Are Age Misreporting Patterns Universal? Evidence and Methodological Solutions for Sub-Saharan Africa

Marília R. Nepomuceno*[1], Benjamin-Samuel Schlüter[1], and Yempabou Bruno Lankoandé[2]

[1] *Max Planck Institute for Demographic Research, Rostock, Germany*

[2] *Institut Supérieur des Sciences de la Population at the Joseph Ki-Zerbo University, Ouagadougou, Burkina Faso*

October 30, 2025

Abstract

Self-reporting of age is still the main way to learn about individuals' age across the globe. However, such reports are often subject to errors. Self-reported ages may be systematically understated, overstated, or distorted due to digit preference. Although age heaping has been extensively documented worldwide, little is known about whether systematic under- or over-reporting of age is similarly universal. To date, systematic over- or under-reporting of age has been documented in some high-income countries such as the United States of America and Costa Rica, as well as in some low- and middle-income countries like Brazil and India. However, methods for addressing these types of errors remain scarce. Evidence and methodological solutions for sub-Saharan Africa are particularly limited, leaving open questions about whether patterns in this region resemble those documented elsewhere and how this data problem can be addressed. This study takes an important step toward filling this gap by providing new insights into patterns of systematic age misreporting in Burkina Faso using record-linkage data. Using this data, we identify patterns of age misreporting across ages and compare them with those documented in other countries. Then, we develop a transition matrix of age misreporting to adjust distortions in population age distributions driven by both systematic under- or over-reporting of age and age heaping. Finally, this study assesses the impact of systematic age misreporting on demographic rates and summary demographic measures. Our preliminary findings show that systematic under- or overreporting of age is not universal. Country- or region-specific strategies are necessary to address this data problem.

*Corresponding author. E-mail: nepomuceno@demogr.mpg.de

1 Introduction

Across the globe, the main way to learn about individuals' ages is still through self-reporting. However, these reports are often subject to errors. Self-reported ages may be systematically understated, overstated, or distorted due to digit preference. While age heaping has been extensively documented worldwide, little is known about whether the systematic under- or over-reporting of age is similarly universal.

The two main types of age misreporting are digit preference, which produces age heaping in age distributions, and the systematic over- or under-reporting of age. The former is the most common and occurs when individuals report ages ending in zero or five. This has been widely documented in low-, middle-, and high-income countries, and there are several methodological solutions to address this data issue (Caldwell and Igun 1971; Shryock and Siegel 1976; Myers 1940, 1954; Spoorenberg and Dutreuilh 2007; Camarda et al. 2008; Randall and Coast 2016; Camarda et al. 2017; Ouedraogo 2020; Carnevali and Gerland 2021). A less explored form of age misreporting is the systematic over- or under-reporting of age, where individuals tend to exaggerate or reduce their age (Rosenwaike and Preston 1984; Elo and Preston 1994; Preston et al. 1999; Nepomuceno and Turra 2020; Masquelier et al. 2021; Lankoandé et al. 2022). Unlike age heaping, these age errors are rarely visible to the naked eye, are much harder to detect and correct, and, to date, only a few country- or region-specific methodological solutions have been proposed (Bhat 1990; Palloni et al. 2021).

Systematic over- or under-reporting of age may arise from multiple factors, some of which are context-specific. Some age misreporting is unintentional and occurs due to memory loss, uncertainty about birth dates, a lack of official documents stating the birth date, or delay in birth registration (Ewbank 1981; Preston et al. 2003; Turra et al. 2023; Reid et al. 2024; UNICEF 2025). Such issues are especially marked in low- and middle-income countries, where civil and vital registration systems remain weak (Lyons-Amos and Stones 2017; UNICEF 2025). In other cases, people deliberately misreport their age to meet eligibility criteria for age-dependent services, such as school enrollment, military service, employment, or social welfare programs (Preston et al. 2003). Others do it purely for the prestige associated with being older or younger. High levels of illiteracy are also associated with errors in reporting age (Preston et al. 2003; Fayehun et al. 2020; Reid et al. 2024). Thus, due to the wide range of factors that can lead to age misreporting, including cultural practices and levels of socioeconomic development, understanding whether these patterns are universal is a crucial step toward developing effective solutions.

To date, systematic over- or under-reporting of age has been documented in some high-income countries such as the United States of America and Costa Rica (Preston et al. 2003; Palloni et al. 2021), as well as in some low- and middle-income countries like Brazil and India (Bhat 1990; Nepomuceno and Turra 2020). But methods for addressing these types of errors remain scarce. This is mainly because the type of data required to identify patterns of age misreporting requires detailed data, such as record linkage data. There is a particular lack of evidence and methodological solutions in sub-Saharan Africa, leaving open important questions about whether patterns in this region mirror those documented elsewhere and how such data issues can be addressed. This study takes an important step toward filling this gap by providing new insights into systematic patterns of age misreporting in Burkina Faso using record linkage data. More specifically, this study *(i)* identifies patterns of age misreporting across ages in Burkina Faso and compare them with those documented in Costa Rica, *(ii)* develops a transition matrix of age misreporting to adjust distortions in population age distributions driven by both systematic under- or over-reporting of age and age heaping; and finally, *(iii)* assesses the impact of systematic age misreporting on demographic rates and summary demographic measures.

2 Data and Methods

We used a record linkage data from Burkina Faso (Lankoandé et al. 2022). 58% of individuals enumerated in the 2006 census (41,589 individuals) were matched to those in the Health and Demographic Surveillance Sites (HDSS) in Nouna, a town in the Kossi province of Burkina Faso. HDSS studies monitor the entire population of a defined geographic area over time, regularly collecting information on births, deaths, marriages, and migration (Pison 2005; Lankoandé et al. 2022). HDSS data is considered of high quality due to the rigorous data collection procedures and multiple consistency checks (Pison 2005). Therefore, HDSS data is much more reliable than census data, enabling comparisons of reported ages between the two sources.

A total of 38,136 individuals aged between 0 and 70 in census were considered, of whom which nearly 50% were females. Of these individuals, 46% declared the same age in the HDSS and census, while 27% over-reported and 27% under-reported their age.

We use a logistic model to estimate the probabilities of age misreporting in the census, using the age reported in the HDSS (true age) and sex as covariates. We create a binary variable whose value is set to one when there is over/under- statement and zero otherwise. Then, we use a multinomial logistic model to estimate conditional probabilities of over- or under-report age by n years, where n is the number of years by which an age is misreported. We consider

9 categories for $n = 1, 2, ..., 10+$, with $n = 1$ being the reference category. Age reported in the HDSS and sex are included as covariates, consistent with the binary logistic model. These probabilities are then used to construct a transition matrix of age misreporting.

3 Preliminary Findings

Figure 1 gives a broad overview of the variation of age misreporting by age. It presents the average number of years misreported in census, by age and sex. It shows that the pattern of age misreporting in Burkina Faso considerably change over age. Up to the age of 10, there is a net age overstatement for both sexes. The magnitude of this age error is greater at first ages and reduces with increasing age. However, this pattern reversed at young adult ages and old ages, where a net age understatement is observed for most ages. Although age patterns are similar for females and males, age understatement is more pronounced among females as they age.

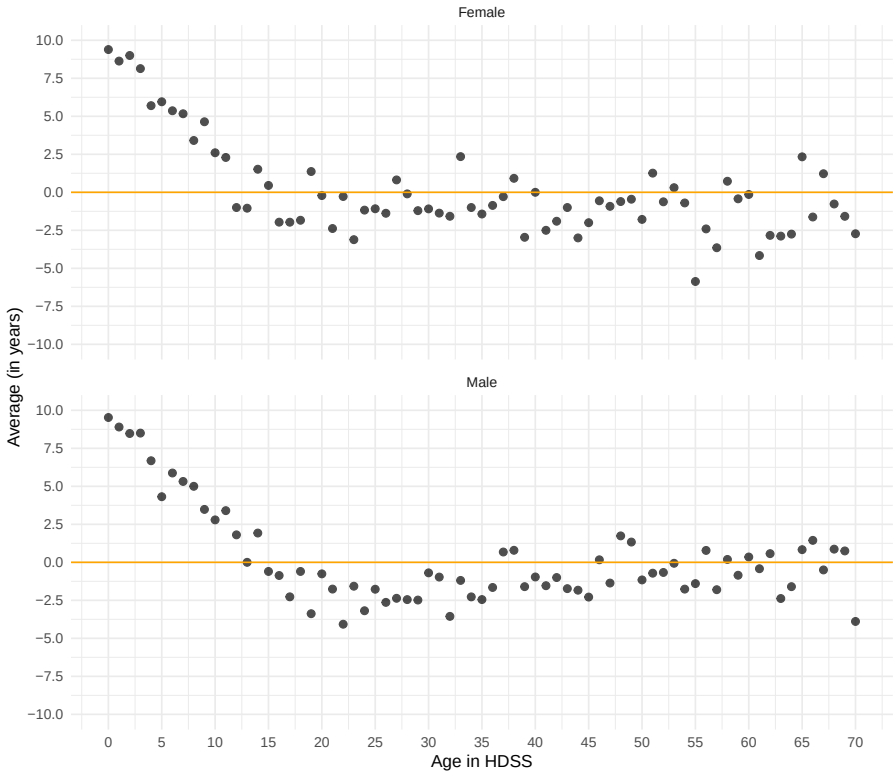

Figure 1: Average number of years misreported in census by age and sex.
Source: Authors' calculations.

Figure 2 compares the ages reported in census and in HDSS. The main diagonal represents the proportion of individuals who reported the same age in both data sources. Values above the main diagonal shows proportion of individuals who overreport their age in census, while

values below the main diagonal show the proportion of individuals who underreport their age in census. By looking at the main diagonal, Figure 2 shows that the accuracy of age reporting changes markedly by age. At ages 0 and 1, around 75% of individuals reported the same age in both databases. However, this proportion notably decreases for ages between 10 and 30, falling to between 20% and 40%. From the age of 40, sex differences become more pronounced, with males reporting their age with more accuracy than females.

Figure 2 also shows age patterns of age misreporting. The dispersion of age misreporting around the true age increases with age. Between ages 10 and 30, although the proportion of age misreporting is relatively high, most cases of under- or overreporting are by 1 or 2 years around the true age. However, this pattern changes as age increases with a greater dispersion of age misreporting around the true age. It is important to mention here that the small number of cases among the oldest ages, can lead to the very high proportions of age misreporting, as observed among females who reported 70 years old in the census and 80 in the HDSS. A preference for the digits 0 and 5 in census can also be observed in Figure 2, particularly among females.

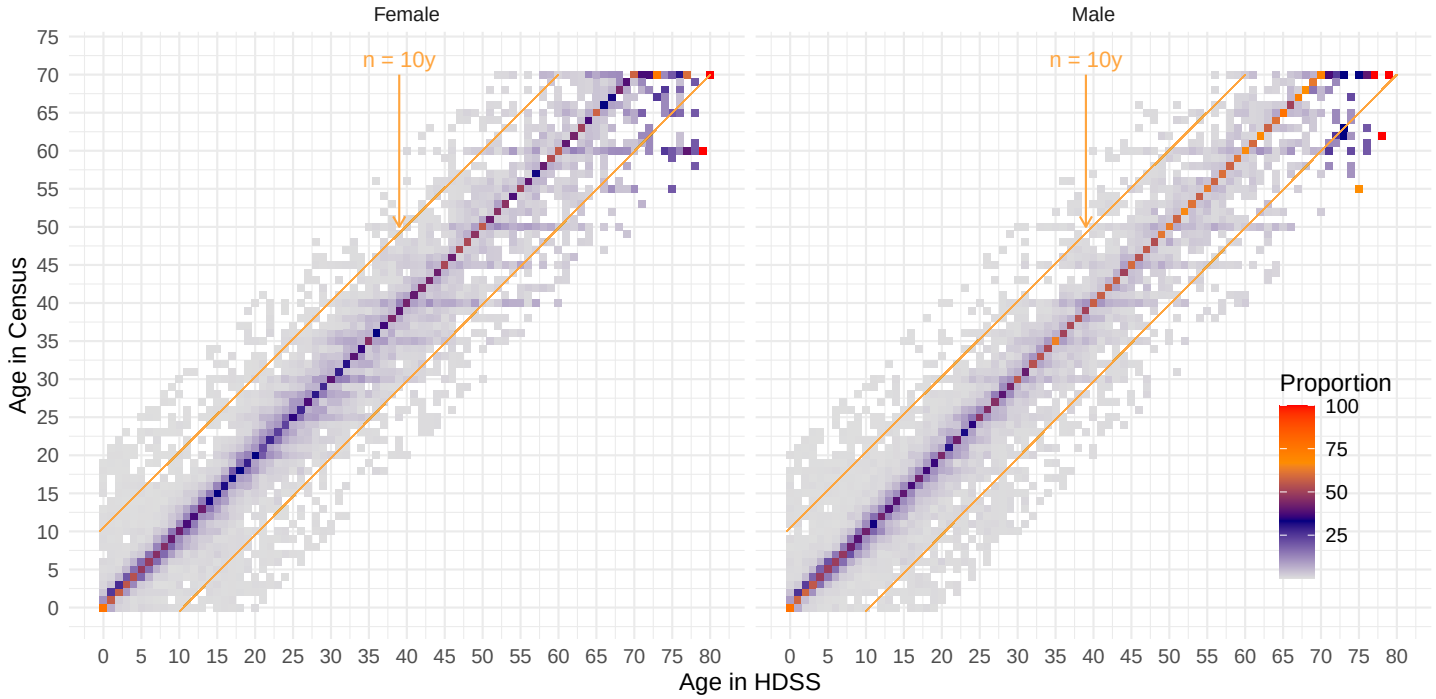

Figure 2: Comparison of reported age (in proportions) between census and HDSS, females and males. *Source: Authors' calculations.*

Figure 3 shows the estimated probabilities to over- and underreport the age in Burkina Faso, along with those estimated for Costa Rica by Palloni and colleagues (Palloni et al. 2021). For comparability, analyses for Burkina Faso are restricted to the age range 45–80. Figure 3

confirms what was already shown in Figures 1 and 2 that in Burkina Faso, the probabilities of age understatement are larger than the probabilities of age overstatement among older adults, with higher probabilities for females. However, when compared to findings for Costa Rica, the differences are striking. Unlike in Burkina Faso, the probabilities of age overstatement are larger than the probabilities of age understatement in Costa Rica. In other words, while there is a net age overstatement among older adults in Costa Rica, there is a net age understatement in Burkina Faso.

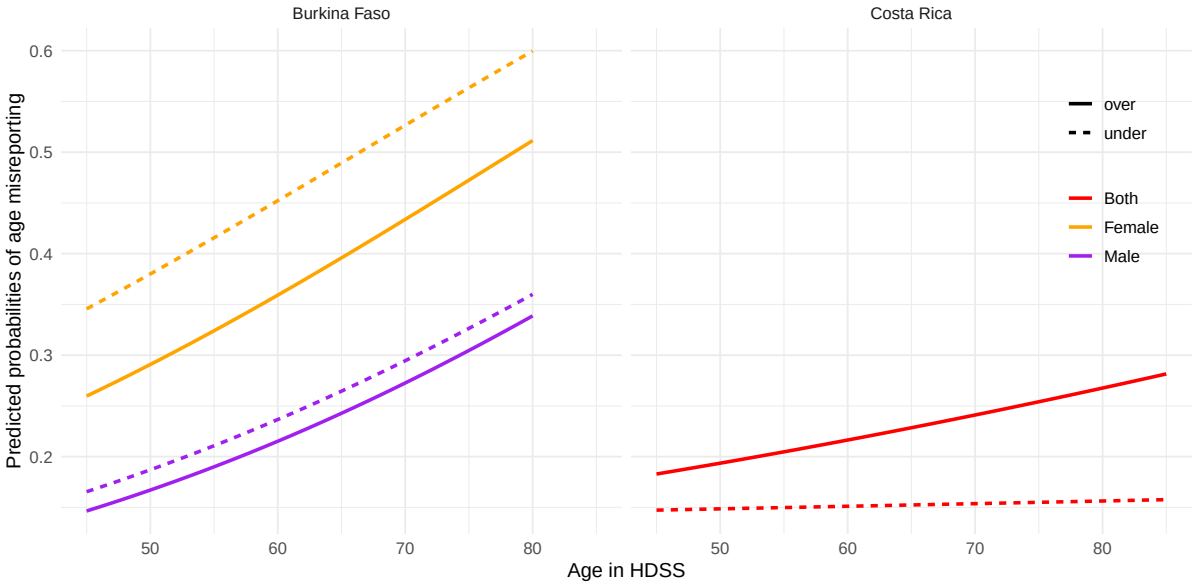

Figure 3: Probabilities of under and over report the age in Burkina Faso and Costa Rica.
Source: Authors' calculations.

4 Conclusions and Next Steps

This study shows that the pattern of age misreporting changes considerably with age in Burkina Faso. The accuracy of age reporting is high for very young ages, but it declines among young adults and begins to improve after age 40. However, this recovery among old adults is only observed among men. The direction of age misreporting in Burkina Faso also varies markedly by age: at younger ages, there is a net overstatement of age, while at adult ages, the pattern reverses and a net understatement of age is observed. This study also shows that patterns of systematic under- and over-reporting of age are not universal. There are substantial differences in the magnitude and direction of age errors between Burkina Faso and Costa Rica, a pattern that has also been documented in several other Latin American countries.

The next steps of this study involves the development of the transition matrix of age misreporting from age 0 to 70 to adjust distortions in population age distributions driven by both

systematic under- or over-reporting of age and age heaping, and a simulation analysis to assess the impact of these age errors on demographic rates and summary demographic measures.

References

Bhat, P. N. M. (1990). Estimating transition probabilities of age misstatement. *Demography*, 27(1):149–163.

Caldwell, J. C. and Igun, A. A. (1971). An experiment with census-type age enumeration in Nigeria. *Population Studies*, 25(2):287–302.

Camarda, C. G., Eilers, P. H., and Gampe, J. (2008). Modelling general patterns of digit preference. *Statistical Modelling*, 8(4):385–401.

Camarda, C. G., Eilers, P. H. C., and Gampe, J. (2017). Modelling trends in digit preference patterns. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(5):893–918.

Carnevali, d. O. R. and Gerland, P. (2021). Analyzing the Quality of Age Reporting from Population Censuses for All Countries of the World: 1950-2020 Census Rounds. In *IUSSP International Population Conference 2021*, Hyderabad, India.

Elo, I. T. and Preston, S. H. (1994). Estimating african-american mortality from inaccurate data. *Demography*, 31(3):427–458.

Ewbank, D. (1981). *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. National Academies Press, Washington, D.C. Pages: 19649.

Fayehun, O., Ajayi, A. I., Onuegbu, C., and Egerson, D. (2020). Age heaping among adults in nigeria: evidence from the nigeria demographic and health surveys 2003–2013. *Journal of Biosocial Science*, 52(1):132–139.

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

Lyons-Amos, M. and Stones, T. (2017). Trends in Demographic and Health Survey data quality: an analysis of age heaping over time in 34 countries in Sub Saharan Africa between 1987 and 2015. *BMC Research Notes*, 10(1):760.

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

Myers, R. J. (1940). Errors and bias in the reporting of ages in census data. *Transaction of the Actuarial Society of America*, 41:395–415.

Myers, R. J. (1954). Accuracy of age reporting in the 1950 united states census. *Journal of the American Statistical Association*, 49(268):826–831.

Nepomuceno, M. R. and Turra, C. M. (2020). The Population of Centenarians in Brazil: Historical Estimates from 1900 to 2000. *Population and Development Review*, 46(4):813–833.

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

Palloni, A., Beltrán-Sánchez, H., and Pinto, G. (2021). Estimation of older-adult mortality from information distorted by systematic age misreporting. *Population Studies*, 75(3):403–420.

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

Preston, S. H., Elo, I. T., Hill, M. E., and Rosenwaike, I. (2003). *The Demography of African Americans 1930–1990*. The Springer Series on Demographic Methods and Population Analysis. Springer Netherlands, Dordrecht.

Preston, S. H., Elo, I. T., and Preston, S. H. (1999). Effects of age misreporting on mortality estimates at older ages. *Population Studies*, 53(2):165–177.

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

Reid, A., Kabagenyi, A., Ojiambo Wandera, S., Beinomugisha, J., and Walters, S. (2024). Age reporting by and for older people in Uganda: relationships with frailty, human capital and population registration. *Journal of Global Ageing*, 1(1):27–51.

Rosenwaike, I. and Preston, S. H. (1984). Age Overstatement and Puerto Rican Longevity. 56(3):503–525.

Shryock, H. S. and Siegel, J. S. (1976). The methods and materials of demography. In Shryock, H. S. and Siegel, J. S., editors, *The Methods and Materials of Demography*, Studies in Population, page ii. Academic Press, San Diego.

Spoorenberg, T. and Dutreuilh, C. (2007). Quality of age reporting: Extension and application of the modified whipple's index. *Population (English Edition, 2002-)*, 62(4):729–741.

Turra, C. M., Fernandes, F., Almeida Calazans, J., and Nepomuceno, M. R. (2023). Age reporting for the oldest old in the Brazilian COVID-19 vaccination database: What can we learn from it? *Demographic Research*, 48:829–848.

UNICEF (2025). Birth Registration in Sub-Saharan Africa: Current levels and trends. Technical report, New York.