Are estimated age data from sample surveys of poor quality? Interrogating interviewer-respondent conversations

Emmanuel Souza, Department of Sociology and Population Studies, University of Malawi

Abstract

Due to incomplete civil registration records, many low- and middle-income countries rely heavily on sample surveys to obtain current demographic information. However, due to lower literacy levels than in developed countries, the retrospective data obtained from such surveys may be of questionable quality. For instance, the accuracy of age-related information heavily relies on respondents' accurate knowledge and recollection of such information. This paper employs a mixed-methods approach to appraise age-related information obtained from a mobile phone survey, examining its quality and investigating the process through which such data is obtained. The findings indicate that age-related data obtained through respondent or enumerator estimates involves a significant amount of guesswork and is very erratic. Specifically, estimated age data are marked by substantial noise, leading to notable discrepancies in the summary statistics derived from such data when compared to correctly reported age data. The implications of these findings are discussed.

1. Introduction

Unlike sampling errors, non-sampling errors are often neglected in survey research (McNabb, 2014). In some instances, non-sampling errors may result in incomplete or inaccurate data, which may distort the overall picture of a social phenomenon under study. In the social sciences, demography in particular, age-specific indicators heavily rely on the accurate reporting of age-related information, such as age or date of birth and age at or date of death. In mortality research, this information is usually reported by proxies. As a result, mortality data are highly susceptible to missing information and recall errors (Masquelier, 2021), especially in surveys conducted in developing countries. These data problems can introduce significant biases in estimated mortality indicators (Coale & Shaomin, 1991). In part, the accuracy of survey data depends on the rapport between the enumerator and respondent (Horsfall et al., 2021; Sun et al., 2021). Conceivably, due to poor enumerator-respondent rapport in mobile phone surveys (MPS), these data problems may be heightened (Holbrook et al., 2003; Heerwegh & Loosveldt, 2008). Indeed, Helleringer et al. (2023) show that MPSs produce noisy self-reported age information compared to other traditional household surveys, such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS).

Incomplete or incorrect reporting of age-related information is a function of many factors, including low literacy levels and limited civil registration of vital events, especially in low- and middle-income countries (LMICs). Although things are slowly improving, obtaining accurate age information is still challenging (Malik, 2021), particularly because birth and death registration is absent and the knowledge regarding the age at death or date of birth, is often lacking. Survey researchers working in LMICs expect age data, for example, to contain a significant amount of missing information and often use statistical methods to overcome such problems during analysis (Khan et al., 2020; Musil et al., 2002), with an assumption that the reported data is modestly accurate. For example, in DHSs some of the age information is imputed, the amount of which varies within and between countries (see Appendix Table 1).

Furthermore, the accuracy of age data is dependent on whether the data is self- or proxy-reported. Proxy reporting is common in many household surveys, where the head of the household or someone more knowledgeable about events of the household reports on behalf of all individuals. Proxy reporting of age data can result in serious data gaps or misreporting if the respondent is unacquainted with the required information.

In mortality surveys, proxy-reporting is inevitable because the dead cannot tell. As such, survey researchers heavily rely on reported information about the deceased. If the proxy knows the details of the event in question, then the reported information is fairly accurate. However, in circumstances where the proxy is unacquainted or unsure about the required information, reasonable estimations are permitted. This is recommended in survey research to reduce the amount of missing information and non-response rates, consequently reducing the cost of data collection. However, the quality of such data is questionable. So, how good are the proxy-estimated (through a probing process) age data relative to the 'fairly accurate' proxy-reported (non-estimated) information?

This study, therefore, interrogates the quality of proxy-reported age data collected through the Rapid Mortality Mobile Phone Survey (RaMMPS) that was conducted in Malawi during COVID-19. It compares age-heaping problems in proxy-reported and proxy-estimated age data. Further, the study sheds light on how proxy-estimated age information is solicited, especially by examining the process of probing for such information. To comprehend the quality of data collected through proxies, we first compare the quality of demographic surveillance data to its proxy-reported equivalent collected in the RaMMPS project.

2. Data

Data for this study comes from two sources: 1) the Karonga Health and Demographic Surveillance System (HDSS) in the northern part of Malawi (see Crampin et al., 2012 for a full description of the site) and 2) the Malawi RaMMPS project for the period between January 2022 and May 2023. The RaMMPS was designed to investigate mortality patterns during the COVID-19 pandemic in Malawi (LSHTM, 2022) due to the limited capacity of the CRVS system in the country and the inability to conduct traditional face-to-face interviews at the height of the pandemic.

Sampling for the RaMMPS was done via Random Digit Dialing of numbers that were verified against the Home Location Register. Interviews were conducted by enumerators who had experience with mobile phone interviews. All enumerators participated in a one-week training before the start of the interviews, and a one-day refresher after each trimester of fieldwork. Enumerators used the SurveyCTO platform to navigate through the questionnaire and enter data. The sample for this analysis was 10,490, where all cases with missing values on both mother and father's survival statuses were dropped and only limited to interviews that were completed.

A random sample of 2-3% of the interviews was recorded via the audio audit tool in SurveyCTO. Neither the enumerator nor the respondent knew whether the interview was selected for recording. During the consent process, the respondent was informed that the interview might be recorded, and the recording itself only started after consent was given. As such, the audio recordings did not include identifying information such as the respondent's name or mobile phone number. However, for purposes of this analysis, necessary information like age and location was retrieved from the anonymized interview data and linked to the audio recordings using unique respondent identity numbers.

To examine the probing process for age information, a total of 97 audio-recorded interviews were transcribed and analyzed. These interviews comprised three main sections: the socioeconomic profile of the respondent's household, as well as questions on mortality and fertility. The mortality section included four modules covering household deaths, child mortality, as well as the survival status of siblings and parents. In all these modules, age-related data were collected.

3. **Analytical approach**

The analytical procedure involves a mixed methods approach. First, we use statistical and demographic data quality assessment procedures to examine data from the Karonga HDSS and those reported through RaMMPS. Thus, we compare death data recorded in the surveillance site to survey data reported by household members from which these events occurred. Likewise, we appraise the quality of estimated data by comparing proxy-reported and proxy-estimated age data collected through the RaMMPS. On the one hand, we employed measures of central tendency and dispersion, such as the mean and frequency distribution of the age data to appraise its general quality. On the other hand, we employed age-heaping indices to assess the reporting quality of the age data. Second, we use a qualitative approach to review and examine transcribed audio-recorded interviews to explore the techniques employed in eliciting estimated age information.

4. Preliminary Findings

4.1 Sample characteristics of respondents

Table 1 provides an overview of the respondents' characteristics and the reporting status of parental information. The mean age for the respondents is 31 years, ranging from 18 to 64 years. The distribution by sex is even, aligning somewhat with the 2018 Malawi Population and Housing Census (MPHC). As expected, the majority of respondents (75%) reside in rural areas. In terms of education, the sample leans towards higher educational attainment compared to the 2018 MPHC and other traditional surveys. Half of the respondents had completed secondary education, whereas 26% had education beyond the secondary level, contrasting with the 2015-16 Malawi DHS where the majority had primary education or lacked formal education. Geographically, the sample distribution across the three regions corresponds to that of the 2018 MPHC, with the highest proportion of respondents residing in the South, followed by the Center and North, respectively.

The reporting of parental survival status by the respondents corresponds with expected mortality patterns, with 73% reporting their mothers as alive, whereas only 56% reported their fathers as alive. Proxy age reporting for parents varied markedly based on their survival status and only slightly based on their sex. For both mothers and fathers, a higher proportion of respondents reported age information for their living than dead parents. Roughly 32% of living parents' ages were estimated, compared to 28% for deceased parents. However, a notable proportion of respondents did not know their parents' ages at the time of death, with figures standing at 28% and 35% for mothers and fathers, respectively.

Table 1: Characteristics of respondents and parental information reporting status

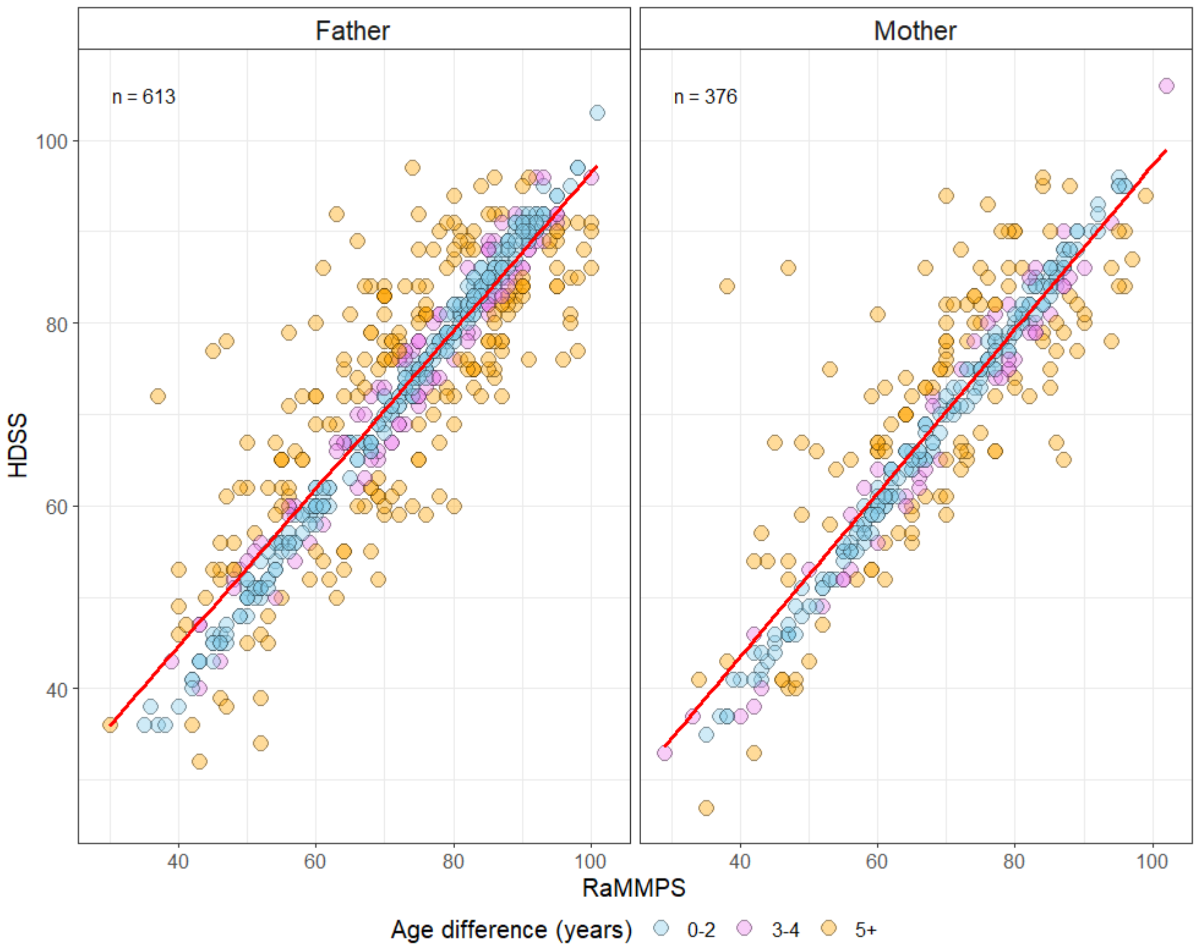
Characteristics	Mean/Prop	*n*	*sd*	*min/max*
Respondent information				
Age	31.3	10490	10	18/64
Sex		10490		
Male	50.3			
Female	49.7			
Type of rersidence		10490		
Rural	74.8			
Urban	25.2			
Education		10433		
None or Primary	22.4			
Secondary	51.1			
Higher	26.5			
Region		11305		
North	15.6			
Central	39.7			
South	44.6			
Parental survival and age reporting status				
Mother alive	73.4	10486		
age reporting (mother alive)		7695		
Reported	57.7			
Estimated	32.4			
DK	9.9			
age reporting (mother dead)		2791		
Reported	43.1			
Estimated	28.4			
DK	28.5			
Father alive	55.7	10404		
Age reporting (father alive)		5796		
Reported	53.3			
Estimated	32.2			
DK	14.4			
Age reporting (father dead)		4606		
Reported	37.2			
Estimated	27.9			
DK	34.8			

sd=standard deviation; DK=don't know

4.2 Quality of recorded demographic surveillance versus survey-reported age data

To start appraising the quality of our estimated data, we first looked at the differences in ages at death between the recorded data from the HDSS and those reported via RaMMPS. Figure 1 shows that the ages reported in the survey do not entirely match the ones recorded through the HDSS. Thus, reporting in the RaMMPS is less accurate relative to the HDSS recorded data. For both mothers and fathers, over 80% of the survey reported ages at death did not match the surveillance data records, with 47% of mothers' ages having a difference of over 2 years and 27% having a difference of 5 years and over. Likewise, 50% of fathers' ages have a difference of 2 years or more, while 34% have a difference of 5 years or more.

Figure1: Quality of parental age reporting: Age at death, RaMMPS vs HDSS

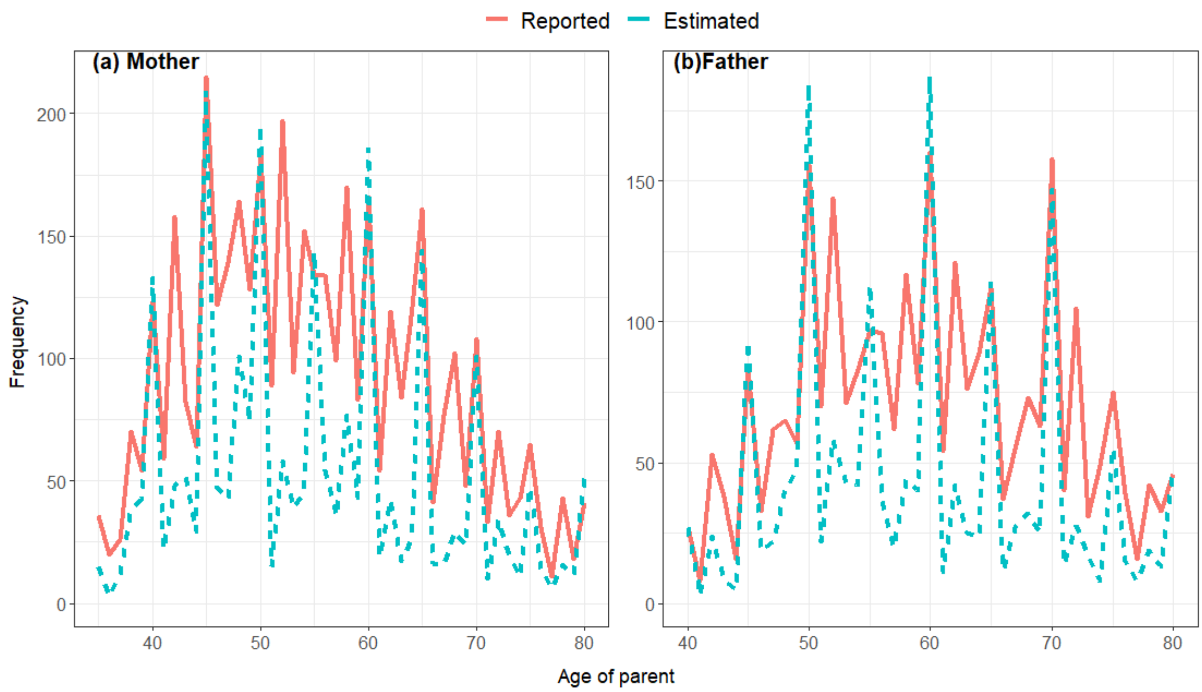


4.3 Quality of proxy-estimated vs proxy-reported parental age data in the RaMMPS

4.3.1 Quality of parental age at last birthday data

Figure 2 presents information on parental ages at the last birthday (mothers' age in panel A and fathers' age in panel B). It differentiates this age data based on whether the information was reported or estimated (either by the respondent or enumerator). From Figure 2, it is evident that there was age heaping for both the reported and estimated data and for mothers and fathers. However, it is also clear that the age heaping is more pronounced for the estimated rather than the reported age data. There are larger spikes in the estimated than reported data for ages ending in 0s and 5s, which are relatively larger for information pertaining to fathers compared to mothers. Notably, extensive age heaping can be observed for ages 40 and 50 in the estimated data compared to the reported data.

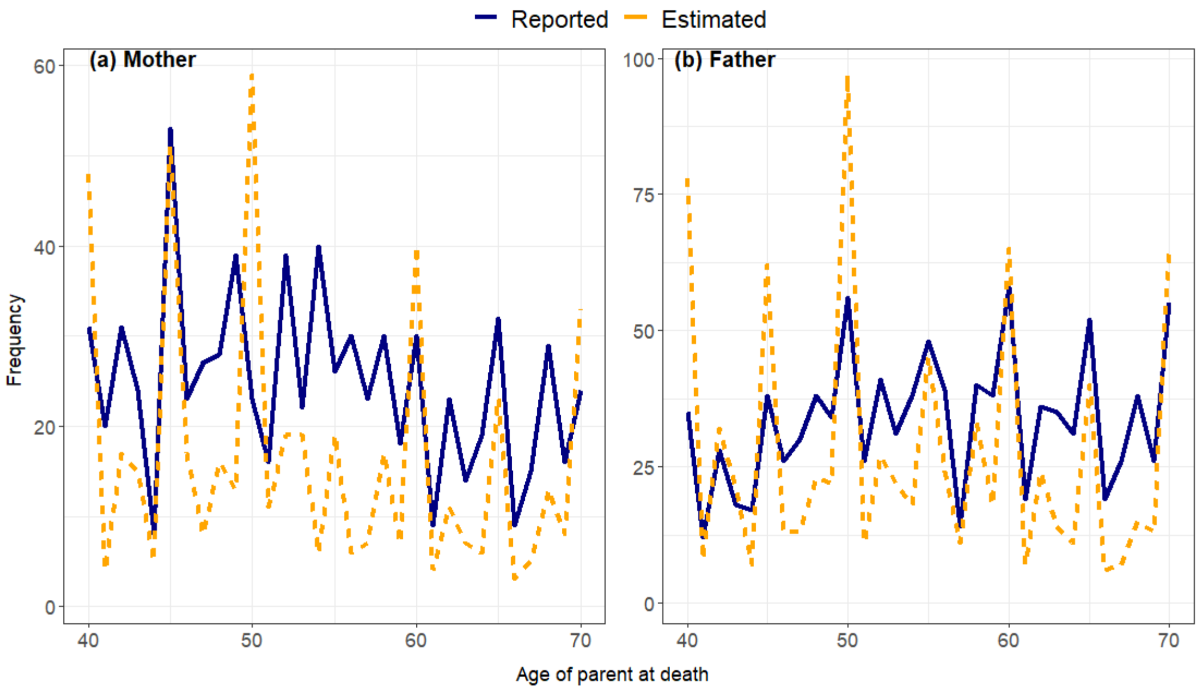
Figure 2: Parental age distribution by status of age reporting and sex of parent, RaMMPS



4.3.2 Quality of parental age at death data

Figure 3 presents information on ages at death for parents in the RaMMPS study. As previously observed, age-at-death data is noisy with substantial heaping. In particular, there are large spikes for ages 40, 45 (for mothers, mostly), 50, and 60. Although this pattern can be observed for both the reported and estimated data, it is more pronounced for the estimated than the reported age data. Notably, for both mothers and fathers, there is particularly more heaping at age 50. We delve into how this pattern may have emerged in the qualitative data section below.

Figure 3: Frequency distribution of parental ages at death, RaMMPS



In Table 2, we present summary statistics for parental age at last birthday and age at death, including demographic indicators that measure the quality of age data. The mean ages at the last birthday for both the reported and estimated data are consistent. However, the mean ages at death for the reported and estimated data are significantly different. In both cases, there is a difference of 5 years between the reported and estimated age data, although the spread is the same.

Age heaping indicators, the Whipple's and Myers' blended indices, show that the estimated age data is rougher than the reported data. Table 2 shows that age heaping is a problem for both datasets. However, the problem is more serious for the estimated age data than reported data. According to the UN's classification of the quality of age data based on the Whipple's index, the quality of the estimated data can be categorized as 'very rough', whereas the reported data is 'rough' (Johnson 2022). Myers' blended index portrays similar information. The quality of age data reporting is the same for both mothers and fathers.

Table 2: Summary measures for parental age data, RaMMPS

	Mother		**Father**	
	Reported	estimated	Reported	estimated
	Age at last birthday (living parents)			
Mean (SD)†	56.1 (11.9)	55.7 (12.0)	61.4 (11.5)	60.2 (11.5)
Whipples index	1.4	2.51	1.49	2.68
Myers blended index	12.41	29.38	15.04	31.57
	Age at death of parent			
Mean (SD)†	56.0 (15.7)	50.9 (15.4)	61.3 (15.2)	55.7 (15.9)
Whipples index	1.28	2.55	1.5	2.41
Myers blended index	12.66	30.65	12.43	30.86

SD = standard deviation; † max age=95

Table 3 presents mortality indicators estimated using 1) proxy-reported data only and 2) all data combined (reported+estimated). We present probabilities of dying between ages *x* and *x+n* (*nqx*) for the 5-year period before the survey. Results show that mortality indicators for the reported data only are higher than those for all data combined, suggesting that combining proxy-reported and proxy-estimated data underestimates mortality indicators. For instance, the probabilities of dying between the ages of 30 and 60 (*30q30*) are 0.189 and 0.398 for females and males respectively when using the non-estimated (reported) data, while the corresponding probabilities are 0.169 and 0.318 for females and males, respectively. The underestimation appears to be more pronounced for males than females.

Table 3: Adult mortality indicators by the reporting status of age information, RaMMPS

	Reported		**Total**	
Indicator	Female	Male	Female	Male
20q30	0.078	0.254	0.068	0.164
30q30	0.189	0.398	0.169	0.318
20q40	0.164	0.285	0.154	0.251
15q45	0.137	0.255	0.134	0.225
30q50	0.400	0.634	0.399	0.574

4.4 The process of obtaining age-related information from mortality surveys

Having observed the noise in the age data, especially the estimated age information, we turned to audio recordings of interviews to examine the potential sources of error, paying particular attention to the probing process for age information. This analysis identified two factors as potential explanations for the observed age data distortions. First, some interviewers lacked effective probing skills, especially when the respondents lacked the knowledge to provide the required information. Despite training interviewers on how to elicit accurate age and date information, in some situations, the probing quality fell short of expectations. In many such instances, a few more probes or reference to a historical calendar of events to locate an event in time could have yielded more plausible estimates than those recorded. To illustrate, we present a conversation between a male interviewer and a male respondent who struggled to provide the age of his mother but ended up providing a rough estimate.

Interviewer [I]*: Is your biological mother alive?*
Respondent [R]*: What?*
I: *Is your biological mother alive?*
R*: Yes*
I*: How old is she?*

R*: Eeh that I don't know.*
I*: [laughs] But, according to how you see her, what do you think is her age?*
R*: More than 50*
I*: 50?*
R*: Yes*
I*: Alright*

This approach to age probing presents a couple of challenges. First, using subjective criteria such as "according to how you see her" can be problematic as appearance alone can be misleading when determining someone's age. Second, the response of "more than 50" is not equivalent to the specific age of 50. It could range from 51, which is a minor difference, to a much older age.

In another scenario, a different interviewer promptly recorded a date that might not be entirely accurate, as observed in the following conversation:

I*: Is your biological father alive?*
R*: No, he died a long time ago.*
I*: When did he die?*
R*: When I was very young, but around 90s or 80s – probably in the 90s.*
I*: 1990, we should just record 1990, right?*
R*: Yes*
I*: Thank you …*

In this situation, recording 1990 as the year of death might be problematic and may lead to misplacing the event in the timeline. To mitigate this problem, the interviewer could have probed further by referencing specific historical events from the decade, thereby facilitating a more precise estimation of the year of death.

The second factor pertains to a sociocultural norm prevalent among Malawians, and perhaps Africans, that hinders the disclosure or inquiry of age-related information within families, especially between parents and children. This results in many children not knowing the ages of their parents. As a result, despite excellent probing skills, interviewers often struggle to obtain reliable age information from respondents. This phenomenon is not exclusive to older individuals; even younger respondents with young parents exhibit this sociocultural disconnect that hinders them from seeking such information. In part, this can be attributed to the absence of the tradition of birthday celebrations, especially for parents and particularly in rural areas. Consequently, across all demographics, most respondents lack knowledge of their parents' birthdates and, consequently, current ages or at the time of death. This can be illustrated by an interview with an 18-year-old male rural respondent, who indicated that he has never actively sought out age-related information from his parents.

I: *You said that you are staying with both your parents. How old is your mother?*
R: *I have never asked her*
I: *How old does she look like?*
R: *Uhm, I don't know*

Likewise, another 19-year-old female urban respondent expressed uncertainty regarding the age of her young adult mother which further highlights the sociocultural disconnect between parents and their children in seeking or revealing age-related information.

I: *Is your biological mother alive?*
R: *Yes*
I: *How old is she?*
R: *I don't know, but it should be around thirty-something*
I: *What should I record, thirty-what?*
R: *Just put 39, iih not 39, maybe 36 or 39, something around that range*

The socio-cultural disconnect highlighted above can be exacerbated by the low literacy levels in countries such as Malawi, where many parents are unsure or unaware of their ages. In such contexts, despite excellent probing skills or children's attempts to inquire, obtaining accurate age data may remain challenging. As a result, there is a significant reliance on guesswork when reporting age information. This is especially noticeable when respondents are proxy-reporting age information of their family members, often resorting to rough estimations when uncertain. Such guessed age information usually leads to age heaping, especially in the preferred digits of 5 and 0's. We have shown in the above analysis that, for the estimated information, there is substantial age heaping in the ages 40 and 50. The following conversations with a 35-year-old female urban respondent and a 45-year-old male respondent, respectively, underscores this phenomenon within this cultural context.

Interview with a 35-year-old female urban respondent:

I*: Is your biological mother alive?*
R*: She died*
I*: I am so sorry. When did she die?*
R*: 2008*
I*: How old was she when she died?*
R*: Iih I cannot know.*
I*: As an estimate, you saw your mother, give me an estimated age*
R*: Between 40 and 50*
I*: I should record forty-what?*
R*: Record 45, yea put 45*

Interview with a 45-year-old male urban respondent:

I*: Is your biological mother alive?*
R*: Yes, she is alive*
I*: How old is she*
R*: I don't know about her age*
I*: Just give me an estimate*
R*: It could be sixty-something, sixty-five*
I*: What?*
R*: 65*
I*: Ok, alright*

5. Discussion and conclusions

Ensuring the quality and accuracy of data is crucial in survey research. However, it is essential to acknowledge that survey data are susceptible to errors. Non-response errors, stemming from both the interviewer and respondent, are virtually inevitable. For example, age information obtained through proxy respondents in surveys is subject to rough estimates, sometimes influenced by enumerators. This study delves into the discrepancies between proxy-estimated and proxy-reported age data obtained from a mobile phone survey conducted in Malawi during COVID-19. It distinguishes the quality of 'accurately' reported data and those obtained through estimation. Furthermore, the study leverages audio-recorded interviews to unearth potential sources of errors in proxy-estimated age information by scrutinizing the interactions between enumerators and respondents.

Our analysis reveals important disparities between proxy-estimated age data and the presumably accurate proxy-reported age data. Proxy-estimated data are very erratic and produce distinct data quality summary measures compared to proxy-reported data. Through graphic analysis and age-heaping indices, we observe a pronounced age-heaping problem in the proxy-estimated age data, especially in ages ending in 0's and 5's, relative to the proxy-reported age data. Specifically, we find substantial heaping at the ages 40 and 50 in proxy-estimated ages of living parents and ages at death, which is consistent across the sex of parents.

In light of the observed disparities, if a significant portion of the sample constitutes estimated data, then this may have serious implications for demographic estimation. Typically, when age data contain a considerable amount of missing information, the standard practice involves the use of statistical methods to impute from the available non-missing data (Khan et al. 2020; Musil et al. 2002). However, if the data used for imputation is flawed, the imputed data will inevitably be flawed as well (a.k.a. garbage in, garbage out). Hence, researchers must exercise caution when incorporating such data into their analyses. Furthermore, if not properly accounted for, the estimated data may distort the quality of the correctly reported data, producing biased age-specific demographic indicators. Therefore, researchers should indicate whether specific age information collected during sample surveys was obtained through an estimation process to ensure accuracy in demographic analyses.

Upon examination of audio-recorded interviews, we identified two issues contributing to highly inaccurate age reporting: 1) inadequate enumerator probing skills and 2) sociocultural disconnect surrounding the disclosure or children's inquiry of parental age information. Consequently, there is a heavy reliance on guesswork when proxy-estimating the ages of family members. Nevertheless, it is important to acknowledge that proxy-collecting age-related information, especially parents, is challenging in settings like Malawi. Therefore noise in the data should be anticipated, and caution should be exercised when handling such data, especially proxy-estimated information. We, therefore, advocate for an improvement in probing skills, especially in low-resource settings characterized by high illiteracy rates and sociocultural barriers to information sharing, compounded by the absence of civil registration systems for family records. This underscores the importance of closely monitoring audio-recorded interviews and continuously learning from previous interview experiences. Indeed, the primary purpose of the RaMMPS project in monitoring audio-recorded interviews was to offer ongoing feedback to enumerators regarding their interviewing practices, including their probing skills. In addition, this highlights the need for LMICs to accelerate the efforts in implementing nationwide birth and death registration systems and foster a culture of timely registration of deaths. These measures are essential for the production of timely, accurate, and reliable vital statistics.

We, however, acknowledge that the analysis of proxy-estimated ages is based on a small sample size.

References

Akinnusi, D. M., & Molosi, L. D. (2008). Low rate of reporting of confirmed AIDS-related deaths using BI-1663 forms by private medical practitioners in the Mafikeng, North West Province. Southern African Journal of Epidemiology and Infection, 23(4), 20-24.

Coale, A. J., & Li, S. (1991). The Effect of Age Misreporting in China on the Calculation of Mortality Rates at Very High Ages. Demography, 28(2), 293–301. https://doi.org/10.2307/2061281

Crampin, A. C., Dube, A., Mboma, S., Price, A., Chihana, M., Jahn, A., Baschieri, A., Molesworth, A., Mwaiyeghele, E., Branson, K., Floyd, S., McGrath, N., Fine, P. E., French, N., Glynn, J. R., & Zaba, B. (2012). Profile: the Karonga Health and Demographic Surveillance System. *International journal of epidemiology*, *41*(3), 676–685. https://doi.org/10.1093/ije/dys088

Heerwegh, D., & Loosveldt, G. (2008). Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly*, 72(5), 836-846. https://doi.org/10.1093/poq/nfn045

Helleringer, S., Lau, S. W., Luhar, S., Banda, J., Lankoande, B., Tlhajoane, M., & Reniers, G. (2023). Increased Age Heaping in Mobile Phone Surveys Conducted in Low-Income and Middle-Income Countries. Socius, 9, 23780231231158766.

Horsfall, M., Eikelenboom, M., Draisma, S., & Smit, J. H. (2021). The effect of rapport on data quality in face-to-face interviews: Beneficial or detrimental? *International Journal of Environmental Research and Public Health*, *18*(20), 10858.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, *67*(1), 79-125.

Johnson, P., Spoorenberg, T., Hertog, S., & Gerland, P. (2022). Method protocol for the evaluation of census population data by age and sex. Population Division, United Nations Department of Economic and Social Affairs. New York.

Huy, T. Q., Johansson, A., & Long, N. H. (2007). Reasons for not reporting deaths: a qualitative study in rural Vietnam. World health & population, 9(1), 14-23.

Khan, S. I., & Hoque, A. S. M. L. (2020). SICE: an improved missing data imputation technique. Journal of Big Data, 7(1), 1-21.

LSHTM. (2022) Rapid Mortality Mobile Phone Survey (RaMMPS). London School of Hygiene and Tropical Medicine (LSHTM). https://www.lshtm.ac.uk/research/centres-projects-groups/rapid-mortality-mobile-phone-survey

Malik, M. A. (2021). Age heaping pattern and data quality: evidence from Indian Household Survey Data (1991–2016). *Communications in Statistics: Case Studies, Data Analysis and Applications*, *7*(3), 382-393.

Masquelier, B., Kanyangarara, M., Pison, G., Kanté, A. M., Ndiaye, C. T., Douillot, L., ... & 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.

McNabb, D. E. (2014). Nonsampling error in social surveys. Sage publications.

Musil CM, Warner CB, Yobas PK, Jones SL. (2002). A Comparison of Imputation Techniques for Handling Missing Data. Western Journal of Nursing Research, 24(7):815-829.

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology*, *9*(3), 429-448.

Appendix

Table 1: Percentage of age information that is imputed in DHS conducted in selected SSA countries

			Women			Men		
Country	Survey type	Survey year	Urban	Rural	Total	Urban	Rural	Total
DRC	DHS	2013-14	4.7	11.3	8.8	2.4	10.2	7.3
Kenya	MIS	2015	7.7	14.4	11.7	na	na	na
Malawi	DHS	2015-16	1.9	7.4	6.4	2.0	7.6	6.6
Mozambique	IDS	2011	4.3	4.2	4.2	1.2	2.0	1.7
Senegal	DHS	2015	16.7	45.8	32.0	6.6	21.0	13.6
Swaziland	DHS	2005	0.8	2.2	1.8	2.1	4.9	4.1
Tanzania	DHS	2015-16	3.6	7.8	6.2	5.9	9.4	8.2
Zimbabwe	DHS	2015	0.1	0.7	0.5	0.2	0.6	0.5

Source: Authors' own compilation using variable m/v014