Sara Randall and Ernestina Coast
The quality of demographic data on older Africans

Article (Accepted version) (Refereed)

Original citation: Randall, Sara and Coast, Ernestina (2016) The quality of demographic data on older Africans. Demographic Research, 34 . pp. 143-174. ISSN 1435-9871
DOI: 10.4054/DemRes.2016.34.5

© 2016 The Authors

This version available at: http://eprints.lse.ac.uk/64834/

Available in LSE Research Online: September 2016

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.
The quality of demographic data on older Africans

Sara Randall
Department of Anthropology, University College London, Gower St, London WC1E 6BT
s.randall@ucl.ac.uk

Ernestina Coast
Department of Social Policy, LSE, Houghton St, WC2A 2AE London
e.coast@lse.ac.uk

Statement of ethical approval:
No ethical approval required for this research, which used publically available datasets.
There are no conflicts of interest.

Corresponding author: Sara Randall  s.randall@ucl.ac.uk

Acknowledgements

The research that stimulated the writing of this paper was financed by ESRC (UK) under the collaborative ESRC-ANR research programme (RES 062-33-0007). Thanks to Ousmane Ida Ibrahima for Niger census data, Philippe Antoine for Senegal census data, Bilampoa Gnoumou for Burkinabe census data and Richard Marcoux and ODSEF for the Malian census data and IPUMS for sample census data for Kenya and Tanzania. Thanks also to Emily Freeman for comments on an earlier draft and to Caroline Ugglå and Heini Vaïsanen for preparatory work on DHS data. The Living Standards Measurement Surveys (LSMS) data were obtained via the World Bank’s Microdata Catalog: the Ethiopian ERSS 2011-12 data were collected by the Ethiopia Central Statistical Agency (CSA) and the Niger 2011 ECVMA by the Niger Institut National de la Statistique (INS). We would like to thank both sources for access to these data sets. An earlier version of this paper was presented in French at the AIDELF conference in November 2012, Ouagadougou, Burkina Faso. We would also like to thank four anonymous reviewers for their comments and the editor for helping reconcile the required improvements.

1 (http://microdata.worldbank.org/index.php/catalog/lsms)
ABSTRACT

BACKGROUND

Developing appropriate and equitable policies for older people in Africa requires accurate and reliable data. It is unclear whether existing data can accurately assess older African population structures, let alone provide the detailed information needed to inform policy decision making.

OBJECTIVE

To evaluate the quality of nationally representative data on older Africans through examining the accuracy of age data collected from different sources.

METHODS

To measure the accuracy of age reporting overall we calculate Whipple’s Index, and a modified Whipple’s Index for older adults, using the single year age-sex distributions from (a) the household roster of 17 Demographic and Health Surveys (DHS) (b) the censuses of 12 of these countries and (c) the Living Standards Measurement Study (LSMS) for Ethiopia and Niger. We compare reported sex ratios by age.

RESULTS

The quality of age data is very poor for most countries outside Southern Africa, especially for older adults. In some Sahelian countries DHS surveys appear to omit a considerable proportion of older women. Data on population structure of older people by age and sex produced by the DHS and the census are inconsistent and contradictory.

CONCLUSIONS

Different field methodological approaches generate contradictory data on older Africans. With the exception of Southern Africa, it is impossible to assess accurately the basic demographic structure of the older population. The data available are so problematic that any conclusions about age-related health and welfare and their evolution over time and space are potentially compromised. This has ramifications for policy makers and practitioners who demand, fund and depend on large scale demographic data sources.

Contribution of the paper

The paper highlights a number of problems with data on older Africans beyond the well-known issues of age heaping. In doing so it contributes to general understanding of the limitations of existing demographic data for any detailed analysis of the situation or characteristics of older Africans. The heterogeneity of data quality and data problems for older Africans across the continent suggest that considerable care should be taken in (a) drawing conclusions from comparative studies using internationally standardised data sets and (b) analyses which combine different sources of data. Particular data problems with surveys in the Sahelian countries could be addressed in future data collection exercises.
Introduction

Age, together with sex, is one of the key “demographic variables” (Bogue 1969 p.147), and is the “centrepiece” of much demographic theory (Poston Jr and Micklin 2006 p.23). In order to accurately understand population change – its scale, speed and variants – we therefore need accurate data on the age of a population’s members. Historically, errors of age reporting in censuses have received considerable attention from demographers because “errors are readily apparent” and “measurement techniques can be more easily developed for age data” (Shryock and Siegel 1976 p.115). Techniques to improve age reporting and recording include the use of “landmark” events with which to anchor life history calendars (Axinn, Pearce et al. 1999), and many census enumerator’s manuals include event calendars to help enumerators determine ages. In low income countries, whilst levels of knowledge of age or date of birth are increasing, especially for younger cohorts due to increased education and administrative requirements for date of birth, there remain substantial problems of age reporting and recording for older people and for those who are unschooled. These problems are particularly acute in sub-Saharan Africa where high levels of illiteracy and low levels of development are combined with a widespread social irrelevance of knowing absolute age until recently, although relative age has always been important.

The percentage of the population aged 60+ has remained constant for sub-Saharan Africa at 4.8% since 1970. Over the same period the percentage aged 60+ increased from 4.8% to 5.1% for all low income countries, from 6.1% to 10.5% for middle income countries and from 13.8% to 20.3% for high income countries (UN 2015). Despite these low and stable proportions the absolute numbers of older people in sub-Saharan Africa are increasing (Velkoff and Kowal 2007) because of rapid population growth. It is unclear whether adequate data exist to assess the numbers of older people accurately, let alone provide the detailed information needed to inform age-appropriate policy decision making. Whereas in Europe and in many countries in Asia and South America, specific surveys focus on older populations, in much of Africa most statistical information on older people comes from more general data sources such as censuses, Demographic and Health Surveys (DHS), Living Standards Measurement Surveys (LSMS) and other nationally representative sample surveys as well as a number of geographically limited demographic surveillance systems (DSS). The relatively small proportions of older people in Africa contributes to this lack of surveys focusing on this demographic sub-group. There are a few detailed studies on aging such as the SAGE (WHO Study on global AGEing and adult health) studies (WHO n.d.-a) , which, at a national level in Africa only covers Ghana (Biritwum, Mensah et al. 2013) and South Africa but which is also implemented within demographic surveillance studies (DSS) in Kenya (Nairobi), Ghana (Navrongo), South Africa (Agincourt) and Tanzania (Ifakara). However these localised studies do not allow a general profile of the older population at national level – for which a census or nationally representative sample survey is needed.
Data on older Africans may be deficient in two ways: limited scope and inaccuracy. In censuses, where data on a range of variables are available for the whole population, the scale and speed of the exercise, and the temporary employment of inexperienced enumerators probably mean that little attention is paid to accurate age reporting of older people, especially in contexts where many older people are illiterate, few know their ages and trying to obtain accurate ages using event calendars is time consuming. One would therefore expect age accuracy to be the main problem in censuses whereas the scope (such as economic activity and migration) of census data is the same for older and younger adults.

Since the mid 1980s the DHS have become an extremely widely used demographic data source for Africa because of their standardised methodology, the fact that they are, generally, nationally representative, and because the data are in the public domain. The DHS are primarily designed to collect detailed data on reproduction and health and focus on issues affecting the large majority of the population: children and younger women and men. Most data are collected on women and men of reproductive age (generally, 15-49) and women’s children (DHS 2015). Data for the entire population are, however, recorded on the DHS household questionnaire; although the primary purpose of these household roster data is to determine eligibility for anthropometry, biometry and the individual questionnaires, they also provide the household context for reproductive and health analyses. DHS standardised format, open access and ease of acquisition has meant that these DHS roster data have been used for a wide range of analyses which go well beyond the original DHS focus. In the DHS only age, sex, education, relationship to head of household and thus residential arrangements - themselves very constrained by household definition (Randall, Coast et al. 2011) - are collected for older people; marital status has been added to the new DHS model household questionnaires and recent surveys also include this variable. Thus DHS data for older people are extremely limited in scope, but they can be used to study age structure and residential arrangements as well as a number of household level characteristics, living conditions, and poverty (via household assets) of older people.

LSMS data have much wider scope than DHS with much more information on health and economic activities collected at individual level for older members of the population. LSMS surveys are available for a smaller range of countries than the DHS, are less standardised and comparable, and less frequently used for secondary analyses. Other data sources on issues around health and wellbeing of older people in Sub-Saharan Africa include the World Health Surveys which were undertaken in the early 2000s in 18 African countries (WHO n.d.-b); because the questionnaire allows for those who do not know their age to be recorded in age ranges, the quality of age reporting in this survey cannot be evaluated.

A report on the nationally representative Ghanaian SAGE study (Biritwum, Mensah et al. 2013) does examine the quality of age reporting, having highlighted the importance of good age data in a study where ‘age is a central focus’. Using Myers’ Blended Index they conclude that age heaping and incorrect aging is very low in this survey but this conclusion is drawn from calculating the index across the whole age range for a youthful population which is dominated by younger people for whom age reporting appears adequate. The single year age distribution of those over 60 shows not only considerable heaping on ages ending in 0
and 5, there also appears to be a deflation of age to below 50 for men. This may well be an enumerator mediated response to the focus of the survey on older people and the fact that those aged 50+ had to complete a number of questionnaires. Thus even in a study whose prime focus is the older population there are apparent problems with the quality of the age data.

There are therefore several questions around the quality of data for older people in sub-Saharan Africa. Are data inaccurate because many older Africans do not know their ages? Are there gender differences in age data quality? Is there any evidence that different methodological approaches (survey vs. census) generate different patterns of age and other errors for older people? A key issue is whether data are accurate enough to provide meaningful denominators for rates. Age reporting from censuses and surveys will also indicate whether ages provided by other sources (e.g. administrative or hospital data) are good enough to provide meaningful numerators.

This paper aims to evaluate the quality of nationally representative data on older Africans for a range of African countries through examining the quality and comparability of age data collected from different sources. These are the fundamental data which should allow us to establish the age-sex structure of the older population, essential for most of the indicators highlighted in the “Minimum Data Set on Ageing and Older Persons in Sub-Saharan Africa” (Ferreira and Kowal 2006). We investigate (a) whether patterns of age reporting / misreporting are similar or different using different data sources for the same country and (b) the evidence around completeness of enumerating older people. Whilst there are smoothing techniques to correct for age misreporting (Carrier and Farrag 1959, Arriaga, Johnson et al. 1994), we do not employ them here because our objectives are focused around the quality of reporting. We have limited our comparison to the widely available data of the census and the DHS, and to a lesser extent, the LSMS. These data collection exercises use different approaches to data collection, although data on all members of a household are usually provided by one respondent. Since the DHS use a well-tested cluster sampling approach and field methodology, have refined their data collection techniques over nearly 30 years and use experienced interviewers who have undergone a rigorous training programme, it is assumed that, despite the focus on children and younger adults, the quality of age reporting in this sample survey should be similar to, and possibly better than, that of other nationally representative sample surveys. For two countries for which data are available, Niger and Ethiopia, we also compare the LSMS profiles of age reporting.

**Older people and welfare in sub-Saharan Africa**

There is considerable concern that old age poverty will become a growing problem for developing countries (Barrientos, Gorman et al. 2003) and studies have highlighted the increasing stresses that younger adults face in caring for elderly parents whilst themselves confronting unemployment and demands from their own children (Apt 2002, Aboderin 2004a, Aboderin 2004b, Antoine 2007, Roth 2010). Other difficult policy issues relate to the provision of health and social care to populations with increasing absolute numbers of older people (McIntyre 2004, Mudege and Ezeh 2009) even when the proportion of the older
population remains stable. Aboderin highlights a number of policy and evidence gaps associated with the health of older African populations, many of which require better data and measures of individual level status as well as access to services (Aboderin 2013).

As both census and survey data have been used for international comparative studies of older adults’ living arrangements it is appropriate to investigate the quality and deficiencies of the data on the older people included in these analyses. Studies focus on living arrangements for a number of reasons: in the absence of widespread pensions or social protection in Africa, family are the primary source of support for older people who cannot care for themselves and co-residence gives some, albeit limited, information on the likely provision of family support. Such support could be in either direction: older people supporting their (often unemployed) children (Antoine 2007, Roth 2010) or adult children supporting their older relatives (Lloyd-Sherlock 2004). In the era of the HIV epidemic there has been considerable focus on the role of older people in caring for orphaned grandchildren, and the impact of such care on the children in terms of poverty and schooling, labour market participation and other dimensions of child well-being. The scale of such living arrangements in different contexts and the characteristics of these households, has been investigated, with multi-country studies usually using DHS data (Zimmer and Dayton 2005, Zimmer 2009, Kautz, Bendavid et al. 2010). However the focus on living arrangements is probably also a function of the fact that these are really the only comparative data available on the older population for large cross-national analyses. To the extent that the household units used in surveys and censuses are able to capture co-residence and support, such data do provide some evidence for different patterns of organisation both world-wide and in sub-Saharan Africa.

Ruggles and Heggeness’ (2008) study of intergenerational co-residence using census data shows little evidence of a decline in intergenerational co-residence in developing countries, in contrast to Europe where intergenerational co-residence declined rapidly during industrialisation and urbanisation. Only 4 African countries are included: South Africa, Kenya, Rwanda and Uganda. The study uses 65+ as the cut off for defining older people to avoid problems posed by men in their 60s married to younger wives and with co-resident dependent children. The authors criticise DHS data for their limited time depth and small sample sizes although they make no comments on the quality of the data collected on older populations either in their census analyses or the DHS (Ruggles and Heggeness 2008).

Using DHS data to analyse the household structures and living arrangements of older adults (defined by the authors as 65+) across 43 developing countries of which 23 were African Bongaarts and Zimmer (2002) note the data limitations enforcing a focus on co-residence and the absence of data on exchanges between older adults and their children both within and outside the household. They justify analysing co-residence with adult children because this is a frequent source of support for older adults in the developing world. They conclude that, despite substantial variation world-wide, most older adults tend to live in large households and often live with an adult child who is more likely to be male than female. They do not reflect on the quality of the DHS data for older people either in terms of age reporting nor the completeness of the enumeration of older people.
Zimmer and Dayton’s study of co-residence of older adults (defined as 60+) with children and grandchildren in 24 Sub-Saharan African studies also uses DHS household roster data, justified because comparable DHS data exist for the widest range of countries (Zimmer and Dayton 2005). Their regression analyses suggest quite different residential patterns for those aged 60-64 compared to those aged 65+ which implies that age is an important factor – yet they do not consider the quality of the age data used. Zimmer and Das (2013) use DHS data and a cut-off of age 60 for older sub-Saharan African households to examine whether there is a link between household composition of older persons and household wealth and concluded that there were substantial inequalities in wealth by household composition. The construction of the household typology was partially dependent on the ages of the members and thus quality of age data was an important underlying issue for analysis and conclusions although no assessment of the quality of age reporting data was made.

Velkoff and Kowal use UN projections to look at the dynamics and projected growth in absolute numbers of older people in sub-Saharan Africa in the coming decades (Velkoff and Kowal 2007). They argue that in sub-Saharan Africa where life expectancy at birth is relatively low, 50 may be a more appropriate cut-off for defining older age. 50 is also the suggested cut-off for the Minimum Data Set on Ageing and Older Persons in Sub-Saharan Africa (Ferreira and Kowal 2006) based on the relatively low life expectancies in Africa. The definition of old age is dependent on the issues being studied. For example, authors (cited above) who focus on living arrangements point out that where there are large spousal age differences and high fertility, both men and women in their 50s will be co-residing with dependent young children.

Two issues emerge: firstly what should be considered ‘old’ in Sub-Saharan Africa and secondly whether existing data are accurate enough to establish the ‘real’ ages of people. There is little reflection on the extent to which any age cut-offs actually represent the ‘real’ ages of the persons concerned, and even less on the impact of age misreporting (age inflation\(^2\) or age heaping\(^3\)) on understanding patterns of living arrangements.

A third issue related to data quality is that of completeness of enumeration of older people. Data from censuses and surveys are predicated upon identifying individuals in households; if older people have ambiguous status in a household or receive support from several households they might be less likely to be enumerated. Methodological differences in data collection may influence the likelihood of some older people being included as well as the quality of age data. In African censuses housing units are identified and numbered; enumerators move from housing unit to housing unit in the enumeration area for which they are responsible, obtaining information on unoccupied units and making call-backs until all the units in their enumeration area have been covered, including housing which might have been missed by earlier cartography phases. Census enumerators are instructed to speak to the household head or another competent adult household member to collect information on the

---

\(^2\) The tendency to report ages which are higher than their actual age.

\(^3\) Also known as digit preference. Age avoidance – a lack of heaping – is another manifestation of error typically found in single-year-of-age data.
whole household. Older men and women are more likely to be at home than younger working age adults and they will provide much of the information for household members including ages. Where younger adults are asked for ages of older adults it is likely that, in environments where birth certificates are rare and many older people are illiterate, younger adults will be unable to provide accurate data – and event calendars are unlikely to be much help.

Surveys use a different approach. Clusters are selected from a national sampling frame and lists of households in those clusters are updated using administrative records and information from local informants. Households are then selected for inclusion in the survey. In some contexts widows may be more likely to be excluded from such lists: for example, if they are partially supported by a son’s household, the survey administration may not consider the widow to be a separate household and she may not be listed – and therefore cannot be selected. However when the household to which they are partially attached is enumerated she may be forgotten, not considered to be a household member by the respondent, or excluded from being a household member by a household definition which requires that they sleep under one roof and/or eat together (Randall, Coast et al. 2015). Furthermore surveys collect much more detailed data than a census and the respondent is usually required to be an adult who knows details about the household economy. This means that information will rarely be provided by the older person for themselves but probably by a younger adult, which will contribute to age misreporting.

Age misreporting and/or under-enumeration of older people may differ considerably between populations, with implications not only for interpreting comparative analyses but also for the policies that these data and analyses might inform. Coale and Kisker showed for a number of populations that age misreporting at older ages generated misleading mortality rates (Coale and Kisker 1986). Poor age reporting in many sub-Saharan African populations has implications for the interpretation of a number of development-related indices such as the dependency ratio.

Data and methods

In order to address our questions we undertake secondary analyses of nationally representative datasets (census, DHS). For two countries (Niger and Ethiopia) we use the LSMS as a third source. We define the older population as those aged 60+. This cut-off was chosen for several reasons: a number of other studies using DHS data for sub Saharan Africa have used 60 as the cut off (Zimmer and Dayton 2005, Kautz, Bendavid et al. 2010, Zimmer and Das 2013); the selection of a specific age is somewhat arbitrary since we are just comparing reporting from different data sources; the digit preference for 0 over 5 is likely to be very strong and thus it seemed prudent to include 60; the selection of women aged 49 or under for the DHS individual questionnaires has contributed to age inflation over the 49/50 boundary and therefore it seemed appropriate to avoid age 50 for the analysis of age heaping – although the impact of this DHS age inflation for women is highlighted below in analyses of sex ratios.
Single year age-sex distributions from the household roster were obtained from DHS data for the mid 2000s for 17 countries representing different regions in Africa (West Coastal, Sahelian, East, Southern Africa)⁴ and census data (usually for the 1990s) were obtained for twelve countries⁵. Niger (2002)⁶ and Senegal (2002) census data were obtained by the authors and their collaborators from national statistics offices. Uganda (2002) census data are from the Integrated Public Use Microdata Series (IPUMS) 10% sample⁷. In order to make comparisons over time we used single year age-sex distributions from the household roster from all DHS undertaken in Mali, Burkina Faso, Tanzania, Kenya and Zambia and from these countries’ censuses available in a 10% sample via IPUMS or, in the case of Mali (1987, 1998, 2009) the full census data from the national statistics office. These countries were chosen because preliminary analyses showed that they had very different levels of accuracy in age reporting. Unfortunately census data from Zambia have an open upper age group of 80+ rendering them inadequate for our comparisons.

There are several possible approaches to measuring age data quality: Whipple’s Index, Myers’ Blended Method and the United Nations Age-sex accuracy index (UN 1955, Hobbs 2004). We chose to use Whipple’s Index which is a method to measure the tendency for individuals to inaccurately report (or interviewers to inaccurately record) their age (Hobbs 2004)⁸. The index can be used to assess the extent to which age data show systematic heaping on certain ages as a result of either rounding or digit preference. Although Whipple’s Index measures only the impact of heaping for ages ending in 0 or 5, the United Nations notes that “such digit preference is usually connected with other sources of inaccuracy in age statements” and therefore the Whipple’s Index represents “a fair measure of the general reliability of the age distribution” (UN 1986 p.20). We chose Whipple’s Index because it was straightforward to compute a modified Whipple Index just for the older population and thus have a consistent way of comparing age misreporting for younger and older adults. Whilst the United Nations age-sex accuracy index would permit comparison across datasets and by sex (because it estimates both an age ratio score and a sex ratio score and is therefore useful for comparative analyses between countries), we did not use it because it does not take into account any changes in the sex ratio with increasing age. Whipple’s Index is calculated by summing the number of persons aged 23-62 years (inclusive) who report their ages ending in 0 or 5, dividing that sum by the total population aged 23-62 years old, and multiplying by 5 to produce an index which is related to the quality of the age data. This Index can range between 100 (no evidence for age preference) and 500 (everyone

---

⁴ Senegal, Ghana, Ivory Coast, Benin, Cameroon, Angola, Mali, Burkina Faso, Niger, Chad, Ethiopia, Kenya, Tanzania, Uganda, Mozambique, Zambia, Zimbabwe
⁶ Niger census data were generously provided by Ousmane Ida Ibrahima
⁷ IPUMS [https://international.ipums.org/international/](https://international.ipums.org/international/): data provided by Stephen Wandera.
reports their age as ending in either 0 or 5). In our analyses we follow the general interpretation of data quality indicated by Whipple indices:

- <105 Highly accurate data
- 105-109.9 Fairly accurate data
- 110-124.9 Approximate data
- 125-174.9 Rough data
- 175+ Very rough data

We examined the quality of the age reporting using Whipple’s Index calculated for both sexes separately. In order to examine whether the age reporting for older people was substantially poorer than that for younger adults we calculated a modified Whipple’s Index (hereafter Whipple60) for those aged 60+, again for both sexes separately. Here we summed the number of persons aged 60 to 94\(^8\) whose ages were reported ending in 0 or 5 and divided that sum by the total population reported as being aged 60+. In order to investigate gender-related reporting we calculated the sex ratios by age (10 year age groups) for older people and the proportions of the adult population (by sex) aged above 60. Where we have both DHS and census data we compare them to establish consistency, and explore reasons for inconsistencies between data sources.

**Results: Quality of age data on older people in Africa**

We calculated the Whipple and Whipple60 index for DHS data for 17 sub-Saharan African countries (Figure 1). For the population aged 23-62, only Zambian women have ‘highly accurate’ age data, and data for males in Zambia, Zimbabwe and Mozambique are ‘fairly accurate’. All the other countries fall into the ‘rough’ or ‘very rough’ category. While poor for younger ages, age reporting is even less accurate at older ages (Whipple60). Only Mozambique had ‘approximate’ data for older ages. Southern African countries have relatively better age reporting in general; data for Angola, Ghanaian, and Burkinabe men are better than elsewhere but still remain ‘rough’ quality.

---

\(^8\) Those aged 95 and above were excluded from the numerator because some censuses classified the oldest group as 95+.
Figure 1: Country comparison of Whipple Index and Whipple60 Index, by sex, DHS data

AO=Angola, BJ=Benin, BF=Burkina Faso, IC=Ivory Coast, CM=Cameroon, ET=Ethiopia, GH=Ghana, KE=Kenya, ML=Mali, MZ=Mozambique, NE=Niger, SN=Senegal, TD=Chad, TZ=Tanzania, UG=Uganda, ZM=Zambia, ZW=Zimbabwe
Figure 2: Country Comparison of Whipple Index and Whipple60 Index, by sex, census data

BJ=Benin, BF=Burkina Faso, IC=Ivory Coast, ET=Ethiopia, KE=Kenya, ML=Mali, MZ=Mozambique, NE=Niger, SN=Senegal, TZ=Tanzania, UG=Uganda, ZM=Zambia, ZW=Zimbabwe

Figure 2 repeats the same analyses, comparing Whipple with Whipple60, for census data. Southern African countries have the best census age reporting at both younger and older ages although it is less accurate than their DHS data. At the other end of the scale the Sahelian countries have better age reporting in the census than in the DHS; Ethiopia and Benin have the poorest quality age data at both younger and older ages in both data sets.

DATA SOURCE COMPARISON

Sample surveys are assumed to generate better quality data than a census because more time can be spent on interviewer training and data quality control. We compare the quality of age data reporting for older people in census and DHS data (Figure 3). The Whipple60 suggests poorer data quality (i.e.: more age heaping) in the DHS compared to the census for men and women in Mali, Burkina Faso and Ivory Coast and also for men in Kenya and women in Niger.
FIGURE 3: Comparison of Whipple60 from census and DHS data, by sex, selected African countries

From these Whipple’s Index comparisons we observe that:

1. In most countries except those in Southern Africa, DHS data for the younger adult population are ‘rough data’. For 5 countries (Mali, Burkina Faso and Niger, Benin and Ethiopia) they are ‘very rough’.
2. Age reporting for older adults is substantially worse than that for younger adults, with the exception of Mozambique.
3. The age reporting for older women is worse than for older men with the exceptions of Southern African countries, Kenya, Tanzania and Ivory Coast.
4. DHS data are more accurate than the census for the Southern African countries and Ethiopia, but the censuses have better age reporting than DHS in the Sahelian countries (Mali, Burkina Faso and Niger).

In order to unpack in more detail the ways in which the quality of age data vary between datasets, we selected two countries – Niger and Ethiopia – for which there were three sources of data (census, DHS and LSMS). These countries were selected on the basis of data availability and accessibility. Figure 4 shows a comparison, by sex, of Whipple indices disaggregated by sex.
In Niger the quality of age reporting appears to be largely related to gender, with higher quality age data reported for men, irrespective of age or data source. In Ethiopia, by contrast, there appears to be an association between data source and age data quality, with LSMS producing the highest and the census producing the lowest quality data, although the gender differentiation remains, with male age reporting usually better than that for women, the exception being younger women in the DHS. These findings are surprising, as the Ethiopian LSMS excluded Addis Ababa, and therefore excluded many of the highly educated population for whom one would expect age reporting to be best.

*Trends over time*
Since low levels of education and general low development are frequently cited as explanations of poor age data in African surveys and censuses one would expect the quality of age reporting to improve considerably over the period from 1990-2010 when earlier investments in education mean that each cohort has more schooling than the previous one, and has been exposed to more urbanisation and civil administration. We selected 5 countries at different points along the Whipple spectrum and compared the trends in both DHS and census reporting.

Over time there is a clear improvement in age reporting (reduction of Whipple and Whipple60) in all the censuses (Figure 5b), with the exception of stable “rough” levels for younger Kenyan adults. However, there is no clear trend of improvement in age reporting in DHS surveys (Figure 5b), with the exception of younger Tanzanian adults, and possibly younger Burkinabe adults.

Enumeration of older men and women

Based on widespread observations of better female survivorship at adulthood in most of the world (Rajaratnam, Marcus et al. 2010) we would expect sex ratios at older ages to decrease with age. If younger female adult mortality is particularly high, due to maternal mortality or HIV/AIDS, then there might be similar numbers of older men and women but we would not expect substantially more men than women (Knodel and Ofstedal 2003). Such expectations are confirmed by examining the estimated life expectancies for men and women at ages 20, 40 and 60, generated for each 5 year period 1980-2015 (UN 2015). Females have higher life expectancy than males in the vast majority of cases. In Africa the only exceptions are a few southern African countries badly affected by HIV such as Lesotho and Zimbabwe where from 1995-2010 women aged 20 had a slightly lower life expectancy than men. However the difference is usually less than a year and is absent for those aged 40+ or 60+. The one non HIV related exception was Mali where over the period 1985-2015 women’s life expectancy
at age 20 is slightly lower (usually less than a year) than men, by age 40 women have the advantage and at age 60 there is little difference between the sexes.

Using DHS data sex ratios\(^9\) were calculated for 5 age groups: under 50, 50-59, 60-69, 70-79, and 80+. In every country the sex ratio for those aged 50-59 was less than 100 indicating more women than men reported in this age range. This reported excess of women in their 50s is certainly a function of the DHS data collection schedule which requires individual questionnaires for women under 50. Increasing women’s ages to 50 or higher may be either a conscious or subconscious response of interviewers when confronted with making decisions about age in uncertain situations (Randall, Coast et al. 2013), knowing that 49 or less would involve much more work. We include this age group for completeness but the broad patterns that we identify and discuss below are based on the sex ratios for those aged 60+.

Three main patterns of sex ratios at older ages could be identified amongst the study countries: those with an excess of older men with all sex ratios above 100, those with sex ratios around 100, and those with an excess of older women with most sex ratios below 100.

**Figure 6: High Sex ratios at older ages, DHS data: Chad (TD), Niger (NE), Angola (AO), Ethiopia (ET), Mali (ML) and Burkina Faso (BF)**

With the exception of Angola, all countries with sex ratios above 100 at older ages (Figure 6) also had very poor quality age reporting (see Figure 1 above), and are all in the bottom 15 countries of the HDI ranking (UNDP 2015). Past sex-differentiated mortality is unlikely to have generated such extreme old age sex ratios (see discussion of UN estimates above) and it

---

\(^9\) Sex ratio= 100*(number of men aged x to (x+10)/number of women aged x to (x+10). Thus a sex ratio of 100 means there are the same number of men and women of that age group.
seems likely that in these populations the very high DHS sex ratios at older ages are a consequence of omission of older women from the household questionnaire (particularly in Niger and Mali).

Five countries had DHS older age sex ratios around 100 although with somewhat irregular ratios (Figure 7). They include three coastal francophone west African countries (Senegal, Ivory Coast, Benin), Tanzania and Zambia, and cover a spectrum of quality of DHS age reporting from accurate in Zambia, rough in Tanzania, Ivory Coast and Senegal, to very rough in Benin. With the exception of rising sex ratios above age 50 in Zambia and Senegal, no country demonstrates a clear pattern which could hint at whether these observed patterns are real or a consequence of poor data or omissions of older women.

**FIGURE 7: Sex ratios at older ages, DHS data: Senegal (SN), Tanzania (TZ), Zambia (ZM), Ivory Coast (IC) and Benin (BJ)**
The third group of countries (Ghana, Cameroon, Togo, Kenya, Zimbabwe, Mozambique and Uganda) are closer to the anticipated sex ratios declining with age with more older women than older men (FIGURE 8).

**FIGURE 8: Sex ratios at older ages, DHS data: Ghana (GH), Cameroon (CM), Togo (TG), Kenya (KE), Zimbabwe (ZW), Mozambique (MZ) and Uganda (UG)**

![Sex ratios at older ages](image)

All these countries have better quality age reporting at both younger and older ages, and have higher HDI than the countries in the high sex ratio group. With the exception of the DHS induced excess of women aged 50-59, Cameroon, Zimbabwe, Ghana and Kenya have the expected patterns of sex ratios diminishing with age, unlike the more irregular patterns in the other countries.

*Sex ratios at older ages: comparing data sources*

There are three possibilities to explain these variations between countries in sex ratios at older ages shown by DHS data: (a) these sex ratio patterns may be real, (b) they may be a product of underlying social conditions leading to difficulties in data collection which would be experienced whatever the data collection method, or (c) they may be a product of different approaches to data collection. In the case of (a) and (b) we would expect similar patterns to emerge from different data sets. In the case of (c) different data sets could generate different patterns. We are able to examine this for the subset of countries where we have single year age distributions from the census. Although the DHS data were collected between 2005 and 2010 whereas the census data are largely from the 1990s, patterns of old age mortality, survivorship and reporting are unlikely to have changed dramatically in the intervening decade.
Figure 9 compares the sex ratios derived from DHS and census for Mali, Niger, Burkina and Ethiopia, all of which had high sex ratios at older ages using DHS data. For Ethiopia and Niger the LSMS sex ratios are also shown. Census data generate old-age sex ratios just above 100 for Niger and Mali and below 100 for Burkina which seem more plausible than the DHS results for these countries (Figure 4) and suggest that in all three Sahelian countries there was under-enumeration of older women in the DHS.

Figure 9: Sex ratios at older ages, for countries with high sex ratios, comparing data sources (DHS, census, LSMS\textsuperscript{10}): Burkina Faso (BF), Ethiopia (ET), Mali (ML), Niger (NE)

For Ethiopia the similar general trend in all three data sets using different methodological approaches (sampling vs total coverage) does suggest that in that country there may genuinely be substantially more older men surviving than older women – which might be a consequence of previous very high maternal mortality and young adult female mortality 30-50 years ago although this is contradicted by the UN estimates of male and female life expectancy at ages 20,40 and 60 for each five year period from 1950 until 2015 (UN 2015)\textsuperscript{11}.

\textsuperscript{10} LSMS = Living Standards Measurement Survey
\textsuperscript{11} None of the three Ethiopian data sources have full national coverage. The census omits nomadic populations from Afar and Somalie provinces, which were enumerated separately at a later date, and for which data are not available. The DHS also excludes much of Afar and Somalie, and excludes all nomadic populations. The LSMS excludes Addis Ababa and is conducted in small towns and rural areas.
However, Ethiopia also had the poorest age reporting in the census of all the countries, and other sources do not suggest better survivorship of males (Fantahun, Berhane et al. 2009).

The situation in Niger is harder to interpret: all the data sets tell the same story of more older men than older women, again contradicted by the UN estimates for 1950-2015. The fact that the census generates the lowest sex ratios suggests that there may be a problem with the omission of older women in sample surveys (see below). It is clear that an assessment of the numbers of men and women at older ages would differ substantially according to data set.

**Figure 10: Sex ratios at older ages – comparison of DHS and census data for Ivory Coast (IC), Senegal (SN), Tanzania (TZ) and Zambia (ZM)**

For those countries with DHS elderly sex ratios around 100, sex ratios at ages 60+ are similar in both data sources (Figure 10) with the exception of Zambia. This suggests that there is no under-enumeration of older women in the DHS for Senegal, Ivory Coast and Tanzania and that women have similar old-age survivorship as men.

Countries where DHS data generated the expected pattern of elderly sex ratios declining with age also have better age-reporting. Here the census sex ratios are similar but slightly higher most notably for Uganda (Figure 11) where census sex ratios are mostly above 100. Nevertheless, with the exception of Kenya (and ignoring the DHS induced vagaries of 50-59) there are still considerable differences in the sex ratios between data sources and unexplained irregularities in the patterns of sex ratios suggesting that, even where the data on older people are better, they remain somewhat inaccurate.
Figure 11: Sex ratios at older ages, comparison of DHS and census data, for countries with low sex ratios: Kenya (KE), Zimbabwe (ZW), Mozambique (MZ) and Uganda (UG)

Proportions of the population over 60

A third way of examining consistency between data sets is to focus on the proportion of the population recorded as older. How do such proportions differ between different countries and over time and are the data consistent between data sets? We use proportions of the adult population (aged 20 and above) who are aged 60+, treating the sexes separately (Figure 12).
Figure 12: Proportion of adult population (20+) aged over 60 years for selected SSA countries: comparison of DHS and census data

Just two countries – Tanzania and Uganda – show similar proportions of older people in both the DHS and census. In every country and data set, with the exception of Mozambique (DHS) and Kenya (census), a greater proportion of the male population is recorded as elderly than female. This may be a consequence of under-enumeration of (often very mobile) younger men in both data collection exercises; it may also reflect under enumeration of older women. The proportions of elderly males are much higher in the DHS than in the census in every country (except Tanzania) and this is particularly marked for Burkina Faso, Niger and Mali. In every country except Mali, the proportions of elderly females are higher in the DHS than in the census.

Discussion

These comparisons suggest that, with the exception of some Southern African countries, the age data for older Africans are very poor with the quality of age reporting at older ages being ‘rough’ or ‘very rough’. Author observations of data collection on age from older people in Mali and Burkina Faso suggest that there is considerable age inflation by older people themselves, their adult children and by enumerators.
The DHS seems to under-enumerate older women in Sahelian countries compared to men, and compared to the census. This implies that the survey methodology itself may contribute to the omission of older women, although we must remember that DHS surveys are largely focused around reproductive aged women so there is little pressure on interviewers to pursue assiduously potential omitted older people.

Reflections on why older women might be omitted in Sahelian surveys suggest a number of pathways. In West Africa there is a widespread assumption that the elderly are cared for by their offspring (Apt 2002, Oppong 2006, Antoine 2007) even if detailed research has shown that this support system may be disintegrating under conditions of urbanisation, migration and general modernisation (Aboderin 2004a, Roth 2010). Such assumptions were confirmed by a series of in-depth interviews undertaken in Ouagadougou, Burkina Faso in 2011 with civil servants who use and produce statistics in government ministries, and non-governmental organisation (NGO) personnel and academics involved in research using demographic data. Although this research project is not reported on in detail here (for more details see www.householdsurvey.info), in these interviews with educated, professional Burkinabé, the overwhelming response to our questions about potential vulnerability of older people and their possible invisibility in statistics, was that in Burkina Faso older people have few problems because they are always looked after by their children. The assumption that older people are cared for by offspring incorporates a general notion that they will also be co-resident with offspring. Such assumptions may contribute to the under-enumeration of older women in sample surveys because of a number of contributory factors outlined below.

Three decades of demographic and qualitative research in rural Mali and Burkina Faso have generated numerous observations around listing and identifying households in different communities. In the process of listing households for inclusion in the sampling frame for DHS it is quite likely that elderly women are omitted because of their ambiguous household status. They may be partially supported by younger adults (kin or otherwise) but omitted from the roster of household members because their affiliation is ambiguous and they are not perceived to be part of the household; this might include decisions to exclude them by interviewers following the criteria in their manuals and training (Randall, Coast et al. 2013). At the same time older widows may not be included in the original listing of household heads that forms part of the sample frame construction because they are not perceived by the leaders involved in preparing the listing to form independent households. These Sahelian countries are strongly patriarchal and gerontocratic, meaning that older men are less likely to be omitted during household membership listing, older men are more likely to be married than women, more likely to be recognised as household heads and less likely to have ambiguous household affiliation. Substantial spousal age differences and polygamy are features of Sahelian marriage (Barbieri and Hertrich 2005) which, in conjunction with probable higher male than female mortality at older ages, mean that few older men are ever widowed or live alone (Oppong 2006, Gning and Antoine 2015).

Residently isolated older women are much more likely to be enumerated by a research methodology that identifies respondents through mapping and visiting all residential units, as is done in the census, rather than through listing households in clusters. It seems likely that
this is the explanation for the extremely high sex ratios in the DHS compared to the census, but it has serious implications for interpreting the DHS based studies of living arrangements (Bongaarts and Zimmer 2002, Zimmer and Dayton 2005, Kautz, Bendavid et al. 2010, Zimmer and Das 2013). In the Sahelian countries there are probably many more older single women households than were identified in the DHS.

These statistically invisible older women are probably among the most vulnerable and the poorest in society. Their exclusion from the data will contribute to reinforcing the stereotypes that most older Sahelians are cared for by their children and are usually co-residential with them because such situations are those that are best captured by the data that are collected. Accusations of sorcery and expulsions of isolated elderly women from communities may also contribute to the statistical invisibility of older women in surveys in many African countries (Adinkrah 2004, Oppong 2006).

A number of factors could contribute to the substantial variation in proportions elderly observed in the different data sets: age inflation, missing older women, omitted younger men. There are so many unknowns it is difficult to conclude where the errors lie, but it is clear that data are poor and different data sources for the same country are usually inconsistent. The high variability in the errors between countries undermines the value of comparative analyses of data on older people, or at least gives serious pause for thought about the validity of such comparative analyses. It is interesting to note that in introducing the Ghana SAGE study the authors cite both the census and the DHS as the sources for the proportion of the Ghanaian population aged 60+, in a context where they say ‘little information exists in Ghana regarding the situation of older people’ (Biritwum, Mensah et al. 2013) thus demonstrating the importance of both these data sources for countries where there are few other data on older people, and rendering the inconsistencies between them an issue of considerable concern.

An interview with senior personnel from Help Age International (HAI), an international NGO dealing with advocacy and interventions for older people, indicated a major concern about the invisibility of older people in data from less developed countries “this is a driver of invisibility......which again drives discrimination”. This appears to be confirmed both in terms of absence of data but also the quality of the data which do exist.

Changes in social protection and an increasing international recognition that even in young, high fertility environments older people have specific social and health needs (Aboderin 2013) mean that there is a need for data to both identify those needs and monitor transformations. This emerging demand for good, detailed data to inform policies and interventions was reflected in the push for a minimum data set on aging and older persons for Africa (Ferreira and Kowal 2006) by WHO a decade ago. However a search of the WHO website suggests that that project has gone no further, possibly because of the huge gaps in the data identified for three out of the four participant countries (the exception being South Africa) (WHO 2003). The minimum data set project seems to have only attempted to establish whether data were available in the appropriate form; it did not assess the quality of those data and we question whether they are adequate to meet such demands. Despite the
demand for disaggregated age data, our analyses of DHS and census data suggest that disaggregating existing data may just provide misleading evidence. In the majority of SSA countries examined here the different datasets give very different results about proportions elderly, sex ratios of the older population and age reporting is often so poor that it would be impossible to determine age specific patterns or measure accurately age-related changes over time.

Conclusion

The population of older Africans remains miscounted, badly represented, and at times invisible, in national statistics in many countries. The most vulnerable elderly people are probably totally statistically invisible, since the stereotypes around the situation of the elderly which permeate all stages of the data collection process mean that they are likely to be omitted altogether, at least in the Sahelian countries. Clearly such stereotypes must be challenged. Moreover, the reality that poor, old people live alone and without the support of their families must be recognised. In order to do this, better data are needed and maybe, despite scarce financial resources, nationally representative surveys that focus specifically on older people should be commissioned. As a first step, improving the focus of census and survey questions and developing ways to overcoming barriers to accurate data collection would lead to better data for informing policy directions which respond to the specific, yet largely unknown, needs of older Africans.

In practical terms the DHS and other sample surveys could improve enumerator training to highlight specific issues around data collection on older people: in Sahelian countries, and other environments where it is likely that older widows are under-enumerated, it would be straightforward to add specific supplementary questions in the cartography and listing phases about whether there might be widows living alone who had been omitted from household lists. In the same way that enumerators are instructed to ask about babies who have not yet been named when completing the household roster, they could also be trained to ask specifically about whether there are any elderly dependents who occasionally eat with the household, and who might have been omitted.

Many surveys and censuses use event calendars to assist with age reporting. Often, however, the respondent is younger than the older person for whom they are providing data and it may not be socially acceptable to clarify such facts with a senior. Some of the issues around poor reporting of age could be improved by instructing interviewers to actually ask the individual concerned and work with them and the calendars; because older people are frequently present around the dwelling this should not be difficult. Furthermore authors’ observations of interactions between interviewers and older people indicate that there is a tendency for interviewers to inflate ages especially for someone who appears to be very old, with wrinkled skin and somewhat stooped. There appears to be an assumption that this person must be very, very old – at least 90 - and the logics of the age relationships are not explored further. If interviewers are provided with a series of prompts about the age of the oldest and youngest child, then, for women at least, age reporting can be substantially improved. For older men event calendars are probably the most appropriate route.
Age data are critical for understanding not only population structure and change, but also population socio-economic characteristics and for calculating indicators of health and age-related disease. Our analyses show that, despite improvements in age data quality at younger ages, older age data for sub-Saharan Africa are generally of poor quality. Analyses that use age data on older people rarely acknowledge, or (where appropriate) use techniques to correct or smooth age data. However smoothing the denominators does not help in understanding age related changes if the age data contributing to numerators are highly inaccurate. Age data in sub-Saharan Africa, particularly for older people, are subject to error, irrespective of the source of data. The implications of these errors will depend on the uses to which the data are put. Our analyses give a sense of the variation in the magnitude of these errors, and provide a reminder to analysts using age data that not only are errors likely to be present, but they should be acknowledged, and might need to be adjusted for.

References