DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 51, ARTICLE 45, PAGES 1429–1470
PUBLISHED 17 DECEMBER 2024
http://www.demographic-research.org/Volumes/Vol51/45/
DOI: 10.4054/DemRes.2024.51.45

Research Article

Improving old-age mortality estimation with parental survival histories in surveys

Bruno Masquelier **Ashira Menashe-Oren**

Benjamin-Samuel Schlüter **Atoumane Fall**

Stéphane Helleringer

This publication is part of the Special Collection on "Innovations in measuring adult mortality in countries with incomplete civil registration," organized by Guest Editors Ashira Menashe Oren, Nandita Saikia, and Stéphane Helleringer.

Contents

Demographic Research: Volume 51, Article 45
Research Article

Improving old-age mortality estimation with parental survival histories in surveys

Bruno Masquelier[1]

Ashira Menashe-Oren[2]

Benjamin-Samuel Schlüter[3]

Atoumane Fall[4]

Stéphane Helleringer[5]

Abstract

BACKGROUND
In many low- and middle-income countries, the mortality of adults over 50 years of age is poorly monitored because death registration systems are deficient. Nationally representative surveys currently focus on the survival of children or adults aged 15 to 49 years.

OBJECTIVE
We propose to measure adult survival beyond age 50 via parental survival histories, in which survey respondents provide data on their parents' ages at the time of the survey, and if deceased, their age at death and date of death. We evaluate the magnitude of possible selection bias in parental survival histories and quantify the sample sizes needed to estimate mortality at ages 50 to 79 with varying levels of precision.

METHODS
We created a population with known parental survival histories using the 2013 national census of Senegal augmented with microsimulations. Using a stratified two-stage sampling procedure, we then conducted household surveys of this artificial population. We compared reference mortality levels in the artificial population of adults aged 50 to 79 years with those inferred from parental survival histories. We also analyzed selection bi-

[1] Center for Demographic Research, Louvain University, Louvain-la-Neuve, Belgium.
Email: bruno.masquelier@uclouvain.be.
[2] Center for Demographic Research, Louvain University, Louvain-la-Neuve, Belgium.
[3] Department of Statistical Sciences, University of Toronto, Toronto, Canada.
[4] Agence Nationale de la Statistique et de la Démographie, Dakar, Senegal.
[5] Program on Social Research and Public Policy, New York University – Abu Dhabi, Abu Dhabi, United Arab Emirates.

ases in simulated populations where mortality above age 50 is correlated with the number of adult children.

RESULTS
The inclusion of modules on parental survival in large-scale surveys, such as the Demographic and Health Surveys, could provide accurate and precise estimates of old-age mortality and capture time trends and age patterns. Estimates derived from parental survival histories are affected by an upward bias when mortality is positively correlated with fertility, and vice versa, but the bias is modest and can be partially corrected.

CONTRIBUTION
Parental survival histories are a promising method to fill important data gaps around mortality at older ages, although more research is needed on possible reporting errors.

1. Introduction

In low- and middle-income countries (LMICs), a growing share of deaths occurs at older ages, thanks to the decline in premature mortality and the gradual population ageing. In 2022, already 74% of deaths occurred over the age of 50 in LMICs, and this proportion is expected to reach 85% by 2050 (United Nations 2022). Yet there is a critical lack of data to measure mortality at these ages because death registration systems are not sufficiently developed in LMICs (Mikkelsen et al. 2015; Ouedraogo 2020). Survey programs such as the Demographic and Health Surveys (DHS) or Multiple Indicator Cluster Surveys (MICS) fill important data gaps when collecting full birth histories and sibling survival histories, but these modules measure mortality among children under the age of 5 and among adults of reproductive age. In sibling survival histories, adults aged 15 to 49 are asked about each maternal sibling: their age at the time of the survey, or, age at death and date of death. Unfortunately, sibling survival histories are inappropriate for measuring mortality in older adults, because siblings are approximately the same age as respondents.

National censuses and large-scale surveys enquiring about recent household deaths are also used to compensate for the lack of death registration data. The resulting estimates are very uncertain, however, as they are subject to systematic age misstatement, underreporting of deaths, and selection bias due to the dissolution of some households following the death of an adult member (Timæus 1991; Lankoandé et al. 2022). Sampling errors around mortality rates obtained from recent household deaths are also quite wide, especially when retrospective surveys rather than censuses are used. For example, Bendavid, Seligman, and Kubo (2011) analyze data collected in eight nationally representative surveys in Africa on mortality over age 60, and their estimates were based on only 1,233 deaths or just over 75 deaths per country for each sex. None of the existing instruments

to measure mortality in surveys and censuses are thus satisfying in covering older adult deaths, and new approaches should be developed.

In this study, we evaluate whether including a module on parental survival histories in nationally representative surveys would help improve the estimation of old-age mortality. Data on parental survival are already collected in DHS and MICS, but the questions are limited to children less than 18 years old, and no information is collected on the ages of surviving parents or, for the deceased parents, their age at death or the timing of death. Deriving mortality estimates from the summary data currently collected requires tabulating proportions of surviving parents by age of respondents and converting these proportions into mortality indices using the indirect orphanhood approach (Timæus 1992; Saikia, Bora, and Luy 2019). The main advantage of this indirect approach is that data are collected quickly using two questions about the survival of mothers and fathers, making it inexpensive and suitable for censuses as well. The indirect approach has several disadvantages, however; the duration of exposure is approximated by the age of the respondents, the resulting estimates refer to a relatively distant past, and they must be dated assuming a regularity of mortality trends (Timæus 2013). With parental survival histories, adults of reproductive age are invited to report on the survival status of their mother and father, their current age or their age at death when deceased, and the number of years since their death. An example of a module collecting parental survival histories is provided in Figure 1. A direct calculation is possible, with no assumption on trends or age patterns of mortality. If implemented in nationally representative surveys, these questions could complement the module on sibling survival to cover mortality in the elderly.

Figure 1: Example of a module on parental survival histories for adult respondents

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
PS1	What is the name of your biological mother?	Name: _____	
PS2	Is your biological mother still alive?	YES ..... 1 NO ..... 2 DON'T KNOW 8	 →PS4 →PS6
PS3	How old is she? RECORD AGE IN COMPLETED YEARS	YEARS ..... ☐☐ DON'T KNOW 9998	→PS6 →PS6
PS4	How old was your biological mother when she died?	YEARS ..... ☐☐ DON'T KNOW 9998	
PS5	How many years ago did the death of your mother occur?	☐☐	
PS6: Repeat questions PS1 to PS5 for the biological father of the respondent.			

The idea of collecting parental survival histories is not entirely new. As early as 40 years ago, Chackiel and Orellana (1985) propose that dates of death of parents be collected so that the reference period would not have to be estimated. Their analyses, conducted on a survey in Honduras in the 1980s, indicate that this approach gave more plausible results than those obtained indirectly. Unfortunately, their work does not lead to revisions to survey questionnaires. Two notable exceptions are the 1987 DHS survey in Burundi, in which questions are asked on the timing of parental deaths (Makinson 1993), and the Vietnam Life History Survey in 1991, in which information is also elicited on dates of birth of parents, allowing a direct calculation. The parental survival histories collected in Vietnam provide estimates that are consistent with life tables available from other sources (Hirschman, Preston, and Loi 1995). More recently, parental survival histories are explored again, in attempting to measure excess mortality related to the COVID-19 pandemic through mobile phone surveys (Chasukwa et al. 2022; Adjiwanou et al. 2020). Based on this limited experience, it is still difficult to conclude on the strengths and limitations of parental survival histories.

Before advocating for the integration of parental survival histories into large-scale national surveys, it is crucial to examine two types of errors inherent to this instrument.

First, as with the indirect orphanhood method, there is a risk of sample bias (Menashe-Oren et al. 2024). For example, in families with lower levels of education, fertility is typically higher, and adult mortality may also be higher (Masquelier and Garbero 2016). This could lead to an overestimation of mortality because a greater number of respondents may report on the experience of their parents who faced higher mortality. Conversely, a higher number of children in adulthood within a family could lower the risk of death at older ages due to the provision of social and material support. This was observed in Senegal, where a negative association between the number of children and mortality among women aged 50 to 70 has been noted, particularly with male children (Duthé et al. 2016). Potential links between parental and offspring mortality risks might also emerge from shared socioeconomic status, ecological environments, disease transmission, cultural influences, or genetic inheritance (Lu and Vogl 2023). Selection biases in indirect orphanhood estimates have already been evaluated by Palloni, Massagli, and Marcotte (1984). These authors conclude that biases remain minor and tend to cancel each other out. It is possible, however, that the biases are more pronounced in parental survival histories when used to measure old-age mortality, as they could develop over a longer period.

Second, we need to assess the extent of random errors around the resulting estimates to establish the sample sizes required to measure mortality at ages 50 to 79 given the prevailing demographic parameters in LMICs. Unlike sibling survival histories, for which the respondent provides information on a fairly large group (four to eight siblings in high fertility countries), each respondent provides data on his or her father and mother only, so the random errors could be wider, despite the higher mortality levels at older ages.

Because parental survival histories have rarely been collected, only a few existing data sources allow us to explore these two sources of errors. We recently investigated selection bias in old-age mortality estimates that would be obtained from parental histories using three health and demographic surveillance systems in Senegal (Menashe-Oren et al. 2024). We observed that older adults without surviving children faced a higher risk of mortality, but this did not result in systematic bias. In this study, we revisit this issue using microsimulations. We also combine microsimulations with data from the 2013 Senegalese census to assess whether parental survival histories can provide sufficiently precise estimates of mortality at ages 50 to 79 nationwide.

2. Data and methods

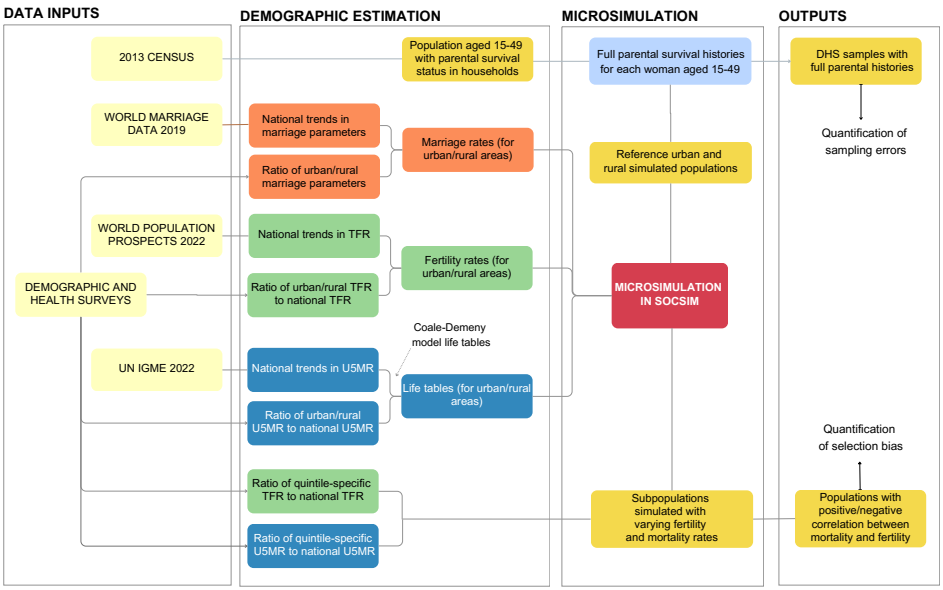
2.1 Microsimulations reflecting the population dynamics of Senegal

To explore selection bias and assess the precision of estimates derived from parental survival histories, we use different types of samples, all involving microsimulated populations reflecting the demographic patterns in Senegal. These simulations are generated using SOCSIM, an open-source program developed at the University of California, Berkeley (Wachter, Blackwell, and Hammel 1997; Zagheni 2011; Verdery et al. 2020; Alburez-Gutierrez, Mason, and Zagheni 2021). In SOCSIM, the individual is the unit of analysis and is exposed to a range of demographic events (having children, getting married, dying, etc.). The waiting time until each event is stochastically determined using a competitive risk model that takes into account predefined demographic rates. SOCSIM is a 'closed' model, meaning that individuals can only enter the simulation by birth and exit by death, which allows for keeping track of all parental IDs.

A flowchart presenting input data, processing steps, and outputs is in Figure 2. We first reconstructed a series of mortality, fertility, and marriage rates for the urban and rural areas since 1950, as detailed in Appendix A. This reconstruction was based on the under-5 mortality rates from the UN Inter-agency Group for Child Mortality Estimation (UN IGME), fertility rates from the World Population Prospects (WPP), and parameters extracted from the World Marriage Data for nuptiality (IGME 2024; United Nations 2019, 2022). We also analyzed all available DHS for Senegal, which are available in the public domain (https://dhsprogram.com/). The DHS were used to capture differences between urban and rural areas in the main parameters of interest and to estimate under-5 mortality and fertility by wealth quintiles to introduce additional heterogeneity in some simulations, as detailed below. To generate complete life tables from under-5 mortality trends, we also used standard age-specific mortality models, retaining the 'South' pattern of Princeton model life tables (Coale, Demeny, and Vaughan 1983). This pattern is used by the United Nations for modeling mortality in Senegal (United Nations 2022).

Compared with the other three families of the Princeton tables, the South pattern corresponds to lower adult mortality for a given level of under-5 mortality, which is in line with age schedules observed in health and demographic surveillance systems in Senegal (Masquelier, Reniers, and Pison 2014).

Figure 2: Flowchart detailing input data, microsimulations, and outputs to examine selection bias and sampling errors

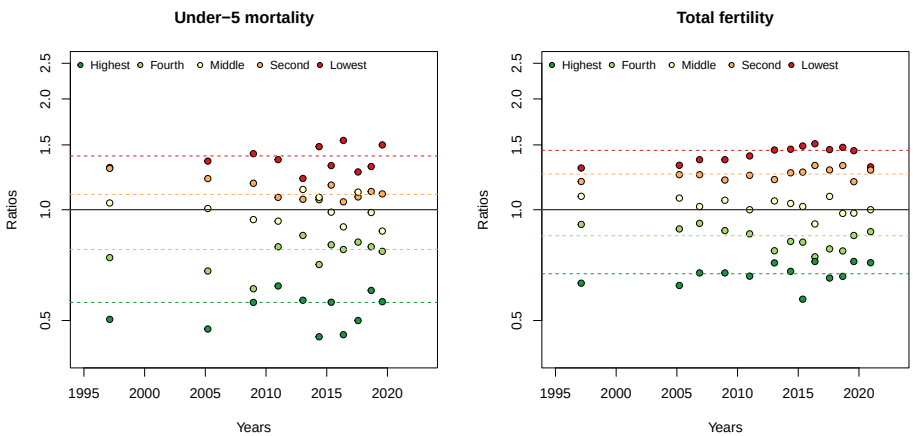


The demographic rates are introduced in SOCSIM to produce two sets of microsimulated populations:

1. In the first set, demographic rates vary only by age, sex, and place of residence. The simulations start in January 1800 and run until 2020. From 1800 to 1950, all rates are fixed, and this long period allows the population pyramid to match the age structure of the stable population determined by mortality and fertility rates. From 1950 onward, simulations are conducted in five-year intervals, with mortality, fertility, and marriage rates changing over time. The starting population is set so that the population reached at the end of the simulations has around 364,000 individuals (urban and rural areas combined). The age structures in 2013 closely match those of the populations enumerated in the 2013 census of Senegal (Figure A-4). We will refer to this first set as our 'reference' microsimulated population.

2. In the second set, we create five subpopulations exposed to specific demographic rates and combine these subpopulations together to introduce some correlation between fertility and mortality. To work with plausible mortality and fertility differentials, we extracted from all Senegalese DHS the under-5 mortality and total fertility rates for each quintile of the wealth index, a metric used to measure households' living standards (Rutstein 2008).
   As depicted in Figure 3, the ratio of mortality in each quintile to the total mortality remains stable across successive surveys in Senegal. The same trend is observed for fertility, with the middle quintile ratios approximating 1. By scaling the under-5 mortality rates with the average ratios for quintiles 1–2 and 4–5 (scaled by factors of 1.4, 1.1, 0.78, and 0.56) and similarly adjusting fertility rates (multiplied by 1.45, 1.25, 0.85, and 0.67), we generate four distinct population strata, assuming again that age-specific mortality conforms to the South pattern of Princeton life tables within each subpopulation.
   Combining these strata with our reference simulation, reflecting the middle quintile, we create a heterogeneous population wherein mortality correlates positively with fertility. Conversely, by inverting the ratios – raising mortality while reducing fertility – we generate four additional population strata. When merged with our reference simulation, this configuration results in a negative correlation between mortality and fertility.

Figure 3: Ratios of quintile-specific under-5 mortality and fertility rates to the rates for the total population in 11 successive DHS conducted in Senegal



2.2 Selection bias in simple random samples drawn from simulations

To quantify selection bias, we first examine mortality estimates derived from parental survival histories in simple random samples drawn from microsimulations where there is no correlation between the number of respondents and parental survival. This is possible in our reference urban and rural microsimulations when analyzed separately, as all individuals are exposed to the same age-specific demographic rates, ensuring that risks of dying in older adults do not depend on the number of their children eligible for an interview. Because maternal and paternal IDs are maintained in the simulations, we can reconstruct complete parental histories because we know, for each simulated individual, his or her age, sex, and their mother's and father's survival status, ages, and possibly age at death and date of death. We compare the mortality rates recalculated from all deaths and person-years generated through the simulations and those derived from parental histories among a simple random sample of women aged 15 to 49.[6] Under the absence of a correlation between mortality and family size, estimates from parental survival histories should align with the underlying rates.

Second, we draw simple random samples from our second set of microsimulations, where mortality and fertility rates vary by subgroups. We analyze the distribution of person-years and deaths by the number of potential respondents, measuring the possible bias on mortality at older ages.

Third, we test a method to correct for the bias present in parental survival histories due to correlation between family size and mortality, drawing upon insights from the literature examining selection biases affecting sibling survival histories (Obermeyer et al. 2010; Masquelier 2013; Feehan and Borges 2021). Building on Gakidou and King (2006), we propose to divide the person-years of exposure and deaths by the inverse of the number of potential respondents. This approach aims to generate an unbiased estimator for parental mortality in families where at least one adult respondent is present during the survey period. Determining the count of potential respondents can be done either by collecting parental histories after a sibling survival module or by incorporating an additional question on the number of surviving daughters aged 15 to 49 at the time of the survey. With the different simulations, we measure the bias stemming from the absence of information concerning families with no respondents.

[6] Confidence intervals are calculated using Chiang's (1984) method.

2.3 Sampling errors in two-stage stratified cluster samples drawn from the 2013 census

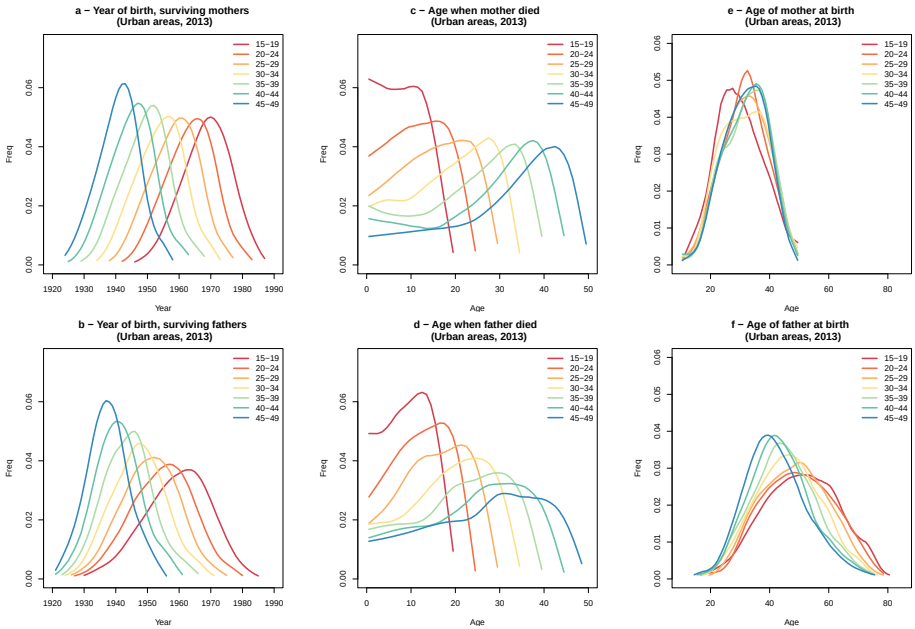
To assess whether parental histories can provide precise estimates of mortality at older ages from national demographic surveys, we need to go beyond simple random draws in simulations and replicate the DHS sampling procedure. Because individuals are not organized by households in SOCSIM, nor are they spatially distributed into clusters, we cannot draw samples comparable to the DHS directly from the simulations. We thus combine our reference simulations with the 2013 census, which served as the sampling frame for several DHS surveys (from the 2014 Continuous Survey to the 2020–2021 Malaria Indicator Survey). We requested and obtained from the National Statistical Office the anonymized data from the full count of the 2013 census, relating to parental survival, age and gender of household members, and recent household deaths, strata and enumeration areas (ANSD 2014).[7] Because no information was collected in the census on the ages of the parents or the dates of their deaths, we need the microsimulations to generate full parental histories. We augment the census by imputing from our reference simulations the dates of births and dates of death for each parent.

We extracted from the 2013 census all records referring to women aged 15 to 49, corresponding to the eligibility age range for the individual woman's questionnaire in DHS. Men aged 15 to 49 (sometimes 15 to 54 or 15 to 59) are also surveyed in the DHS, but the men's questionnaire is typically included in a subsample of the households only. In this study, therefore, we examine only selection bias and sample errors relating to samples of women aged 15 to 49.

For surviving parents, we impute a year of birth by randomly selecting a year within distributions calculated from the microsimulations by age group of the respondent, gender of the parent, and type of residence ($7 \times 2 \times 2 = 28$ distributions) (see Figure 4a and 4b for urban areas). For parents reported as deceased, we impute a year of death by randomly selecting the age of respondents at the death of parents from the simulations, again working with 28 distributions (Figure 4c and 4d). The year of death is calculated as 2013 minus the difference between the age of the respondent at the time of the census and the imputed age at the death of the parent. The year of birth of parents is obtained by sampling from the distributions of the age of parents at the birth of the respondents (Figure 4e and 4f).

[7] The description of the census data and access conditions are detailed on the ANSD website: https://anads.ansd.sn/index.php/catalog/51. A 10% sample of the microdata is also accessible via the IPUMS International project: https://international.ipums.org/international/index.shtml.

Figure 4: Distributions of years of birth of surviving mothers (a) and fathers (b), age of respondents at the death of their mother (c) and father (d), and age of mothers (e) and fathers (f) (among those deceased before 2013) at the birth of respondents in the simulations (urban areas, for 2013, for five-year age groups of respondents)



Note: Age-specific distributions are smoothed using Friedman's super smoother in R. Estimates refer to all parents of women aged 15 to 49.

We then draw samples from the augmented census, as if a DHS survey had been conducted in 2013 and had included a module on parental survival histories. We replicate the sampling procedure used for the 2017 continuous DHS survey in Senegal (see Appendix B). To vary the sample size, we multiply the number of enumeration areas by a factor ranging from 0.25 to 2.5 to obtain between 98 and 1,000 enumeration areas. In total, we construct 100 different samples for each of the 10 sample sizes from our augmented census data, with the number of households ranging from 2,200 to 22,000. By comparison, in standard DHS surveys conducted since 2000, the sample sizes have varied from 3,536 to 636,669 households, with the largest survey being the NFHS in India.[8] The median sample size in standard DHS since 2000 is 11,684 households. We compute sampling

[8] See https://dhsprogram.com.

weights to be applied to the parental survival histories, following procedures detailed in ICF International (2012). We assume that all women who fall into the sample agree to take the survey and calculate only the design weights.

For each of the 1,000 samples, we recalculate mortality rates from the parental survival data and use the recent deaths reported in the households and the population enumerated in the households to generate a second set of mortality estimates. This second set represents the current approach used to estimate old-age mortality from surveys (Bendavid, Seligman, and Kubo 2011). We compute the summary index ${}_{30}q_{50}$, corresponding to the risk for an adult aged 50 to die before reaching age 80. For the parental histories, the age-specific mortality rates calculated for all parents in the full augmented census serve as our benchmark. For rates derived from recent household deaths, our reference is the life table based on all deaths and population counts from the full census dataset. Confidence intervals around the probability ${}_{30}q_{50}$ are computed using Jackknife resampling (Pedersen and Liu 2012).

To evaluate the method's performance, we examine the accuracy, precision, and coverage of the estimates from parental survival histories. Accuracy refers to how close these estimates are to the true values. It is quantified using the mean percentage errors (MPE) for each sample size n, calculated as

$$MPE_n(\%) = 100 \times \frac{\sum_s \frac{q^{\text{PSH}} - q^{\text{true}}}{q^{\text{true}}}}{100},$$

where s iterates over the samples. We divide by 100 as we have 100 samples for each size n, and multiply by 100 to express the metric as a percentage.

Precision characterizes the consistency of the parental histories estimates. It is assessed through two measures: the standard error of our estimator across the 100 samples, indicating the variability of estimates across samples, and the average coefficient of variation, which reflects the width of confidence intervals around the estimates.

Coverage pertains to the proportion of true values falling in the confidence intervals. For each sample s, we define $\gamma_s = 0$ if the 95% confidence interval contains the true probability ${}_{30}q_{50}$, and $\gamma_s = 0$ otherwise. The coverage for each sample size n is then computed as

$$Cov_n(\%) = 100 \times \frac{\sum_s \gamma_s}{100},$$

3. Results

In Subsection 3.1, we assess the magnitude of selection bias in simple random samples drawn from the microsimulated populations, and we explore in Subsection 3.2 the sampling errors based on a stratified two-stage cluster design applied to the augmented census.

3.1 Selection bias

As expected, we find that parental survival histories provide unbiased estimates of the age-specific risks of dying in the absence of correlation between mortality and fertility, even if there are multiple respondents per mother and some mothers are not reported at all. This is illustrated in Figure 5a. This figure presents three sets of age-specific mortality rates ($_nq_x$) for the period 2015–2020: (1) The thick line corresponds to rates introduced in the simulation, (2) the thin black line refers to rates recalculated from the population and deaths that were generated during the simulation, and (3) rates in oranges were computed based on the parental survival histories as if data had been collected in a simple random sample of 10,000 women aged 15 to 49. This example is based on the reference simulation for urban areas and for female mortality, but we find unbiased estimates in all other cases.

Figure 5b shows the probabilities $_{30}q_{50}$ within each population stratum for the second simulation set (with positive and negative correlation between mortality and fertility), as calculated from parental histories and as obtained from all deaths and person-years. Again, when considering each population stratum separately, estimates from parental survival are unbiased. The median ratio of the true estimates to the parental estimates of $_{30}q_{50}$ equals 1.00 for females and 1.02 for males. In all but one case, the true probability $_{30}q_{50}$ is included in the 95% confidence intervals around the estimate obtained from 10,000 respondents.

Figure 5: Mortality rates (‰) introduced in the simulation (for urban areas) and reestimated from parental survival histories (PSH) in a random sample of respondents aged 15 to 49 for the period 2015–2020 (a: age-specific $_nq_x$ in the reference simulation for urban areas, b: probability $_{30}q_{50}$ in each population stratum defined by quintile-specific rates)

Figure 6a shows the distribution of women aged 50 to 79 over the period 2015–2020, by survival status in 2020 and the number of adult daughters aged 15 to 49 who would be eligible for an interview in that year. This distribution is drawn from our reference simulation for urban areas, based on 5,000 women aged 50 to 79 selected at random. Our sample is now drawn from the pool of women aged 50 to 79 (instead of respondents aged 15 to 49) to be able to compute mortality rates for older women without any respondents. All women are exposed to the same set of mortality and fertility rates, and yet the percentage of women who will die between 2015 and 2020 declines as the number of respondents increases. This is because women with fewer respondents are older on average. When controlling for age and calculating the probability $_{30}q_{50}$ by the number of respondents, differences in mortality are small and nonsignificant (Figure 6b). The estimate derived from parental histories is virtually the same as the probability estimated directly from the mortality experience of the 5,000 women (247‰ against 240‰). Women with fewer respondents are also more likely to have died at the beginning of the five–year period when mortality was slightly higher than closer to the survey, but this effect is small here as we are estimating mortality over the last five years. Extending the reference period would

exacerbate the differences in mortality by the number of respondents and the possible selection bias, even in the absence of any correlation between family size and mortality.

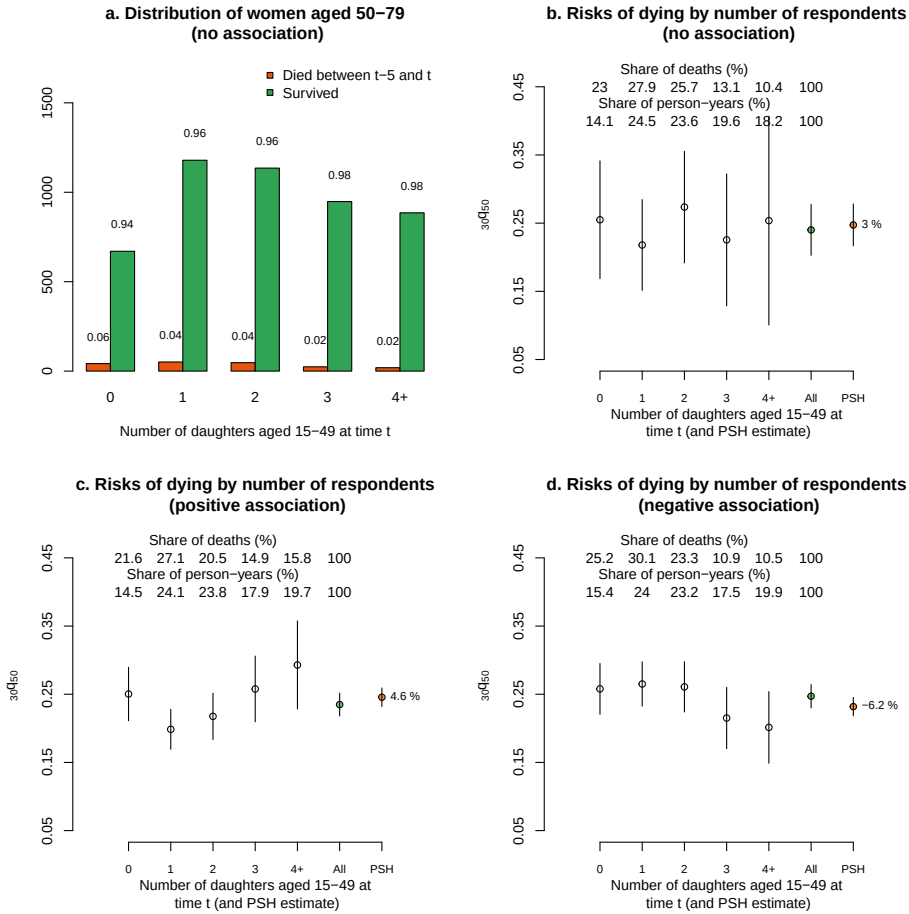
To introduce heterogeneity in our sample, we now combine the different strata presented in Figure 5b to form two large populations, one characterized by a positive correlation between family size and mortality, and another population where such correlation is negative.[9] We recalculate the 'true' mortality of these populations combined to establish the reference probability ${}_{30}q_{50}$.

Mortality varies substantially across strata, as illustrated in Figure 5b. In the pooled population with a positive correlation between mortality and fertility, women in the poorest quintile have a risk of dying between ages 50 and 80 estimated at 321‰, compared to 195‰ in the richest quintile. Such a high ratio in mortality at ages 50 to 79 ($\times$ 1.6) corresponds to larger disparities in under-5 mortality; mortality below the age of 5 in the poorest quintile is 2.5 higher than in the richest quintile, as observed in the DHS (Figure 3). By design, there is also a positive association between mortality and the number of respondents: Women in the poorest quintile have on average 2.8 daughters alive and aged 15 to 49 at the time of the survey, against 1.6 in the richest quintile. Mortality levels, therefore, increase with the number of respondents: The ${}_{30}q_{50}$ probability increases from 250‰ (in women with no respondents) to 293‰ (in women with four respondents or more). Note, however, that the disparities by number of respondents are smaller than by strata, as each stratum contributes to the rates calculated by the number of respondents in varying proportions. The overall bias on the ${}_{30}q_{50}$ probability remains, therefore, very modest (+4.6%) because these differences according to the number of respondents are further amalgamated in the overall estimate. In this simulation, 62.5% of deaths and 65.8% of person-years occur in families with between one and three respondents aged 15 to 49, where mortality levels vary relatively little (from 199‰ to 258‰). In other words, in a simulation with plausible and large mortality differentials and a strong association between mortality and fertility, similar to what is observed in Senegal by wealth quintiles, these disparities do not translate into a significant bias in the estimates derived from parental survival histories.

The same observation can be made with the population created by inverting the ratios presented in Figure 3: Women in the last stratum now have on average 1.5 fewer respondents than those in the last stratum, but they face much higher risks of dying (327 versus 193‰). And yet the bias on the ${}_{30}q_{50}$ probability remains modest: –6.2%. This is again because the disparities in mortality by strata are diluted by differences in mortality by the number of respondents and because most of the deaths and person-years are concentrated in families of intermediate size (one to three respondents).

[9] Because the strata have varying mortality and fertility rates, the final populations also vary, and to give equal weight to each stratum (or quintile), we select from each stratum 5,000 women who will be between 50 and 80 years old over the period 2015–2020.

Figure 6: Distribution of women aged 50 to 79 years 15 years before the survey by survival status and number of adult daughters eligible to an interview in 2020, and corresponding risk of dying between 50 and 80 years (‰, in the positive and negative correlation scenarios, for urban areas)



In the appendices, we report on a second approach, which is less realistic but provides a more diverse perspective. It consists of enforcing a higher correlation between the numbers of respondents and mortality through oversampling families with a surviv-

ing mother among small sibships. It confirms that selection bias should not introduce important distortions.

In addition, these biases can be in part adjusted, building on the literature on sibling survival. For sibling histories, Gakidou and King (2006) propose a method of adjustment, which consists of reweighting the observations according to the size of the sibship. We can apply a similar principle to parental histories, dividing the deaths and person-years of exposure by $1/S$, where S refers to the total number of potential respondents. The remaining bias is caused by the absence of information for parents who have no surviving adult children. In our reference urban simulation, 14% of women aged 50 to 79 have no daughter eligible for an interview in 2020. While the naive estimate leads to overestimating mortality by 4.6% in the presence of a positive correlation (Figure 6c), reweighting deaths and person-years by $1/S$ reduces this bias to –2.3%. Similarly, when imposing a negative correlation (Figure 6c), there is a downward bias of 6.2% with the naive estimate, against 1.9% with the adjusted estimate.

3.2 Sampling errors

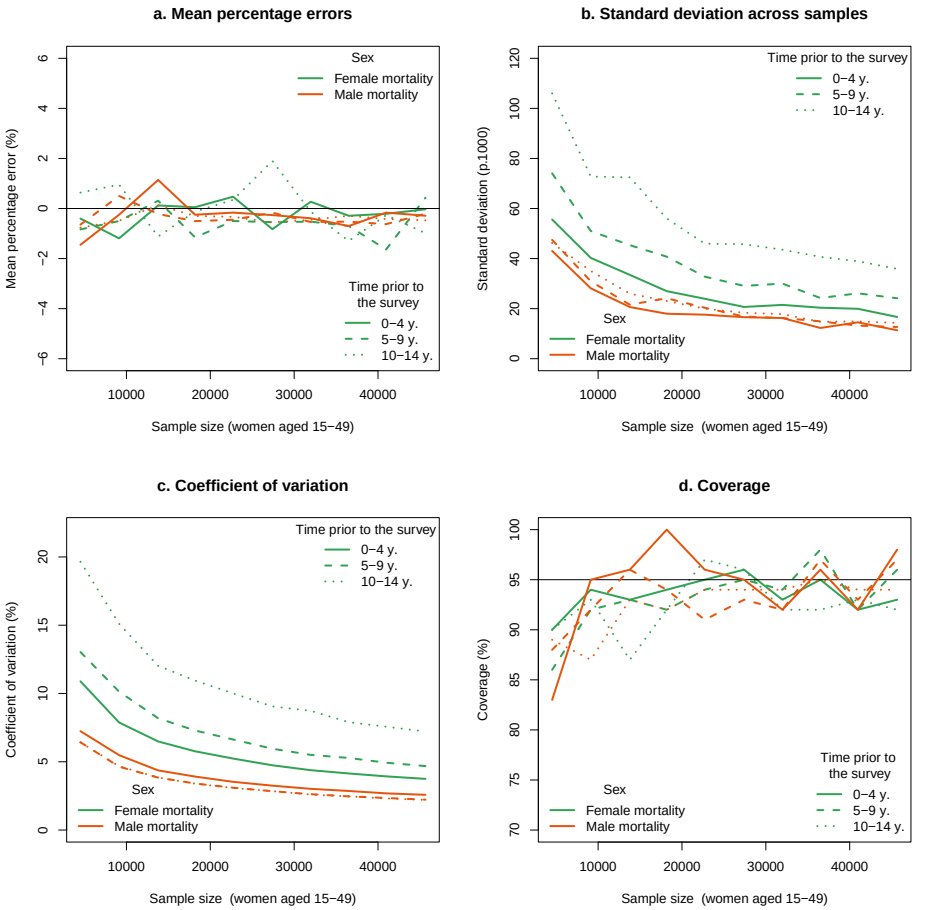
We now turn to the issue of quantifying the sample sizes required to derive precise estimates. Figure 7 presents metrics of accuracy, precision, and coverage computed from 1,000 two-stage stratified samples drawn from the augmented census, with sample sizes varying from about 4,500 to 45,700 eligible women aged 15 to 49. The corresponding values are displayed in Table A-2 in the Appendix, for three five-year periods before data collection. It should be kept in mind here that no reporting error has been introduced, as we are evaluating only sampling errors. Estimates vary around the true value, calculated from the survival of parents of all eligible women in the census, but the relative errors remain small, less than 2% for all three periods before data collection, regardless of sample size and gender of parent. There is some variation across the 100 samples drawn for each size: Standard deviations for the most recent period before the survey are around 20‰ for female mortality when the sample reaches 27,400 respondents, for a reference probability of 471‰. The estimates for the period 10 to 14 years before data collection are less precise. In men, the standard deviations are smaller while risks of dying are higher (from 505‰ to 530‰). Overall, estimates from parental histories can be considered precise for sample sizes comparable to or larger than the 2017 DHS (8,800 households and about 18,200 respondents).[10] Uncertainty around estimates calculated over five-year periods remains modest, with coefficients of variation lower than 8% for men even for the smallest samples (2,200 households and about 4,500 respondents), and lower than 10% for

[10] More precisely, the final sample size in the 2017 DHS was 8,380 households and 16,787 women aged 15 to 49 because, out of the 8,800 households selected, only 8,522 were occupied and 8,380 could be interviewed. Here we assume that all households are occupied and the response rate is 100%.

women for the two periods closest to the survey from samples comprising at least 6,600 households and about 13,800 respondents. Confidence intervals are wider for mothers than for fathers because mothers are younger on average and face lower mortality rates. In terms of coverage, the percentage of confidence intervals containing the true value is close to 95%, for both fathers and mothers, except for the smallest samples ($<$ 13,800 respondents).

The coefficients of variation around the ${}_{30}q_{50}$ probability can be compared to those obtained when calculating mortality for younger adults, aged 15 to 49, from sibling survival histories. These are a widely accepted source of mortality to track changes in premature adult mortality, despite the uncertainty in the estimates, often requiring to model trends and/or age patterns (Timæus and Jasseh 2004). In both types of questionnaire modules, the target probability covers an interval of 30 to 35 years of exposure, but mortality levels captured in sibling histories are lower than in parental histories since siblings are approximately the same age as respondents. This is compensated by the fact that respondents provide information on more relatives in sibling histories (5.6 siblings on average in the 2017 DHS survey in Senegal). As a result, coefficients of variation are broadly similar. In the 2017 DHS survey, the coefficient of variation around the female probability ${}_{35}q_{15}$ was 9% for the period 0 to 4 years before the survey and 13% for the 5 to 9 and 10 to 14 years survey periods, based on a sample of 16,787 women of reproductive age. With a comparable sample size, the 2017 DHS could have measured the ${}_{30}q_{50}$ probability with a similar precision with parental histories. By interpolating within Table A-2, we estimate that with the same number of respondents, the coefficients of variation for female mortality would have been 6% for the 0 to 4 years before the survey, 8% for the 5 to 9 years period, and 11% for the 10 to 14 years period. For men, the estimates derived from fathers between 50 and 80 years of age would have been even more precise than the estimates derived from brothers between 15 and 50 years of age: a coefficient of variation lower than 4%, regardless of the reference period, compared 9% (0 to 4 years prior to the survey) or 15% (10 to 14 years) with sibling histories.

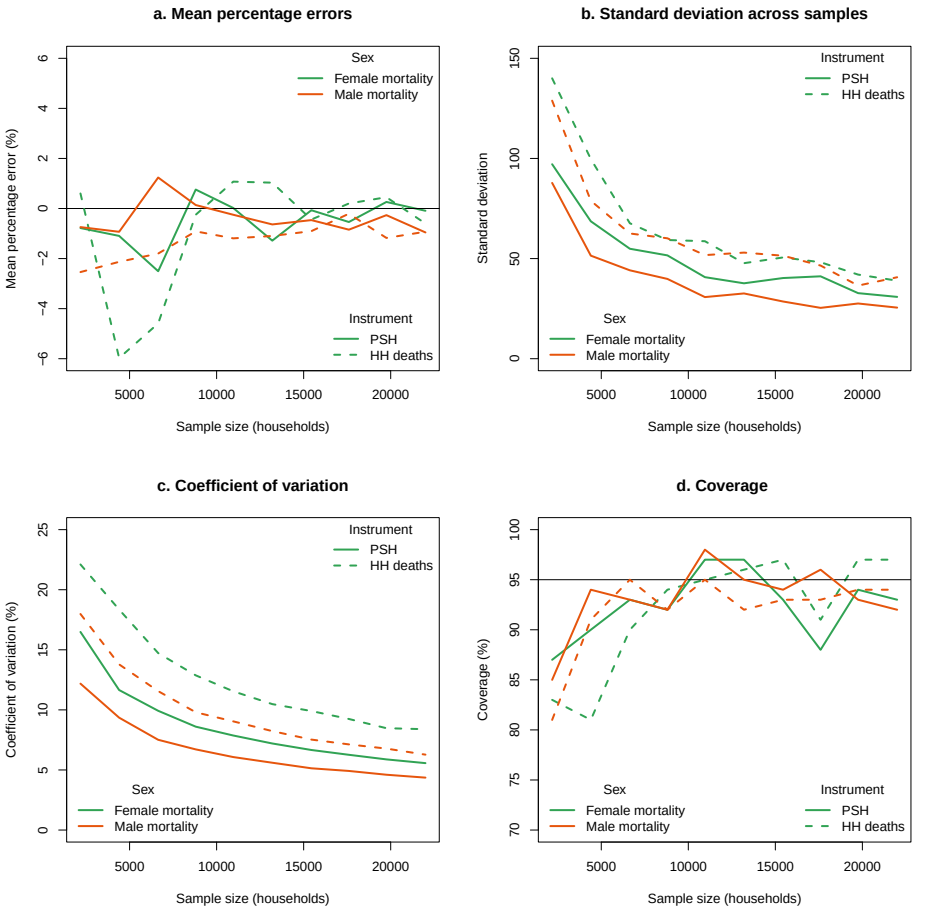
Figure 7: **(a) Mean percentage errors in the $_{30}q_{50}$ probability derived from parental histories, (b) standard deviation across 100 samples, (c) coefficient of variation around the $_{30}q_{50}$ estimate, and (d) coverage of the 95% confidence intervals, by sex and reference period before data collection**



It is also relevant to compare parental histories with the method currently used to measure mortality over 50, namely deaths reported in households in the last 12 months. To make this comparison, we recalculated the $_{30}q_{50}$ probabilities for the 12 months prior

to the census from the parental histories: Our reference becomes the probability calculated for this very short reference period from the parents of all women aged 15 to 49 listed in the census. Estimates derived from recent household deaths in each sample are compared to the life table obtained from household deaths in the full census. Figure 8 shows the mean percentage errors, the standard deviation coverage across 100 samples, the coefficients of variation, and coverage values for estimates derived from parental histories and those obtained from recent household deaths (see also Table A-3 in the Appendix). The sample sizes are presented in terms of the number of households, as recent household death data are typically collected through household questionnaires (e.g., India NFHS 2015–2016 and 2019–2021 surveys). It is important to note that the average number of household members in the 2017 DHS in Senegal was quite high, at 8.7. When considering only the most recent survey in each country, the average household size across all DHS varied from 2.5 (Ukraine 2007) to 8.2 (Senegal 2019), with a median size of 4.5 household members. In smaller samples ($<$ 10,955 households), the mean percentage errors in estimates derived from recent deaths are larger than those from parental histories, and they are similar when the sample sizes increase. Estimates from parental histories are also more precise as the standard deviations across samples are always lower than in estimates from recent household deaths, and the coefficients of variation are smaller. Logically, there is more uncertainty around the mortality calculated over this small reference period, but with samples of more than 6,600 households, the coefficient of variation falls below 10% for both sexes with parental histories. This is usually considered acceptable (Pedersen and Liu 2012). Finally, the coverage values are broadly similar across survey instruments and are close to 95%, except in small samples ($<$ 6,600 households).

Figure 8: **(a) Mean percentage errors in the $_{30}q_{50}$ probability derived from parental histories and recent household deaths, (b) standard deviation across samples, (c) coefficient of variation around the $_{30}q_{50}$ estimates, and (d) coverage of the 95% confidence intervals, by sex for the single year before the census**



4. Discussion

The empirical basis for tracking changes in the mortality of people over the age of 50 in low- and middle-income countries is still highly partial. Innovative methods to measure mortality beyond the age of 50 years in the absence of efficient vital registration systems are urgently needed.

In this study, we combined microsimulations and census data from Senegal to evaluate the potential of parental survival histories for measuring old-age mortality. Beyond this case study on Senegal, our study opens up possibilities for exploiting the large set of parental survival data collected through censuses and surveys since the 1960s in low- and middle-income countries. This would allow refining the sample size calculations in a variety of contexts: for example, where household sizes are smaller or in settings with different fertility and mortality age patterns. This could also allow a greater use of the summary data already collected. Existing indirect methods based on orphanhood rely on standard age patterns of fertility and mortality, and biases are introduced when demographic trends deviate from such patterns: for example, due to the HIV/AIDS epidemic (Timæus and Nunn 1997). When applied to a single inquiry, these methods are also underpinned by a strong assumption that past mortality trends have been regular (Brass and Bamgboye 1981). With microsimulations, detailed parameterizations tailored to specific countries are possible, with much more flexibility and by explicitly modeling the random nature of the processes involved. Microsimulations could therefore improve statistical inferences about the demographic parameters underlying the proportions of surviving parents observed in censuses and surveys. This would entail generating a large number of simulations with varying parameters and adjusting the set of input rates to match the real-world proportions of parents surviving – a process akin to simulation calibration (Zagheni 2010). However, computational limitations may constrain the number of simulations, and reconstructing the long-term demographic trends to inform the simulations can be time-consuming, even for a single country. This remains an avenue to consider to analyze the summary data on orphanhood collected in the past, but future data collection on adult mortality should move toward full parental histories.

Here we assessed the magnitude of selection bias that could plague the estimates derived from parental survival histories and examined the sample sizes needed to produce precise estimates. Our results suggest that the effects of selection bias on the risks of dying between ages 50 and 80 should remain small. This conclusion is based on simulations that reflect the magnitude of child mortality and fertility differentials observed in Senegal at the national level. They are also consistent with earlier results involving indirect estimates (Palloni, Massagli, and Marcotte 1984; Blacker 1984). However, they need to be confirmed with real data. In another study, the extent of the association between mortality and adult daughters aged 15 to 49 who would be eligible for an interview was measured based on data from health and demographic surveillance systems, also in

Senegal (Menashe-Oren et al. 2024). There, it was observed that older adults without known surviving children above age 15 faced higher risks of dying. Mortality estimates based on parental histories were therefore lower than the true levels of mortality. It is possible, however, that the magnitude of the bias on risks of dying was overestimated due to structural limitations in health and demographic surveillance systems. In particular, the total number of respondents eligible to an interview in each family was unknown, as all parents could not be matched to their offspring. Our study also showed that apparent correlations between mortality and the number of respondents can be spurious and introduced by mortality trends or different age structures of the pool of parents being reported for each family size. In simulations, we could not measure the bias directly, as we had to artificially create a correlation between mortality and family size. Nevertheless, when we assessed the possible effect of various patterns of associations between mortality and family size, biases remained small. In addition, if the number of potential respondents is collected, weighting schemes to adjust for selection bias can be used. By dividing the number of person-years and deaths by S, the number of eligible respondents in each family, one will recover an unbiased estimate for families with at least one respondent. To detect any possible bias, this weighted estimate should be systematically compared with rates obtained without weighting, and estimates calculated according to the number of children of eligible age should be compared with each other. An adjustment for families with zero respondents remains to be developed, as is the case for sibling histories when weights based on sibship sizes are applied (Gakidou and King 2006). This is another area for further research.

In terms of sample sizes, we have shown that parental histories could provide estimates of the probability ${}_{30}q_{50}$ that are at least as precise as those derived from reports on deaths that occurred in the past 12 months. Unlike estimates of recent deaths, they allow the reconstruction of mortality trends with good precision. With approximately 16,800 women of reproductive age in the sample, as in the 2017 Senegal DHS survey, the ${}_{30}q_{50}$ estimates would be even more precise than those inferred from sibling survival for the probability ${}_{35}q_{15}$. Parental histories thus appear to be a viable complement to this widely used survey module to extend the perspective above the age of 50. As the number of questions is limited (three or four, depending on the survival status of each parent), the impact of adding this module on the overall cost of the survey and interviewer training time is marginal. The module could also be integrated into the men's questionnaire to reduce confidence intervals and enable additional consistency tests.

Our study has important limitations. First, we introduced heterogeneity in our simulations using quintiles of the wealth index, and other variables could have been considered, such as education. However, we also quantified the selection bias using a large range of correlations between family size and mortality, and our conclusions are similar. Second, we varied mortality and fertility in population subgroups but did not investigate the effects of any correlation in the risks of dying within families. For example, the death

of an older parent during the study period may impact the survival rates of their adult children, potentially introducing additional selection biases. Similarly, the risks of dying among older adults could be influenced by the recent loss of an adult child. However, while there is extensive literature documenting the excess mortality of orphans during childhood (Zaba et al. 2005; Ronsmans et al. 2010; Scott et al. 2017), little is known about the potential excess mortality among adults aged 15 to 49 who lose an older parent, or among older adults who lose an offspring aged 15 to 49, particularly in low- and middle-income countries. Third, internal or international migrations are not modeled in our simulations. Our conclusions on selection bias and the sample sizes required for a nationwide estimate are unlikely to be significantly affected by not accounting for migration. However, using parental survival histories to derive subnational estimates could introduce misclassification errors if respondents' characteristics are used as proxies for their parents' characteristics. In the 2013 census, approximately 15% of Senegal's native resident population were born in a region different from their region of residence at the time of enumeration (ANSD 2014). To obtain more accurate disaggregated estimates, specific questions about where parents lived or died could be included. Fourth, we did not consider misreporting errors that could taint the parental survival histories. One source of reporting error affecting orphanhood data is 'adoption bias,' caused by confusion between biological and foster parents (Blacker and Mukiza-Gapere 1988; Robertson et al. 2008). This error leads to the omission of certain parental deaths and disturbs the indirect estimation of mortality, as this method approximates the duration of exposure to the risk of dying as a parent based on the respondent's age. Because ages at the time of the survey or death are reported in full parental histories, these approximations are unnecessary, and whether the reports refer to a biological parent or a foster parent becomes irrelevant. While the omission of a recent death of a biological parent, replaced by a foster parent in the reports, remains possible, this is likely an infrequent error among adult respondents. A more serious problem is probably a lack of knowledge about the ages of surviving parents, or, in the case of death, the ages at death and dates of death. For example, in the 1987 Burundi survey, about 25% of women could not date their mother's death and there was some strong heaping on round numbers in the distribution of years since death (Makinson 1993). There could also be systematic age errors or transfers of deaths out of the reference period, as observed in reports on siblings (Masquelier et al. 2021). Unfortunately, few studies have collected data on parental survival, and evidence on misreporting errors is scarce. In this context, it seems prudent to use different approaches to measure old-age mortality and triangulate them; future surveys that aim to estimate adult mortality should include questions on recent household deaths in addition to modules on sibling and parental survival.

As mentioned in the introduction, the idea of collecting full parental histories is not new, having been tested in the 1990s during a survey in Vietnam (Hirschman, Preston, and Loi 1995). Since then, the increasing concentration of deaths among adults over 50

has only reinforced the relevance of this survey instrument. The integration of parental histories into DHS surveys had already been recommended by several demographers in 2014 during the review process of the optional special topic questionnaire modules for the DHS-7 round.[11] This call was reiterated by the Population Division of the United Nations in 2019 during the review of tools for DHS-8 round.[12] More recently, parental histories were recommended again at a United Nations Expert Group Meeting on *Innovative methods to measure the impact of COVID-19 on mortality through surveys and censuses*.[13] We hope that our study will strengthen the case for these recommendations and help convince countries, funders, and technical support agencies of the relevance of this module to measure old-age mortality.

5. Acknowledgments

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R01HD088516, and by the National Institute on Aging under award number R03AG070660.[14] This study was also supported by the Bill & Melinda Gates Foundation, as part of the Rapid Mortality Surveillance during COVID-19 (RaMMPS) Project (INV-023211).

We thank two anonymous reviewers for helpful suggestions and comments.

[11] See https://www.ihsn.org/sites/default/files/resources/Suggestions_for_DHS_adult_mortality_module_IMT-PG.pdf.

[12] See https://userforum.dhsprogram.com/index.php?t=tree&goto=17391&.

[13] See https://unstats.un.org/iswghs/events/egm-covid-mortality/Background-paper-adult-mortality-20220817.pdf.

[14] The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Adjiwanou, V., Alam, N., Alkema, L., Asiki, G., Bawah, A., Béguy, D., Cetorelli, V., Dube, A., Feehan, D.M., Fisker, A.B., Gage, A., Garcia, J., Gerland, P., Guillot, M., Gupta, A., Haider, M.M., Helleringer, S., Jasseh, M., Kabudula, C., LeGrand, T., Masquelier, B., Menashe-Oren, A., Moultrie, T.A., Queiroz, B.L., Reniers, G., Soura, A., Timæus, I., and You, D. (2020). Measuring excess mortality during the COVID-19 pandemic in low- and lower-middle income countries: The need for mobile phone surveys. Charlottesville: Center for Open Science. doi:10.31235/osf.io/4bu3q.

Alburez-Gutierrez, D., Mason, C., and Zagheni, E. (2021). The "sandwich generation" revisited: Global demographic drivers of care time demands. *Population and Development Review* 47(4): 997–1023. doi:10.1111/padr.12436.

ANSD (2014). Rapport définitif du RGPHAE 2013. Dakar: Agence Nationale de la Statistique et de la Démographie. https://ireda.ceped.org/inventaire/ressources/sen-2013-rec-o1_rapport-definitif.pdf.

ANSD and ICF (2018). Sénégal: Enquête démographique et de santé continue (EDS-continue) 2017. Dakar: ANSD/ICF. https://www.ansd.sn/sites/default/files/2022-11/Rapport%20Final%20EDS%202017_0.pdf.

Bendavid, E., Seligman, B., and Kubo, J. (2011). Comparative analysis of old-age mortality estimations in Africa. *PLoS ONE* 6(10): e26607. doi:10.1371/journal.pone.0026607.

Blacker, J. (1984). Experience in the use of special mortality questions in multi-purpose surveys: The single-round approach. In: United Nations (ed.). *Data bases for mortality measurement*. New York: United Nations: 79–89.

Blacker, J. and Mukiza-Gapere, J. (1988). The indirect measurement of adult mortality in Africa: Results and prospects. In: International Union for the Scientific Study of Population (ed.). *African Population Conference, Dakar, Senegal, November 7–12, 1988*. Liège: IUSSP: 23–38.

Bongaarts, J. and Casterline, J. (2022). Extramarital fertility in low- and middle-income countries. *Demographic Research* 47(3): 59–72. doi:10.4054/DemRes.2022.47.3.

Brass, W. and Bamgboye, E. (1981). The time location of reports of survivorship: Estimates for maternal and paternal orphanhood and the ever-widowed. London School of Hygiene and Tropical Medicine: Centre for Population Studies (research paper 81-1).

Chackiel, J. and Orellana, H. (1985). Adult female mortality trends from retrospective questions about maternal orphanhood included in census and surveys. In: *International Population Conference, Florence. Volume 4*. Liege: IUSSP: 39–51.

Chasukwa, M., Choko, A.T., Muthema, F., Nkhalamba, M.M., Saikolo, J., Tlhajoane, M., Reniers, G., Dulani, B., and Helleringer, S. (2022). Collecting mortality data via mobile phone surveys: A non-inferiority randomized trial in Malawi. *PLOS Global Public Health* 2(8): e0000852. doi:10.1371/journal.pgph.0000852.

Chiang, C. (1984). *The life table and its applications*. Malabar: Robert E. Krieger.

Coale, A.J., Demeny, P.G., and Vaughan, B. (1983). *Regional model life tables and stable populations (2nd ed.)*. New York: Academic Press.

Coale, A.J. and McNeil, D.R. (1972). The distribution by age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association* 67(340): 743–749. doi:10.1080/01621459.1972.10481287.

Duthé, G., Pison, G., Delaunay, V., and Douillot, L. (2016). L'effet à long terme de la vie reproductive sur la mortalité des femmes en milieu rural sénégalais. *African Population Studies* 30(1): 2135–2148. doi:10.11564/30-1-796.

Feehan, D.M. and Borges, G.M. (2021). Estimating adult death rates from sibling histories: A network approach. *Demography* 58(4): 1525–1546. doi:10.1215/00703370-9368990.

Gakidou, E. and King, G. (2006). Death by survey: Estimating adult mortality without selection bias from sibling survival data. *Demography* 43(3): 569–585. doi:10.1353/dem.2006.0024.

Hajnal, J. (1953). Age at marriage and proportion marrying. *Population Studies* 7(2): 111–136. doi:10.1080/00324728.1953.10415299.

Hirschman, C., Preston, S., and Loi, V.M. (1995). Vietnamese casualties during the American War: A new estimate. *Population and Development Review* 21(4): 783–812. doi:10.2307/2137774.

ICF International (2012). Demographic and Health Survey. Sampling and Household Listing Manual. Calverton: MEASURE DHS/ICF International.

IGME (2024). Levels and trends in child mortality: 2023 Report. Geneva, New York, Washington: United Nations Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, The World Bank, UN DESA/Population Division).

Lankoandé, B., Pison, G., Masquelier, B., Soura, A.B., Zabre, P., Duthé, G., Bangré, H., and Ali, S. (2022). Estimating mortality from census data: A record-linkage study of the Nouna Health and Demographic Surveillance System in Burkina Faso. *Demographic Research* 46(22): 653–680. doi:10.4054/DemRes.2022.46.22.

Lu, F. and Vogl, T. (2023). Intergenerational persistence in child mortality. *American Economic Review: Insights* 5(1): 93–110. doi:10.1257/aeri.20210604.

Makinson, C. (1993). Estimates of adult mortality in Burundi. *Journal of Biosocial Science* 25(2): 169–186. doi:10.1017/S0021932000020472.

Masquelier, B. (2013). Adult mortality from sibling survival data: A reappraisal of selection biases. *Demography* 50(1): 207–228. doi:10.1007/s13524-012-0149-1.

Masquelier, B., Kanyangarara, M., Pison, G., Kanté, A.M., Ndiaye, C.T., Douillot, L., Duthé, G., Sokhna, C., Delaunay, V., and Helleringer, S. (2021). Errors in reported ages and dates in surveys of adult mortality: A record linkage study in Niakhar (Senegal). *Population Studies* 75(2): 269–287. doi:10.1080/00324728.2020.1854332.

Masquelier, B., Reniers, G., and Pison, G. (2014). Divergences in mortality trends in sub-Saharan Africa: Survey evidence on the survival of children and siblings. *Population Studies* 68(2): 161–177. doi:10.1080/00324728.2013.856458.

Masquelier, B. and Garbero, A. (2016). Educational differentials in adult mortality in low- and middle-income countries. *Quetelet Journal* 4(1): 7–28. doi:10.14428/rqj2016.04.01.01.

Menashe-Oren, A., Masquelier, B., Schlüter, B.S., and Helleringer, S. (2024). Estimer la mortalité aux âges élevés dans les pays à données incomplètes: Les estimations basées sur la survie des parents sont-elles affectées par des biais de sélection? *Population* 78(3): 501–522.

Mikkelsen, L., Phillips, D.E., AbouZahr, C., Setel, P.W., de Savigny, D., Lozano, R., and Lopez, A.D. (2015). A global assessment of civil registration and vital statistics systems: Monitoring data quality and progress. *Lancet* 386(10001): 1395–1406. doi:10.1016/S0140-6736(15)60171-4.

Obermeyer, Z., Rajaratnam, J.K., Park, C.H., Gakidou, E., Hogan, M.C., Lopez, A.D., and Murray, C.J.L. (2010). Measuring adult mortality using sibling survival: A new analytical method and new results for 44 countries, 1974–2006. *PLoS Medicine* 7(4): e1000260. doi:10.1371/journal.pmed.1000260.

Ouedraogo, S. (2020). Estimation of older adult mortality from imperfect data: A comparative review of methods using Burkina Faso censuses. *Demographic Research* 43(38): 1119–1154. doi:10.4054/DemRes.2020.43.38.

Palloni, A., Massagli, M., and Marcotte, J. (1984). Estimating adult mortality with maternal orphanhood data: Analysis of sensitivity of the techniques. *Population Studies* 38(2): 255–279. doi:10.1080/00324728.1984.10410289.

Pedersen, J. and Liu, J. (2012). Child mortality estimation: Appropriate time periods for child mortality estimates from full birth histories. *PLoS Medicine* 9(8): e1001289. doi:0.1371/journal.pmed.1001289.

Pison, G., Hill, K., Cohen, B., and Foote, K. (eds.) (1995). *Population Dynamics of Senegal*. Washington D.C.: National Academy Press. doi:10.17226/4900.

Robertson, L., Gregson, S., Madanhire, C., Walker, N., Mushati, P., Garnett, G., and Nyamukapa, C. (2008). Discrepancies between UN models and DHS survey estimates of maternal orphan prevalence: Insights from analyses of survey data from Zimbabwe. *Sexually Transmitted Infections* 84 S(1): 57–62. doi:10.1136/sti.2008.029926.

Rodriguez, G. and Trussell, J. (1980). Maximum likelihood estimation of the parameters of Coale's model nuptiality schedule from survey data. London, Voorburg: World Fertility Survey (Technical Bulletins 7/Tech. 1261).

Ronsmans, C., Chowdhury, M.E., Dasgupta, S.K., Ahmed, A., and Koblinsky, M. (2010). Effect of parent's death on child survival in rural Bangladesh: A cohort study. *Lancet* 375(9730): 2024–2031. doi:10.1016/S0140-6736(10)60704-0.

Rutstein, S.O. (2008). The DHS wealth index: Approaches for rural and urban areas. Calverton: DHS (Working Papers 60). http://dhsprogram.com/pubs/pdf/WP60/WP60.pdf.

Saikia, N., Bora, J.K., and Luy, M. (2019). Socioeconomic disparity in adult mortality in India: Estimations using the orphanhood method. *Genus* 75(7). doi:10.1186/s41118-019-0054-1.

Scott, S., Kendall, L., Gomez, P., Howie, S.R.C., Zaman, S.M., Ceesay, S., D'Alessandro, U., and Jasseh, M. (2017). Effect of maternal death on child survival in rural West Africa: 25 years of prospective surveillance data in The Gambia. *PLoS ONE* 12(2): e0172286. doi:10.1371/journal.pone.0172286.

Timæus, I.M. (1991). Measurement of adult mortality in less developed countries: A comparative review. *Population Index* 57(4): 552–568. doi:10.2307/3644262.

Timæus, I.M. (1992). Estimation of adult mortality from paternal orphanhood: A reassessment and a new approach. *Population Bulletin of the United Nations* 33: 47–63.

Timæus, I.M. (2013). Indirect estimation of adult mortality from orphanhood. In: Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timæus, I., and Zaba, B. (eds.). *Tools for demographic estimation*. Paris: International Union for the Scientific Study of Population: 222–243.

Timæus, I.M. and Jasseh, M. (2004). Adult mortality in Sub-Saharan Africa: Evidence from demographic and health surveys. *Demography* 41(4): 757–772. doi:10.1353/dem.2004.0037.

Timæus, I.M. and Nunn, A. (1997). Measurement of adult mortality in populations affected by AIDS: An assessment of the orphanhood method. *Health Transition Review* 7(2): 23–43.

United Nations (2019). World marriage data 2019. New York: Department of Economic and Social Affairs, United Nations. https://www.un.org/development/desa/pd/data/world-marriage-data.

United Nations (2022). World Population Prospects: The 2022 Revision. New York: UN Department of International Economic and Social Affairs Statistics Division, United Nations. https://esa.un.org/unpd/wpp.

Verdery, A.M., Smith-Greenaway, E., Margolis, R., and Daw, J. (2020). Tracking the reach of COVID-19 kin loss with a bereavement multiplier applied to the United States. *Proceedings of the National Academy of Sciences* 117(30): 17695–17701. doi:10.1073/pnas.2007476117.

Wachter, K., Blackwell, D., and Hammel, E. (1997). Testing the validity of kinship microsimulation. *Mathematical and Computer Modeling* 26(6): 89–104. doi:10.1016/S0895-7177(97)00172-6.

Zaba, B., Whitworth, J., Marston, M., Nakiyingi, J., Ruberantwari, A., Urassa, M., Issingo, R., Mwaluko, G., Floyd, S., Nyondo, A., and Crampin, A. (2005). HIV and mortality of mothers and children: Evidence from cohort studies in Uganda, Tanzania, and Malawi. *Epidemiology* 16(3): 275–280. doi:10.1097/01.ede.0000155507.47884.43.

Zagheni, E. (2010). The impact of the HIV/AIDS epidemic on orphanhood probabilities and kinship structure in Zimbabwe [PhD thesis]. Berkeley: University of California.

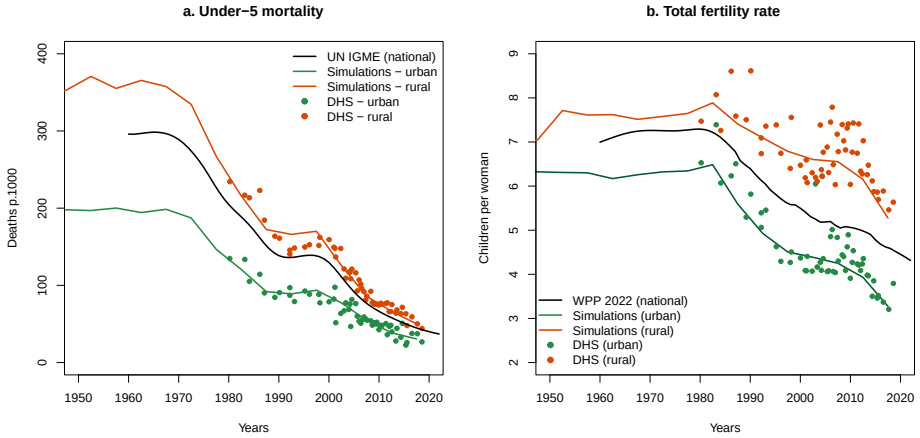
Zagheni, E. (2011). The impact of the HIV/AIDS epidemic on kinship resources for orphans in Zimbabwe. *Population and Development Review* 37(4): 761–783. doi:10.1111/j.1728-4457.2011.00456.x.

Appendices

A. Estimation of demographic rates

To build sector-specific life tables since 1950, we first estimate under-5 mortality rates (U5MR) by type of residence from the 11 Demographic and Health Surveys available in Senegal with full birth histories. Estimates are smoothed using local polynomial regression.[15] The ratio of urban/rural to total U5MR is applied to the national estimates available for each sex from UN IGME since 1960 (IGME 2024) to obtain a series of sex-specific U5MRs for each type of residence. The trends in U5MR recalculated from the simulations are displayed in Figure A-1a. These rates are combined with the South pattern of the Princeton model life tables system to generate mortality rates for all age groups (Coale, Demeny, and Vaughan 1983). Mortality is assumed to be constant before 1960.[16]

Figure A-1: Comparison of trends in (a) U5MR (both sexes combined) and (b) TFR in simulations and Demographic and Health Surveys for urban and rural areas, with UN IGME or WPP estimates for Senegal as reference



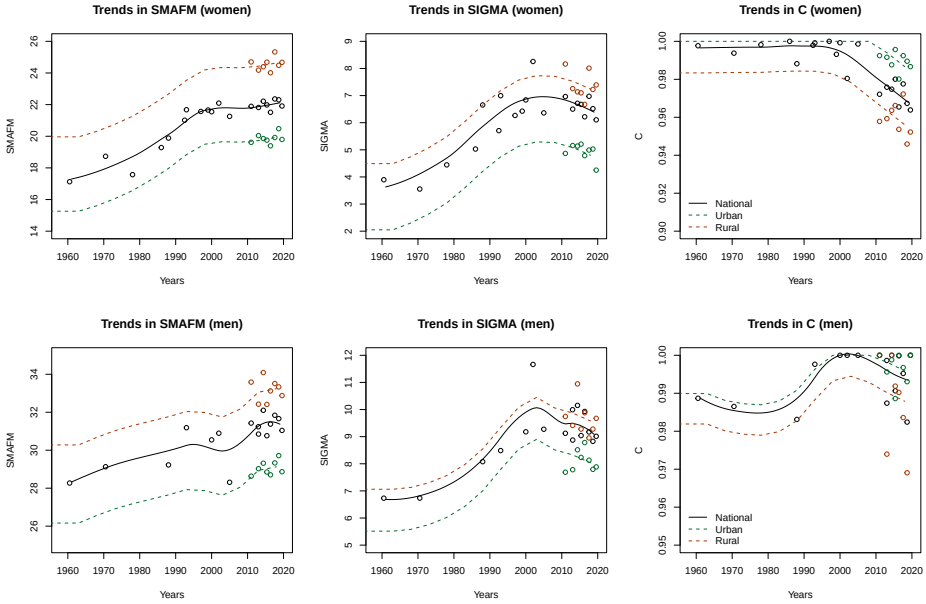
[15] Estimates were computed for five reference periods of three years before each survey. For the local polynomial regression, weights were proportional to the inverse of the variance around estimates, and span parameter was set at 0.75.

[16] Existing data sources for Senegal suggest that mortality stagnated or declined at a relatively slow pace before the 1960s (Pison et al. 1995) In any case, trends before 1960 will have a negligible impact on our results referring to parental histories collected among adult respondents aged 15 to 49 in 2013.

Age-specific fertility rates are estimated for urban and rural areas from the same DHS. Again, fertility rates are smoothed over time using a local polynomial regression, and the ratios of urban/rural fertility to the total fertility are applied to the WPP estimates to obtain rates for urban and rural areas since 1950. These ratios are assumed to be constant before 1983, the first year we can estimate them from DHS, and fertility rates are also kept constant before 1950, the first year available in WPP. Figure A-1b displays the recalculated TFRs from the simulations and compares them to WPP and DHS estimates.

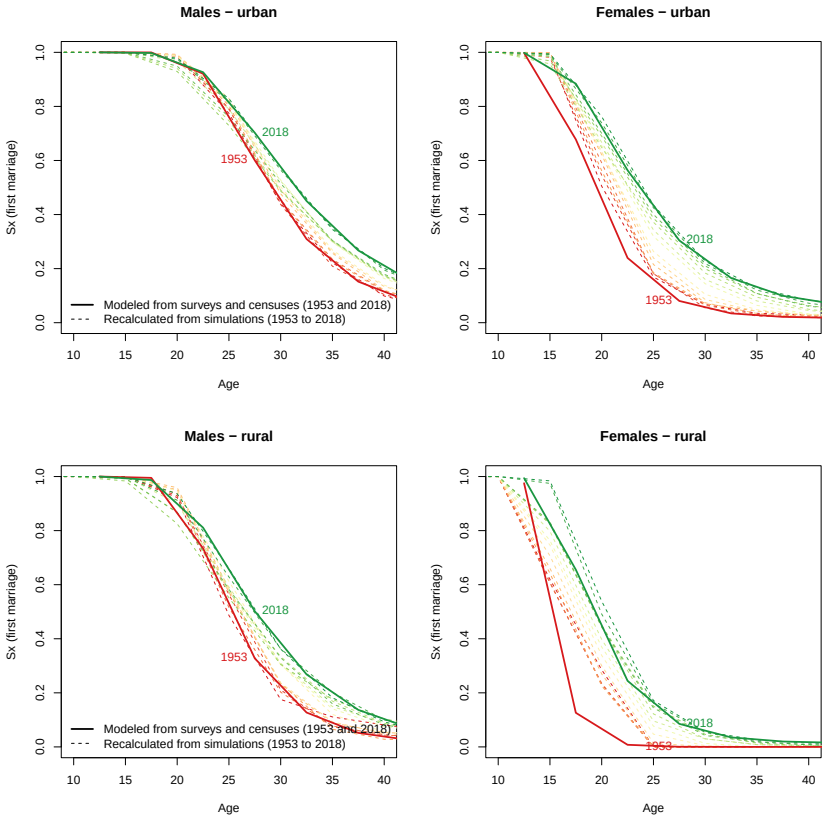
We also reconstruct long-term trends in marriage rates, because fathers are identified through only their union with mothers. We extract from the World Marriage Data three parameters of the Coale marriage model for the national level: c, the proportion of women or men in a cohort who will eventually marry; μ, the mean age at first marriage; and σ, the standard deviation of the age at marriage for those who marry (Coale and McNeil 1972). To distinguish between urban and rural areas, the same indicators are extracted from the DHS for each type of residence (see Figure A-2). The average ratio of values observed in urban/rural areas to the national estimate is used to reconstruct trends in these three indicators over time and to recalculate age-specific marriage rates (Rodriguez and Trussell 1980). To account for polygamy, we estimate the probability of remaining monogamous by age among married men in the DHS, and translate this probability into rates of entering a polygamous union (Hajnal 1953). The modeled proportions of men and women remaining never married and those recalculated from the simulations are presented in Figure A-3. Once nuptiality rates have been estimated, we compute premarital fertility rates, supposed to be 25% of marital fertility, and assume that the share of premarital births is around 10% (Bongaarts and Casterline 2022). For children born out of marriage, fathers are allocated randomly, accounting for expected age differences between parents.

Figure A-2: Trends in the proportion of adults in a cohort who will eventually marry, the mean and standard deviation of age at marriage, by type of residence, in Senegal



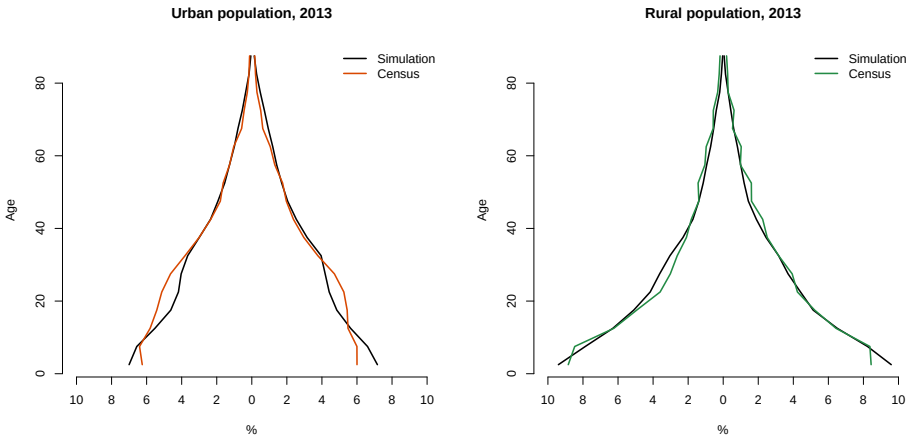
Note: SMAFM is the singulate mean age at first marriage. SIGMA is the standard deviation of age at marriage. C is the proportion of women or men in a cohort who eventually marry.

Figure A-3: Comparison of proportions of men and women never married by age, in the simulations and in censuses and surveys



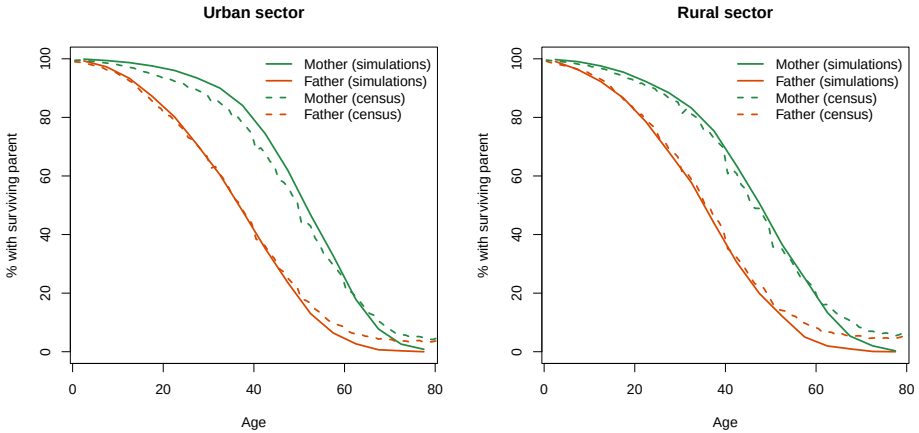
The age structure of the simulated populations in 2013 is close to those in the national census (Figure A-4), although, in urban areas, the census recorded proportionally more people in the 10 to 14 to 25 to 29 age groups, probably because internal migration is not modeled in our simulations.

Figure A-4: Comparison of age structures in the 2013 census and in simulations



The proportions of surviving parents in our simulations are also in agreement with those calculated from the census, although above age 60, proportions of household members with a surviving parent are higher in the census (Figure A-5).

Figure A-5: Parental survival by age in the simulations (in 2013) and the 2013 Senegalese census



B. The DHS 2017 sample design

The 2017 DHS used a stratified random sample drawn in two stages. Each of the 14 regions of the country was separated into an urban and a rural component to form sampling strata. The primary units were the census enumeration areas (EAs), of which there are 17,148 nationwide. These EAs were sorted within each stratum by the department and then by the borough to introduce an implicit stratification below the region. In the first stage, 400 EAs were drawn with a probability that was proportional to the number of households they contained. In the second stage, 22 households were randomly drawn from each selected EA with equal probability. Table A-1 gives the distribution of the sample in terms of EAs and households, as reported in the DHS 2017 survey report (ANSD and ICF 2018).

Table A-1: Distribution of the census enumeration areas and households in the sample by region and type of residence, DHS Senegal 2017

	Census enumeration areas			Households		
	Urban	Rural	Total	Urban	Rural	Total
Dakar	42	2	44	924	44	968
Diourbel	7	21	28	154	462	616
Fatick	8	20	28	176	440	616
Kaffrine	7	19	26	154	418	572
Kaolack	13	12	25	286	264	550
Kédougou	9	15	24	198	330	528
Kolda	11	16	27	242	352	594
Louga	10	20	30	220	440	660
Matam	8	18	26	176	396	572
Saint-Louis	17	13	30	374	286	660
Sédhiou	7	17	24	154	374	528
Tambacounda	11	16	27	242	352	594
Thiès	21	13	34	462	286	748
Ziguinchor	15	12	27	330	264	594
Senegal	186	214	400	4,092	4,708	8,800

Source: ANSD and ICF (2018).

C. Quantifying selection bias through oversampling families with surviving mother

A second, more flexible but less plausible approach to introduce a correlation between the number of respondents and adult mortality consists of oversampling certain families (in other words, certain sibships of adult respondents) according to the number of adults eligible to an interview and the survival of parents. To create a positive correlation, and thus higher mortality in families with more respondents, we oversample sibships of respondents whose parents had died recently among the larger sibships. Conversely, to create a negative correlation, we oversample small sibships when these respondents had recently lost one of their parents. By artificially varying the magnitude of the correlation between the size of the set of respondents and parental mortality, we are able to explore the magnitude of selection biases in a more diverse set of configurations than those informed by quintile-specific rates.

This is illustrated in Figure A-6. In the top panel, we start from our reference urban population with 5,000 women aged 50 to 79 (as in Figure 6a), and we create a subsample by excluding some families with a surviving mother, with a larger share of families being excluded when more daughters are eligible to an interview. We exclude 10% of families with a surviving mother when these mothers have no respondent, 20% of families with one respondent, 30% of families with two respondents, and 40% of families with three respondents. Half of the families with surviving mothers are excluded when there are four

respondents or more. This way, we artificially create a strong and positive correlation between mortality and the number of respondents, beyond what is possible through the combination of our different population strata calibrated with the wealth quintiles. For example, women with no respondents now face a risk of dying between the ages of 50 and 80, which is 1.5 times lower than women with four respondents or more. We then compute the 'true' probability ${}_{30}q_{50}$ in this subsample and compare it to what would be derived from parental histories collected in this subsample. In this example, the parental histories would overestimate the true risks of dying by only 12%, even under these quite extreme conditions. In Figures A-6c and A-6d, we use the same approach, but oversample families with surviving mothers when they have more respondents, to create a negative correlation, and again, we note that the bias on the ${}_{30}q_{50}$ probability remains limited, at –7.6%.

To go beyond these two examples, we create 200 different subsamples, and randomly select, for each family size, the proportion of families with surviving mothers being excluded, from 0% to 100%. For each subsample, we compute the correlation between the probability ${}_{30}q_{50}$ for each family size and the number of respondents available, and we relate this coefficient to the bias in the estimate of the overall ${}_{30}q_{50}$ probability, measured here as a relative difference. The correlation coefficients and relative differences are presented in Figure A-7a. As expected, when the correlation is negative, mortality is underestimated, while it is overestimated when the correlation is positive. The relative difference varies here from –18% to +24%, but in half of the subsamples, it is contained between –3% and +8%. In the 200 different subsamples, applying the weighting scheme proposed in Section 2.2 would substantially reduce the bias, with a relative difference now contained between –2.2% and 2.5% in half of the subsamples (Figure A-7b). The mean percentage error would be reduced from 7.5% to 3.2%. This confirms that selection biases are small and can easily be adjusted for.

Figure A-6: Distribution of women aged 50 to 79 years 15 years before the survey by survival status and number of adult daughters, and corresponding risk of dying between 50 and 80 years (in the positive and negative correlation scenarios)

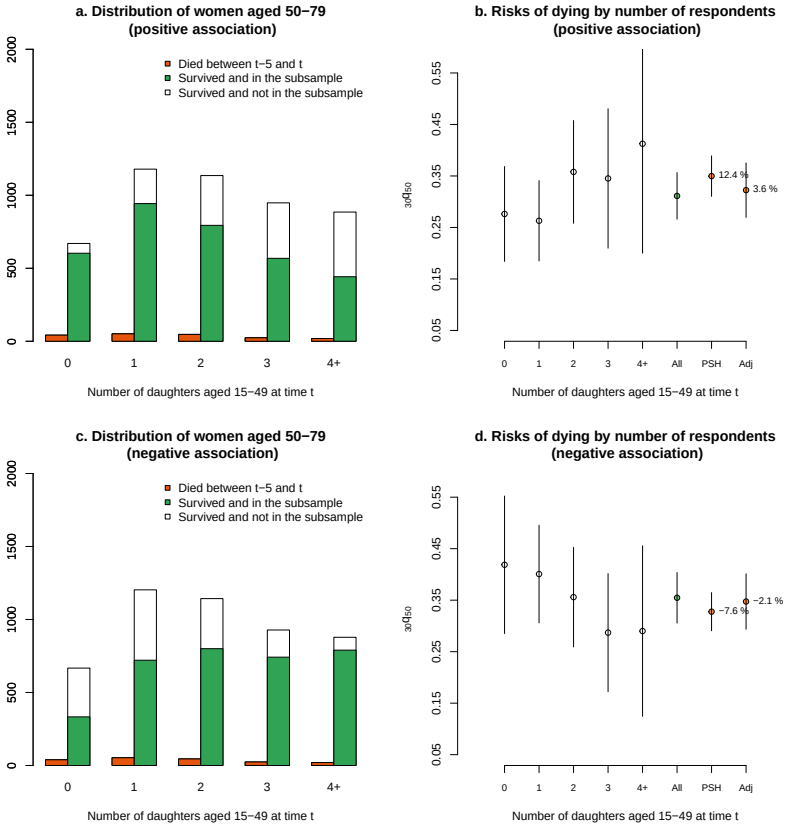
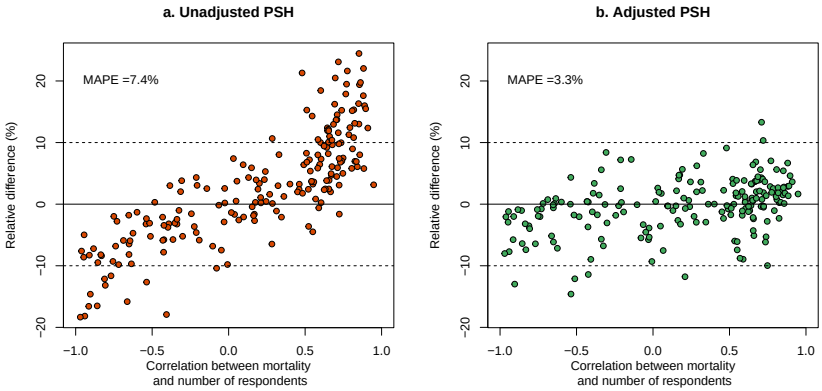


Figure A-7: Relative difference between the 'true' probability $_{30}q_{50}$ and the probability obtained from parental histories, as a function of the correlation between mortality and number of respondents, with and without weighting by $1/S$



D. Additional figures and tables

Figure A-8: Age-specific $_nq_x$ recalculated from parental histories from the augmented census data and estimated from recent household deaths in the 2013 census

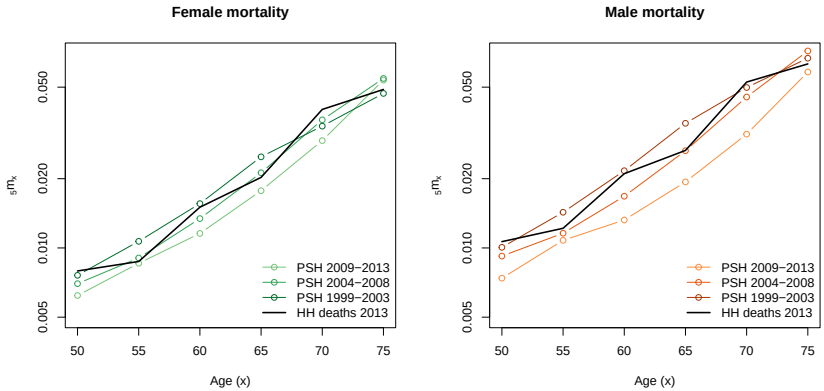


Table A-2: Mean percentage errors (MPE), standard deviation across samples, coefficient of variation, and coverage for the probability $_{30}q_{50}$ derived from parental histories for three five-year periods before data collection (0 to 4, 5 to 9, 10 to 14 years), for various sample sizes

Sample size			MPE (‰)			Standard dev. (‰)			Coeff. of variation (%)			Coverage (%)		
EAs	HHs	Women	0–4	5–9	10–14	0–4	5–9	10–14	0–4	5–9	10–14	0–4	5–9	10–14
Female mortality														
99	2,178	4,518	−0.4	−0.8	0.6	56	74	106	10.9	13.0	19.6	90	86	90
200	4,400	9,124	−1.2	−0.5	1.0	40	51	73	7.9	10.1	15.1	94	92	93
302	6,643	13,783	0.1	0.3	−1.1	33	45	72	6.5	8.2	12.0	93	93	87
400	8,799	18,185	0.1	−1.2	−0.1	27	41	56	5.8	7.3	10.9	94	92	92
498	10,955	22,716	0.5	−0.5	0.3	24	33	46	5.2	6.6	10.0	95	94	97
600	13,199	27,406	−0.8	−0.5	1.9	21	29	46	4.7	6.0	9.1	96	95	96
702	15,442	31,980	0.3	−0.5	−0.1	21	30	44	4.4	5.5	8.7	93	94	92
800	17,598	36,503	−0.3	−0.7	−1.3	20	24	41	4.1	5.3	7.9	95	98	92
898	19,754	40,970	−0.2	−1.6	−0.3	20	26	39	3.9	4.9	7.6	92	92	93
1,000	21,998	45,660	0	0.4	−1.0	17	24	36	3.7	4.7	7.2	93	96	92
Male mortality														
99	2,178	4,518	−1.4	−0.6	−0.8	43	47	46	7.2	6.4	6.4	83	88	89
200	4,400	9,124	−0.2	0.5	−0.5	28	31	35	5.5	4.6	4.7	95	92	87
302	6,643	13,783	1.1	−0.2	0.1	21	22	26	4.4	3.8	3.9	96	96	93
400	8,799	18,185	−0.2	−0.5	−0.3	18	24	23	3.9	3.4	3.4	100	94	92
498	10,955	22,716	−0.2	−0.4	−0.3	18	20	20	3.5	3.1	3.1	96	91	94
600	13,199	27,406	−0.3	−0.2	−0.6	17	17	18	3.2	2.8	2.9	95	93	94
702	15,442	31,980	−0.4	−0.5	−0.5	16	16	18	3	2.6	2.6	92	92	94
800	17,598	36,503	−0.7	−0.5	−0.3	12	15	15	2.9	2.5	2.5	96	97	95
898	19,754	40,970	−0.2	−0.6	−0.4	15	13	15	2.7	2.3	2.3	92	93	94
1,000	21,998	45,660	−0.3	−0.2	−0.5	11	13	14	2.6	2.2	2.2	98	97	94

Note: Estimates presented are obtained by averaging over 100 samples for each sample size.

Table A-3: Mean percentage errors in the $_{30}q_{50}$ probability derived from parental survival histories (PSH) and recent household deaths (HH deaths) for the single year before the census, standard deviation across samples, coefficient of variation around the $_{30}q_{50}$ estimates, and coverage of the 95% confidence intervals, in 1,000 samples drawn from the augmented census using the DHS sampling strategy, for each sex

Sample size			MPE (‰)		Standard dev. (‰)		Coeff. of variation (%)		Coverage (%)	
EAs	HHs	Women	PSH	HH deaths	PSH	HH deaths	PSH	HH deaths	PSH	HH deaths
Female mortality										
99	2,178	4,518	–0.8	0.6	97	140	16.5	22.1	87	83
200	4,400	9,124	–1.1	–6.0	69	100	11.7	18.3	90	81
302	6,643	13,783	–2.5	–4.6	55	68	9.9	14.7	93	90
400	8,799	18,185	0.8	–0.2	52	59	8.6	12.9	92	94
498	10,955	22,716	0.0	1.1	41	59	7.9	11.5	97	95
600	13,199	27,406	–1.3	1.0	38	48	7.2	10.5	97	96
702	15,442	31,980	–0.1	–0.4	40	51	6.7	9.9	93	97
800	17,598	36,503	–0.5	0.2	41	48	6.3	9.2	88	91
898	19,754	40,970	0.3	0.5	33	42	5.9	8.5	94	97
1,000	21,998	45,660	–0.1	–0.6	31	39	5.6	8.4	93	97
Male mortality										
99	2,178	4,518	–0.7	–2.5	88	129	12.2	18	85	81
200	4,400	9,124	–0.9	–2.1	51	79	9.4	13.8	94	91
302	6,643	13,783	1.2	–1.8	44	63	7.5	11.6	93	95
400	8,799	18,185	0.1	–0.9	40	60	6.7	9.8	92	92
498	10,955	22,716	–0.2	–1.2	31	52	6.1	9.0	98	95
600	13,199	27,406	–0.6	–1.1	33	53	5.6	8.2	95	92
702	15,442	31,980	–0.5	–0.9	29	51	5.1	7.5	94	93
800	17,598	36,503	–0.8	–0.2	25	47	4.9	7.1	96	93
898	19,754	40,970	–0.3	–1.2	28	37	4.6	6.8	93	94
1,000	21,998	45,660	–1.0	–0.9	26	41	4.4	6.3	92	94