Estimating Mortality from Census Data: an record linkage study in the Nouna Demographic and Health Surveillance System in Burkina Faso

Bruno Lankoande[(1)(2)], Bruno Masquelier[(1)(3)], Hélène Zida[(4)], Pascal Zabré[(5)], Géraldine Duthé[(1)], Abdramane Soura[(2)], Gilles Pison[(2)(6)], Ali Sié[(5)]

(1) Institut National d'Etudes Démographiques (INED), Paris, France
(2) Institut Supérieur des Sciences de la Population (ISSP), Ouagadougou, Burkina Faso
(3) Centre de recherche en démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgique
(4) Institut national de la statistique et de la démographie (INSD), Ouagadougou, Burkina Faso
(5) Centre de recherche en santé de Nouna (CRSN), Nouna, Burkina Faso
(6) Muséum national d'histoire naturelle (MNHN), Paris, France

Abstract:

Background: Because of the absence of comprehensive death registration in most countries in Sub-Saharan Africa, mortality levels and trends are regularly derived from retrospective reports on deceased relatives in sample surveys and censuses. These data sources are potentially affected by recall errors.

Objective: Using high-quality data collected in 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 2006 census and compared to those obtained from the prospective mortality data. We linked census and HDSS records at the individual level and evaluated the reported ages of household members and those who died in the 12 month preceding the census against those in the HDSS. We decomposed differences in life expectancies between the census and the HDSS into the effects of age errors and omissions of deceased relatives.

Results: Life expectancies derived from recent deaths reported in households pointed to lower mortality than monitored in the HDSS, with a difference of 3.2 years for men and 6.8 years for women. Recall errors related to the population aged 60 and above accounted for about half of these differences. Age errors were limited for the surviving population, but larger for the deceased, especially among women. Their effect on mortality estimates were modest.

Conclusions: Triangulating national census data with demographic surveillance systems can help in assessing mortality rates derived from various estimation methods. Innovative strategies are required to improve data collection on ages and recent household deaths.

Contribution: This record linkage study suggests that mortality estimates from recent household deaths will be biased downwards in Burkina Faso, especially for female mortality. Underreporting of deaths plays a larger role than age errors.

Keywords: Burkina Faso, Demographic Surveillance System, mortality, indirect demographic techniques, national censuses

Background

Counting who is dying and what they are dying from is fundamental to generate accurate statistics for guiding the resource allocation within the health sector and evaluating the effectiveness of programs to reduce 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, the registration of deaths is still inefficient at the national scale in most of Sub-Saharan Africa, and progress has been modest in this area in recent decades [1]. One of the least developed countries, Burkina Faso, located in Western Africa, is no exception. In the absence of a full-fledged CRVS system, mortality rates are currently derived from survey or census data. While large-scale sample surveys have collected birth and sibling survival histories, allowing the direct estimation of mortality [2], national censuses have included questions on the survival of children, parents, and recent household members, mostly for indirect estimation [3].

A first set of questions in censuses aims to collect summary birth histories, that is, the number of children ever born and surviving of women of reproductive age. There is a general consensus in the literature that the proportion dead of children born to women by age (or duration since first birth) provide robust estimates in the absence of abrupt mortality changes [4,**¡Error! No se encuentra el origen de la referencia.**,6]. Indirect estimates of adult mortality can also be derived from reports on orphanhood, but these are considered with more skepticism. The most pervasive problem is commonly known as “the adoption effect” [7,8,9], caused by enumerators not systematically probing whether children observed in households are the true offspring of the adults being interviewed, or by foster parents deliberately or inadvertently claiming adopted orphans as their own children. Thirdly, censuses also regularly collect data on the number of deaths in each household in the year preceding the enumeration, which makes it possible to generate a complete life table. Biases are introduced by underreporting of deaths due to recall errors, dissolution of some households following the death of adults, but also coverage errors, and transfers of some deaths outside of the 12-month reference period [10,11,12]. A series of "death distribution methods" have been developed to adjust upwards 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 also sensitive to systematic age misstatement [13] and under enumeration of some specific populations, especially young adults or the elderly [14].

The magnitude and direction of these errors are difficult to assess in the absence of a mortality gold standard. While census-based estimates have sometimes been evaluated in simulated environments [15,16,17], there have been few attempts to compare them to high-quality data from Health and Demographic Surveillance Systems (HDSS). In HDSS, the population of a locally defined population is followed regularly through repeated household visits. Individuals enter the population through birth or immigration, and exit through emigration, death or censoring (at the time of the last visit). At each visit, interviewers review the list of the household members who were present at the previous visit, and collect information on births, deaths, in- and out- migrations, but also marriages and sometimes pregnancies. This greatly limits the scope for omissions of vital events and age misreporting, making HDSS one the most accurate sources of mortality estimates in countries without vital registration, but referring only to specific geographical areas [18,19].

There are currently 49 HDSS sites in low- and middle-income countries federated in the INDEPTH Network, many of them in operation since several decades [20]. All these sites are included in the frame of national censuses and they 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 in three rural HDSS sites: under-five 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 significantly under-reported in the 2002 census [21].

In this paper, we use the Nouna HDSS, located in the northwest of Burkina Faso, to extend the comparison and evaluate the reliability of mortality indicators derived from the last national census, conducted in 2006. We compare mortality rates at the aggregate level, and link individual records to evaluate the quality of ages reported in the census and their impact on mortality estimates.

Data and methods

Comparisons at the aggregate level

Our reference dataset is made of data collected regularly in the Nouna HDSS since the 1990s, administered by the *Centre de Recherche en Santé de Nouna* (CRSN) in Burkina Faso. The HDSS was set up in 1992 in a rural area, about 300 kilometers from the capital city, Ouagadougou. After an initial census conducted in 1992, household have been visited three times a year [22]. The population under surveillance was estimated at more than 80,000 individuals in 2006 and is spread over 58 villages and a small town.

The 2006 national census of Burkina Faso was conducted from 9 to 23 December. Based on the names of villages, we extracted individual-level data of the population under surveillance in the HDSS from the census database. We compared the size of population of each village and its composition by sex and age according to the two sources. The definition of the 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 in the present study, we consider as resident an individual whose length of stay in a household located in the area is at least 6 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, the census includes migrants who arrived in the area recently and plan to stay, while they are not considered as residents in the HDSS.

In the 2006 national census, household respondents had to report on the number of deaths by age and sex in the last 12 months. Using the census records covering the area of the population under surveillance, we divided these numbers by the population enumerated in December 2006 to generate a complete life table. Since the census took place in mid-December, we adjusted upwards the number of deaths to account for the absence of deaths reported in last two weeks of 2006. We did not adjust for the small effect of population growth on denominators. In the HDSS, we approached as much as possible the census enumeration method in 2006 to compute the population used as the denominator for rates. More precisely, since HDSS data are 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 June 15 of 2006 (that is, 6 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 [24].

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, based on their names and information collected from local informants. Second, an automatic search based on first and last names was performed on a computer program. Each household enumerated in the census was compared to all households identified in the HDSS. This comparison only included the names of household members. The proximity of names was measured 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 in the presence of spelling variations and data entry errors. If half of the household members were considered as having the same name, we tentatively considered that we were in the presence of the same household in the two data sources. Each member of the HDSS has a unique ID number and we recorded this number for the persons listed in the census for which the full names matched.

Secondly, we reconstructed kinship graphs based on census reports from the links with the head of household, and the order of appearance in the census form (e.g. children are usually placed under their mother in the household roster). The same exercise was performed for the HDSS data, based on genealogical data, and the members unique identifiers were visible in these diagrams. Households matched by the automatic procedure described above were identified. Each relationship diagram was then visually examined by a team member. The objective of this work was to check the automatic record linkages and searching for additional matches. At no time during the record linkage did we use information on ages or other demographic variables. The names, household structure and kinship relations were the only information used. Local informants sometimes had to call the HDSS residents to ask questions about other names or nicknames.

We run logistic regressions on the probability to be matched in order to assess how the matched-sample and the non-matched sample differed in terms of socio-demographic characteristics. Restricting the analysis on the matched-sample, we compared ages of the surviving population as well as of the deceased in 2006 across data sources. Linear regressions were performed on age differences in order to assess the effects of socio-economic characteristics on age misstatement.

[Note for the selection committee for the APC conference: It is important to mention that the individual-level results we present here are based on record linkages for about half of the villages only. We are currently finalizing the other record linkages. In addition, we intend to assess empirically the effects of age misstatement in the census on mortality indicators, by computing a life table using ages reported in the HDSS. Results present here are provisional and will be updated before the APC conference with the full set of linkages and further analysis.]

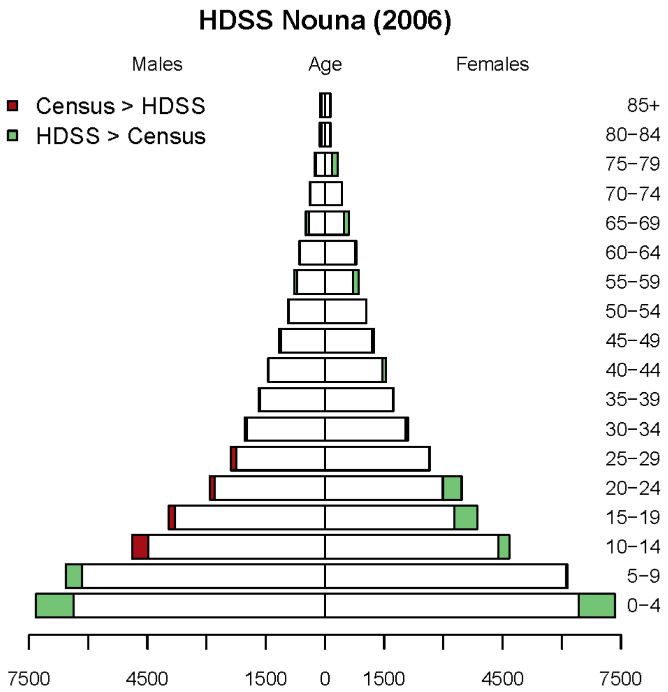
Results

Comparisons at the aggregate level

Population by age and sex in 2006

Figure 1 shows the population pyramid according to the HDSS and the census. Overall, the two sources are highly consistent. The male population is only 2% larger in the HDSS, but the female population is 7% larger in the HDSS, as compared to the census[1]. The most notable difference between the two populations relates to children aged less than 5 years, with a relative difference of - 13% in both sexes, which is 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 to 59, with relative differences lower than 5%.

Figure 1: Population pyramid of the Nouna HDSS according to the demographic follow-up and the 2006 census (December 2006)



[1] Relatives 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.

Number of deaths by age and sex

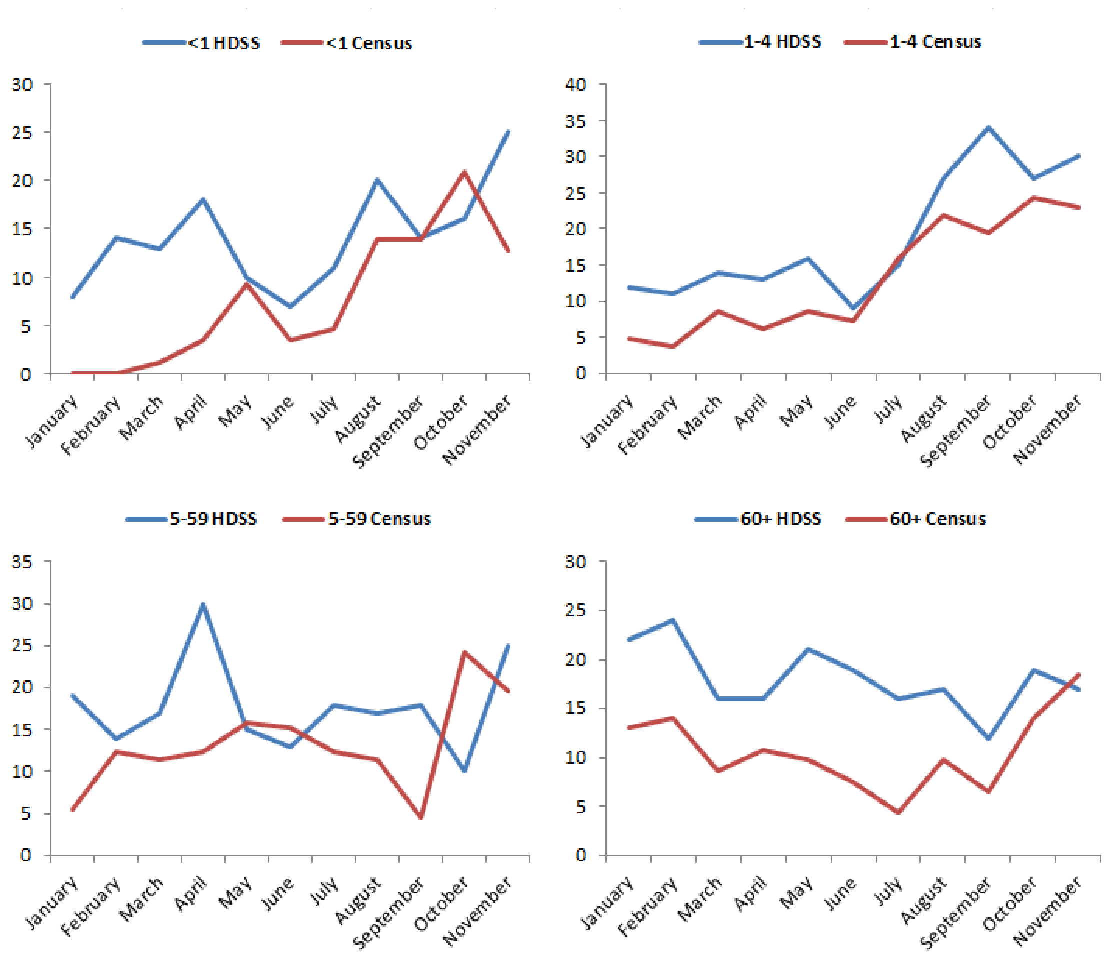
Similar to the population pyramid, table 1 presents the number of deaths reported in 2006 by age group and sex according to the census and the HDSS. Overall, more deaths were reported in the HDSS (827 vs 632), and the underreporting in the census was more pronounced in women. More precisely, compared to the HDSS, 18.1% fewer deaths were collected in men, and 29.6% in women. The underreporting of deaths in the census does not affect all the age groups to the same extent. Relative differences were positive in some age groups (15-29 in both sexes), pointing to more deaths reported in the census, and potentially to age errors transferring some deaths in these age groups. However, they were negative (fewer deaths in the census) below age 15 and above age 60 for both men and women.

Table 1: Number of deaths reported in 2006 by age group and sex according to the census and the HDSS

	Men			**Women**		
Age group	**Census**	**HDSS**	**Relatives differences (%)**	**Census**	**HDSS**	**Relatives differences (%)**
0-1	69	82	-16	44	88	-50
1-4	101	125	-19	86	96	-10
5-14	22	26	-15	18	25	-28
15-29	21	16	31	27	23	17
30-44	21	27	-22	22	25	-12
45-59	46	46	0	12	29	-59
60-74	42	56	-25	44	62	-29
75+	35	55	-36	24	46	-48
Total	357	433	-18	277	394	-29

In addition to table 1, we compare the number of deaths reported in the census by month and age group with the HDSS in 2006 (Figure 2). In both sources, there are more deaths in the recent period preceding the census (August - November) than in the first half of the year, reflecting the seasonality of mortality with a typical excess of child deaths at the end of the rainy season (due to malaria and diarrheal diseases) [25]. However, we note that in all age groups, the ratio between the reported numbers of deaths in the census and in the HDSS is larger in the first half of the year, especially among infants for which the HDSS registered four times more deaths than the census in the period January-July 2006.

Figure 2: 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 the census took place between Dec 9 and 23.

Mortality estimates in 2006 according to the HDSS and the census

Figure 2 compares age-specific mortality rates as estimated from reported number of deaths in the 12 months preceding the census with HDSS estimates for the same year, for each sex. Table 2 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 5% in the probabilities ${}_5q_0$ and ${}_{45}q_{15}$. Still, we observe a difference of 3.2 years in life expectancy at birth, which is largely due to lower mortality among men aged 60 and above. In particular, the risk of dying between ages 60 and 80 among males was 18% lower in the census. Among women, infant mortality is lower in the census, as well as mortality rates in the age group 40-59. There is also a larger difference in mortality rates above age 60 among females. This results in an estimate of life expectancy at birth which is 6.8 years higher in the census for females, when compared to the HDSS. The gap in mortality rates in the age group 60 and above contributes to 3.1 years in this difference in life expectancy at birth.

Figure 3: Age specific mortality rates (ASMR) inferred from the census and the HDSS data, Nouna, 2006

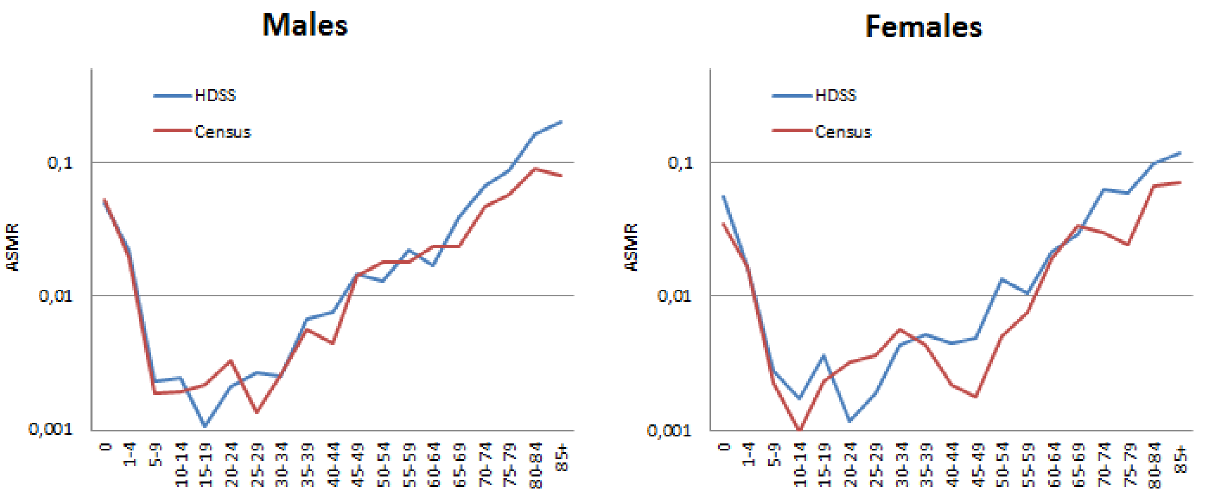


Table 2: Direct estimates of mortality in Nouna according to the HDSS and the reporting of deaths that occurred in households during the last 12 months in 2006 census

	Males				Females			
	Census	HDSS	Relative diff.[1]	Contrib. to the diff.[2]	Census	HDSS	Relative diff.[1]	Contrib. to the diff.[2]
Risk of dying (p. 1000)								
$_5q_0$	123	128	-3%	0.3	96	114	-16%	1.9
$_{10}q_5$	19	23	-19%	0.0	16	22	-28%	-0.8
$_{45}q_{15}$	296	304	-3%	0.9	164	219	-25%	2.5
$_{20}q_{60}$	535	653	-18%	2.1	414	581	-29%	3.1
Life expectancy (in years)								
				Abs. diff				Abs. diff
e0	61	57.8	6%	3.2	68.6	61.8	11%	6.8

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 refers to mortality rates in the open-ended age group 60+ (including mortality above age 80).

Individual-level analysis

Matching rates

Among surviving population, 58% of household members enumerated in the 2006 census could be matched to one individual in the HDSS record. The matching rate was much lower among the deceased individuals (36.30%), because names were missing for a large fraction of the census dataset (21.62%). We investigated how matching rates vary by age group and other demographic characteristics through logistic regression among the surviving population. Table 3 below presents the coefficients of the logistic regression, indicating that matching rates were higher for children aged less than 15, for females, for heads of households and their spouse, and for residents who were actually present at the time of the census.

Table 3: Coefficients of a logistic regression on the probability of an individual enumerated during the census to be matched with an individual in the HDSS in 2006

	Unadjusted coefficients	Adjusted coefficients
Sex: females vs males	0.052^{**}	0.047^{*}
	(0.021)	(0.025)
Age group 15-59 vs <15	-0.136^{***}	-0.337^{***}
	(0.021)	(0.033)
Age group- 60+ vs <15	-0.300^{***}	-0.341^{***}
	(0.045)	(0.052)
Child of HH head vs HH head	0.085^{***}	-0.366^{***}
	(0.028)	(0.040)
Other relationship vs HH head	-0.482^{***}	-0.862^{***}
	(0.039)	(0.045)
Sibling of HH head vs HH head	-0.566^{***}	-0.835^{***}
	(0.074)	(0.076)
Spouse of HH head vs HH head	0.091^{***}	-0.026
	(0.035)	(0.042)
Absent resident vs HH head	-0.243^{***}	-0.190^{***}
	(0.067)	(0.068)
Has attended school vs no education	0.065	
	(0.041)	
Is currently attending vs no education	-0.083	
	(0.087)	
Nb of household members	0.054^{***}	0.065^{***}
	(0.003)	(0.003)
Constant		0.348^{***}
		(0.043)

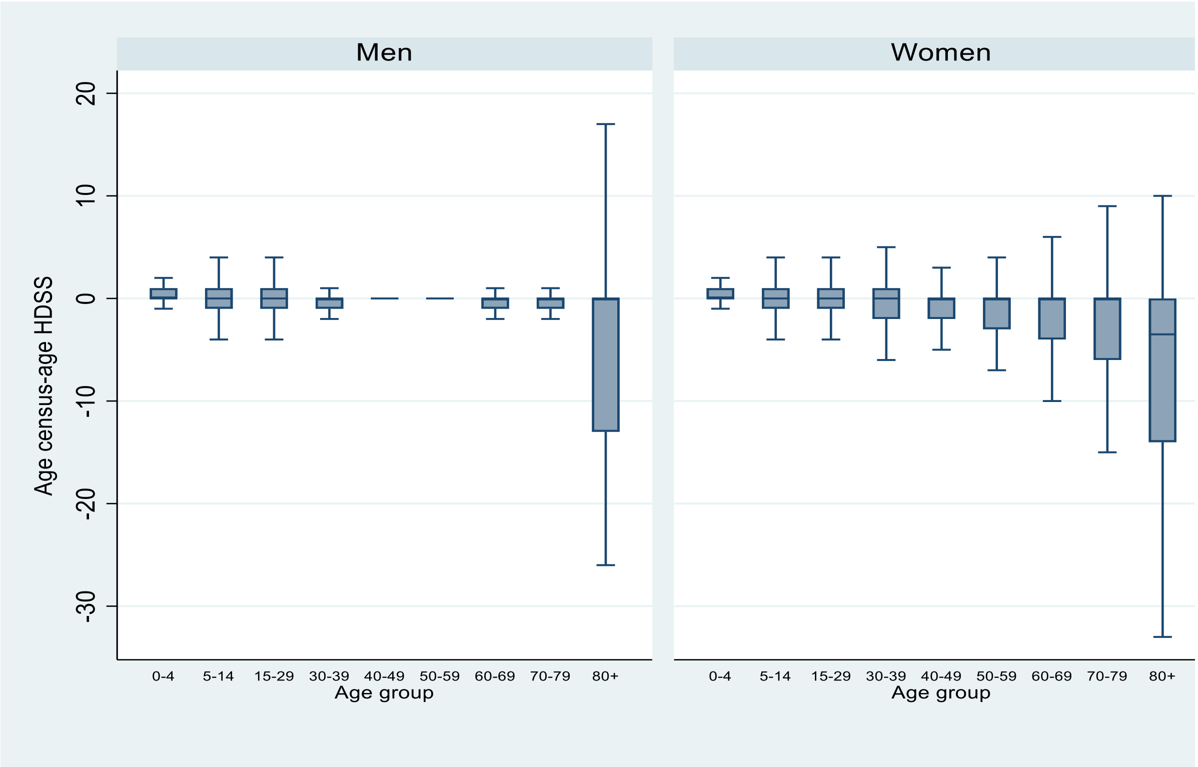
Observations	38,481	38,481

Age misstatement

The information held on ages by the Nouna HDSS is of high quality because particular attention was paid to their collection when setting up the demographic surveillance. Dates of birth of individuals born after the initial census were registered during the follow-up, so they are known with great precision. This is confirmed by a quick evaluation of age reporting in the Nouna HDSS and the census using Myers Index. This index was estimated respectively in men and women at 3.48 and 7.45 in the HDSS, as compared to 10.01 and 14.31 in the census. Whatever the source of information on ages, age reporting was better in men than in women.

To capture the age differences of individuals according to the two sources of data, the distribution of age differences using the HDSS as reference are presented in Figure 4 by sex. In men, ages reported in the census are surprisingly accurate, particularly between age 40 and 60. Small differences appear in children and young adults but these differences are less than 2 years and should not have a large impact on mortality estimation if one relies on age groups for the analysis. Very large age differences are likely due to matching errors.

Figure 4: Age differences in men and women between the census and the HDSS in 2006 using the HDSS as a reference



In women, age differences between the census and the HDSS are salient in the different age groups. Before age 30, the differences are small and centered on zero. As for men, there is no clear pattern of age errors of surviving women reported in the census before age 30. However, above age 30, the age of women tends to be under-estimated, and these age differences tend to increase with age. Irrespective of possible errors on 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 denominators for calculating age-specific rates will be artificially inflated in young women, and reduced in the elderly.

Table 4 shows the adjusted effects of socio-economic characteristics on age differences between the census and the HDSS. In men as in women, the differences tend to increase with age, but their magnitude is higher in women. There is also an association between the relationship to the head of household and age differences between the two sources. The ages of the head of household are more accurate than that of the other members (expect for spouses for which the difference is not statistically significant).

Table 4: Effects of socio-economic characteristics on age differences[2] between the census and the HDSS in men and women in 2006 (Linear regression)

VARIABLES from the census	Men	Women
Education (ref. no educated)		
Educated	0.0365	-0.266
	(0.190)	(0.216)
Don’t know	0.646***	0.0835
Age group (ref. <5 y.o.)		
	(0.205)	(0.206)
Age group 5-14	-1.388***	-1.293***
	(0.228)	(0.238)
Age group 15-29	-2.548***	-2.176***
	(0.248)	(0.272)
Age group 30-39	-3.158***	-3.054***
	(0.364)	(0.367)
Age group 40-49	-3.817***	-3.914***
	(0.413)	(0.401)
Age group 50-59	-3.841***	-4.086***
	(0.479)	(0.449)
Age group 60-69	-4.931***	-5.695***
	(0.540)	(0.531)
Age group 70-79	-5.766***	-5.916***
	(0.679)	(0.702)
Age group 80+	-18.00***	-21.22***
	(1.197)	(1.169)
Status in household (ref. head)		
Spouse of HH head	-3.691	-0.779
	(7.954)	(0.578)
Child of HH head	-1.474***	-1.900***
	(0.318)	(0.612)
Sibling of HH head	-1.881***	-0.751
	(0.466)	(0.986)
Other relationship to HH head	-0.623*	-0.409
	(0.324)	(0.587)
Constant	3.042***	3.394***
	(0.353)	(0.627)

[2] Age in the census-age in the HDSS

Observations	10,855	11,153

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

We conducted the same analysis presented above on the matched sample of those deceased in 2006 (Appendix A2). The matched sample and the non-matched sample did not differ in terms of age and sex. 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 for those age above 80 (Figure A2.1). This means that old-age mortality estimates will most likely be under-estimated due to this age misstatement.

Discussion

In low-income countries, censuses remain a fundamental data source for assessing population dynamics, in the absence of a comprehensive system 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 [3]. Comparison of different data sources is often used, but it is difficult to reach solid conclusions in the absence of a reference that can also be affected by errors. In this paper, we used high-quality mortality data from one HDSS in Burkina Faso to evaluate estimates inferred from one census. We found a good agreement in population counts, despite some deviations in children aged less than 5 and in young adults. The larger number of young males in the census could be related to recent arrivals in the area of men who intend to stay but have not yet spent 6 months in the HDSS. However, the undercounting of young women is more difficult to explain. One can speculate that some young women who temporarily left the HDSS area for educational or professional reasons were still considered as resident in the HDSS despite not being counted as such in the census. Indeed, in the site of Nouna, some young females are engaged in seasonal outmigration (during the dry season) towards the towns of Ouagadougou and Bobo Dioulasso. Collecting comprehensive data on seasonal migrations in both national censuses and HDSS is needed to adequately reflect demographic patterns in local areas and reduce biases caused by selective migration.

In terms of mortality rates, 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. Our analysis clearly indicates that this assumption would be violated in the census data, even when restricting the analysis to adults.

A number of reasons may be put forward to explain these deviations with HDSS estimates. First, deaths could be underreported in the census, either 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). As the census included a question on months of deaths, we compared the distribution of reported deaths by month prior to the census date, and observed that the number of reported deaths declines as the number of months between their occurrence and the census increases (Figure 2), pointing to omissions. Second, like in many other censuses, there is a mismatch between the reference period used for recent household deaths in the census (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, the ages of the deceased and the surviving population are affected by misreporting, which is more pervasive among women. Finally, 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 kind of selection bias will be more prevalent for deaths that occurred in the more distant past, as compared to deaths immediately preceding the census.

When the quality of data on recent household deaths is called into question, demographers regularly turn to indirect techniques for mortality estimation. For example, reports on the numbers of children ever born and surviving provide under-five mortality estimates, and information on parental survival is used to generate adult mortality rates. We used these methods and compared with the underlying mortality rates in the HDSS (See Appendix A.1). 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 necessarily more reliable than those derived from recent household deaths.

The patterns observed here are slightly different from previous comparisons made in Senegal, where direct estimates based on recent household deaths were plausible beyond the age of five, while indirect estimates of under-five mortality too low in the censuses [21]. This reinforces the need to undertake similar analysis in other HDSS sites in Burkina Faso and other countries where vital registration systems are incomplete before generalizing our results.

Finally, this study also highlights the need for renewed efforts to improve data quality by seeking to limit the use of proxy respondents in censuses, and developing innovative ways to improve the reporting of age and demographic events, such as historical calendars.

It is important to point out some limitations of this research. First, ages derived from the HDSS cannot be taken at face value, even if dates of events are known with great precision. Age misreporting may affect some groups of individuals in the HDSS; this is particularly the case for in-migrants, and individuals who were present at the initial census. Even if particular attention was paid to age reporting, the dates of birth collected at the time of first enumeration remain of poor quality. The results for the population aged 60 and above should therefore be interpreted with cautious. Second, as we could not link all individuals, the analysis conducted on age reporting was based only on the matched sample. However, the probability of an individual enumerated during the census to be matched with an individual in the HDSS depends on various criteria including omission, migration, age and sex. We can expect that age errors to be larger among individuals we failed to matched compared to those who were successfully linked. This can lead to an underestimation of age differences. For example, as matching rate was lower among males than among females, it is possible that age misreporting in men was underestimated compared to women.

Conclusions

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 in children as well as in adults were underestimated in the 2006 census in this area, especially among women. A key result of this analysis is also that omissions of deaths play a larger role than age errors in explaining these gaps.

Appendix

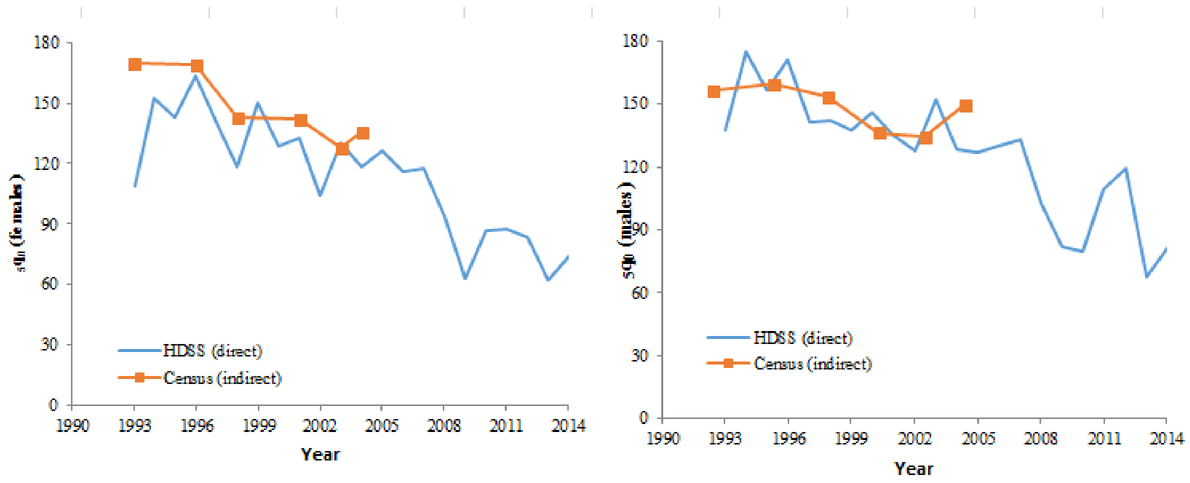
A.1. An evaluation of indirect mortality indicators derived from the census

Indirect estimation of under-five mortality

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-five mortality rates were estimated indirectly from these numbers, using the North pattern of model life tables [4]. This method uses a set of standard coefficients capturing variations in mortality and fertility patterns to convert proportions dead of children born to women by age group into the probability $_5q_0$, that is, the risk of a newborn dying before his or her fifth birthday. 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-five mortality derived from census data on children ever born and surviving are fairly consistent with the HDSS estimates (FigureA1.1). As expected, census estimates are smoothed, because the time-location of indirect estimates is based on an assumption of linear and unidirectional mortality decline. If we discard the period 1992-3, which was the start of the follow-up, and may have been affected by underreporting in the HDSS, census estimates for girls are on average 8% higher than estimated from the demographic 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. Considering that census and HDSS estimates do not necessarily refer to the same children (e.g. 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), this consistency is remarkable.

Figure A1.1: Comparison of indirect estimates of child mortality ($_5q_0$) inferred from the census with HDSS estimates

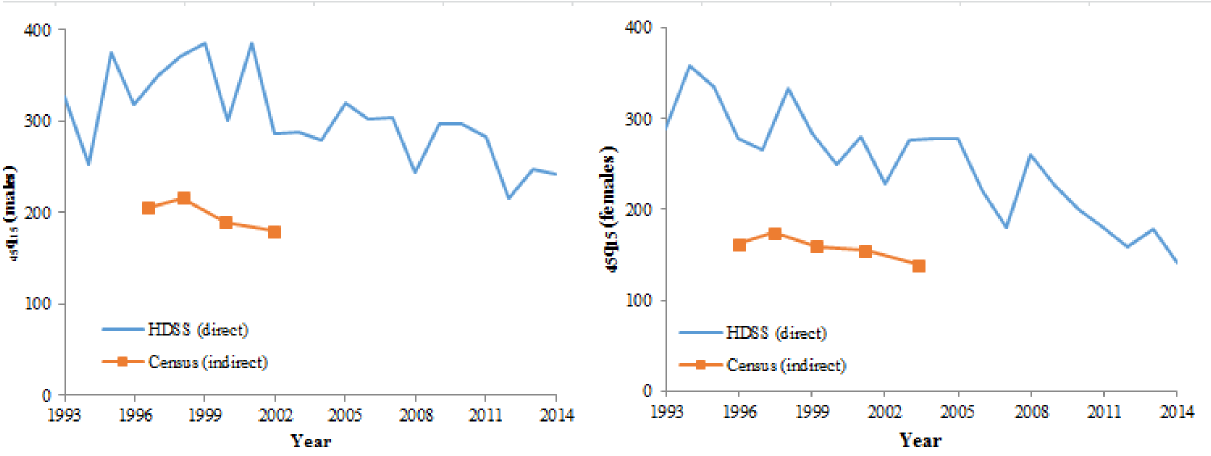


<u>*Indirect estimation of adult mortality*</u>

The census also asked about the survival of parents of individuals aged less than 30. We applied the standard orphanhood method [23]. Estimates obtained for the different age groups of respondents were converted into the probability $_{45}q_{15}$ (the risk of an individual aged 15 dying before reaching 60 under the mortality rates prevailing in a given year), again using the North model life table.

Figure A1.2 indicates that adult mortality estimates derived from orphanhood data collected in the census are implausibly low, when compared to HDSS estimates. These are 43% lower on average in women during the period 1996-2003 and 40% lower on average in 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 live 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. 78% 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 to 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 deaths of parents, for both maternal and paternal orphanhood.

Figure A1.2: Comparison of indirect estimates of adult mortality ($_{45}q_{15}$) inferred from the census with HDSS estimates



A2. Age differences among the deceased persons in 2006

Figure A2.1: Age differences of deceased persons between the census and the HDSS in 2006 using the HDSS as a reference

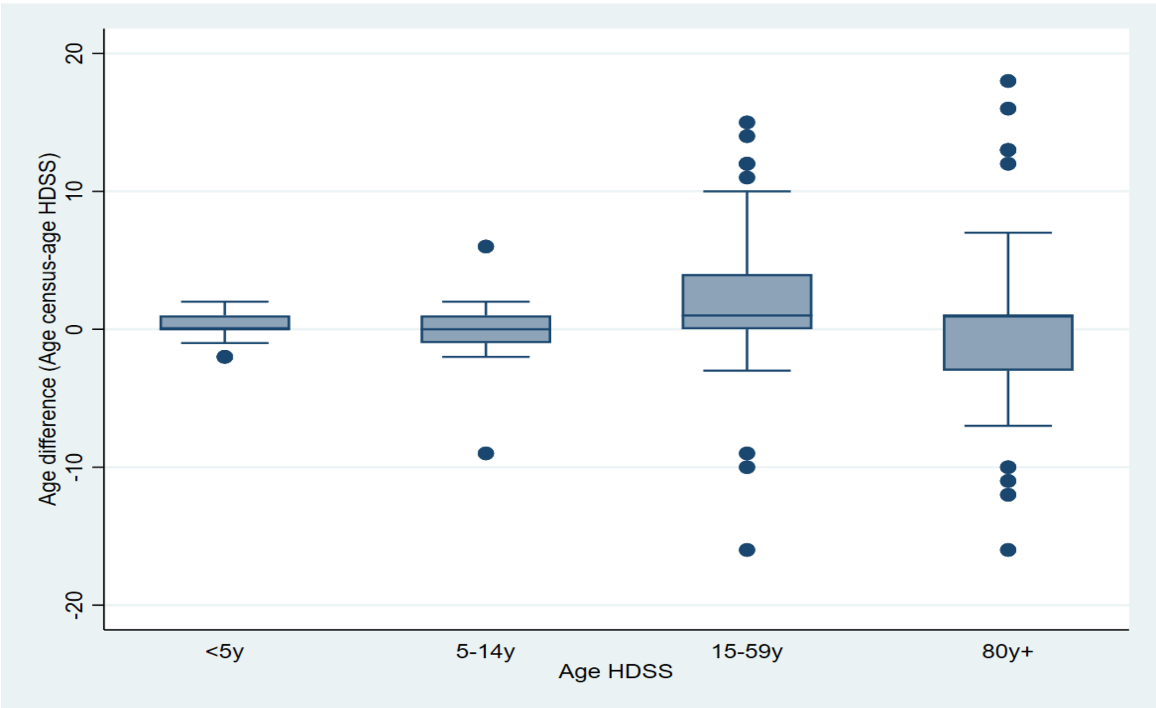


Table A2.1: Logistic regression on the probability of a death reported during the census to be matched with a deceased person in the HDSS in 2006

VARIABLES	Adjusted odd ratios
Age group 5-14	1.350
	(0.479)
Age group 15-59	0.908
	(0.197)
Age group 60+	1.125
	(0.244)
Gender	0.871
	(0.150)
Constant	0.601***
	(0.0865)
Observations	594

se in parentheses ; *** p<0.01, ** p<0.05, * p<0.1

Table A2.2: Effects of gender and age group on age differences between the census and the HDSS among deceased persons in 2006

VARIABLES	Age difference
Gender	-0.295
	(0.571)
Age group 5-14	0.129
	(1.100)
Age group 15-59	0.794
	(0.724)
Age group 60-79	-0.467
	(0.780)
Age group 80+	5.393***
	(1.207)
Constant	0.313
	(0.483)
Observations	214

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

List of abbrevations

HDSS – Health and Demographic Surveillance Systems

Declarations

Ethical approval: We did not seek ethical approval for this study because it is based on secondary analysis of aggregate census and HDSS data, with no record linkage involved. Ethical approval was the responsibility of the institutions which collected the data.

Availability of data and materials: This study used an extract of the complete database of the 2006 national census of Burkina Faso. While this extract cannot be made publicly available, a representative sample of the national census is available through the IPUMS program (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 analyzed the census data, PZ and BL analyzed the HDSS data. BM, HZ and BL wrote the first draft of the paper. All authors read and approved the final manuscript.

References

[1] L. Mikkelsen, D. E. Phillips, C. AbouZahr, P. W. Setel, D. de Savigny, R. Lozano, and A. D. Lopez. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet*, May 2015.

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

[3] T. . Moultrie, R. Dorrington, A. Hill, K. Hill, I. Timæus, and B. Zaba. *Tools for demographic estimation*. Paris: International Union for the Scientific Study of Population, 2013.

[4] K. Hill. Indirect estimation of child mortality. In T. Moultrie, R. Dorrington, A. Hill, K. Hill, I. Timaeus, and B. Zaba, editors, *Tools for Demographic Estimation*, chapter 16. UNFPA/IUSSP, 2013.

[5] J. Rajaratnam, T. Linda, A. Lopez, and C. Murray. Measuring under-five mortality: Validation of new low-cost methods. *PLoS Med*, 7 (4), 04 2010.

[6]R. Silva. Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories. *PLoS Med*, 9 (8): e1001296, 08 2012.

[7]J. Blacker. The Estimation of Adult Mortality in Africa from Data on Orphanhood. *Population Studies*, 31 (1): 107–128, 1977.

[8] K. Hill and J. Trussell. Further developments in indirect mortality estimation. *Population Studies*, 31(2): 313–334, 1977.

[9]L. Robertson, S. Gregson, C. Madanhire, N. Walker, P. Mushati, G. Garnett, and C. Nyamukapa. 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, 2008.

[10] N. Bennett and S. Horiuchi. Mortality estimation from registered deaths in less developed countries. *Demography*, 21(2): 217–234, 1984.

[11] K. Hill. Estimating Census and Death Registration Completeness. *Asian and Pacific Population Forum*, 1(3): 8–13, 1987.

[12] I. Timæus. Measurement of adult mortality in less developed countries: A comparative review. *Population Index*, 57 (4): pp. 552–568, 1991.

[13] D. Ewbank. Age misreporting and age-selective underenumeration: Sources, patterns and consequences for demographic analysis. Committee on Population and Demography, Report No. 4, 1981.

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

[15] K. Hill, D. You, and Y. Choi. Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data error. *Demographic Research*, 21: 235–254, 2009.

[16] C. Murray, J. Rajaratnam, J. Marcus, T. Laakso, and A. Lopez. What Can We Conclude from Death Registration? Improved Methods for Evaluating Completeness. *PLoS Med*, 7 (4): e1000262, 04 2010.

[17] A. Verhulst. Child mortality estimation: An assessment of summary birth history methods using microsimulation. *Demographic Research*, 34 (39): 1075–1128, June 2016. URL https://ideas.repec.org/a/dem/demres/v34y2016i39.html.

[18] P. K. Streatfield, W. A. Khan, A. Bhuiya, N. Alam, A. Sié, A. B. Soura, B. Bonfoh, E. K. Ngoran, B. Weldearegawi, M. Jasseh, A. Oduro, M. Gyapong, S. Kant, S. Juvekar, S. Wilopo, T. N. Williams, F. O. Odhiambo, D. Beguy, A. Ezeh, C. Kyobutungi, A. Crampin, V. Delaunay, S. M. Tollman, K. Herbst, N. T. Chuc, O. A. Sankoh, M. Tanner, and P. Byass. Cause-specific mortality in Africa and Asia: evidence from INDEPTH health and demographic surveillance system sites. *Global Health Action*, 7 (1): 25362, 2014.

[19] G. Pison. Population observatories as sources of information on mortality in developing countries. *Demographic Research* 13 (2005): 301-334.

[20] O. Sankoh and P. Byass. The INDEPTH Network: filling vital gaps in global epidemiology. *Int J Epidemiol*, 41 (3): 579–588, Jun 2012.

[21] B. Masquelier, C. Ndiaye, G. Pison, B. Dieme, S. Helleringer, O. Ndiaye, and V. Delaunay. Evaluation des estimations indirectes de mortalité dans trois observatoires de population au Sénégal . *African Population Studies*, 30 (1), 2016.

[22] A. Sie, V. R. Louis, A. Gbangou, O. Muller, L. Niamba, G. Stieglbauer, M. Ye, B. Kouyate, R. Sauerborn, and H. Becher. The Health and Demographic Surveillance System (HDSS) in Nouna, Burkina Faso, 1993-2007. *Glob Health Action*, 3, Sep 2010.

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

[24] E. E. Arriaga. Measuring and explaining the change in life expectancies. *Demography*, 21 (1): 83–96, Feb 1984.

[25] G. Kynast-Wolf, G. Hammer, O. MÃ¼ller, B. KouyatÃ©, and H. Becher. Season of death and birth predict patterns of mortality in Burkina Faso. *International Journal of Epidemiology*, 35 (2): 427–435, 2006.