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

DEMOGRAPHIC RESEARCH

VOLUME 46, ARTICLE 22, PAGES 653–680
PUBLISHED 5 APRIL 2022
https://www.demographic-research.org/Volumes/Vol46/22/
DOI: 10.4054/DemRes.2022.46.22

Research Article

Estimating mortality from census data: A record-linkage study of the Nouna Health and Demographic Surveillance System in Burkina Faso

Yempabou Bruno Lankoandé

Bruno Masquelier **Abdramane B. Soura**

Pascal Zabre **Gilles Pison**

Hélène Bangré **Sié Ali**

Géraldine Duthé

Contents

Demographic Research: Volume 46, Article 22
Research Article

Estimating mortality from census data: A record-linkage study of the Nouna Health and Demographic Surveillance System in Burkina Faso

Yempabou Bruno Lankoandé[1]
Bruno Masquelier[2]
Pascal Zabre[3]
Hélène Bangré[4]
Géraldine Duthé[5]
Abdramane B. Soura[6]
Gilles Pison[5]
Sié Ali[3]

Abstract

BACKGROUND
In low- and middle-income countries, mortality levels are commonly derived from retrospective reports on deceased relatives collected in sample surveys and censuses. These data sources are potentially affected by recall errors.

OBJECTIVE
Using high-quality data collected by the Nouna Health and Demographic Surveillance System (HDSS) in Burkina Faso, we evaluate the reliability of mortality estimates based on the 2006 national census.

METHODS
We extracted from the census database all records referring to the population under surveillance in the HDSS. Life tables were estimated from recent household deaths reported in the census and compared to those obtained from the prospective mortality data. To evaluate age errors and assess their impact on mortality, we linked census and HDSS records at the individual level for the surviving population and the deceased. Indirect estimates of mortality were also calculated based on the reported survival of children and parents.

[1] Institut Supérieur des Sciences de la Population (ISSP), Université Joseph Ki-Zerbo, Ouagadougou, Burkina Faso. Email: blankoande@issp.bf.
[2] Université catholique de Louvain, Belgium.
[3] Centre de Recherche en Santé de Nouna, Burkina Faso.
[4] Institut national de la statistique et de la Démographie, Ouagadougou, Burkina Faso.
[5] Institut National d'Études Démographiques, Paris, France.
[6] Université Joseph Ki-Zerbo, Ouagadougou, Burkina Faso.

RESULTS
Life expectancies at birth derived from recent household deaths pointed to a lower mortality than monitored in the HDSS, with a difference of 4 years for men and 8 years for women. Underreporting of deaths among the population aged 60 and above accounted for more than half of these differences. Age errors were small for the surviving population and larger for the deceased, but their effects on mortality estimates were modest. Indirect estimates of child mortality were consistent with the HDSS data, but orphanhood-based estimates were implausibly low.

CONCLUSION
Additional elicitation questions should be asked during the census interviews to improve the collection of data on recent household deaths.

CONTRIBUTION
Mortality rates derived from recent household deaths can seriously underestimate mortality. In Burkina Faso the downward bias in the 2006 census was larger among females and was mostly attributable to underreporting of deaths.

1. Introduction

Counting who is dying and what they are dying from is fundamental for priority setting within the health sector and evaluating the effectiveness of programmes to reduce premature mortality. The ideal data source is a comprehensive system of civil registration and vital statistics (CRVS) providing information on deaths by age and sex, with causes of death certified by a medical practitioner. However, in most countries in sub-Saharan Africa the registration of deaths remains incomplete at the national level, and progress has been modest in this area in recent decades (Mikkelsen et al. 2015). Burkina Faso, a landlocked country located in West Africa, is no exception. In the absence of an efficient CRVS system, mortality rates are derived from survey or census data. While large-scale sample surveys have collected birth and sibling survival histories, allowing the direct estimation of mortality, national censuses have included questions on the survival of children, parents, and recent household members, mostly for indirect estimation (Moultrie et al. 2013).

A first set of census questions aims at collecting summary birth histories; that is, the number of children ever born and surviving of women of reproductive age. The proportion of dead children born to women by age (or time elapsed between the first birth and the census) provides reliable estimates in the absence of abrupt mortality changes (Rajaratnam et al. 2010; Silva 2012). Indirect estimates of adult mortality can also be derived from reports on orphanhood, but these are considered with more scepticism. The most pervasive problem is commonly known as the 'adoption effect', caused by enumerators not systematically probing whether children observed in households are the true offspring of the adults being interviewed and parents claiming adopted orphans as their own children (Blacker 1977; Robertson et al. 2008). Censuses regularly collect data on the number of deaths in each household in the year preceding the enumeration. Unlike other sources of data, reports on recent deaths allow the computation of complete life tables, including for subnational areas or specific populations. Other data sources provide only summary indices of mortality for specific age groups, and statistical modelling is often required to disaggregate the estimate at small geographical scales due to sample size limitations (Mercer et al. 2015).

Reports on recent household deaths are not exempt from bias, however. The underreporting of deaths due to recall errors, the dissolution of some households following the death of adults, coverage errors, and transfers of some deaths outside of the 12-month reference period have been identified as the main sources of errors (Timæus 1991; Moultrie et al. 2013). A series of death distribution methods have been developed to upward adjust mortality levels obtained from recent household deaths, but they are based on strong assumptions, such as a constant underreporting of deaths over a certain age limit. Finally, all estimation approaches are sensitive to systematic age misstatement and under-renumeration of some specific populations, especially infants, young adults, and the elderly (Ewbank 1981; Pison and Ohadike 2006; Randall and Coast 2016).

The magnitude and direction of these errors are difficult to assess in the absence of a mortality gold standard. Census-based estimates have been evaluated in simulated environments (Hill, You, and Choi 2009; Murray et al. 2010; Verhulst 2016) but there is little evidence from direct evaluation studies. In this study we use high-quality data from a Health and Demographic Surveillance System (HDSS) in Burkina Faso as our gold standard. In an HDSS a locally defined population is followed up regularly through repeated household visits. Individuals enter the population through birth or immigration and exit through emigration or death. At each visit, interviewers review the list of household members who were present at the previous visit and collect information on births, deaths, in- and out- migrations, marriages, and selected health outcomes. This greatly limits the scope for omission of vital events and age misreporting, making the HDSS the most reliable source of mortality estimates in countries without vital registration, despite not being nationally representative (Pison 2005; Streatfield et al. 2014). These HDSS provide an ideal environment for assessing data quality. Several validation studies have been conducted on child or sibling survival (Helleringer et al. 2014; Masquelier et al. 2021; Rerimoi et al. 2019), but there have been few attempts to analyse census data.

There are currently more than 50 HDSS in low- and middle-income countries, many of them in operation for several decades (Sankoh and Byass 2012). All these sites are included in the frame of national censuses and can be used to evaluate the reliability of census estimates. In Senegal, mortality data from the 2002 and 2013 censuses were compared to those collected by three rural HDSS at the aggregate level (Masquelier et al. 2016); under-5 mortality rates from censuses were lower than expected based on demographic surveillance. Estimates inferred from parental survival were implausibly low. By contrast, age-specific mortality rates based on recent household deaths were consistent with the follow-up data, except for infant mortality, which was vastly underreported in the 2002 census. Similar comparisons conducted in Ghana in 2010 showed that the HDSS and the census yield consistent mortality estimates before the age of 50 (Wak et al. 2017). Above this age, higher mortality levels were observed in the HDSS, particularly for women.

In this study we use the Nouna HDSS, located in northwest Burkina Faso, to evaluate the reliability of mortality indicators derived from the national census, conducted in 2006.[7] First, we compare mortality rates at the aggregate level. Second, we link individual records to evaluate the quality of ages reported in the census and their impact on mortality estimates. The individual record linkages help to distinguish the effects of omissions and age errors on mortality rates.

2. Data and methods

2.1 Comparisons at the aggregate level

Our reference dataset is made up of data collected regularly by the Nouna HDSS, administered by the Centre de Recherche en Santé de Nouna (CRSN). The HDSS was set up in 1992 in a rural area about 300 kilometres from the capital city, Ouagadougou (Sié et al. 2010). The population, followed longitudinally, was estimated at about 80,000 individuals in 2006 and is spread over 58 villages and the small town of Nouna. Procedures are in place to ensure data quality and completeness in the collection of vital events and migrations. For example, a key informant independently records vital events in each village, and the interviewers can compare such records with the information collected during the household interviews to limit omissions (Sié et al. 2010). Each household is visited at least three times a year by the HDSS interviewers, and between 5% and 10% of households are re-interviewed by a supervisor. Consistency tests are also implemented on the basic variables in the database. The HDSS database has been extensively used to monitor mortality trends (Kynast-Wolf et al. 2002), analyse the seasonality of death (Kynast-Wolf et al. 2006), and describe cause-of-death patterns

[7] Individual data from the last census, conducted in 2019, were not available at the time of this study.

(Würthwein et al. 2001). It also serves as a platform to conduct randomised controlled trials.

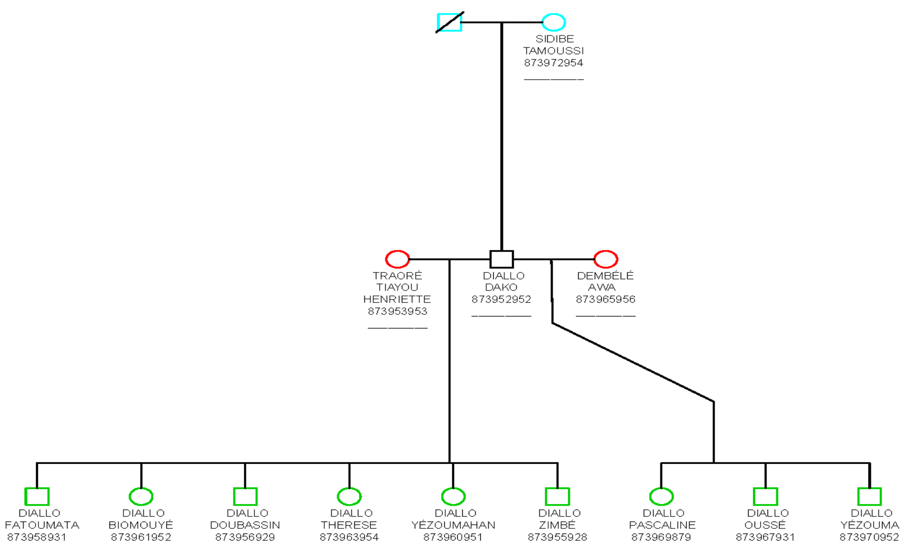
The 2006 national census of Burkina Faso was conducted from 9 to 23 December. In addition to the household roster, questions were asked about recent household deaths, children ever born and surviving, and orphanhood. Based on the names of villages, we extracted individual-level data for the population under surveillance in the HDSS from the census database. We compared the population size of each village and the age and sex structure according to the two sources. The definition of residency used in the HDSS is more restrictive than the definition retained in the census. The concept of residence has changed in the Nouna HDSS over the years, but we considered as resident an individual whose length of stay in a household located in the area is at least six months. In the 2006 national census a resident is any person who has been living in the enumerated household for at least six months or who intends to do so. As a result, while migrants who arrived in the area recently and plan to stay are considered as residents by the census, they are not considered as residents under the HDSS.

Using the census records covering the area of the population under surveillance, we divided the number of deaths reported for the 12 months prior to data collection by the population enumerated in December 2006 to generate a complete life table. We did not adjust for the small effect of population growth on denominators. In the HDSS we paralleled the census enumeration method from 2006 as well as we could in order to compute the population used as the denominator for rates. More precisely, since HDSS data can be formatted for event history analysis, the population in the HDSS corresponding to the one which should have been enumerated in the 2006 census was obtained by considering those present since 15 June 2006 (that is, six months before the date of enumeration). We compared summary indices of mortality between census and HDSS estimates and decomposed the differences in life expectancies at birth into contributions of the major age groups (Arriaga 1984).

2.2 Record linkages and analysis at the individual level

To conduct the record linkages we proceeded in three steps. First, villages included in the HDSS were associated with villages enumerated in the census. Second, an automatic search based on first and last names of all household members was performed (using the R statistical software). Each household enumerated in the census was paired with each household identified in the HDSS in the same village, and the proximity of names of all household members was assessed using the Jaro–Winkler distance. This distance reflects the length of the names and the number of characters which need to be transposed to switch from one name to another. This allowed us to associate names despite spelling variation and data entry errors. If half of the household members were considered as having the same name, we tentatively considered that the two entities from the census and HDSS databases referred to the same household. Each member of the HDSS has a unique ID number, and we recorded this number for the persons listed in the census for whom the full names matched. Third, we reconstructed kinship diagrams to identify the position of individuals in the household, using the kinship2 R package (Sinnwell, Therneau, and Schaid 2014). The identifiers of the parents of each individual had been collected during field visits (when the parents live in the HDSS area), and we used this information to identify each known child, parent, or sibling of the household members. Each individual was represented alongside their unique individual identifier from the HDSS database. An example of a kinship diagram is given in Figure 1.

Figure 1: Kinship diagram for a household in the Nouna HDSS based on genealogical data



Notes: Names have been changed. Males are represented with a square and females with a circle. The square with no name corresponds to a relative who was deceased at the time of the census but was needed to reconstruct the diagram.

Based on the census data, we drew similar diagrams using the information on the relationship with household head and the order of appearance in the census form (e.g., children are usually placed under their mother in the household roster). Households matched by the automatic procedure described above were identified. Each relationship diagram was then visually examined by a team member. The objective was to check the automated record linkages and to search for additional matches. At no time during the record linkage process did we use information on age or other demographic variables, as these variables are the focus of our study. The names, household structure, and kinship relations were the only types of information used. This manual matching procedure is laborious, and recent studies in other HDSS have shown that probabilistic matching can achieve similarly good results with limited human intervention (Kabudula et al. 2014; Masquelier et al. 2021). However, here we opted for a manual approach because we used only names and relationships to the household heads to link records. In addition, we could not limit searches within a well-defined subset using blocking fields because the households identified in the census are not necessarily identical to those in the HDSS (the average size was 5.42 in the census and 7.18 in the HDSS). Visual examination of kinship diagrams makes it easy to identify relative positions in the household, even when the head of the household in the census is not identified as such in the HDSS.

We ran logistic regressions on the probability of being matched in order to assess how the matched sample and the non-matched sample differed in terms of socio-demographic characteristics. Restricting the analysis to the matched sample, we then compared across data sources the ages of the surviving population as well as ages at death for those deceased in 2006. We assessed the effects of age misstatement in the census on mortality indicators by recalculating mortality rates using the ages reported in the HDSS. For the matched cases, ages at the census or at death were replaced with those found in the HDSS. For the non-matched cases, age was corrected using the age deviation between the census and the HDSS from the matched cases, by sampling with replacement, separately for each age group and sex.

Finally, we evaluated whether questions on the survival of children and parents in the census provided more reliable mortality estimates than those on recent household deaths.

3. Results

3.1 Comparisons at the aggregate level

3.1.1 Population by age and sex in 2006

Figure 2 shows the population pyramid according to the HDSS and the census. Overall, the two sources are highly consistent. In the HDSS the male population is only 2% larger and the female population 7% larger than in the census.[8] The most notable difference between the two sources relates to children aged less than 5 years, with a relative difference of −13% in both sexes, most likely due to under-enumeration of infants in the census. The census also enumerated fewer young women aged 10–24 (with a relative difference of −11%), but a larger number of males aged 10–29 (+10%). There was a good congruence in the numbers of adults aged 30–59, with relative differences lower than 5%.

Figure 2: Population pyramid of the Nouna HDSS according to the demographic follow-up and the 2006 census (December 2006)

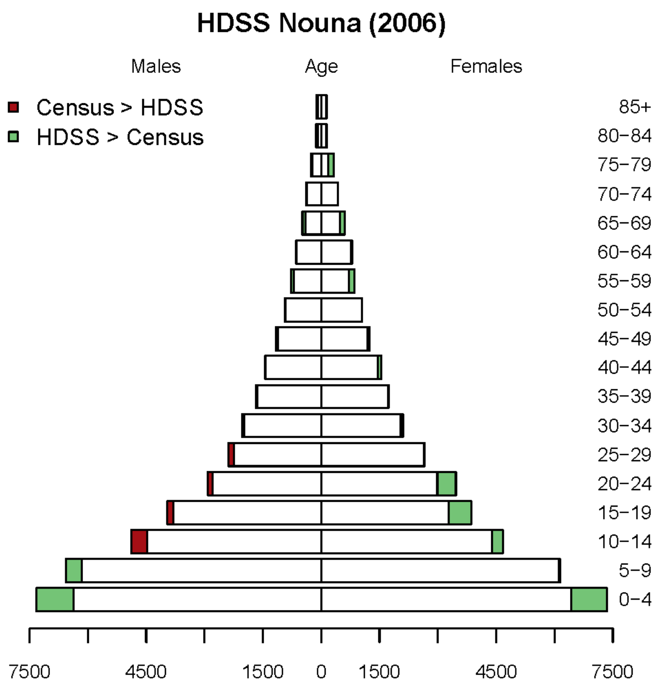


Table 1 presents the values of the Myers index, a measure of digit-preference in ages (Siegel and Swanson 2004). In the HDSS database this index is very low for the ages of

[8] Relative differences are computed as $(n_{census} - n_{hdss})/n_{hdss}$ where n_{census} refers to the population of the census and n_{hdss} to the population of the HDSS.

the surviving populations (<3%) and higher for the ages at death (8%), but this could reflect irregularities related to small numbers. The corresponding values from the census data (6% for the enumerated population and 15% for the deceased) point to a lower quality of data. In both sources there was less heaping on the ages reported for males.

Table 1: Myers index of the surviving population and the deceased in the census and the HDSS

		Male	Female	Both sexes
Surviving population	Census	5.2	7.1	6.2
	HDSS	2.1	3.6	2.6
Deaths in the last 12 months	Census	14.1	15.1	14.6
	HDSS	7.6	12.1	8.2

Sex ratios are also useful in assessing data quality. According to the HDSS data, the male-to-female sex ratio of the enumerated population declines with age, as expected from higher survival in females (Figure 3). The trough observed in young adults is due to sex differences in net migration. Overall, in- and out-migration rates are higher in females than in males in the area, but net migration rates are relatively similar across sexes, except for the age group 10–14 (when more females leave the area) and in youth (15–24), when net migration is negative among males and positive among females (Ginsburg et al. 2016). Sex ratios in the census are higher than in the HDSS from age 10 to 69, and again after 75 years of age. This pattern points to the under-enumeration of older women and to sex differentials in the classification of recent migrants as residents or non-residents.

Table 2 presents the number of deaths reported in 2006 by age group and sex according to the census and the HDSS. Overall, more deaths were reported in the HDSS (821 vs. 604), and the underreporting in the census was more pronounced for women. Compared to the HDSS, the census collected 22% fewer deaths for men, and 32% fewer deaths for women. The underreporting of deaths in the census did not affect all age groups to the same extent. Relative differences were positive in the 15–29 age group for both sexes, pointing to more deaths reported in the census, presumably due to age misstatement causing transfers across age groups. Relative differences were negative (i.e., fewer deaths in the census) in all other age groups, for both men and women.

Figure 3: Sex ratio at birth among the surviving population and the deceased in the census and the HDSS; number of deaths by age and sex

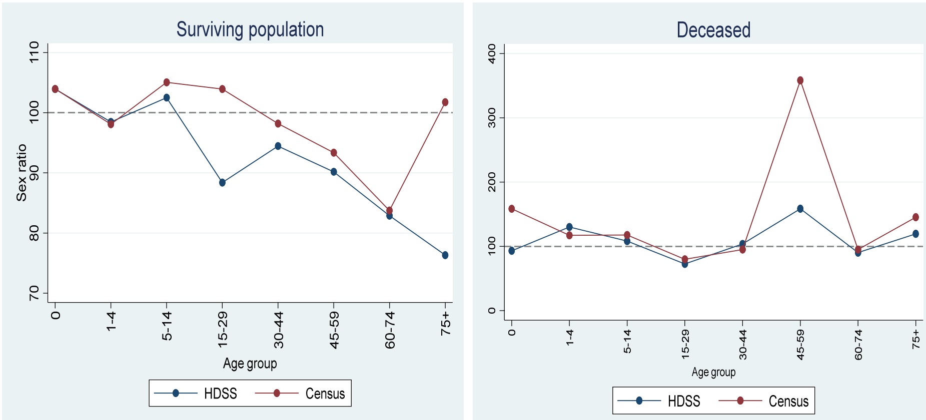


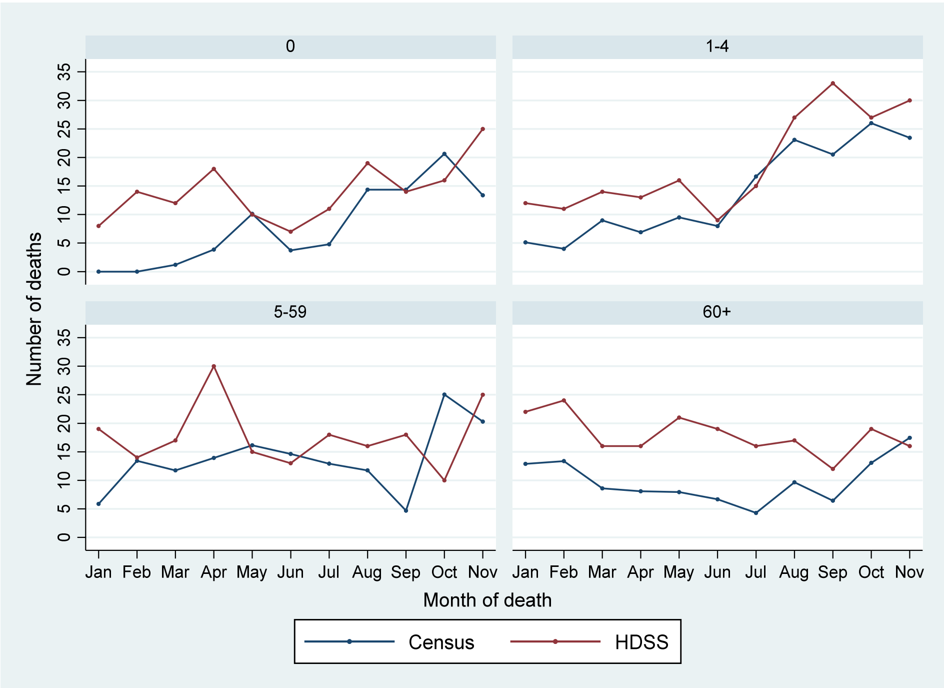
Table 2: Number of deaths reported in 2006 by age group and sex according to the census and the HDSS

Age group	Men			Women		
	Census	HDSS	Relative differences (%)	Census	HDSS	Relative differences (%)
0–1	65	80	−19%	42	88	−52%
1–4	97	125	−22%	82	95	−14%
5–14	20	26	−23%	17	24	−29%
15–29	20	16	25%	26	22	18%
30–44	19	27	−30%	22	26	−15%
45–59	43	46	−7%	12	28	−57%
60–74	39	55	−29%	42	62	−32%
75+	33	55	−40%	23	46	−50%
Total	338	430	−21%	266	391	−32%

In Figure 4 we compare the number of deaths reported in the census by month and age group with the HDSS in 2006. In both sources there were more deaths in the four-month period preceding the census than in earlier months, reflecting the seasonality of mortality, with a typical excess of child deaths at the end of the rainy season (due to malaria and diarrhoeal diseases) (Kynast-Wolf et al. 2006). However, in all age groups the ratio between the reported numbers of deaths in the census and in the HDSS was smaller in the first half of the year, especially among infants. Only a quarter of infant deaths registered in the HDSS were reported in the census in the period January–July

2006. This pattern suggests that recent deaths were more likely to be reported in the census than those that occurred at the beginning of the year.

Figure 4: Number of deaths reported by month in 2006 in Nouna according to the HDSS and the census, by age group



Note: the reported numbers for December are not shown because it was not possible to compare deaths that occurred in December 2005 and December 2006. The census took place between December 9 and 23 and the questions on the number of deaths that occurred in the household covered the twelve months before the census.

3.1.2 Mortality estimates in 2006 according to the HDSS and the census

Figure 5 compares the age-specific mortality rates estimated from deaths reported in the 12 months preceding the census with HDSS estimates for the same year. Table 3 presents the corresponding summary indices of mortality. The consistency between both series of estimates clearly varies by sex. Mortality levels in boys aged less than 5 and men aged 15–59 matched fairly well with HDSS rates, with relative differences lower than 10% in the probabilities ${}_5q_0$ and ${}_{45}q_{15}$. Still, we observe a difference of 4.1 years in life expectancy at birth for men, which is largely due to lower mortality at age 60 and above. The risk of dying between ages 60 and 80 among males was 21% lower in the census. Among females, infant mortality was 19% lower in the census, and the probability $_{45}q_{15}$ was 27% lower. There was also a substantial difference in mortality rates above age 60 among females (−31%). This resulted in an estimate of life expectancy at birth which was 7.6 years higher for females in the census than in the HDSS. The gap in mortality rates in the age group 60 and above supplied 4.2 years (55%) of this difference in life expectancy at birth (Table 3). It is worth noting that mortality rates are not systematically lower in the census: among women aged 20–34, as well as among men aged 15–24 and at older ages (50–54, 60–64), mortality is higher according to the census. This could be explained by age misstatement, transferring some deaths and population at risk from one age group to another.

Figure 5: Age-specific mortality rates (ASMR) inferred from the census and the HDSS data, Nouna, 2006

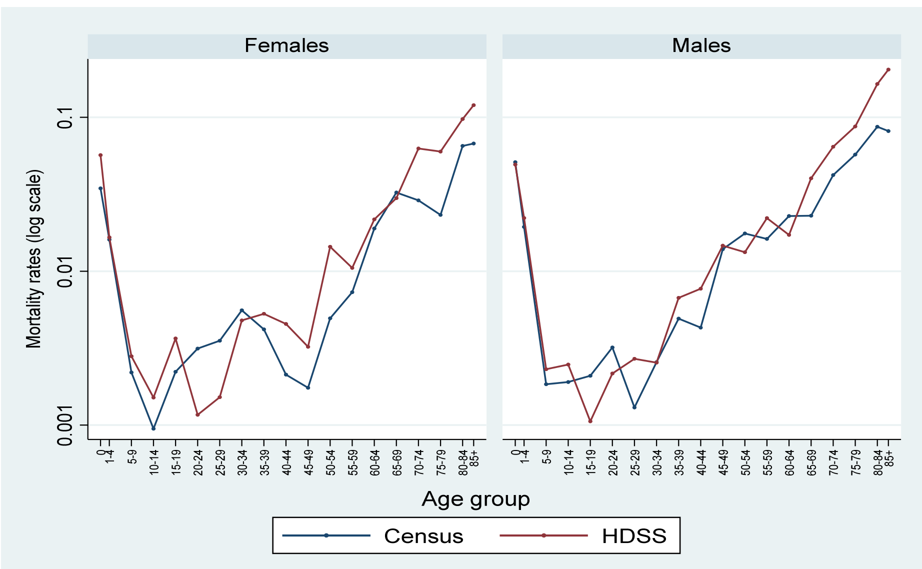


Table 3: Direct estimates of mortality in Nouna according to the HDSS and the deaths reported in households during the last 12 months in 2006 census

	Males				Females			
	Census	HDSS	Relative diff.[1]	Contrib. to the diff.[2]	Census	HDSS	Relative diff.[1]	Contrib. to the diff.[2]
Probabilities of dying (p. 1000)								
$_{5}q_{0}$	120	128	−6%	0.4	93	115	−19%	1.6
$_{10}q_{5}$	19	24	−21%	0.3	16	21	−27%	0.3
$_{45}q_{15}$	282	306	−8%	0.6	160	218	−27%	1.5
$_{20}q_{60}$	518	652	−21%	2.8	405	584	−31%	4.2
Life expectancy (in years)								
				Abs. diff				Abs. diff
e_0	61.9	57.8	7%	4.1	69.4	61.8	12%	7.6

Notes: (1) Relative difference = $\frac{rate_{census}-rate_{HDSS}}{rate_{HDSS}}$; (2) contributions of mortality difference in the age group to the difference in life expectancy at birth (in years). The last estimate of the contribution refers to mortality rates in the open-ended age group 60+ (including mortality above age 80, while $_{20}q_{60}$ is the probability of dying from age 60 to 80).

3.2 Individual-level analysis

3.2.1 Matching rates

Among the surviving population we were able to manually match 58% of household members enumerated in the 2006 census to individuals in the HDSS record. Variation in names used in the HDSS and reported in the census, misclassification of migrants, and differences in household size and composition across sources were the main obstacles to record linkage. We investigated how matching rates varied by age group and other demographic characteristics through logistic regression (Table 4). Matching rates were significantly higher for children aged less than 15, for heads of households and their spouses, and for residents who were present at the time of the census. Household members who had attended school or were attending at the time of the census were also more likely to be matched. Matching rates increased with household size.

Table 4: Odds ratios of a logistic regression of the probability of matching an individual alive at the time of the 2006 census with a record in the HDSS

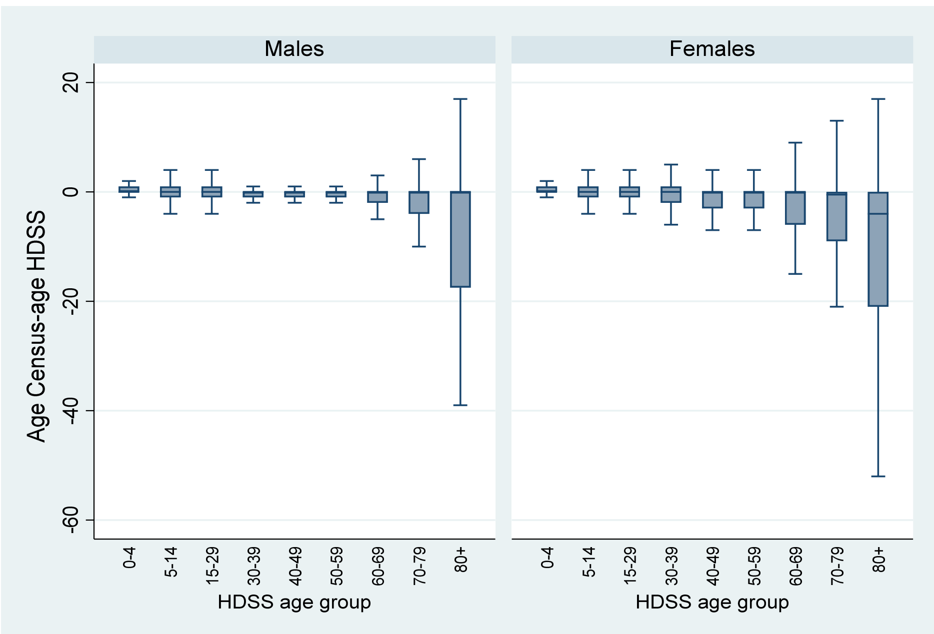
Variable		Unadjusted odds ratios (95% CI)	Adjusted odds ratios (95% CI)
	Males	Ref.	Ref.
Sex	Females	1.008	0.977
		(0.978–1.038)	(0.943–1.013)
	0–4	Ref.	Ref.
	5–14	0.858	0.765
		(0.821–0.898)	(0.719–0.815)
	15–29	0.620	0.553
		(0.591–0.649)	(0.516–0.592)
	30–39	0.942	0.722
		(0.889–0.999)	(0.661–0.787)
	40–49	1.124	0.847
Age group		(1.051–1.202)	(0.771–0.931)
	50–59	0.933	0.775
		(0.863–1.008)	(0.700–0.857)
	60–69	0.761	0.721
		(0.696–0.833)	(0.647–0.805)
	70–79	0.657	0.696
		(0.583–0.740)	(0.608–0.796)
	80+	0.693	0.787
		(0.600–0.800)	(0.675–0.916)
Relationship to head	Household head	Ref.	Ref.
	Spouse	1.074	1.105
		(1.023–1.129)	(1.040–1.175)
	Child	1.082	0.836
		(1.039–1.127)	(0.783–0.893)
	Sibling	0.640	0.598
		(0.579–0.708)	(0.537–0.666)
	Other	0.580	0.474
		(0.549–0.612)	(0.443–0.508)
	Present	Ref.	Ref.
Residency status	Absent	0.729	0.749
		(0.667–0.795)	(0.685–0.819)
	No education	Ref.	Ref.
	Has attended	0.965	1.084
		(0.920–1.013)	(1.030–1.141)
Education	Attending	1.080	1.227
		(1.033–1.126)	(1.169–1.289)
	DK	1.181	0.955
		(1.126–1.240)	(0.893–1.022)
Household size	Size	1.048	1.056
		(1.043–1.052)	(1.051–1.061)
Constant			1.368
			(1.260–1.485)
Observations		71,706	71,706

3.2.2 Age misstatement

The information on ages in the Nouna HDSS is of good quality, thanks to particular attention having been paid to their collection when setting up the demographic surveillance, through numerous consistency checks. The birth dates of individuals born after the initial census were registered during the follow-up, so they are known with great precision. This is confirmed by the values of the Myers index presented above.

To capture the differing ages of individuals provided by the two data sources, Figure 6 presents the distribution of age differences using the HDSS as reference. In men, ages reported in the census are surprisingly accurate, including in adults aged 40–59. Small differences appear in children and young adults (deviations smaller than 2 years), but these should not have a large impact on mortality rates. Very large age differences are observed in the older age group, but they are likely due to matching errors.

Figure 6: Age differences in the census v. the HDSS in 2006, men and women, using the HDSS as the reference



Age differences between the census and the HDSS are more salient in women. Before age 30 the differences are small and centred around zero. Like for men, there is no clear pattern of age misstatement for women alive at the time of the census before they reach age 30. Above age 30 there seems to be a systematic underestimation of age, and absolute age differences increase. Irrespective of possible errors in ages at death, this underestimation of ages of women enumerated in the census is likely to result in downward bias in mortality in young adults and upward bias in mortality in the elderly. This is because the denominators for calculating age-specific rates will be artificially inflated in young women and reduced in the elderly.

We conducted the same analysis on the matched sample of those deceased in 2006 (see Appendix). The matching rate was lower among the deceased individuals (36.3%) because names were missing for 21.6% of the deceased in the census dataset. The matched sample and the non-matched sample did not differ in terms of age and sex (Table A-1). Ages at death reported in the census and in the HDSS were close for children aged less than 15. However, ages at death were significantly higher in the census above age 80 (Figure A-1). Misstatement of ages at death will therefore result in the underestimation of old-age mortality.

3.3 Effects of age misstatement on mortality estimates

To isolate the effects of age errors and omissions or transfers out of the reference period, we computed mortality estimates from the census data after correcting the reported ages based on HDSS data, either through selecting the age encoded in the HDSS database or through sampling age deviations between the HDSS and the census. For men, life expectancy at birth declined only slightly, from 61.9 to 61.1 years. For women, a similarly modest decline of 0.7 years in life expectancy at birth was observed, mainly due to an upward correction of premature adult mortality (the probability ${}_{45}q_{15}$ rose by a third). After this correction, the risk of dying from age 15 to 60 extracted from the census was very close to the HDSS estimate. These results confirm that age errors in the census mainly affect adults and the elderly, but the bias they introduce in life expectancy is smaller than that introduced by underreporting due to omissions of deaths or transfers out of the reference period.

Table 5: Mortality estimates in the census corrected for age misstatement

Index	Census	Census (after correction for age misstatement)	HDSS	Relative difference (1)	Relative difference (2) (after correction)
Males					
Probability of dying (p. 1000)					
$_5q_0$	120	121	128	−6%	−6%
$_{10}q_5$	19	21	24	−21%	−12%
$_{45}q_{15}$	282	291	306	−8%	−5%
$_{20}q_{60}$	518	526	652	−21%	−19%
Life expectancy at birth (years)					
e_0	61.9	61.1	57.8	7%	6%
Females					
Probability of dying (p. 1000)					
$_5q_0$	93	92	115	−19%	−20%
$_{10}q_5$	16	16	21	−27%	−25%
$_{45}q_{15}$	160	214	218	−27%	−2%
$_{20}q_{60}$	405	449	584	−31%	−23%
Life expectancy at birth (years)					
e_0	69.4	68.7	61.8	12%	11%

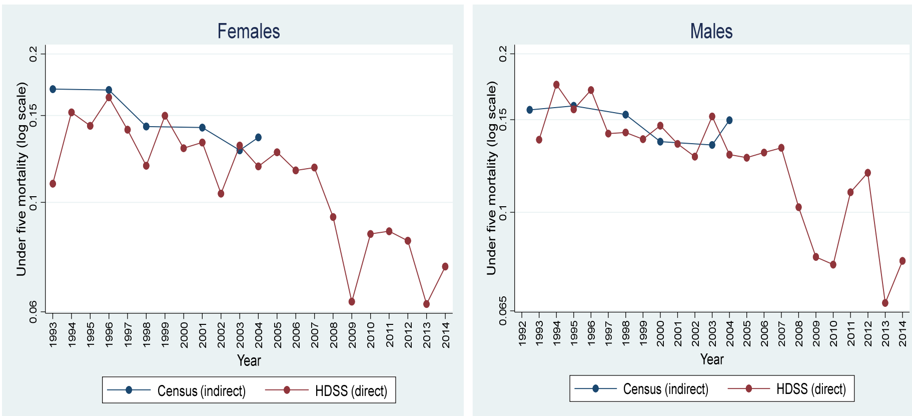
Notes: (1) Relative difference with no corrected estimates
(2) Relative difference with corrected estimates

3.4 Alternative mortality estimates derived from the survival of relatives

The census included a question on the number of children ever born to women aged 10 years and above, and the number of these children surviving at the time of the census. Under-5 mortality rates were estimated indirectly from these numbers, using the North pattern of model life tables (Moultrie et al. 2013). This indirect approach uses a set of standard coefficients capturing variations in mortality and fertility patterns to convert proportions of dead children born to women by age group into the probability $_5q_0$. We discarded estimates derived from women aged 15 to 19 because these are usually plagued by selection biases associated with higher mortality of first-born children.

In both sexes, levels and trends in under-5 mortality derived from census data on children ever born and surviving are fairly consistent with the HDSS estimates (Figure 7). Census estimates are smoother, because the time-location of indirect estimates is based on an assumption of linear and unidirectional mortality decline. If we discard the period 1992–1993, which was the start of the follow-up, census estimates for girls are on average 8% higher than estimated from the longitudinal follow-up during the period 1996–2004. Among boys, census estimates are on average 1% higher than estimates derived from the HDSS during the period 1995–2004. Census and HDSS estimates do not necessarily refer to the same children because all children born to women residing in the area at the time of the census will be considered for indirect estimates, whereas only children who reside in the area should be included in HDSS estimates. Considering this limitation, this consistency across sources is remarkable.

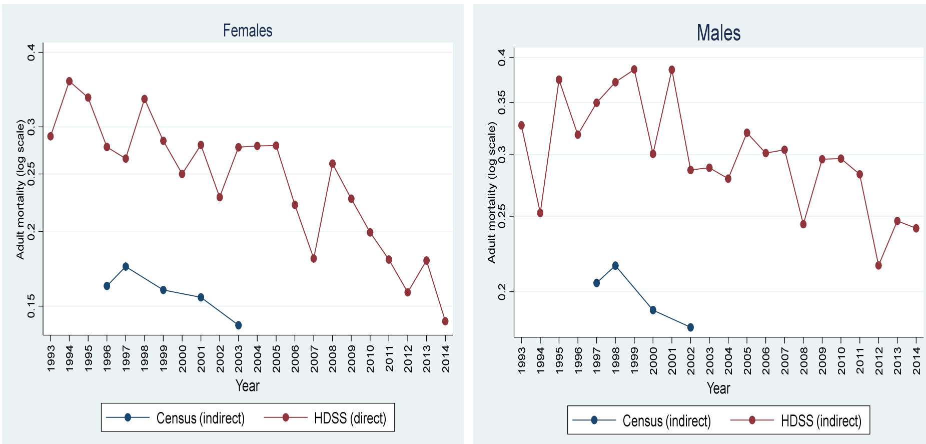
Figure 7: Comparison of indirect estimates of child mortality ($_5q_0$) inferred from the census and direct estimates from the HDSS



The census also asked about the survival of parents of individuals aged less than 30. We applied the standard orphanhood method (Timæus 1992). Estimates obtained for the different age groups of respondents were converted into the probability $_{45}q_{15}$, again using the North model of Coale–Demeny life tables.

Figure 8 indicates that adult mortality estimates derived from orphanhood data collected in the census are implausibly low when compared to HDSS estimates. They are 43% lower on average for women during the period 1996–2003 and 40% lower on average for men during the period 1997–2002. Again, these two series do not necessarily refer to the same individuals, because census estimates are obtained from residents of the HDSS, whose parents do not necessarily live in the area, while the HDSS estimates refer to mortality experienced in Nouna. A question in the census on the place of residence of surviving parents showed that the percentage of children whose parents lived in the same household declined rapidly by age, from 78% among adolescents aged 10–14 living with their mothers to 13% among 25 to 29 year olds. Seventy-eight per cent of children aged 10–14 whose father was alive at the time of the census also lived with their father, against only 16% of young adults aged 25–29. Adult mortality rates inferred from parental survival in the census are therefore a mix of local mortality conditions and conditions prevailing outside of the HDSS. However, the discrepancies between census and HDSS estimates are so large that they indicate a considerable amount of underreporting of parental deaths, for both maternal and paternal orphanhood.

Figure 8: Comparison of indirect estimates of adult mortality ($_{45}q_{15}$) inferred from the census with direct HDSS estimates



4. Discussion

In low- and middle-income countries, censuses remain a fundamental data source for assessing population dynamics, in the absence of comprehensive systems of vital registration. Many demographic methods have been developed to detect and adjust for recall errors in censuses, in particular by examining the external validity or the internal consistency of the estimates (Moultrie et al. 2013). Triangulating different data sources is common, but it is difficult to reach solid conclusions in the absence of a reference that is not itself affected by errors. In this study we used high-quality mortality data from an HDSS in Burkina Faso to evaluate mortality rates inferred from a national census. We found good agreement in population counts, despite some deviations in children aged less than 5 and in young adults. The larger number of young men in the census could be related to recent arrivals in the area of men who intended to stay but had not yet spent six months, a group not included in the HDSS. The undercounting of young women is more difficult to explain. One can speculate that some young women who temporarily left the HDSS area to marry out of the area or for educational/professional reasons were still considered as resident in the HDSS despite not being counted as such in the census. In the Nouna HDSS some young women were engaged in seasonal out-migration (during the dry season) towards the towns of Ouagadougou and Bobo-Dioulasso. Collecting comprehensive data on seasonal migrations in both national censuses and HDSS would help to adequately reflect demographic patterns in local areas and reduce biases caused by selective migration.

In terms of mortality estimation, we noted substantial discrepancies between HDSS and census estimates of female mortality from recent household deaths, with a better congruence for males. As indicated earlier, one key assumption underpinning the demographic methods designed to adjust data on recent household deaths is that underreporting is invariant by age (over a certain age limit). Our analysis clearly indicates that this assumption would be invalidated in the census data, even when restricting the analysis to adults.

Several reasons may be put forward to explain the deviations from HDSS estimates. First, the ages of the deceased and the surviving population are affected by misreporting, which is more pervasive among women. However, correcting these ages based on the deviations between sources observed in the matched sample shows that these errors had little impact on mortality estimates. This is particularly true for under-5 mortality, which has a strong influence on life expectancy at birth in this context of high mortality. Second, like in many other censuses, there is a mismatch between the reference period used for recent household deaths (12 months) and the period used to define the resident population (6 months). In cases of large flows of seasonal migration, this mismatch could distort mortality estimates. Third, deaths could be underreported in the census because enumerators did not systematically ask about household deaths (which remain a relatively rare event), because respondents were unwilling to talk about their deceased relatives, or because some deceased were not clearly identified as members of one specific household (e.g., recent migrants). Furthermore, some households may disperse and recompose after the death of one of their members, making it more likely that this death goes unreported in the census. This selection bias will be more prevalent for deaths that occurred in the more distant past, as compared to deaths immediately preceding the census. As the census included a question on month of death, we compared the distribution of reported deaths by month prior to the census date and observed that the number of reported deaths declined as the number of months between their occurrence and the census increased, pointing to more frequent omissions. In summary, our study suggests that underreporting of deaths is the main source of bias in mortality rates in the census.

When the quality of data on recent household deaths is called into question, demographers regularly turn to indirect techniques of mortality estimation. For example, women's reports on the numbers of children ever born and surviving provide under-5 mortality estimates, and information on parental survival is used to generate adult mortality rates. We used these methods and compared the resulting indirect estimates with the underlying mortality rates in the HDSS. Child mortality rates obtained indirectly were broadly consistent with HDSS estimates, especially among males. By contrast, indirect estimates of adult mortality obtained from reports on orphanhood in the census were implausibly low. This suggests that indirect mortality estimates are not more reliable than those derived from recent household deaths.

The patterns observed here differ from previous comparisons made in Senegal, where direct estimates based on recent household deaths were plausible beyond the age of 5, while indirect estimates of under-5 mortality were too low in the censuses (Masquelier et al. 2016). In this study of Nouna, all estimates except for indirect child mortality rates were affected by a downward bias. This reinforces the need to undertake similar analysis of other HDSS sites in Burkina Faso and other countries where vital registration systems are incomplete before generalising our results.

Finally, this study also highlights the need to improve data quality by limiting the use of proxy respondents in censuses and developing innovative ways to improve the reporting of age and demographic events, such as historical calendars. A comparable study on sibling survival histories suggests that data quality can be improved using additional recall cues and prompts (Helleringer et al. 2014).

This research has some limitations. First, ages derived from the HDSS could also be affected by misreporting, even if dates of events are known with greater precision than in the census. This is particularly the case for in-migrants and for individuals – especially older adults – who were present at the initial census. Even if particular attention is paid to age reporting in the HDSS, the values of the Myers index were still high (3.5 for males and 7.5 for females). Second, as we could not link every individual, the analysis of age reporting was based only on the matched sample. However, the probability of an individual enumerated during the census being matched with an individual in the HDSS depends on various criteria including omission, migration, age, and sex. Age errors are likely to be larger among individuals that we could not match than among those who were successfully linked. This can lead to an underestimation of age differences. As our matching rates were lower for males than for females, it is possible that age misreporting was underestimated for men compared to women. Finally, errors measured in the HDSS may be smaller than those that would be observed in the general population because HDSS residents are used to being asked about demographic events.

5. Conclusion

Triangulating national census data with demographic surveillance systems can help in assessing mortality rates derived from various estimation methods. Given that the HDSS puts heavy emphasis on the collection of accurate demographic data with regular visits, it is likely that mortality rates for both children and adults were underestimated in the 2006 census in this area, especially among women. Omissions of deaths play a larger role than age errors in explaining these gaps. Caution should be exercised when using standard death distribution methods, as the extent of underreporting varies by age.

6. Declarations

Ethical approval: We did not seek ethical approval for this study because it is based on secondary analysis of census and HDSS data. Ethical approval was the responsibility of the institutions that collected the data. Nominative data were only accessible to the project team to conduct the record linkages, based on extracts from the files. The analysis was then carried out on an anonymous data file resulting from the processing of nominative data. As a result, no third party could have had access to the identifiable data containing both census information and information from the Nouna HDSS. Participants were not solicited during this matching process. The ethics committee of the Nouna HDSS confirmed that ethical approval was not required.

Availability of data and materials: This study used an extract of the complete database of the 2006 national census of Burkina Faso conducted by INSD. While this extract cannot be made publicly available, a representative sample of the national census is available through the IPUMS programme (https://international.ipums.org/international/). The datasets referring to the HDSS in Nouna analysed during the current study are available from the last author on reasonable request.

Funding: The project leading to this work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690984: DEMOSTAF project.

Authors' contributions: BL, HZ, and BM analysed the census data. PZ and BL analysed the HDSS data. BM, PZ, and BL wrote the first draft of the paper. All authors contributed to the design and implementation of the project, revised the first draft of the paper, and approved the final manuscript.

References

Arriaga, E.E. (1984). Measuring and explaining the change in life expectancies. *Demography* 21(1): 83–96. doi:10.2307/2061029.

Blacker, J.G. (1977). The estimation of adult mortality in Africa from data on orphanhood. *Population Studies* 31(1): 107–128. doi:10.1080/00324728.1977.10412750.

Ewbank, D. (1981). Age misreporting and age-selective underenumeration: Sources, patterns and consequences for demographic analysis. Report 4, Committee on Population and Demography. Washington, DC: National Academies Press.

Ginsburg, C., Bocquier, P., Béguy, D., Afolabi, S., Augusto, O., Derra, K., Odhiambo, F., Otiende, M., Soura, A., Zabré, P., White, M.J., and Collinson, M.A. (2016). Human capital on the move: Education as a determinant of internal migration in selected INDEPTH surveillance populations in Africa. *Demographic Research* 34(30): 845–884. doi:10.4054/DemRes.2016.34.30.

Helleringer, S., Pison, G., Masquelier, B., Kanté, A., Douillot, L., Duthé, G., Sokhna, C., and Delaunay, V. (2014). Improving the quality of adult mortality data collected in Demographic surveys: Validation study of a new siblings' survival questionnaire in Niakhar, Senegal. *PLoS Medicine* 11: e1001652. doi:10.1371/journal.pmed.1001652.

Hill, K., You, D., and Choi, Y. (2009). Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors. *Demographic Research* 21(9): 235–254. doi:10.4054/DemRes.2009.21.9.

Kabudula, C.W., Clark, B.D., Gómez-Olivé, F.X., Tollman, S., Menken, J., and Reniers, G. (2014). The promise of record linkage for assessing the uptake of health services in resource constrained settings: A pilot study from South Africa. *BMC Medical Research Methodology* 14(71). doi:10.1186/1471-2288-14-71.

Kynast-Wolf, G., Hammer, G.P., Müller, O., Kouyaté, B., and Becher, H. (2006). Season of death and birth predict patterns of mortality in Burkina Faso. *International Journal of Epidemiology* 35(2): 427–435. doi:10.1093/ije/dyi150.

Kynast-Wolf, G., Sankoh, O.A., Gbangou, A., Kouyaté, B., and Becher, H. (2002). Mortality patterns, 1993–98, in a rural area of Burkina Faso, West Africa, based on the Nouna demographic surveillance system. *Tropical Medicine and International Health* 7(4): 349–356. doi:10.1046/j.1365-3156.2002.00863.x.

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(1): 1–19. doi:10.1080/00324728.2020.1854332.

Masquelier, B., Ndiaye, C.T., Pison, G., Dieme, N.B., Diouf, I.S., Helleringer, S., N'Diaye, O., and Delaunay, V. (2016). Evaluation des estimations indirectes de mortalité dans trois observatoires de population au Sénégal. *African Population Studies* 30: 2227. doi:10.11564/30-1-803.

Mercer, L.D., Wakefield, J., Pantazis, A., Lutambi, A.M., Masanja, H., and Clark, S. (2015). Space-time smoothing of complex survey data: Small area estimation for child mortality. *The Annals of Applied Statistics* 9(4): 1889–1905. doi:10.1214/15-AOAS872.

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. *The Lancet* 386(10001): 1395–1406. doi:10.1016/S0140-6736(15)60171-4.

Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., and Zaba, B. (2013). *Tools for demographic estimation*. Paris: International Union for the Scientific Study of Population.

Murray, C.J., Rajaratnam, J.K., Marcus, J., Laakso, T., and Lopez, A.D. (2010). What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Medicine* 7(4): e1000262. doi:10.1371/journal.pmed.1000262.

Pison, G. (2005). Population observatories as sources of information on mortality in developing countries. *Demographic Research* 13(13): 301–334. doi:10.4054/DemRes.2005.13.13.

Pison, G. and Ohadike, P.O. (2006). Errors and manipulations in age assessment. In: Sauvain-Dugerdil, C., Leridon, H., and Mascie-Taylor, N. (eds.). *Human clocks. The bio-cultural meaning of age*. Berne: Peter Lang: 313–336.

Rajaratnam, J.K., Tran, L.N., Lopez, A.D., and Murray, C.J. (2010). Measuring under-five mortality: Validation of new low-cost methods. *PLoS Medicine* 7(4): e1000253. doi:10.1371/journal.pmed.1000253.

Randall, S. and Coast, E. (2016). The quality of demographic data on older Africans. *Demographic Research* 34(5): 143–174. doi:10.4054/DemRes.2016.34.5.

Rerimoi, A., Jasseh, M., Agbla, S., Reniers, G., Roca, A., and Timæus, I. (2019). Under-five mortality in The Gambia: Comparison of the results of the first demographic and health survey with those from existing inquiries. *PLoS One* 14(7): e0219919. doi:10.1371/journal.pone.0219919.

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(Suppl. 1): i57–i62. doi:10.1136/sti.2008.029926.

Sankoh, O. and Byass, P. (2012). The INDEPTH Network: Filling vital gaps in global epidemiology. *International Journal of Epidemiology* 41(3): 579–588. doi:10.1093/ije/dys081.

Sié, A., Louis, V., Gbangou, A., Müller, O., Niamba, L., Stieglbauer, G., Yé, M., Kouyaté, B., Sauerborn, R., and Becher, H. (2010). The Health and Demographic Surveillance System (HDSS) in Nouna, Burkina Faso, 1993–2007. *Global Health Action* 3(1): 5284. doi:10.3402/gha.v3i0.5284.

Siegel, J. and Swanson, D. (2004). *The methods and materials of demography*. 2nd ed. Elsevier Academic Press.

Silva, R. (2012). Child mortality estimation: Consistency of under-five mortality rate estimates using full birth histories and summary birth histories. *PLoS Medicine* 9(8): e1001296. doi:10.1371/journal.pmed.1001296.

Sinnwell, J.P., Therneau, T.M., and Schaid, D.J. (2014). The kinship2 R package for pedigree data. *Human Heredity* 78(2): 91–93. doi:10.1159/000363105.

Streatfield, P.K., Khan, W.A., Bhuiya, A., Alam, N., Sié, A., Soura, A.B. et al. (2014). Cause-specific mortality in Africa and Asia: Evidence from INDEPTH health and demographic surveillance system sites. *Global Health Action* 7(1): 25362. doi:10.3402/gha.v7.25362.

Timaeus, 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.

Verhulst, A. (2016). Child mortality estimation: An assessment of summary birth history methods using microsimulation. *Demographic Research* 34(39): 1075–1128. doi:10.4054/DemRes.2016.34.39.

Wak, G., Bangha, M., Azongo, D., Oduro, A., and Kwankye, S. (2017). Data reliability: Comparison between census and Health and Demographic Surveillance System (HDSS) outputs for Kassena-Nankana East and West districts, Ghana. *Population Review* 56(1): 31–45. doi:10.1353/prv.2017.0001.

Würthwein, R., Gbangou, A., Sauerborn, R., and Schmidt, C.M. (2001). Measuring the local burden of disease: A study of years of life lost in sub-Saharan Africa. *International Journal of Epidemiology* 30(3): 501–508. doi:10.1093/ije/30.3.501.

Appendix

Age differences among deceased persons in 2006

Figure A-1: Age differences in the census v. the HDSS in 2006, deceased persons, using the HDSS as a reference

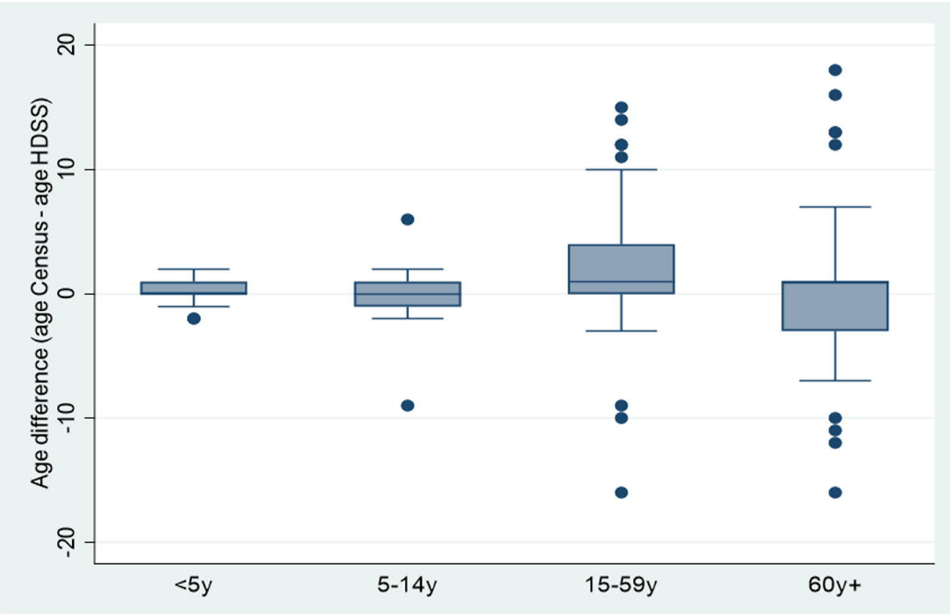


Table A-1: Logistic regression of the probability of a death reported in the census being matched with a deceased person in the HDSS in 2006

Variable		Adjusted odds ratio
Age group	0–4	Ref.
	5–14	1.335
		(0.474)
	15–59	0.918
		(0.199)
	60–79	1.223
		(0.297)
	80+	0.913
		(0.332)
Sex	Male	Ref.
	Female	0.86
		(0.148)
Constant		0.611***
		(0.0881)
Observations		589

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1