(9 \times 20) + (10 \times 45)] = 14.3$  per 1,000, whereas a vital status error in which a respondent misclassifies one deceased sister as alive would yield  ${}_{45}M_{15} = 9 / [(9 \times 20) + (11 \times 45)] = 13.3$  per 1,000.

errors, can be “replaced” by misreporting of more-distant deaths (e.g.,  $D_4, D_5$ ) or deaths having occurred outside of the adult age range (e.g.,  $D_2, D_5$ ).

### Current Diagnostics of Reporting Errors in Siblings’ Survival Histories

To detect reporting errors, demographers perform internal consistency checks on siblings’ data sets (Stanton et al. 2000). These verify that reported SSH data replicate trends in fertility and mortality known from other sources. For example, because the number of siblings of a respondent corresponds to her mother’s lifetime fertility, older respondents should report larger sibships than younger respondents in settings where fertility has been declining. Checks also include testing for imbalances in the sex ratio at birth (SRB) in reported sibships. In the 2010 Senegal DHS, the SRB was 108, thus suggesting a possible underreporting of sisters (ANSD and ICF 2012).

Other diagnostics have consisted of external comparisons between a test SSH data set—that is, whose quality is to be evaluated—and an independent reference data set. SSH data have been compared with model life tables, VR, census data on household deaths in the past 12 months, and orphanhood data (e.g., Feeney 2001). Most recently, several studies have compared the reporting of deaths in consecutive SSH surveys conducted several years apart in the same country (e.g., Obermeyer et al. 2010; Reniers et al. 2011). If SSH are collected in, say, 2005 and 2010, then adult deaths having occurred in 2004 should be reported both by respondents in the 2005 and the 2010 surveys, but recall should be more complete in the 2005 data set. Using this strategy, Reniers et al. (2011) estimated that the recall of deaths having occurred 6–9 years before a survey was only 83 % complete. Obermeyer et al. (2010) incorporated this approach into a regression-based method aimed at correcting estimates of  ${}_{45}q_{15}$  (the corrected sibling survival method, or CSS).

These diagnostics have important limitations, however. First, they frequently assume that respondents completely recall the recent (i.e., less than three years before a survey) events that have affected their siblings. This is the case, for example, of the CSS method, which assumes that a respondent’s recall of recent deaths is complete but subsequently declines linearly with time between the death and the survey. Studies in cognitive psychology suggest that this is a particularly strong assumption: even the most recent events can be forgotten, whereas more-distant events may be more easily recalled (Wixted 2004). Furthermore, some omissions of deaths may not be due to forgetting. Respondents may deliberately omit to report a death, for example, if it is due to a stigmatized cause, such as AIDS. Other omissions may be due to interviewers’ behaviors (as described earlier). SSH may thus also be incomplete for the most recent reporting periods, and current diagnostics may underestimate the extent of reporting errors in SSH.

Second, current diagnostics do not control for differences in sample universe between reference and tests data sets. As a result, they can identify only the joint effects of sample selection bias and reporting errors on SSH estimates. For example, a recent adult death may not be reported on a census question about household deaths in the past 12 months if the death precipitated household dissolution. It may, however, be reported in a SSH survey if the deceased has at least one surviving adult sibling residing elsewhere. In this context, differences in estimates of  ${}_{45}q_{15}$  between the reference (census) data set and the test SSH

data set can be attributed to (1) the incomplete coverage of both the reference and test data sets, and/or (2) reporting errors in both the reference and test data sets.

Finally, current diagnostics include only aggregate comparisons of adult mortality indicators, which do not permit identifying which type of reporting errors is most common in SSH. This is the case because different types of reporting errors have potentially similar and/or offsetting effects on adult mortality rates.<sup>4</sup> This is an important limitation because a detailed understanding of the different types of reporting errors that affect SSH is required to guide improvements in data collection and/or develop analytical adjustment strategies.

### An Alternative Approach to Diagnosing Reporting Errors

A more precise approach to diagnosing reporting errors requires linking the SSH report of the survival of a particular sibling to a record of that same sibling's survival in a high-quality reference data set (Fig. 2). Such record linkage studies have been used, for example, to estimate the quality of old-age mortality data collected in the United States (e.g., Hill et al. 2000; Preston et al. 1996).

For list errors in SSH to be ascertained in a record linkage study, individuals must be grouped into sibships in the reference data set. Unfortunately, most available data sets on adult mortality do not permit constituting reasonably complete lists of the maternal siblings of a SSH respondent. VR and census data, for example, are rarely available at the individual level because of confidentiality issues. When available, VR data include only the name of the mother on the birth certificate, which is not sufficient to establish a sibship list: a fairly large number of individuals who are not siblings may have a mother with the same name. Census/survey data on recent household deaths do not collect information on the biological mother of recently deceased household members. They thus permit identifying siblings only if they reside in the same household as their live mother at the time of the census/survey. Even then, reports of household relationships may be affected by “adoption bias”: that is, the caretaker of a household member may be mistakenly classified as his/her “mother” by the household informant.

### Using Health and Demographic Surveillance Data as Reference

Health and demographic surveillance systems (HDSS) constitute another potential source of information about adult mortality that could be used to evaluate SSH. HDSS consist of monitoring over time an entire population located in a small geographic area (Pison 2005). They include a baseline census, followed by continuous registration of demographic events (i.e., births, deaths, marriages, migrations) affecting this population. Event registration happens yearly or more frequently. HDSS interviewers visit every household and ask household informants to provide information on recent demographic events among household members (e.g., Oduro et al. 2012). In some HDSS, data collection relies on key

---

<sup>4</sup>For example, in a cohort of 20 women in which 10 deaths occur, all at age 35, the “true” death rate at adult ages  $45M_{15}$ , obtained from a complete and accurate reference data set, is  $10 / [(10 \times 20) + (10 \times 45)] = 15.4$  per 1,000. If one of these women is omitted by respondents during SSH (i.e., list error), but respondents make no age errors, then the estimated death rate  $45M_{15}$  would be  $9 / [(9 \times 20) + (10 \times 45)] = 14.3$  per 1,000. On the other hand, the estimated death rate  $45M_{15}$  would also be 14.3 per 1,000 if respondents reported all their deceased sisters (i.e., no list errors) but overestimated the age at death of their sisters by an average of five years (age errors).



informants at the community level (e.g., Jahn et al. 2007): a small number of village residents provide information on the demographic events that have occurred among households in their vicinity.

HDSS were initiated to remedy the lack of VR data in a number of sub-Saharan and South Asian countries (Sankoh et al. 2006). HDSS estimates of adult mortality may be affected by many factors, including being predominantly rural (Sankoh and Byass 2012), small sample sizes, and selective in- and out-migrations (Clark et al. 2007). HDSS nonetheless can help address limitations of other reference data sets because on one hand, the biological mother of each member of the population is often identified in HDSS data sets: that is, she is attributed a unique ID number. This is accomplished at the time of the initial census, at the time of birth, or when an individual migrates into the HDSS population. Sibship lists are then formed by looking up all the individual members of the HDSS population who have the same mother ID number. On the other hand, HDSS data also provide potentially more accurate information on vital status, ages, and dates of death of siblings than other data sources: they are collected prospectively, whereas other data sets are primarily retrospective.

In this study, we used data from the Bandafassi HDSS (Senegal) as a reference data set. We identified sibships included in that reference data set that could also be included in a SSH survey (hereafter, “reference sibships”). We then conducted a SSH survey (our “test data set”) among the adult female members of these reference sibships. After SSH data collection, we linked the reference and test data sets *at the sibling level* (Fig. 2). Finally, we measured the prevalence and correlates of each type of reporting error in the SSH data set and then compared measures of adult survival obtained from SSH and HDSS data.<sup>6</sup>

## Data

### The Bandafassi Health and Demographic Surveillance System

The Bandafassi HDSS is located in southeastern Senegal (Online Resource 1, section 1). With 12,770 residents (all rural) as of August 1, 2010, the area is divided in three ethnic groups (the Bedik, the Mandinka, and the Fulani) who live in separate villages. In the period 2000–2006, the total fertility rate was high (at about seven children per woman). Life expectancy at birth was low (<55 years) despite declines in child mortality (Desgrees du Lou et al. 1995; Kanté and Pison 2010). Out-migrations from the study area are frequent: adults often migrate to urban areas of Senegal or abroad for work. In-migrations are less common and happen primarily at the time of marriage (e.g., when a woman from a neighboring village marries a male resident of the HDSS area).

<sup>5</sup>VR data are also collected prospectively in theory, but events may be registered late (e.g., children are registered when they reach school age) so that VR data are affected by significant age/date errors.

<sup>6</sup>Several comparisons of SSH and HDSS data have already been conducted, but these have compared DSS data for small areas with SSH estimates from national data sets (Obermeyer et al. 2010). In these comparisons, differences in estimates of adult mortality between SSH and HDSS may have been due to idiosyncrasies of the local HDSS populations rather than issues with SSH data quality. Some comparisons were conducted within the same geographic areas but did not control for differences in sampling universe between HDSS and SSH data sets (e.g., Ngom et al. 1999). To the best of our knowledge, only one study comparing SSH and HDSS was based on a linkage design (Shahidullah 1995). This study, however, was limited to assessing the reporting of maternal deaths.

The Bandafassi HDSS started in 1970 with a census of the local population. This census was followed by detailed oral genealogies<sup>7</sup> of each population member (Pison 1987), which permit identifying with precision the maternal siblings of individuals who were already born at the time of the census.<sup>8</sup> During the census, innovative procedures were used to determine the age of population members whose exact year of birth was unknown (Pison 1980; Pison and Langaney 1985). Since the initial census, key informants in each village have been visited every year. During every visit, the lists of residents are updated; and births, marriages, migrations, and deaths that occurred since the last visit are registered. For births and in-migrations, information on the identity of the child's or the immigrant's biological parents is collected from the key informants. This information is used to update the list of maternal siblings of every population member.

The Bandafassi HDSS provides reference information on (1) lists of maternal siblings of a respondent, (2) age and vital status of each sibling at the time of the SSH survey, and (3) age and date of death of deceased siblings. We classified each sibling as either a child (i.e., individuals who have never reached age 15) or an adult (i.e., individuals who have ever reached age 15). The age at death of adult siblings was classified in three groups (<25 years old, 25–44 years old, and 45 years old). Deaths were classified as having occurred during an eight-year reference period (2002–2009<sup>9</sup>) or before that reference period. HDSS data also provide information on characteristics of siblings and sibships used in multivariate analyses (see the Results section), including a categorical variable indicating whether a sibling was still alive, dead, or lost to follow-up at the last HDSS visit; a dummy variable indicating whether a sibling has the same father as the SSH respondent; and a categorical variable identifying the ethnic group of the sibship.

### The 2010 SSH Survey

In 2010, we conducted a SSH survey among the population of the Bandafassi HDSS. Our key objective was to estimate the extent of reporting errors in SSH. A secondary objective entailed testing whether reported adult deaths were correctly classified by SSH respondents as pregnancy-related (Helleringer et al. 2013). Because SSH are most commonly collected among women during DHS, we included only female respondents. Unlike most DHS, which include respondents aged 15–49, we also included women aged 50–59. The sample of the SSH survey was not designed to be representative of the HDSS population but rather to control for differences in sampling universe between reference and test data sets. We thus reviewed HDSS data and identified sibships in which at least one known sister was alive at the time of the SSH survey and met eligibility criteria (reference sibships). To ensure adequate numbers of adult deaths in reported SSH, we focused on reference sibships in which at least one sister had died at age 15–59 in the Bandafassi HDSS since 1980.

---

<sup>7</sup>In oral genealogies, multiple informants are interviewed to confirm the existence and nature of kinship ties between population members. See, for example, Pison (1987), Quinlan and Hagen (2008), and Silagan (1986).

<sup>8</sup>The Bandafassi HDSS differs from other HDSS in that respect given that oral genealogies are not necessarily conducted at the time of the census in each HDSS. In other HDSS, the data on maternal sibships of individuals who were already present in the population at the time of the initial census may be missing or may be obtained from one single informant. Lists of maternal siblings may thus be less complete/accurate in other HDSS.

<sup>9</sup>The SSH survey took place before the 2010 HDSS data became available; as a result, the reference period and all results herein do not include deaths and person-years from 2010.



We then drew a list of all the female members of these reference sibships who were born between 1950 and 1995 and were potentially still alive at the time of the SSH survey. Among those, some were still resident of the Bandafassi HDSS area, and others had emigrated. All eligible HDSS residents were contacted for participation. Because recall of SSH may be affected by separation from one's siblings after migration, we also traced<sup>10</sup> and interviewed emigrants who had moved to Tambacounda or Kédougou, the two cities that are largest closest to Bandafassi. Migrants who had moved to other regions of Senegal were not contacted.

We found 573 deaths among women aged 15–59 in Bandafassi since 1980. Among those, we had no information on the mother of 40 deceased women. For 256 other deaths, there was no known eligible sister, according to the HDSS. Finally, the eligible sisters of 80 women who died between 1980 and 2010 could not be interviewed because they resided in regions not covered by the survey, they had temporarily migrated to small agricultural hamlets around Bandafassi (which were inaccessible by car because of floods), or they refused to participate in the SSH survey ( $n = 2$ ). We interviewed at least one of the sisters of 197 women who had died at adult ages between 1980 and 2010. In some instances, multiple members of the same sibship were interviewed. In total, there were 268 respondents in the SSH survey. Further details on the constitution of the SSH sample and survey response rates are available in sections 2 and 3, respectively, of Online Resource 1.

We used a SSH questionnaire similar to the 2005 Senegal DHS instrument.<sup>11</sup> It provides test data on lists of maternal siblings as well as the vital status, ages, and dates of death of these siblings. It also provides information on respondent characteristics used in multivariate analyses of reporting errors. These include a dummy variable indicating whether the respondent was interviewed in the migratory situations (i.e., in Tambacounda or Kédougou); a categorical variable indicating the age group of the respondent (<25 years old, 25–34 years old, 35–44 years old, and 45 years old); and a dummy variable indicating whether the respondent had ever attended school.

To link SSH reports of a sibling's survival to HDSS records of the survival of that same sibling, the basic SSH instrument was first augmented by questions about the last known residence of the sibling (including village and name of the head of compound), the name of the sibling's spouse(s), and other names/nicknames the sibling may be known by. To maximize comparability with the standard SSH instrument, these questions were asked for each sibling only after completion of the entire standard SSH module. Second, each SSH report was independently linked to the HDSS by two members of the study team.<sup>12</sup> We used HDSS information only about the expected sibship of a given respondent, including the names, sex, and residence of each sibling to establish linkages. We did not use HDSS information on vital status, age, and date of death because they constituted study outcomes.

---

<sup>10</sup>To trace migrants, we first conducted a short migration survey with a member of the last known HDSS residence of the migrant. This survey included information on destination, new address (if available), phone number, and other information (open-ended) that may be useful for migrant tracing.

<sup>11</sup>The full questionnaire, in French, is available online (<http://www.columbia.edu/~sh2813/BandafassiQuest.pdf>).

<sup>12</sup>Pairs of reviewers were randomly selected among the study's coauthors. Each team member was asked to select an ID number from the HDSS files for each SSH-reported sibling based on the available information.

Finally, we assessed the concordance of the two independently assigned linkages.<sup>13</sup> For discordant linkages, or linkages that were made by only one of the two team members, the linkage was reviewed by both team members and another investigator. Discrepancies were resolved. Full linkage results are presented in Online Resource 1, section 4. For simplicity, we call “matches” the SSH reports of siblings that were linked to a HDSS record for that sibling; we call “additions” the SSH report of siblings that could not be linked to a HDSS record; and we call “omissions” the siblings included in the DSS who were not reported during the SSH survey (Fig. 2).

## Results

### Descriptive Statistics

The characteristics of SSH survey respondents are presented in Table 1. Forty-five respondents (16.8 %) were migrants. Most respondents had never been to school (81.0 %). Two-thirds of respondents belonged to sibships in which more than one sister was interviewed. Basic indicators of data quality and internal consistency checks for SSH are available in section 5 of Online Resource 1.

### How Complete Are the Reference and Test Data Sets?

Table 2 presents the extent of overlap between the sibship lists obtained in SSH and HDSS data. A total of 993 unique sisters of all ages are listed in either the HDSS or SSH data sets, with 80.1 % appearing in both data sets, 10.8 % appearing only in HDSS, and 9.1 % appearing only in SSH. Only 62.9 % of children appeared in both data sets versus 88.2 % among adult sisters.<sup>14</sup>

Using capture-recapture techniques, we estimated the completeness of each data source for the reference sibships (Amstrup et al. 2005; Sutherland 2006). We considered that each sibling was first marked through HDSS and was then recaptured by SSH. We assumed that being marked in the HDSS data set did not affect the probability that a sibling would be recaptured during the SSH survey.<sup>15</sup> This yielded estimates of completeness of 89.8 % for HDSS and 88.1 % for SSH. However, we estimated that HDSS was 96.3 % complete at adult ages versus 91.3 % for SSH. Among sisters deceased at adult ages, the HDSS was virtually complete (99.1 %).

### List Errors in SSH Data

SSH respondents reported 626 adult sisters, among which 26 (4.2 %) could not be linked to the HDSS record of a woman from her maternal sibship (i.e., additions); 5.2% of adult sisters reported as alive were additions versus approximately 1% of adult sisters reported as

<sup>13</sup>The concordance was assessed by comparing the ID number attributed by each member.

<sup>14</sup>Some siblings were reported as adults by the SSH and recorded as children by the HDSS, and vice versa.

<sup>15</sup>In Bandafassi, different persons report demographic events during demographic surveillance and SSH: for example, key informants may be health workers posted in the rural villages or heads of compounds (predominantly older males), whereas SSH respondents were women aged 15–59. However, the HDSS and SSH data may not be entirely independent because, for example, key informants may have heard about the death of an individual by relatives of the deceased who may be interviewed during SSH surveys. Similarly, some deaths may be less likely to be reported in both data sources because of the cause of death or the location where the death took place. For example, if a DSS resident seeks health care in a city and dies while away, his/her death may be less likely to be reported both in the DSS and in the SSH data sets. Capture-recapture estimates thus likely overestimate the completeness of each data source.

deceased ( $p < .01$ ). Respondents omitted 3.8% of their adult sisters who were alive at the most recent HDSS visit (Table 3), 9.1% of their sisters who had died at adult ages, and 16.6% of their adult sisters who had migrated outside the HDSS area prior to the most recent HDSS visit ( $p < .01$ ). We conducted multivariate analyses of the likelihood of omitting an adult sister during SSH. In these analyses, sisters observed in the HDSS data set constituted the unit of analysis and were nested by respondent and sibship.<sup>16</sup> To account for this hierarchical structure, we used multilevel logistic regression models (Gelman and Hill 2007). We estimated the following model:

$$\ln\left(\frac{y_{i,r,m}}{1 - y_{i,r,m}}\right) = \beta'_1 \mathbf{X}_{i,r,m} + \beta'_2 \mathbf{X}_{r,m} + \beta'_3 \mathbf{X}_m + U_{r,m} + U_m \quad U_{r,m} : N(0, \sigma_{r,m}^2) \quad U_m : N(0, \sigma_m^2), \quad (1)$$

where  $y_{i,r,m}$  is a categorical variable coded 1 if sibling  $i$  of respondent  $r$  from sibship  $m$  is omitted in SSH, and 0 otherwise;  $\mathbf{X}_{i,r,m}$ ,  $\mathbf{X}_{r,m}$ , and  $\mathbf{X}_m$  are vectors of sister, respondent, and sibship characteristics, respectively; and  $\beta$  s are vectors of coefficients. The terms  $U_{r,m}$  and  $U_m$ , assumed to be normally distributed, refer to the respondent-level and sibship-level random effects, respectively. Results are presented in the rightmost column of Table 3. Even after we controlled for characteristics of sibships and respondents, sisters who were deceased at the time of the survey remained more likely to be omitted than live sisters (OR = 2.95), as were sisters lost to HDSS follow-up (OR = 11.6). Omissions were also more common among respondents interviewed in migratory situations (OR = 5.18), among younger respondents (<25 years old), and among sibships of the Bedik ethnic group (OR = 7.28). Finally, respondents were less likely to omit sisters who had the same father as them (OR = 0.19).

Omissions were less common among deaths having occurred during the reference period (2002–2009) than among deaths having occurred before the reference period (Fig. 3a;  $p < .01$ ). Omissions of deaths were also associated with age at death (Fig. 3b;  $p = .05$ ): they were most common among sisters who had died before age 25 (10.8 %) or at ages 45 and older (18.8 %).

### Vital Status Errors in SSH Data

Among matches between the HDSS and SSH data sets, only one adult sister was reported as alive in SSH but classified as deceased in HDSS, yielding a sensitivity<sup>17</sup> of 99.6 % for SSH reports of vital status. There were no SSH reports of death among adult sisters who were alive at the time of the last HDSS visit (specificity<sup>18</sup> = 100 %).

### Age Errors in SSH Data

SSH respondents underestimated the ages of their adult sisters. Among the 252 live matches, the average difference between the current age recorded in the HDSS and the SSH-reported value was 1.74 years (standard deviation (SD) = 4.6). Among the 264 deceased matches, the average difference was 3.79 years (SD = 7.93). The distributions of age errors were

<sup>16</sup>This is the case because multiple sisters were interviewed by sibship (see Table 1).

<sup>17</sup>Sensitivity refers to the proportion of deaths among adult sisters recorded in the HDSS who are reported as deceased in the SSH survey.

<sup>18</sup>Specificity refers to the proportion of alive sisters recorded in the HDSS who are reported as alive in the SSH survey.

overdispersed because of a small number of large errors, possibly resulting from linkage or interviewer errors.<sup>19</sup>

We analyzed the determinants of age errors using a multilevel model comparable to Eq. (1). The dependent variable was a continuous variable representing the difference between the HDSS-observed and SSH-reported values of the current age/age at death. In the model for errors in age at death, we also included a dummy variable taking a value of 1 if the death had occurred during the reference period (2002–2009). We considered only two levels in these models—siblings and sibships—because only a very small number of respondents had multiple adult female deaths in their sibships. The results from these analyses are reported in Table 4.

The largest age errors concerned the oldest adult sisters of respondents: those older than 45 ( $\beta = 6.28$  for deceased sisters,  $\beta = 4.22$  for living sisters). Migrants underestimated their live sisters' age more frequently than residents of the HDSS ( $\beta = 2.04$ ). Older respondents overestimated the age of their living sisters (e.g.,  $\beta = -2.37$  for respondents  $\geq 45$ ) but not the age at death of their deceased sisters. Finally, errors in reported age at death also increased among deaths having occurred *during* the reference period ( $\beta = 1.86$ ).

### Estimates of Adult Survival in Reference Sibships

We compared life-table estimates of adult survival among the reference sibships according to both HDSS and SSH data (see Fig. 4). We found that differences in estimates of adult survival between HDSS and SSH emerged primarily after age 50. For example,  ${}_{25}q_{15}$  was 0.5 in HDSS versus 0.47 in SSH;  ${}_{35}q_{15}$  was 0.61 in HDSS versus 0.59 in SSH; but  ${}_{45}q_{15}$  was 0.78 in HDSS versus 0.63 according to SSH. We further detail these patterns using Cox regressions in section 6 of Online Resource 1. The estimates presented in Fig. 4 do not consider the potential impact of date errors because there were also too few adult deaths in recent years to restrict estimates of  ${}_{45}q_{15}$  to a reference period.

### Date Errors in SSH Data

To explore the potential effects of date errors, we first measured whether respondents correctly classified their sister's death as having occurred before/during an eight-year reference period. This was not the case: only 65 of 85 deaths recorded during the reference period by the HDSS were reported as having occurred during the reference period by SSH, whereas 20 of the 124 deaths recorded before the reference period by the HDSS were reported as having occurred during the reference period by SSH. This yielded estimates of 76.5 % and 83.9 % for the sensitivity<sup>20</sup> and specificity<sup>21</sup> of SSH date reports, respectively.

<sup>19</sup>It appears that in a few cases, interviewers mistakenly recorded the date of death in the space where the age at death should have been recorded, and vice versa. For example, if a respondent reported that her sibling died three years ago at age 37, the interviewer could have mistakenly recorded that the death occurred 37 years ago at age 3. This generates very large age and date errors that cannot be detected by data editors because they leave the reported birth intervals between siblings unaffected.

<sup>20</sup>Sensitivity refers to the proportion of deaths observed during the reference period (2002–2009) that were reported by SSH respondents as having occurred during that time frame.

<sup>21</sup>Specificity refers to the proportion of deaths observed before the reference period (before 2002) that were reported by SSH respondents as having occurred during that timeframe.

We then investigated the correlates of date errors, using a multilevel model similar to the model used to investigate age errors. The dependent variable was a continuous variable measuring the difference between HDSS-observed and SSH-reported duration between survey and death (Table 5). Respondents made larger date errors for deaths observed during the reference period than for deaths having occurred before the reference period ( $\beta = -3.36$ ). On average, they reported that deaths during the reference period occurred *earlier* than their HDSS-observed date. To account for missing data on dates of deaths (21.3 % of female adult deaths had missing date of death; Online Resource 1, section 5), we conducted a sensitivity analysis using Heckman sample selection models (Heckman 1979). In these models, we used a set of dummy variables identifying the interviewer having conducted the SSH interview as instruments (for a similar approach to accounting for missing data in DHS data, see Barnighausen et al. 2011). Results were similar to those obtained in a model based solely on the complete case data. Further details about the Heckman models are available in section 7 of Online Resource 1.

## Discussion

In this article, we improved on previous assessments of reporting errors in SSH through an innovative record linkage study with HDSS data. This permitted isolating reporting errors from sample selection biases; it also fostered a better understanding of the diversity and complexity of reporting errors in SSH and their implications for estimates of adult survival.

Respondents frequently omitted some of their adult sisters in SSH. They also added others who were not their maternal siblings, but this was less common. Such list errors were associated with the vital status of siblings: omitted sisters were more likely to be deceased or to have migrated away from the HDSS area, whereas added sisters were more likely to be alive at the time of the survey. Age and date errors were common: respondents underreported the current age of their living sisters and the age at death of their deceased sisters; they also misreported the date of death of their sisters by several years. On the other hand, virtually no respondent reported their living sisters as deceased and vice versa.

In our reference sibships in Bandafassi, SSH yielded an estimate of  ${}_{45}q_{15}$  that was 20 % lower than the HDSS estimate (Fig. 4). The differences between SSH and HDSS estimates emerged largely above age 50. Prior to that age, SSH and DSS yielded remarkably similar estimates of adult survival (e.g.,  ${}_{25}q_{15} \approx 0.5$  according to both SSH and HDSS data). This congruence between SSH and DSS below age 50, however, does not imply that respondents accurately report the survival of their younger siblings: list errors and age errors were common even for siblings younger than 50 (e.g., see Fig. 3b). Rather, it is the result of a complex interaction between different types of reporting errors. For siblings below age 50, list errors are offset by age errors: the omitted deaths are replaced with deaths that have in fact occurred at older ages but are shifted downward (in age) by the respondent. Above age 50, however, this substitution does not operate because (1) list errors are more common (Fig. 3b), and (2) there are too few deaths above age 60 to replace the omitted deaths in the age group 50–59.

There were several unexpected findings. First, the quality of SSH data was not uniformly better during a recent reference period. Only list errors were reduced (Fig. 2a), whereas age and especially date errors were actually more common (Tables 4 and 5). Because of date errors, a number of deaths were shifted in and out of the reference period. This has potentially troublesome consequences for estimates of levels of, and trends in, adult mortality. In contexts of adult mortality declines (where most adult deaths in respondents' sibships will have occurred prior to the reference period), date errors as observed in our study will lead to overestimates of  ${}_{45q15}$  in the reference period (Fig. 5). In contexts of increasing adult mortality rates (e.g., HIV epidemics, wars), the opposite is true. This may limit the validity of program evaluations based on SSH data and a difference-in-differences econometric framework (Bendavid et al. 2012). In such evaluations, deaths must be precisely dated as having occurred before or after the implementation of a public health program.

Second, contrary to expectations (e.g., Stanton et al. 2000), older respondents provided more complete sibship lists than younger respondents (Table 3) but made only slightly larger age errors about their living sisters than other respondents (Table 4). Our findings thus support the call by Obermeyer et al. (2010) to extend the collection of SSH data to older age groups during the DHS and other surveys. This would not deteriorate the quality of data on SSH, but it would increase available sample size and possibly allow using SSH to estimate mortality at older ages (e.g., >60 years old).

### Limitations

There are several important limitations to our study. First, some of the discrepancies between HDSS and SSH data observed in our study may be attributable to limitations of the HDSS data. Key informants may not be aware of specific events at the time of the annual visit and may thus report them with delays and imprecisions. Loss to follow-up in the HDSS may be associated with mortality if, for example, healthier individuals leave the HDSS area to look for work in cities. The association between list errors and vital status may thus not be as strong as we estimated in Table 3. Finally, the procedures used to estimate the age of population members born prior to the initial census may also be imprecise. This may explain why age errors are more prevalent among siblings aged 45 and older (i.e., born prior to the initial census).

Second, the generalizability of our analyses may be limited because they concern only one small western African population. Patterns of reporting errors may be different in other populations where SSH are used to estimate adult mortality rates—for example, in sub-Saharan or south Asian countries. SSH may be more accurate in settings with uxori-local marriage, where sisters coreside for much longer periods of time than in Bandafassi (where marriage is virilocal). The reporting of SSH may also be affected by local attitude toward death or specific causes of death. Deaths resulting from HIV/AIDS, in particular, may be underreported because of stigma.

The impact of a given pattern of reporting errors on estimates of adult mortality may also vary across epidemiological contexts. For example, because respondents are more likely to omit deaths occurring at young (i.e., <25 years) and older (i.e., >45 years) adult ages (see



Fig. 3a), biases in SSH estimates may be larger in settings where causes of death that follow a U-shaped age pattern (e.g., maternal mortality) are prevalent. Comparative studies are thus needed to provide a better understanding of reporting under a variety of sociocultural and epidemiological contexts.

Third, our SSH survey may not have been representative of typical SSH surveys. It deviated from the DHS study protocol in a number of respects. We included respondents older than 50, whereas most DHS focus on respondents aged 15–49. Robustness tests (Online Resource 1, section 8) indicate that this inclusion does not alter our main findings. We also included only a few ( $n = 10$ ) sociodemographic questions prior to the SSH module, whereas the DHS include numerous other modules (e.g., fertility history and sexual behaviors). Respondent and interviewer fatigue may thus be significantly more common in the DHS than in our SSH survey. Finally, after being studied for such an extended period of time, the HDSS population may have gained a better knowledge of ages and date of events than similar neighboring populations not undergoing demographic surveillance. The extent of reporting errors identified in our study could thus be an underestimate of the likely extent of underreporting in other non-HDSS populations.

Fourth, our SSH survey had significantly more missing data on dates of death than typical DHS, possibly because of differences in the questionnaire editing process. Our analyses of the correlates of date errors, however, were not affected by these missing data patterns (Table 5; see also section 8 of Online Resource 1).

Finally, our analyses are based on small sample sizes that do not permit investigating more complex patterns of reporting errors. For example, we could not calculate  ${}_{45q15}$  separately for the person-years that occurred during a recent reference period, nor could we investigate interactions between different types of reporting errors. Future investigations of reporting errors in SSH data should be based on larger samples.

## Implications

Our results also have important implications. They suggest that recently proposed methods to correct for reporting errors (i.e., the CSS method of Obermeyer et al. 2010) are unlikely to produce unbiased estimates of adult mortality. These regression-based adjustments assume that reporting errors are limited to the forgetting of deaths and that such forgetting is a simple linear function of time between a death and the survey. Our results show that reporting errors are much more diverse. They also have complex effects that may offset each other under specific conditions. Such effects are not well captured by a simple correction factor. Our results also suggest new strategies for improving data quality for SSH. Omissions from sibship lists could be prevented by adding probes (Brewer and Garrett 2001; Brewer et al. 2005) in SSH interviews focused on siblings that are frequently omitted (e.g., migrants, siblings deceased at young or old adult ages). Age and date errors could be reduced by adapting calendar tools previously used by demographers for the collection of data on contraception, migration, or sexual behaviors (Freedman et al. 1988; Luke et al. 2011; White et al. 2008). Finally, our study highlights the potential use of data from HDSS to help improve demographic measurement in countries with limited VR. More than 40 HDSS are now in the INDEPTH network (Sankoh and Byass 2012), an international

organization that fosters collaboration between research centers operating HDSS in low and middle-income countries. We hope our study will help establish HDSS as “laboratories” where demographers can refine their data collection tools and analytical techniques.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank Patrick Gerland, Bruno Masquelier, and Samuel Preston for comments on previous drafts of this article. The project described was supported by Award Numbers R24HD058486 to the Columbia Population Research Center and R03HD071117 to S. Helleringer, both from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and by grant ANR-11-BSH1-0007 from the Agence Nationale de la Recherche to G. Pison/Institut National d'Études Démographiques. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

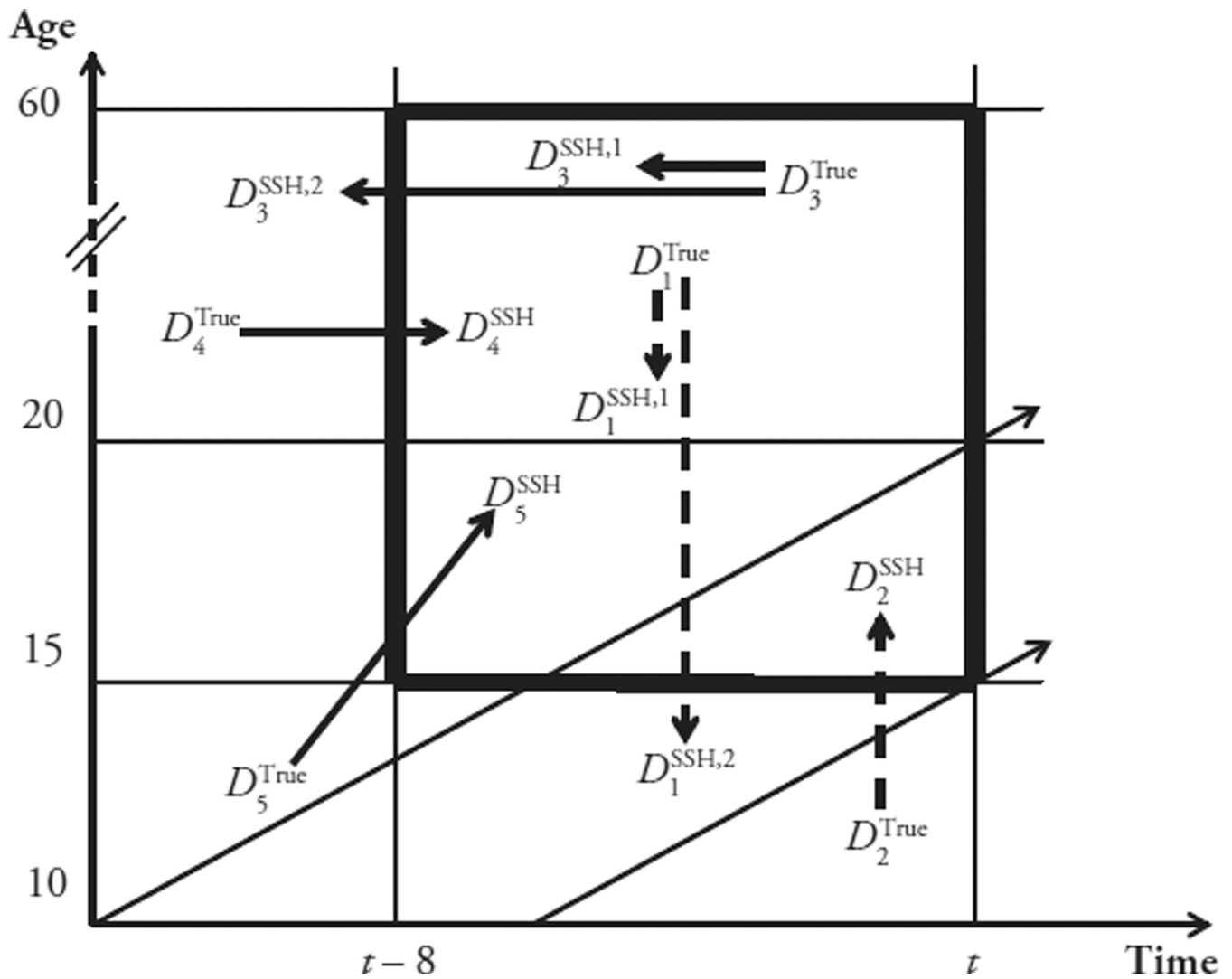
## References

- Ainsworth M, Dayton J. The impact of the AIDS epidemic on the health of older persons in northwestern Tanzania. *World Development*. 2003; 31:131–148.
- Amstrup, SC.; McDonald, TL.; Manly, BFJ. *Handbook of capture-recapture analysis*. Princeton, NJ: Princeton University Press; 2005.
- ANSD, & ICF. *Enquête Démographique et de Santé à Indicateurs Multiples au Sénégal (EDS-MICS) 2010–2011 [Senegal DHS, 2010–2011—Final report]*. Calverton, MD: ANSD and ICF International; 2012.
- Barnighausen T, Bor J, Wandira-Kazibwe S, Canning D. Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. *Epidemiology*. 2011; 22:27–35. [PubMed: 21150352]
- Bendavid E, Holmes CB, Bhattacharya J, Miller G. HIV development assistance and adult mortality in Africa. *Journal of the American Medical Association*. 2012; 307:2060–2067. [PubMed: 22665105]
- Brewer DD, Garrett SB. Evaluation of interviewing techniques to enhance recall of sexual and drug injection partners. *Sexually Transmitted Diseases*. 2001; 28:666–677. [PubMed: 11677390]
- Brewer DD, Potterat JJ, Muth SQ, Malone PZ, Montoya P, Green DL, Cox PA. Randomized trial of supplementary interviewing techniques to enhance recall of sexual partners in contact interviews. *Sexually Transmitted Diseases*. 2005; 32:189–193. [PubMed: 15729158]
- Case A, Ardington C. The impact of parental death on school outcomes: Longitudinal evidence from South Africa. *Demography*. 2006; 43:401–420. [PubMed: 17051820]
- Chao LW, Pauly M, Szrek H, Pereira NS, Bundred F, Cross C, Gow J. Poor health kills small business: Illness and microenterprises in South Africa. *Health Affairs*. 2007; 26:474–482. [PubMed: 17339676]
- Clark SJ, Collinson MA, Kahn K, Drullinger K, Tollman SM. Returning home to die: Circular labour migration and mortality in South Africa. *Scandinavian Journal of Public Health*. 2007; 35:35–44. [PubMed: 17676501]
- de Walque D, Filmer D. Trends and socioeconomic gradients in adult mortality around the developing world. *Population and Development Review*. 2013; 39:1–29.
- de Walque D, Verwimp P. The demographic and socio-economic distribution of excess mortality during the 1994 genocide in Rwanda. *Journal of African Economies*. 2010; 19:141–162.
- Desgrées du Lou A, Pison G, Aaby P. Role of immunizations in the recent decline in childhood mortality and the changes in the female/male mortality ratio in rural Senegal. *American Journal of Epidemiology*. 1995; 142:643–652. [PubMed: 7653475]
- Duthé G, Pison G. Adult mortality in a rural area of Senegal: Non-communicable diseases have a large impact in Mlomp. *Demographic Research*. 2008; 19:1419–1448. (article 37).

- Evans DK, Miguel E. Orphans and schooling in Africa: A longitudinal analysis. *Demography*. 2007; 44:35–57. [PubMed: 17461335]
- Feeney G. The impact of HIV/AIDS on adult mortality in Zimbabwe. *Population and Development Review*. 2001; 27:771–780.
- Freedman D, Thornton A, Camburn D, Alwin D, Young-Demarco L. The life history calendar: A technique for collecting retrospective data. *Sociological Methodology*. 1988; 18:37–68.
- Gakidou E, King G. Death by survey: Estimating adult mortality without selection bias from sibling survival data. *Demography*. 2006; 43:569–585. [PubMed: 17051828]
- Garrib A, Herbst AJ, Hosegood V, Newell ML. Injury mortality in rural South Africa 2000–2007: Rates and associated factors. *Tropical Medicine & International Health*. 2011; 16:439–446. [PubMed: 21284789]
- Gelman, A.; Hill, J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press; 2007.
- Goosby E. The President's Emergency Plan for AIDS Relief: Marshalling all tools at our disposal toward an AIDS-free generation. *Health Affairs*. 2012; 31:1593–1598. [PubMed: 22778350]
- Graham W, Brass W, Snow RW. Estimating maternal mortality: The sisterhood method. *Studies in Family Planning*. 1989; 20:125–135. [PubMed: 2734809]
- Heckman JJ. Sample selection bias as a specification error. *Econometrica*. 1979; 47:153–161.
- Helleringer S, Duthé G, Kanté AM, Andro A, Sokhna C, Trape JF, Pison G. Misclassification of pregnancy-related deaths in adult mortality surveys: Case study in Senegal. *Tropical Medicine & International Health*. 2013; 18:27–34. [PubMed: 23130912]
- Hill K, Choi Y, Timaeus IM. Unconventional approaches to mortality estimation. *Demographic Research*. 2005; 13:281–300. (article 12).
- Hill K, El Arifeen S, Koenig M, Al-Sabir A, Jamil K, Raggars H. 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*. 2006; 84:173–180. [PubMed: 16583075]
- Hill K, You D, Inoue M, Oestergaard MZ. Technical Advisory Group of United Nations Inter-agency Group for Child Mortality Estimation. Child mortality estimation: Accelerated progress in reducing global child mortality, 1990–2010. *PLoS Medicine*. 2012; 9(8):e1001303. [PubMed: 22952441]
- Hill ME, Preston SH, Rosenwaike I. Age reporting among white Americans aged 85+: Results of a record linkage study. *Demography*. 2000; 37:175–186. [PubMed: 10836175]
- Hogan MC, Foreman KJ, Naghavi M, Ahn SY, Wang M, Makela SM, Murray CJ. Maternal mortality for 181 countries, 1980–2008: A systematic analysis of progress towards Millennium Development Goal 5. *Lancet*. 2010; 375:1609–1623. [PubMed: 20382417]
- Jahn A, Crampin AC, Glynn JR, Mwinuka V, Mwaiyeghele E, Mwafilaso J, Zaba B. Evaluation of a village-informant driven demographic surveillance system in Karonga, Northern Malawi. *Demographic Research*. 2007; 16:219–248. (article 8).
- Kanté AM, Pison G. Maternal mortality in rural Senegal. The experience of the new Ninesescha Hospital. *Population-E*. 2010; 65:653–678.
- Kautz T, Bendavid E, Bhattacharya J, Miller G. AIDS and declining support for dependent elderly people in Africa: Retrospective analysis using Demographic and Health Surveys. *British Medical Journal*. 2010; 340:c2841. [PubMed: 20554660]
- Lagarde E. Road traffic injury is an escalating burden in Africa and deserves proportionate research efforts. *PLoS Medicine*. 2007; 4(6):e170. [PubMed: 17593893]
- Lozano R, Wang H, Foreman KJ, Rajaratnam JK, Naghavi M, Marcus JR, Murray CJ. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: An updated systematic analysis. *Lancet*. 2011; 378:1139–1165. [PubMed: 21937100]
- Luke N, Clark S, Zulu EM. The relationship history calendar: Improving the scope and quality of data on youth sexual behavior. *Demography*. 2011; 48:1151–1176. [PubMed: 21732169]
- Masquelier B. Adult mortality from sibling survival data: a reappraisal of selection biases. *Demography*. 2013; 50:207–228. [PubMed: 23055235]

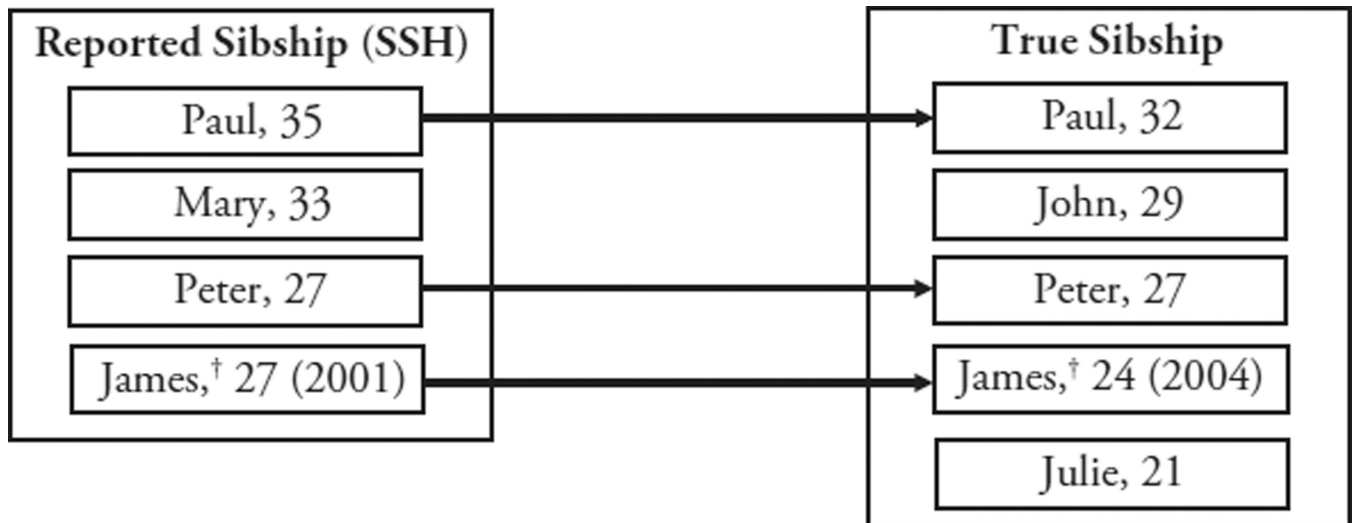
- Mathers CD, Boerma T, Fat DM. Global and regional causes of death. *British Medical Bulletin*. 2009; 92:7–32. [PubMed: 19776034]
- Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006; 3(11):e442. [PubMed: 17132052]
- Murray CJL, Ferguson BD, Lopez AD, Guillot M, Salomon JA, Ahmad O. Modified logit life table system: Principles, empirical validation, and application. *Population Studies*. 2003; 57:165–182.
- Ngom P, Akweongo P, Adongo P, Bawah AA, Binka F. Maternal mortality among the Kassena-Nankana of northern Ghana. *Studies in Family Planning*. 1999; 30:142–147. [PubMed: 16617548]
- Obermeyer Z, Murray CJ, Gakidou E. Fifty years of violent war deaths from Vietnam to Bosnia: Analysis of data from the World Health Survey Programme. *BMJ*. 2008; 336:1482–1486. [PubMed: 18566045]
- Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD, Murray CJ. Measuring adult mortality using sibling survival: A new analytical method and new results for 44 countries, 1974–2006. *PLoS Med*. 2010; 7(4):e1000260. [PubMed: 20405004]
- Oduro AR, Wak G, Azongo D, Debpuur C, Wontuo P, Kondayire F, Binka F. Profile of the Navrongo Health and Demographic Surveillance System. *International Journal of Epidemiology*. 2012; 41:968–976. [PubMed: 22933645]
- Pison G. Calculating age without asking for it: Method of estimating the age and agestructure of the Peul-Bande (Eastern Senegal). *Population*. 1980; 35:861–892.
- Pison G. Le recueil de généalogies orales: intérêt et limites pour l'histoire démographique de l'Afrique [Collecting oral genealogies: Potential and limitations for the demographic history of Africa]. *Annales de démographie historique*. 1987:67–83.
- Pison G. Population observatories as sources of information on mortality in developing countries. *Demographic Research*. 2005; 13:301–334. (article 13).
- Pison G, Langaney A. The level and age pattern of mortality in Bandafassi (Eastern Senegal)—Results from a small-scale and intensive multi-round survey. *Population Studies*. 1985; 39:387–405.
- Preston SH, Elo IT, Rosenwaike I, Hill M. African-American mortality at older ages: Results of a matching study. *Demography*. 1996; 33:193–209. [PubMed: 8827165]
- Quinlan RJ, Hagen EH. New genealogy: It's not just for kinship anymore. *Field Methods*. 2008; 20:129–154.
- Reniers, G.; Masquelier, B.; Gerland, P. Adult mortality in Africa. In: Rogers, R.; Crimmins, E., editors. *International handbook of adult mortality*. New York: Springer; 2011. p. 151-170.
- Sankoh O, Byass P. The INDEPTH network: Filling vital gaps in global epidemiology. *International Journal of Epidemiology*. 2012; 41:579–588. [PubMed: 22798690]
- Sankoh, OA.; Ngom, P.; Clark, SJ.; de Savigny, D.; Binka, F. Levels and patterns of mortality at INDEPTH Demographic Surveillance Systems. In: Jamison, DT.; Feachem, RG.; Makgoba, MW.; Box, ER.; Baingana, FK.; Hofman, KJ.; Rogo, KO., editors. *Disease and mortality in sub-Saharan Africa*. 2nd ed.. Washington, DC: World Bank; 2006. p. 75-86.
- Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S, AbouZahr C. A scandal of invisibility: Making everyone count by counting everyone. *Lancet*. 2007; 370:1569–1577. [PubMed: 17992727]
- Shahidullah M. The sisterhood method of estimating maternal mortality: The Matlab experience. *Studies in Family Planning*. 1995; 26:101–106. [PubMed: 7618193]
- Silagan MF. The Genealogy of Mukama: The methodology of oral tradition. *Dialectical Anthropology*. 1986; 10:229–247.
- Stanton C, Abderrahim N, Hill K. An assessment of DHS maternal mortality indicators. *Studies in Family Planning*. 2000; 31:111–123. [PubMed: 10907277]
- Stover J. Projecting the demographic consequences of adult HIV prevalence trends: The Spectrum Projection Package. *Sexually Transmitted Infections*. 2004; 80(Suppl. 1):i14–i18. [PubMed: 15249694]
- Sutherland, WJ. *Ecological census techniques: A handbook*. Cambridge, UK: Cambridge University Press; 2006.

- Timaeus IM. Estimation of mortality from orphanhood in adulthood. *Demography*. 1991a; 28:213–227. [PubMed: 2070895]
- Timaeus IM. Measurement of adult mortality in less developed countries: A comparative review. *Population Index*. 1991b; 57:552–568. [PubMed: 12284917]
- Timaeus IM, Jasseh M. Adult mortality in sub-Saharan Africa: Evidence from Demographic and Health Surveys. *Demography*. 2004; 41:757–772. [PubMed: 15622953]
- Trussell J, Rodriguez G. A note on the sisterhood estimator of maternal mortality. *Studies in Family Planning*. 1990; 21:344–346. [PubMed: 2075625]
- United Nations. *World Population Prospects: the 2008 revision*. New York: United Nations; 2010.
- White MJ, Muhidin S, Andrzejewski C, Tagoe E, Knight R, Reed H. Urbanization and fertility: An event-history analysis of coastal Ghana. *Demography*. 2008; 45:803–816. [PubMed: 19110898]
- Wixted JT. The psychology and neuroscience of forgetting. *Annual Review of Psychology*. 2004; 55:235–269.
- Yamano T, Jayne TS. Measuring the impacts of working-age adult mortality on smallscale farm households in Kenya. *World Development*. 2004; 32:91–119.

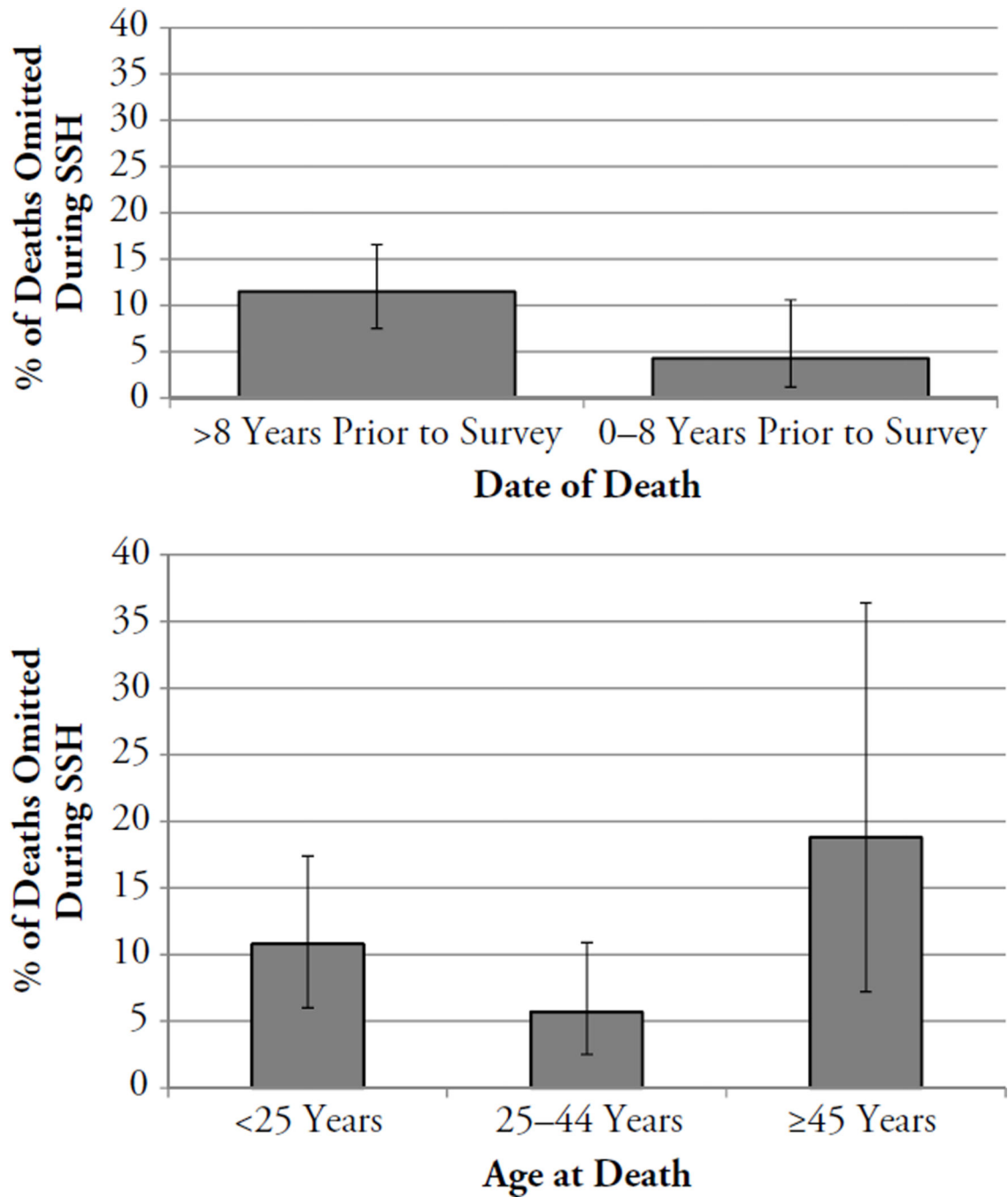


**Fig. 1.** Lexis diagram representation of date/age errors in SSH data on siblings’ deaths. On this diagram, the survey takes place at time  $t$ . We hypothesize a reference period of eight years, which is represented in thicker vertical lines. The thicker horizontal lines represent the age interval between ages 15 and 60 (i.e., “adulthood”). The area inside the thicker lines is the “reference area” containing all the person-years and events that are to be included in the analysis of siblings’ data sets derived from SSH data. Diagonal dotted arrows represent the aging of birth cohorts. Dashed black arrows represent age errors: namely, deaths for which the respondent reports the age at death as either higher or lower than the true age at death of her sister. Solid black arrows represent date errors: namely, deaths for which the respondent reported the date of death as either more recent or more distant than the true date of death of her sister. All arrows originate from an event, noted  $D$  with a superscript noted “true,” and end at another event  $D$  with a superscript “SSH.” Age and date errors may lead to misclassifications of death as having occurred during/outside the reference area (i.e., if the solid or dashed arrow cross any of the thicker lines).

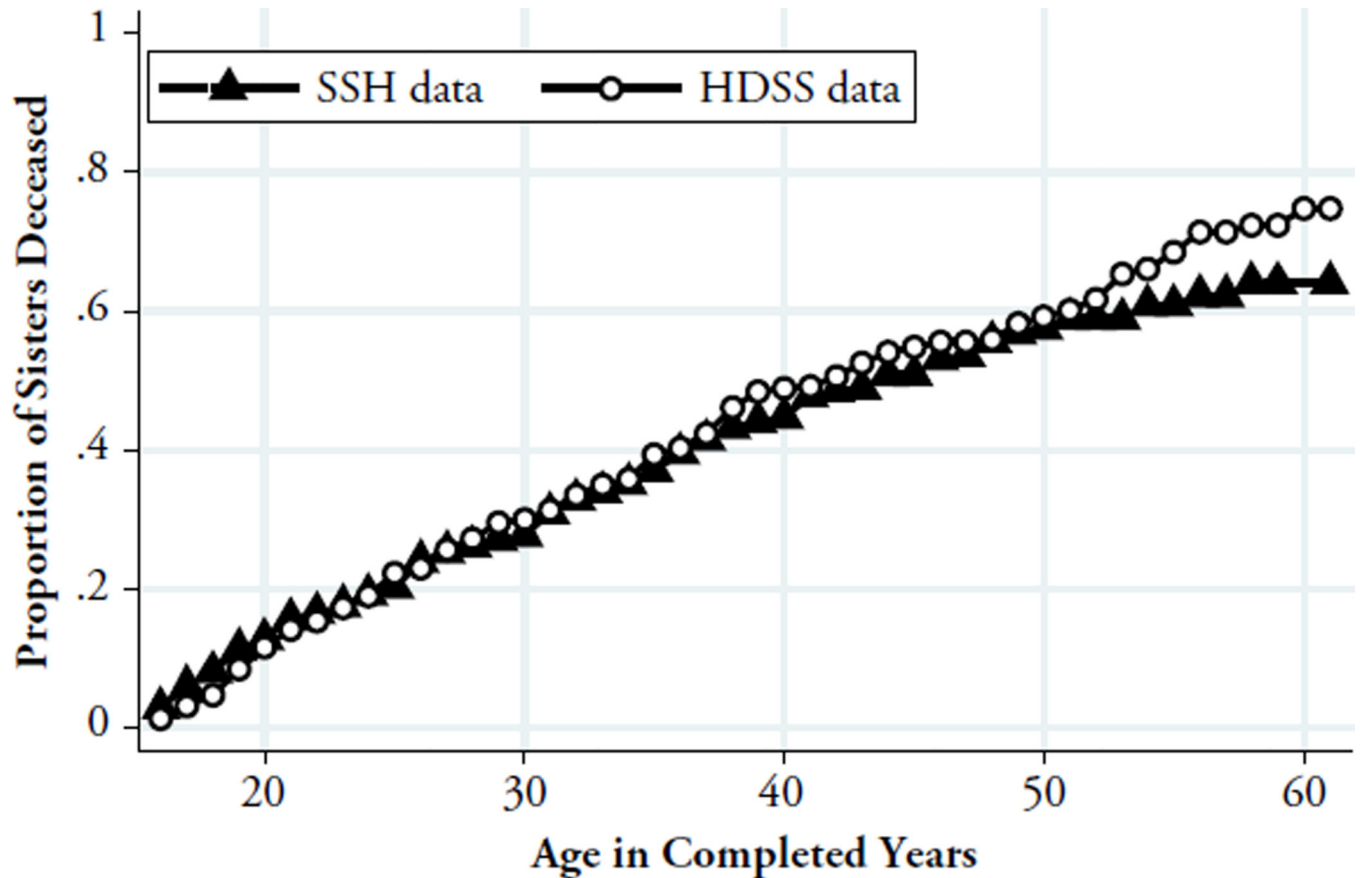




**Fig. 2.** Study design and data linkages. The figure represents the design of a record linkage study conducted at the sibling level. It focuses on deaths and person-years potentially included both in the reference and test data set (i.e., “reference sibships”) and establishes linkages at the individual level. In that example, we compare a list of the maternal siblings of a respondent obtained from SSH with a hypothetical reference data set. Paul, Peter, and James are “matches” (represented by the unidirectional arrows), Mary is a potential “addition,” and John and Julie are potential “omissions.” The age of Paul is misreported during SSH (35 vs. 32, “age error”), whereas the age of Peter is correctly reported. James’ age at death is misreported during SSH (27 vs. 24, “age error”) as is his date of death (2001 vs. 2004, “date error”). This latter date error leads James’ death to be misclassified as having occurred outside the eight-year reference period based on SSH data (when in fact it occurred in 2004, within the reference period).

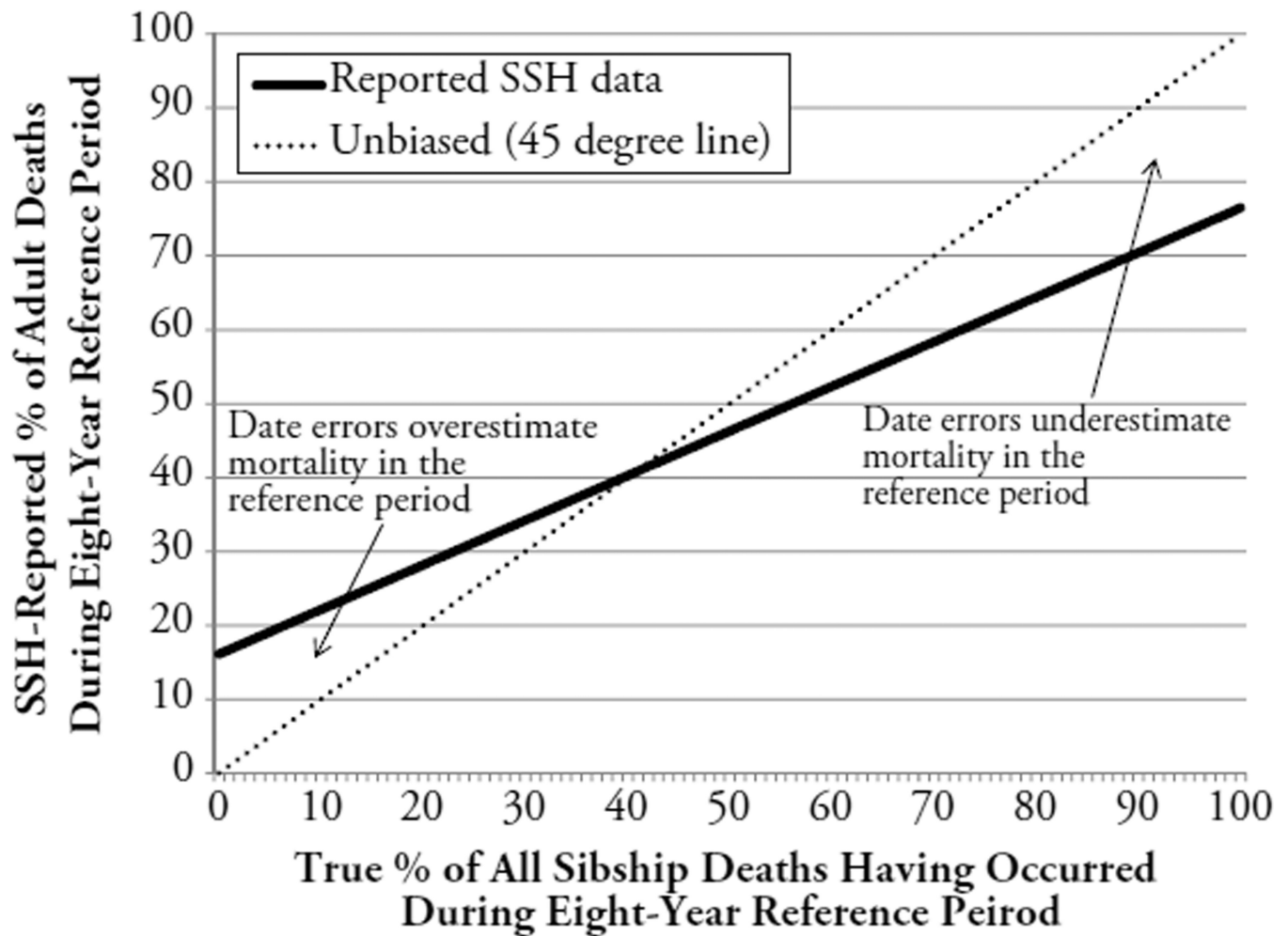


**Fig. 3.** Proportions of deaths omitted from SSH by date at death (time between death and survey) (panel a, top) and by age at death (panel b, lower) ( $n = 308$ ). Error bars represent 95 % confidence intervals; the difference in the probability of omission by age at death was significant at the  $p < .05$  level (two-sided chi-square test); the difference in the probability of omission by time since death was significant at the  $p < .05$  level (two-sided chi-square test).



**Fig. 4.**

Estimates of adult survival between ages 15 and 59 according to DSS and SSH data. The survival curves represent the cumulative probability of death for adult women (i.e., having ever reached age 15) who were members of the reference sibships included in the study. Respondents were excluded from these calculations. In the HDSS estimates, observations of sisters lost to follow-up were censored at the time of exit from the HDSS population. In SSH estimates, these observations were censored at the time of the survey given that the respondent reported the vital status of siblings at the time of survey. Similar results were obtained when SSH data on siblings lost to HDSS follow-up were censored at the time of exit from the HDSS rather than at the time of the survey. In this figure, missing ages were not imputed in the SSH data. In a robustness test, however, missing data on current age/age at death in the SSH data were imputed using the same procedure adopted during the DHS. Using this approach, we obtained similar results to the ones presented here. The age at death of four siblings could not be imputed because of missing data on both the age and date of death.



**Fig. 5.**

The impact of date errors in SSH data on estimates of  ${}_{45}q_{15}$ . The solid line represent the SSH-reported proportion of adult deaths occurring in an eight-year reference period prior to the SSH survey under the assumptions that (1) there is a fixed number of deaths in the sibships of respondents (e.g.,  $n = 100$ ), (2) only the timing of these deaths (before/during the reference period) varies between scenarios plotted along the  $x$ -axis, (3) SSH respondents make date errors in SSH data according to sensitivity (83.9 %) and specificity (76.5 %) estimates reported in the main text, and (4) SSH respondents make only date errors in reports of siblings' survival; there are no list errors and/or age errors. The dotted line represents equality between true and reported proportions of adult deaths during the reference period. If more than 40 % of all sibship deaths occurred during an eight-year reference period, SSH will underestimate mortality because of date errors.

**Table 1**

## Characteristics of study respondents

	<b>Respondents N(%)</b>
<b>Characteristics of the Respondent</b>	
Place of interview	
Bandafassi area	223 (83.2)
Tambacounda/Kédougou	45 (16.8)
Age	
<25 years old	52 (19.4)
25–34 years old	64 (23.9)
35–44 years old	64 (23.9)
45 years old	88 (32.8)
Education	
Never been to school	217 (81.0)
Ever been to school	51 (19.0)
<b>Characteristics of the Sibship</b>	
Ethnic group	
Fulani	147 (54.9)
Bedik	121 (45.1)
Mandinka	
All siblings have the same biological father?	
No	89 (33.2)
Yes	179 (66.8)
Number of sisters interviewed during SSH	
1	100 (37.3)
2 or more	168 (62.7)
2	94 (35.1)
3	51 (19.0)
4	8 (3.0)
5	15 (5.6)
<i>N</i>	268

*Notes:* Column percentages are in parentheses.

*Source:* 2010 SSH Survey.

**Table 2**

Estimates of the completeness of the DSS and SSH data sources

	Observed Data			Capture-Recapture Estimates		
	Only in SSH Data (“additions”) (A)	Only in HDSS Siblings Lists (“omissions”) (B)	In both data sets (“matches”) (C)	Number Missed by Both Sources <sup>a</sup> (D)	Completeness of HDSS Data <sup>b</sup>	Completeness of SSH Data <sup>c</sup> (%)
All Sisters <sup>d</sup>	91 (9.2)	107 (10.8)	795 (80.0)	12	89.8	88.2
Sisters Under Age 15	67 (22.6)	43 (14.5)	186 (62.9)	15	73.6	81.3
Adult Sisters ( 15)	26 (3.4)	64 (8.4)	675 (88.2)	2	96.3	91.3
Sisters deceased at adult ages	3 (1.0)	28 (8.6)	295 (90.4)	<1	99.1	91.4

<sup>a</sup>This is calculated as  $(A \times B) / C$ .

<sup>b</sup>This is calculated as  $(B + C) / (A + B + C + D)$ .

<sup>c</sup>This is calculated as  $(A + C) / (A + B + C + D)$ .

<sup>d</sup>These tabulations ignore discrepancies in reports of vital status and age at death/current age between SSH and HDSS data. For example, one sister was reported as deceased by the SSH respondent but was recorded as alive in the HDSS data set: she is counted in column C here, not in column A.



**Table 3**

Multilevel logistic regression of the probability of list errors during SSH interviews

	<i>N (%)<sup>a</sup></i>	<b>Odds Ratios</b>
Sibling's Follow-up Status		
Alive	265 (3.8)	Ref.
Deceased	308 (9.1)	2.95*
Lost to follow-up	157 (16.6)	11.6***
Sibling's Characteristics		
Same biological father as respondent?		
No	127 (18.1)	Ref.
Yes	603 (6.8)	0.19**
Respondent Characteristics		
Place of interview		
Bandafassi DSS area	603 (7.1)	Ref.
Tambacounda/Kédougou	127 (16.5)	5.18*
Age		
<25 years old	149 (16.1)	Ref.
25–34 years old	180 (7.2)	0.20 <sup>†</sup>
35–44 years old	177 (7.9)	0.18 <sup>†</sup>
45 years old	224 (5.8)	0.15 <sup>†</sup>
Education		
Ever been to school	146 (15.8)	Ref.
Never been to school	584 (7.0)	1.10
Sibship Characteristics		
Ethnic group		
Fulani	484 (7.6)	Ref.
Bedik	130 (15.4)	7.28*
Mandinka	116 (6.0)	2.04
<i>N<sup>b</sup></i>	730	

<sup>a</sup>The proportion of sisters omitted during the SSH survey appears in parentheses.

<sup>b</sup>Twelve adult sisters (1.8 %) are omitted from these models because of missing data on various covariates.

<sup>†</sup> $p < .10$ ;

\* $p < .05$ ;

\*\* $p < .01$ ;

\*\*\* $p < .001$

**Table 4**

Multilevel linear regression of the magnitude of age errors during SSH interview

	Live Sisters		Deceased Sisters	
	Median Error (IQR)	$\beta$	Median Error (IQR)	$\beta$
Sibling's Characteristics				
Age <sup>a</sup>				
<25 years old	0.27 (-1, 2)	Ref.	1.94 (0, 4)	Ref.
25–44 years old	1.58 (-1, 4)	2.50**	3.39 (0, 7)	1.24
45 years old	2.56 (-1, 6)	4.22***	7.52 (0, 14)	6.28***
Same biological father as respondent?				
No	1.86 (-2, 5)	Ref.	4.84 (0, 5)	Ref.
Yes	1.72 (-1, 4)	0.23	6.77 (0, 6)	-0.76
Death during eight-year reference period?				
No	—	—	2.52 (0, 5)	Ref.
Yes	—	—	4.08 (0, 6)	1.86*
Respondent Characteristics				
Place of interview				
Bandafassi DSS area	1.44 (-1, 4)	Ref.	3.11 (0, 6)	Ref.
Tambacounda/Kédougou	3.49 (1, 6)	2.04*	2.73 (0, 5)	0.18
Age				
<25 years old	1.75 (-1, 4)	Ref.	2.13 (0, 3)	Ref.
25–34 years old	2.41 (0, 4)	-0.40	3.29 (0, 5)	0.77
35–44 years old	1.26 (-2, 5)	-2.47*	2.56 (0, 6)	-0.11
45 years old	1.53 (-2, 5)	-2.37*	3.79 (-1, 7)	0.08
Education				
Never been to school	1.70 (-1, 5)	Ref.	2.92 (0, 6)	Ref.
Ever been to school	1.93 (0, 3)	0.41	3.09 (0, 6)	-0.09
Sibship Characteristics				
Ethnic group				
Fulani	1.95 (-1, 4)	Ref.	2.34 (0, 5)	Ref.
Bedik	1.53 (-2, 4)	-0.02	5.37 (0, 10)	2.43*
Mandinka	1.18 (-1, 2)	-0.50	3.15 (0, 5)	1.04
Constant <sup>b</sup>	—	0.04	—	1.58
<i>N<sup>c,d</sup></i>	252		264	

<sup>a</sup>Current age or age at death, as observed in HDSS.<sup>b</sup>Model includes only a sibship-level random effect because there are too few deaths per respondent to estimate a respondent-level random effect.<sup>c</sup>Five outliers for which the difference between observed and reported age at death was >30 years (in absolute value) were excluded; we chose to exclude these outliers because such large errors may be due to linkage errors rather than true age-reporting errors. They may also be due, however,

to interviewer errors in which the answers to the questions on age and date of death were mistakenly switched; estimates of the fixed effects are unaffected by the inclusion of these outliers, but estimates of the random-effect nearly double if the outliers are included.

<sup>d</sup> Eleven cases were excluded because of missing data on age at death, and three cases were excluded because of missing data on current age.

<sup>†</sup>  $p < .10$ ;

\*  $p < .05$ ;

\*\*  $p < .01$ ;

\*\*\*  $p < .001$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5**

Multilevel linear regression of the magnitude of date errors during SSH interviews

	Median Error (IQR)	$\beta$ (complete case)	$\beta$ (Heckman)
Sibling's Characteristics			
Age at death (observed)			
<25 years old	0.11 (-1, 3)	Ref.	Ref.
25–44 years old	-0.33 (-2, 3)	0.47	0.45
45 years old	-2.31 (-7, 4)	-1.53	-1.18
Same biological father as respondent?			
No	-1.72 (-4, 1)	Ref.	Ref.
Yes	-0.12 (-2, 3)	1.54	2.09
Death during eight-year reference period?			
No	0.86 (-2, 4)	Ref.	Ref.
Yes	-2.10 (-2, 1)	-3.36***	-2.67**
Respondent Characteristics			
Place of interview			
Bandafassi DSS area	-0.47 (-2, 3)	Ref.	Ref.
Tambacounda/Kédougou	0.36 (0, 3)	0.39	0.46
Age			
<25 years old	-0.37 (-1, 1)	Ref.	Ref.
25–34 years old	0.39 (-1, 2)	-0.84	0.25
35–44 years old	-1.40 (-5, 3)	-2.29	-1.33
45 years old	-0.03 (-4, 4)	-1.54	-0.08
Education			
Never been to school	-0.35 (-2, 3)	Ref.	Ref.
Ever been to school	-0.29 (-1, 2)	-0.24	-0.03
Sibship Characteristics			
Ethnic group			
Fulani	0.52 (-1, 3)	Ref.	Ref.
Bedik	-1.92 (-7, 3)	-2.21†	-2.29
Mandinka	-2.50 (-7, 1)	-3.45*	-2.92†
Constant <sup>a</sup>	—	1.74	-0.38
N <sup>b</sup>	209	209	280

<sup>a</sup>The model includes only a sibship-level random effect because there are too few deaths per respondent to estimate a respondent-level random effect.

<sup>b</sup>One outlier for which the difference between observed and reported date of death was >30 years (in absolute value) was excluded; we chose to exclude this outlier because such large errors may be due to linkage errors rather than true date-reporting errors. They may also be due to interviewer errors in which the answers to the questions on age and date of death were mistakenly switched. Estimates of the fixed effects are unaffected by the inclusion of this outlier, but estimates of the random effect nearly double if the outlier is included.