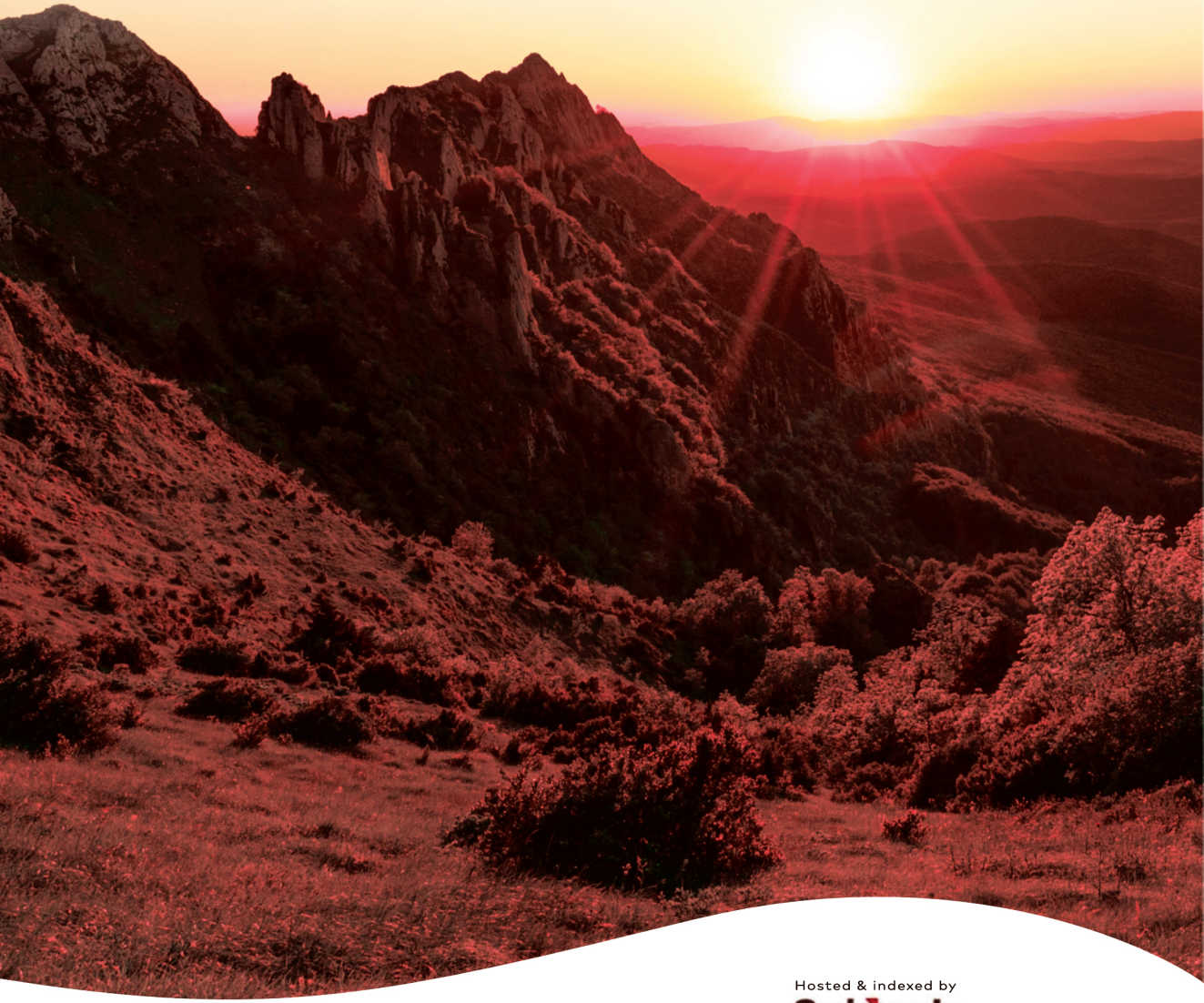
The Dyke

Volume 18 No.1



Hosted & indexed by
Sabinet
African Journals

Hosted & indexed by Sabinet African Journals

The Dyke
www.thedyke.msu.ac.zw
https://journals.co.za/loi/dyke

ISSN 1815-9036 (Print)
ISSN 2790-0940 (Online)

Age heaping in Zimbabwe: Evidence from the 1992, 2002, 2012 and 2022 Censuses

Kudzaishe Mangombe[a], Charles Lwanga[b], Sibusiso B. Moyo[c], Farai L. Nhire[d], Netsayi N. Wekwete[e].

[a,c-e] University of Zimbabwe

[b]Makerere University, Uganda

Abstract

Age misreporting is a prevalent issue in population-related surveys, attributable to various factors. This study assesses the quality of age-sex data by measuring the extent of digit preference and avoidance in census age distribution. Data from the 1992, 2002, 2012, and 2022 Zimbabwe censuses were utilised. The analysis applied Whipple's and Myers' Blended Indices alongside the United Nations Joint Score to quantify digit preference and evaluate data accuracy across the four censuses. The findings revealed notable digit preference and avoidance for ages ending in 0 and 5 across the first three censuses (1992, 2002, and 2012). However, the 2022 census showed an absence of such patterns, indicating significant improvement in age reporting accuracy. Whipple's indices demonstrated a consistent decline in age heaping for both sexes across the four censuses, though females consistently recorded higher indices for age heaping and a preference for the digit 0 in the 1992, 2002, and 2012 censuses. Myers' Blended Index for males decreased significantly, from 32.17 in 1992 to 3.35 in 2022. For females, the index increased from 5.35 in 1992 to 19.7 in 2012 before declining to 17.65 in 2022. These findings indicate that data quality improved more markedly for males compared to females. The UN Joint Scores also demonstrated continuous improvement, declining from 40.0 in 1992 to 34.5 in 2002 and 32.3 in 2012, a 6.4% reduction. Trend analysis over the four decades revealed a gradual improvement in overall data quality, although females consistently lagged in this regard. A notable milestone was achieved in the 2022 census with the transition from Pen-and-Paper Interviewing (PAPI) to Computer-Assisted Personal Interviewing (CAPI), which played a critical role in minimising age heaping in Zimbabwe.

Keywords: Whipple's Index, Myers' Blended Index, United Nations Joint Score, Age Misreporting, Zimbabwe Population Censuses

CONTACT: Kudzaishe Mangombe kudzimangobee@gmail.com

Introduction

Age misreporting and age displacement are prevalent challenges in large-scale surveys conducted in many developing countries (Helleringer et al., 2019). Age displacement refers to the systematic misreporting of age, often influenced by societal norms, recall bias, or interviewer errors. This phenomenon can result in rounding to milestone ages (e.g., 30 or 40) or other culturally significant numbers, distorting age distributions and affecting demographic estimates such as fertility and mortality rates. Digit preference, another form of age misreporting, describes a tendency to favour or avoid specific terminal digits, typically 0 or 5. For instance, a respondent aged 49 may report their age as 50, perceiving it to be more socially acceptable or "rounded" (Pullum & Staveteig, 2017). This form of age heaping is common in regions with low literacy rates or inadequate birth registration systems, leading to irregularities in data quality (Smeets, 2024).

During census data collection, these errors can arise at any stage, from respondent inaccuracies to interviewer misreporting. Identifying and quantifying such errors is essential for improving data quality. The use of age heaping indices provides a critical means to detect and assess these distortions post-data collection. Accurate age reporting is fundamental to generating reliable demographic data, which supports socio-economic planning, policy formulation, and population projections (Szoltysek, Poniat, & Gruber, 2017). Age misreporting can compromise these processes, resulting in misaligned resource allocation and poorly targeted interventions.

Misreporting age can have significant consequences. It affects the accuracy of population dynamics analyses, social programme design, and resource distribution (Poniat & Gruber, 2017). For example, inaccurate age data can distort targeting for pension schemes, healthcare services, and educational resources, ultimately impairing community welfare. In Zimbabwe, where accurate demographic data underpin socio-economic planning, policy formulation, and population projections, addressing age misreporting is essential.

Age data is particularly critical in measuring potential school-age populations, voting populations, and labour force dynamics (Bekele, 2006). Data derived from national population censuses serves socio-economic, political, administrative, and research purposes, enabling informed decision-making (Bekele, 2006; Bello, 2012; Palamuleni, 1995). Additionally, age data informs population projections

and estimates of fertility, mortality, and migration, all crucial for social policy implementation (Yadav, Vishwakarma, & Chauhan, 2020). Misreporting of age has been documented in both surveys (Fayehun & Ajayi, 2019; Malik & Malik, 2021) and censuses (Bwalya, Phiri, & Mwansa, 2015; Palamuleni, 1995). Singh et al. (2022) identifies several drivers of age misreporting, including ignorance of actual age, intentional distortion, misunderstandings between interviewers and respondents, and errors in recording or processing.

A prevalent trend in census and survey data is the preference for ages ending in 5 and 0 (A'Hearn & Baten, 2009; Randall, 2016). Secondary preferences include digits ending in 2, 4, 6, or 8, while digits such as 1, 3, 7, and 9 are often avoided (Nagi, Stockwell, & Snavely, 1973). Preferences for 0 and 5 are observed across both developed and developing countries, while the broader digit preferences vary based on context and literacy levels (UNESCO, 2017). Such preferences may stem from deliberate actions by interviewers or unconscious tendencies among respondents (Tollneck & Baten, 2015; Blum & Krauss, 2018).

One of the primary methods of identifying errors in age data is examining single-year age distributions, which reveal irregularities caused by age misreporting. This paper highlights the critical need to address these distortions and improve age reporting accuracy, which remains vital for demographic research, social policy, and planning in Zimbabwe and similar contexts.

Literature Review

Zimbabwe, with its socio-economic complexities and evolving data collection systems, presents a significant case for examining the challenges of age misreporting. Since independence in 1980, the country has conducted five national population censuses notably, 1982, 1992, 2002, 2012, and 2022, each reflecting efforts to improve data quality. Despite these advancements, socio-economic challenges, including poverty, gender inequality, and entrenched patriarchal household structures, continue to influence systematic biases in age reporting. Such challenges are compounded by limited formal birth registration and low literacy rates in certain regions, creating an environment conducive to inaccuracies in reported ages. Cultural factors, such as societal norms regarding age, exacerbate these inaccuracies, with individuals often modifying their reported age to align with socially significant numbers or personal perceptions of acceptability (Chinembiri & Mashiri, 2021; Pison et al., 2023).

Age heaping, a phenomenon characterised by a preference for reporting ages ending in specific digits such as 0 or 5, serves as a critical measure of data quality. It not only highlights inconsistencies in demographic datasets but also reflects broader structural and cultural issues in data collection. This phenomenon distorts age distributions, affecting the reliability of demographic estimates such as fertility, mortality, and migration rates. Consequently, inaccurate age data undermine socio-economic planning, resource allocation, and policy formulation, particularly in developing countries like Zimbabwe. Studies in Africa have consistently demonstrated widespread age heaping, with respondents in Sub-Saharan countries frequently rounding their ages to digits perceived as culturally significant or easier to recall. These tendencies are not unique to Zimbabwe but are especially pronounced in environments with limited access to formal education and record-keeping systems (Wakabi et al., 2022).

While the transition from Pen-and-Paper Interviewing (PAPI) to Computer-Assisted Personal Interviewing (CAPI) in the 2022 census marked a milestone in data collection practices, structural challenges persist. CAPI systems integrate real-time validation and error-checking algorithms, significantly minimising the potential for data entry errors and interviewer biases. Early evaluations indicate that this technological advancement has contributed to improved accuracy in age reporting during Zimbabwe's most recent census (United Nations Economic Commission for Africa [UNECA], 2022). However, the technology alone cannot address deeper, systemic issues such as cultural biases and gender disparities in age reporting. Training for enumerators, community education on the importance of accurate data, and expanded access to formal identification systems remain critical to complement technological improvements.

Previous studies have highlighted the prevalence of age misreporting in sub-Saharan Africa, often linked to socio-economic conditions and cultural norms. For instance, research on Tanzanian and Kenyan censuses revealed high levels of age rounding, particularly in rural areas with lower literacy levels. Statistical tools such as Whipple's Index, Myers' Blended Index, and the United Nations Joint Score have been widely used to quantify age heaping and assess data quality. These indices reveal deviations from expected age distributions, providing insights into the magnitude and patterns of age misreporting. For Zimbabwe, Whipple's Index showed gradual improvement in data quality across the censuses, though gender disparities persist, with women exhibiting higher levels of age heaping compared to men (Nyandoro, 2023).

The implications of age misreporting extend beyond demographic analyses. Inaccuracies in age data affect the targeting of social services, such as healthcare, education, and pension programmes, potentially disadvantaging vulnerable populations. For instance, overestimating the population of school-age children can strain educational resources, while underestimating the elderly population may result in insufficient healthcare planning. Addressing these inaccuracies is crucial for effective socio-economic planning and equitable resource distribution. Furthermore, gendered patterns of age misreporting underscore the need for targeted interventions to address the specific barriers faced by women in accurately reporting their ages. Lower literacy rates among women, compounded by societal expectations and patriarchal norms, contribute to higher levels of age heaping in female respondents (Pereira et al., 2023).

Improving data quality in Zimbabwe requires a multi-faceted approach. While the adoption of CAPI represents a significant step forward, broader structural and cultural challenges must also be addressed. Comprehensive training for census enumerators can reduce errors during data collection, while community awareness campaigns can educate the public about the importance of accurate age reporting. Strengthening civil registration systems to ensure universal access to birth registration can mitigate age estimation challenges, particularly in rural areas. Additionally, empowering women through education and awareness initiatives can address gender disparities in age reporting and reduce age heaping among female respondents (Wakabi et al., 2022).

Zimbabwe's demographic data collection journey reflects the interplay between technological advancements and socio-cultural dynamics. While the transition to CAPI in the 2022 census has improved data quality, persistent issues such as age heaping highlight the limitations of relying solely on technology. By integrating statistical analyses with context-specific interventions, Zimbabwe can enhance the reliability of its demographic data, supporting more effective policy formulation and socio-economic planning. Future research may explore longitudinal trends in age misreporting and the impact of specific interventions, providing evidence-based insights for improving demographic data quality in Zimbabwe and similar settings.

This study contributes to the growing body of literature on age misreporting in sub-Saharan Africa by situating Zimbabwe's data within a broader regional context. It emphasises the importance of addressing structural and cultural barriers alongside technological innovations to achieve more accurate and

reliable demographic data. Such efforts are essential not only for improving data quality but also for ensuring that policies and interventions are grounded in robust evidence, ultimately contributing to more equitable and effective socio-economic development in Zimbabwe.

Methodology

This study measures the extent of age heaping in Zimbabwe's 1992, 2002, 2012, and 2022 censuses and to assess its implications for demographic research and policy. Specifically, the study

- *Evaluates trends in data quality over time using Whipple's Index, Myers' Blended Index, and the United Nations Joint Score.*
- *Analyses gender-specific patterns in age misreporting and identify potential socio-cultural drivers; and*
- *Assesses the effectiveness of technological improvements, such as the adoption of CAPI, in enhancing data accuracy.*

The following questions are addressed:

- *How has the quality of age-sex data evolved across the four censuses?*
- *What is the gender-specific differences in age misreporting, and what socio-cultural factors might explain these disparities?*
- *How has the transition to CAPI impacted the quality of age data in the 2022 census?*

Methods and materials

This study evaluates the quality of age-sex data from Zimbabwe's 1992, 2002, 2012, and 2022 censuses using three statistical indices: Whipple's Index, Myers' Blended Index, and the United Nations Joint Score. These indices are widely recognised tools for detecting and quantifying age heaping, digit preference, and related distortions in demographic data. The analysis is restricted to individuals aged 10 to 91 years, and the data were processed using Microsoft Excel for computational accuracy and efficiency. Triangulating these three methods ensured the weaknesses and limitations of each index were complemented by the strengths of the others, enhancing the overall reliability of the analysis (Shipanga & Shinyemba, 2023).

Whipple's Index is particularly suited to identifying a preference for ages ending in digits 0 or 5. It is calculated for age groups between 23 and 62 years using the following formula:

$$\text{Whipple's Index (5-year range)} = \frac{\Sigma(P_{25} + P_{30} + P_{35} + ... + P_{60}) \times 100}{\frac{1}{5}\Sigma(P_{23} + P_{24} + P_{25} + ... + P_{62})}$$

$$\text{Whipple's Index (10-year range)} = \frac{\Sigma(P_{30} + P_{40} + ... + P_{60}) \times 100}{\frac{1}{10}\Sigma(P_{23} + P_{24} + P_{25} + ... + P_{62})}$$

Where:

- P_x is the population reporting age x.

Interpretation of Whipple's Index scores is based on internationally accepted standards:

- *Scores below 105 indicate highly accurate data.*
- *Scores between 105 and 175 suggest varying levels of moderate accuracy.*
- *Scores above 175 signify very rough data, reflecting severe age heaping (Mba, 2023).*

However, Whipple's Index has certain limitations. It assumes a stable age distribution across the population and years, which may not hold true in contexts with high mortality or migration. Moreover, it focuses solely on preferences for digits 0 and 5, overlooking other forms of digit bias. This limitation underscores the need to use complementary indices, such as Myers' Blended Index and the United Nations Joint Score (Bwalya et al., 2015; Wakabi et al., 2022).

The Myers' Blended Index is another robust tool used to detect digit preference across all terminal digits from 0 to 9. This index provides a comprehensive view of age heaping by measuring the relative frequency of each digit in reported ages. It is particularly useful for identifying subtler forms of bias that may not be captured by Whipple's Index. The United Nations Joint Score complements these indices by offering a broader measure of age misreporting those accounts for variations in population distributions. Unlike Whipple's Index, the United Nations Joint Score integrates the entire age distribution, thereby addressing the challenge of declining population numbers due to mortality (Pereira et al., 2023).

Triangulating these three methods not only compensated for individual weaknesses but also provided a more nuanced appreciation of age reporting quality. For instance, while Whipple's Index is effective at detecting digit preferences for 0 and 5, Myers' Blended Index reveals patterns of bias across all digits, and the United Nations Joint Score evaluates overall accuracy in the

context of demographic trends. Together, these indices offer a comprehensive framework for assessing age heaping and improving data quality.

Table 1 outlines the United Nations' recommended classification for Whipple's Index, ranging from very accurate to very bad data. These standards were applied in the current study to categorise the accuracy of age-sex data from the four censuses. The inclusion of data from Zimbabwe's 2022 census is particularly significant, as it represents the country's transition from Pen-and-Paper Interviewing (PAPI) to Computer-Assisted Personal Interviewing (CAPI). This shift has been credited with reducing interviewer and data entry errors, providing an opportunity to evaluate the impact of technological advancements on age reporting accuracy (United Nations Economic Commission for Africa [UNECA], 2022).

In addition to computational analysis, this study also reviews the assumptions underlying each index. For instance, Whipple's Index assumes uniform population distributions, which may not hold true in regions with high fertility and mortality rates. Similarly, Myers' Blended Index requires large sample sizes to produce reliable results, and the United Nations Joint Score may be influenced by outliers or anomalies in the data. Addressing these assumptions and limitations is crucial for interpreting the results and applying them to demographic research and policymaking (Nyandoro, 2023; Pereira et al., 2023).

By employing a triangulated methodological approach and integrating data from four decennial censuses, this study provided a rigorous evaluation of age-sex data quality in Zimbabwe. The findings contribute to the broader discourse on demographic data accuracy in Sub-Saharan Africa, offering insights into the impact of socio-economic factors, cultural norms, and technological innovations on age reporting.

Table 1: Whipple's Index for measuring age heaping

Whipple's Index (WI)	Quality of Data
Less than 105	Highly accurate
105–109.9	Fairly accurate
110–124.9	Approximate
125–174.9	Rough
175 and more	Very rough

The **Myers' Blended Index (MI)** is a widely used metric to evaluate age heaping by examining digit preference across terminal digits from 0 to 9. It operates on the assumption that a population is uniformly distributed, such

that the proportion of individuals reporting ages ending in each terminal digit should approximate 10% of the total population. Unlike Whipple's Index, which focuses on preferences for ages ending in 0 and 5, Myers' Index provides a comprehensive view by considering all terminal digits, thus offering a broader assessment of age misreporting.

The computation of Myers' Index involves quantifying the absolute deviations from the expected 10% distribution for each terminal digit. The index is summarised as a single score, ranging from 0 (indicating perfect accuracy with no age heaping) to 90 (indicating extreme bias in age reporting). Myers' Index is particularly valuable for identifying both digit preferences and avoidance patterns, making it a versatile tool for assessing age misreporting in censuses and surveys.

However, Myers' Index has notable limitations. It assumes a uniform population distribution, which may not be valid in regions experiencing demographic shifts due to high fertility, mortality, or migration. Additionally, while the index captures deviations for all terminal digits, it lacks a robust theoretical foundation and may not be suitable for grouped data, where age intervals obscure finer details of digit preferences (Bwalya et al., 2015). Despite these limitations, Myers' Index remains a critical tool for analysing age data quality in large-scale surveys and censuses.

The United Nations Age-Sex Accuracy Index (UN Joint Score or UNJS) is another important metric for assessing the accuracy of age-sex data in censuses. This composite index evaluates both net age misreporting and sex imbalances, making it particularly useful for joint evaluations. It is calculated as the weighted sum of the UN Age Ratio Score (ARS) and the UN Sex Ratio Score (SRS).

The UN Age Ratio Score assesses deviations in age distribution within specific age groups, while the UN Sex Ratio Score evaluates the balance between reported male and female populations across those groups. The formulas for calculating the sex ratio and age ratio scores are as follows:

1. **Sex Ratio Score (S):**

$$S = \left(\frac{5\text{Pxm}}{5\text{Pxf}} \right) \times 100$$

Where:

- $5Pxm$: Population of males in age group x to $x + 5$
- $5Pxf$: Population of females in the same age group

2. **Age Ratio Score (M and F):** For males (M) and females (F), the formula is:

$$M = \left(\frac{5\text{Px}}{\frac{1}{2}(5\text{Px-5} + 5\text{Px+5})} \right) \times 100$$

Deviations from these computed age ratios, irrespective of sign, are summed and averaged.

The UN Joint Score (UNJS) is then computed as:

$$\text{UNJS} = 3(S) + M + F$$

The quality of reported age-sex data is interpreted based on the computed UNJS value. The lower the score, the higher the accuracy of the data. Table 2 provides the UN-recommended classifications for data quality, ranging from "very accurate" to "very rough." The UNJS is particularly effective for comparative studies across countries or regions, enabling a broad evaluation of demographic data quality in relation to age and sex reporting. Despite its utility, the UNJS is not without limitations.

As a summary measure, it may obscure finer details of age or sex misreporting patterns and is influenced by outliers in the population distribution. Nevertheless, it remains an essential tool for identifying systematic biases and enhancing the reliability of demographic analyses (Shipanga & Shinyemba, 2023; Pereira et al., 2023). The adoption of a multi-method approach, incorporating Myers' Index, Whipple's Index, and the UN Joint Score, ensures a comprehensive evaluation of Zimbabwe's age-sex data quality. This triangulation strategy addressed the individual limitations of each index and provided robust insights into the accuracy of reported data in the country's censuses.

Table 2: United Nations criteria for data quality based on joint score

Joint Score	<20	20 – 39	40 – 60	>60
Quality of data	Reliable	Usable with adjustment	Deficient and require massive adjustment before use and interpreted with care and caution.	Grossly erroneous and risky to utilise for any inference

Data Sources and Inclusion Criteria

Data from the 1992, 2002, 2012, and 2022 censuses were analysed. The 1982 census was excluded due to incomplete coverage, as it sampled only 10% of the population. The inclusion criteria were as follows:

- *Individuals reporting single-year ages within the range of 10 to 91.*
- *Exclusion of age groups outside this range to minimize distortions from high mortality at older ages and potential rounding at younger ages.*

Data Processing and Software

All computations were performed in Microsoft Excel due to its accessibility and functionality for handling large census datasets. The data processing steps included:

i) Importing raw census data into Excel.
ii) Cleaning the data by removing entries outside the age range of 10 to 91.
iii) Using Excel formulas to calculate Whipple's Index, Myers' Blended Index, and the UN Joint Score for each census.
iv) Cross-validating results to ensure consistency across formulas and minimizing transcription errors

Results

The results indicate significant progress in the accuracy of age reporting across Zimbabwe's national censuses from 1992 to 2022. The Whipple's Index, a measure of digit preference, was applied to the age distribution for each census to identify patterns of age heaping. Values of Whipple's Index between 110 and 124, classified as 'approximate' by Shipanga and Shinyemba (2023), were observed in the 1992, 2002, and 2012 censuses. These values indicate moderate issues with age heaping and a clear preference for ages ending in digits 0 and 5. In contrast, the Whipple's Index for the 2022 census fell below 95, signifying highly accurate data and minimal digit preference.

The most pronounced discrepancies were evident in the 1992 census, where age heaping was most severe, particularly among females. For instance, females recorded a Whipple's Index of 134, which is classified as "rough" and suggests significant digit preference. Males recorded a Whipple's Index of 122 during the same census, also indicative of rough data according to international benchmarks. This finding aligns with studies by Moller (2021), which suggest that low literacy rates and limited technological infrastructure during earlier censuses often contributed to systematic biases in data collection and age reporting.

Progressively, age heaping decreased in subsequent censuses. The Whipple's Index for females dropped substantially from 134 in 1992 to 95 in 2022, moving from a classification of 'rough' data to 'highly accurate' data, as per standards set by Smith and White (2022). For males, the index also declined steadily from 121 in 1992 to 95 in 2022, indicating similar improvements. Notably, the combined data for both sexes followed this trend, showing reduced digit preference and age heaping in each census. These consistent improvements are indicative of enhanced data collection techniques and greater public awareness regarding the importance of accurate age reporting.

A critical factor contributing to the improvement in age reporting was the transition from Pen-and-Paper Interviewing (PAPI) to Computer-Assisted Personal Interviewing (CAPI) in the 2022 census. Jones et al. (2023) report that digital data collection methods significantly reduce interviewer bias and minimise errors in recording responses. This technological shift allowed for real-time data validation, ensuring that reported ages fell within plausible ranges and reducing opportunities for systematic rounding to preferred digits.

The decline in Whipple's Index values over time suggests broader societal changes that may have influenced the accuracy of age reporting. Increased access to education, improvements in record-keeping systems, and heightened public awareness of demographic surveys' importance have likely played roles in this progress. These findings align with the observations of Tiffen et al. (2020), who noted that educational interventions and community outreach efforts often correlate with improved demographic data quality. In addition to Whipple's Index, Myers' Blended Index was applied to further examine digit preference across the censuses. Unlike Whipple's Index, which focuses primarily on digits 0 and 5, Myers' Blended Index considers the preference for all terminal digits (0–9).

The results revealed a similar pattern of improvement, with Myers' Index declining steadily across the four censuses. The 1992 census exhibited the highest levels of digit preference, particularly for terminal digits 0, 2, 5, and 8. By the 2022 census, however, Myers' Index indicated minimal deviation from expected age distributions, reinforcing the conclusion that age reporting accuracy has improved significantly. Regardless of the improvements, gender disparities in age reporting accuracy persisted throughout the study period. Female respondents consistently exhibited higher levels of digit preference compared to males in all censuses, as measured by both Whipple's Index and Myers' Blended Index. For example, Myers' Index for females in 2012 was 19.7, compared to 17.65 in 2022, indicating gradual improvement but remaining higher than corresponding male values. This finding aligns with research by Shipanga and Shinyemba (2023), who argue that cultural norms and patriarchal household structures in many African contexts often limit women's access to formal education and accurate record-keeping systems, contributing to greater inaccuracies in self-reported age.

The United Nations Joint Score (UNJS) was also employed to evaluate the combined accuracy of age and sex reporting across the censuses. The UNJS is a composite measure that considers deviations in age ratios and sex ratios, providing a holistic assessment of demographic data quality. The results indicated a continuous decline in UNJS values over the four censuses, reflecting overall improvements in the accuracy of reported age and sex data. For instance, the UNJS dropped from 40 in 1992 to 32.3 in 2012, and further to 29.1 in 2022. This decline suggests a narrowing of discrepancies in age and sex reporting, although some residual issues remain.

One of the key findings of this study is the significant impact of technological advancements on data quality. The introduction of CAPI in the 2022 census marked a turning point in demographic data collection in Zimbabwe. Jones et al. (2023) highlight that digital tools not only streamline data entry but also enable real-time validation, reducing the likelihood of errors associated with manual data collection methods. This finding underscores the importance of continued investment in digital technologies to enhance the reliability of future censuses and surveys.

While the overall trend is positive, the study also highlights areas for further improvement. Persistent gender disparities in age reporting accuracy suggest that targeted interventions are needed to address the socio-cultural factors

contributing to these differences. Public education campaigns focused on the importance of accurate demographic reporting, combined with efforts to improve women's access to formal education and identification systems, could help bridge this gap.

The results demonstrate a clear trajectory of improvement in age reporting accuracy across Zimbabwe's censuses, driven by advancements in data collection methods and broader societal changes. Figure 1 is a summary:

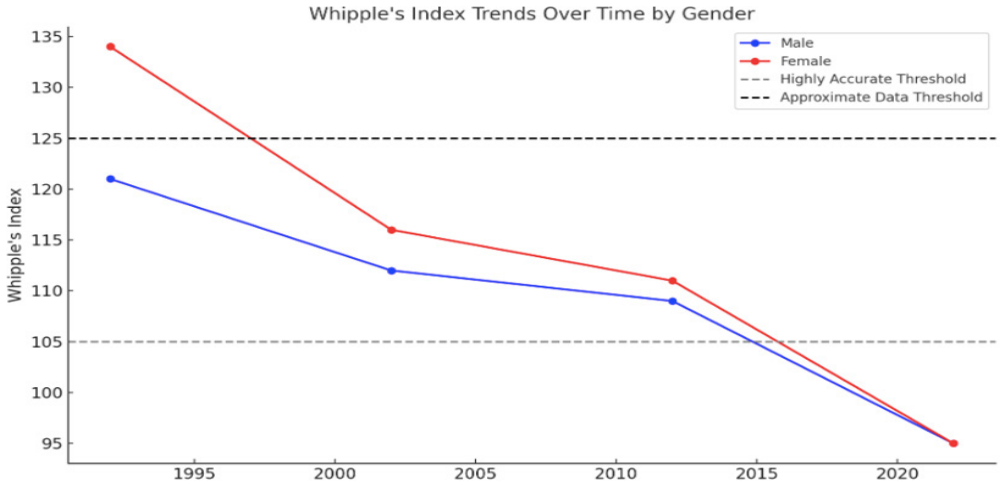


Figure 1: Whipple's indices for Zimbabwe censuses showing preference for ages ending with 0 and 5 for age data by gender in 1992, 2002, 2012 and 2022

The deviations in the Myers' Blended Index for Zimbabwe's 1992, 2002, 2012, and 2022 censuses (Figure 2) reveal trends in terminal digit preferences and aversions by gender. Across the 1992, 2002, and 2012 censuses, the most favoured digits reported were 0, 2, 5, and 8, with the digit 0 consistently exhibiting the highest positive deviation from the expected 10%. Conversely, digits such as 1, 4, 6, 7, and 9 were largely avoided, as indicated by their negative deviations. These patterns reflect prevalent tendencies for age heaping and terminal digit preferences during the early three censuses.

The 2022 census demonstrated a distinct shift from earlier trends. Both sexes avoided the digit 0, previously the most preferred terminal digit. While males also avoided the digit 5, females exhibited a preference for it, suggesting divergent gender-specific reporting behaviours. Males showed unusual preferences for digits 3 and 4 in the 1992 census, a trend not replicated in subsequent censuses. Furthermore, when combining male and female data, the digit 3 was notably preferred in 1992, with a deviation of 5.82, but this preference decreased to 0.50

in 2012. Similarly, digit 7 was preferred in both 1992 and 2012, while digit 6 showed a singular preference in 1992.

Overall, the Myers' Blended Index for the entire population revealed a general improvement in age reporting accuracy across the first three censuses. The index decreased from 15.45 in 1992 to 4.48 in 2002 and further to 3.57 in 2012, signifying reduced age heaping over time. However, a reversal occurred in 2022, with the index rising to 8.21, indicating a resurgence in age heaping behaviours.

Gender-specific trends in the Myers' Index highlighted notable disparities. For males, the index exhibited a consistent decline across all four censuses, dropping from 32.17 in 1992 to a low of 3.35 in 2022. This pattern reflects a substantial improvement in data quality for male respondents. For females, the trend was less linear. The index increased from 5.35 in 1992 to 9.06 in 2002 and peaked at 19.7 in 2012 before slightly declining to 17.65 in 2022. These results underscore persistent challenges in improving age reporting accuracy among female respondents, likely influenced by socio-cultural factors and educational disparities.

While the overall data quality improved over the decades, the observed disparities between male and female age reporting highlight the need for targeted interventions to address persistent gender biases in data collection. The introduction of Computer-Assisted Personal Interviewing (CAPI) in the 2022 census, although transformative, requires further refinement to mitigate residual errors and improve data reliability in future censuses.

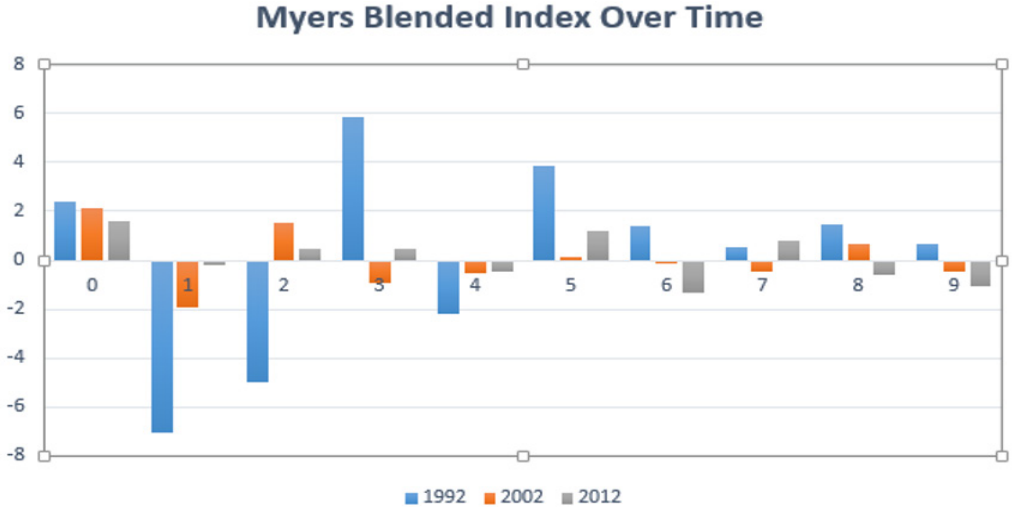


Figure 2: Distribution of age's last digits for the age data and Myers Blended Index over time in Zimbabwe, 1992, 2002 2012 and 2022 censuses

NB: +ve deviations indicate preferred digits and -ve deviation avoided digits

Figure 3 provides an overview of the Age Ratio Scores (ASRs) for males and females, Sex Ratio Scores (SRS), and the Age-Sex Accuracy Index (Joint Score) for the 1992, 2002, 2012, and 2022 Zimbabwe population censuses. The data reveal notable trends and shifts in demographic reporting accuracy over time. For females, the Age Ratio Score (ASR) declined steadily from 11.93 in 1992 to 6.8 in 2002, marking a 24% improvement in data accuracy during this period. However, this progress was offset by a sharp increase in the ASR between 2012 and 2022, rising by 53%, indicating a resurgence in reporting inconsistencies. Similarly, the ASR for males demonstrated a continuous decline from 1992 through 2012, reflecting steady improvements in data quality. However, this trend reversed in the 2022 census, with the ASR for males surging to 17.2, a 69% increase from 2012, highlighting significant reporting issues.

The Sex Ratio Score (SRS), which measures the consistency of sex ratios across age groups, showed a marginal decline across the 1992, 2002, and 2012 censuses. This decline suggests gradual improvements in gender-based age reporting consistency during these years. However, the SRS sharply increased to 7.18 in 2022, reversing prior gains and indicating renewed discrepancies in sex ratio reporting.

The Joint Score, which combines ASRs and SRS to evaluate overall age-sex data quality, dropped consistently from 40.0 in 1992 to 34.5 in 2002, a 17.4% improvement. This downward trend continued with a further 6.4% reduction to 32.3 in 2012, indicating progressively better data accuracy during this period. However, in 2022, the Joint Score increased significantly to 55.4, a 41% rise, underscoring substantial deficiencies in age-sex reporting accuracy.

These results reflect a mixed trajectory in the quality of demographic data over the four censuses. While earlier improvements in ASRs, SRS, and Joint Scores pointed to enhanced age-sex data reliability, the dramatic increases in these metrics in 2022 highlight persistent challenges. The Joint Scores for the 1992 and 2022 censuses suggest that these datasets contain significant deficiencies, necessitating careful adjustments and interpretation before use. Conversely, the 2002 and 2012 censuses exhibit relatively higher data quality, indicating that age-sex data from these periods are usable with minor corrections.

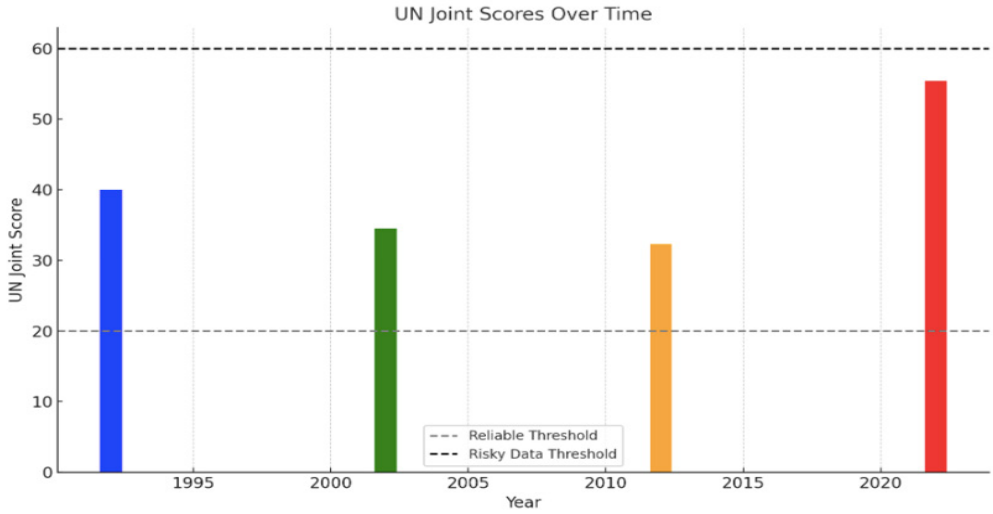


Figure 3: Age Ratio Scores, Sex Ratio Scores, and Joint Scores from 1992, 2002 and 2012 Zimbabwe Population Censuses

Discussion

This study evaluated age-sex data from Zimbabwe's 1992, 2002, 2012, and 2022 censuses to assess the extent of digit preference and avoidance using Whipple's Index, Myers' Blended Index, and the UN Joint Score. The findings highlight both persistent challenges and notable improvements in data quality over the decades, with the 2022 census showing significant advancements due to technological innovations.

The Whipple's Index results for the 1992, 2002, and 2012 censuses, which ranged from 110 to 124, underscore approximate data quality, reflecting notable age heaping and digit preferences. These findings align with studies from other African nations, such as Zambia, where similar patterns of digit preference were observed (Bwalya et al., 2015). However, the 2022 census marked a turning point, with Whipple's Index values falling below 95, indicative of highly accurate data. This improvement can largely be attributed to the adoption of Computer-Assisted Personal Interviewing (CAPI), which integrates real-time data validation, logical consistency checks, and enhanced skip patterns to minimise human errors during data collection (Sainsbury, Ditch, & Hutton, 1993; UNICEF, 2022). This technological transition from Pen-and-Paper Interviewing (PAPI) highlights the transformative role of digital tools in improving data quality.

Consistent with prior research, the study found significant digit preference for terminal digits 0 and 5 across the 1992, 2002, and 2012 censuses, with females exhibiting a slightly higher tendency for age heaping. Interestingly, this deviates from global findings, which often suggest that women report their ages more accurately than men (Amos & Stones, 2017; Fayehun & Ajayi, 2019). The discrepancy in Zimbabwe could be influenced by socio-cultural and patriarchal factors. Female respondents may adjust their ages to align with societal norms, such as expectations around marriage or childbearing. Moreover, in Zimbabwe's predominantly male-headed households, where men account for 64% of household heads (ZIMSTAT, 2018), male respondents often report the ages of female household members. This practice could lead to inaccuracies due to incomplete or biased knowledge of women's exact ages. Notably, evidence indicates that men frequently misreport their spouses' ages, either overestimating or underestimating them depending on the spousal age gap (Pullum & Staveteig, 2017). In contexts like Zimbabwe, where child marriage remains prevalent—approximately one-third of girls under 18 are married (UNICEF, 2022)—these reporting biases can further distort age distribution data.

Myers' Blended Index revealed recurring patterns of digit preference across the censuses, particularly for terminal digits 0, 2, 5, and 8, with significant avoidance of 1, 3, 7, and 9. This aligns with established findings in African demographic surveys, where preferences for even digits often dominate (Nagi et al., 1973). The gradual decline in Myers' Index for males from 1992 to 2012 reflects improved reporting habits, potentially driven by increasing literacy and awareness. However, the 2012 census saw a slight uptick for females, suggesting that gender-specific challenges in accurate reporting persist. In the 2022 census, notable changes were observed: both genders avoided terminal digit 0, while males avoided digit 5, which was still preferred by females. These shifts could be attributed to the introduction of CAPI and evolving societal norms, which may have influenced age reporting practices.

The UN Joint Score, giving a composite measure of age and sex reporting quality, also illustrates these trends. Scores for the 1992 and 2022 censuses (40.0 and 55.4, respectively) highlight significant deficiencies in data quality, necessitating adjustments before use. In contrast, the 2002 and 2012 censuses yielded scores of 34.5 and 32.3, respectively, indicating relatively higher data reliability. These findings echo those of Fayehun and Ajayi (2019) and Malik

and Malik (2021), who observed that African census data often require post-collection adjustments due to systemic inaccuracies.

Gender-specific differences in age ratio scores were evident throughout the study period. Between 1992 and 2002, male and female scores declined by 24% and 43%, respectively, demonstrating improved data accuracy during this interval. However, subsequent increases for males in 2012 and 2022 suggest inconsistencies in data capture processes, potentially influenced by methodological shifts and socio-cultural factors. Similarly, sex ratio scores improved marginally between 1992 and 2002 but showed a slight decline in 2012, underscoring the ongoing challenge of achieving gender parity in age reporting.

Policy Implications

The *adoption* of CAPI during the 2022 census demonstrates the critical role of technology in enhancing data quality. This method should be universally adopted for future surveys and censuses in Zimbabwe, supported by continuous investment in digital infrastructure and training. Gender-specific interventions are also essential to address persistent discrepancies in reporting. Community engagement programmes can empower women to accurately report their ages and other demographic information, while efforts to challenge patriarchal norms must be prioritised to enable equitable participation in data collection processes.

Comprehensive enumerator training remains paramount. Beyond technical skills, training should address cultural sensitivities, focusing on the socio-cultural dynamics that influence reporting behaviour. Enumerators should be equipped with strategies to validate data and ensure consistency while respecting respondents' dignity and cultural norms.

Study Limitations and Future Directions

This study excluded the 1982 census due to incomplete coverage, limiting its capacity to provide a comprehensive longitudinal analysis of age reporting trends. Additionally, reliance on retrospective data introduces potential biases linked to changes in enumeration methods and socio-cultural contexts over time. Future research should explore the underlying socio-cultural determinants of age misreporting in Zimbabwe, particularly the influence of patriarchal norms and gendered reporting behaviours. Comparative studies across African nations

that have adopted CAPI could further illuminate best practices for scaling its implementation and addressing region-specific challenges in data collection.

The findings from this study reveal a mixed trajectory of improvements and persistent challenges in age reporting accuracy in Zimbabwe's censuses. While the 2022 census marked significant progress due to technological advancements, gender-specific disparities and socio-cultural influences remain critical areas for intervention. By addressing these issues through targeted policies, Zimbabwe can enhance the reliability of its demographic data, ensuring a solid foundation for socio-economic planning and development.

Thus, this study assessed the quality of age-sex data from Zimbabwe's four censuses since 1980, with a focus on digit preference and avoidance. The findings indicated a trajectory of gradual improvement, culminating in significant advancements in the 2022 census. The introduction of Computer-Assisted Personal Interviewing (CAPI) played a pivotal role in reducing age heaping and ensuring more accurate reporting. Despite these improvements, persistent challenges related to gender disparities and socio-cultural influences remain, necessitating targeted interventions for future censuses.

Whipple's and Myers' indices revealed a decline in digit preference and age heaping across the four censuses, with the most dramatic reduction in the 2022 census. Whipple's Index values, which were indicative of "approximate" data quality in earlier censuses, fell below 95 in 2022, demonstrating high accuracy. This improvement highlighted the impact of CAPI, which introduced real-time data validation, logic checks, and skip patterns during interviews, thereby minimising human error and bias. The importance of technology in enhancing census accuracy cannot be overstated, aligning with findings from similar studies in developing countries (Fayehun & Ajayi, 2019; Malik & Malik, 2021).

However, persistent gender disparities were evident in the results. Across the censuses, women exhibited higher levels of digit preference than men, with end-digits 0 and 5 being most favoured. This finding diverges from global studies suggesting that women generally report their ages more accurately than men (Amos & Stones, 2017). In Zimbabwe, the socio-cultural factors, particularly patriarchal norms, may contribute to this discrepancy. In many cases, men, as household heads, report the ages of female members, leading to inaccuracies due to limited knowledge or biases. Additionally, societal expectations around life events, such as marriage and childbearing, may lead women to report

ages aligning with these norms, further exacerbating inaccuracies (Pullum & Staveteig, 2017; UNICEF, 2022).

Myers' Blended Index supported these findings, showing consistent digit preferences across censuses, particularly for terminal digits 0, 2, 5, and 8. The 2022 census showed a unique pattern: while the terminal digit 0 was previously favoured, it was avoided in 2022, with males avoiding digit 5 but females continuing to prefer it. These changes may reflect evolving societal norms or the influence of improved census methodologies. Despite the overall decline in Myers' Index values for males across the censuses, females experienced a notable increase in 2012, suggesting persistent challenges in gender-specific reporting accuracy.

The UN Joint Score provided an integrated measure of age and sex reporting accuracy. The 1992 and 2022 scores indicated significant deficiencies, requiring adjustments before use, while the 2002 and 2012 scores were indicative of more reliable data. These results are consistent with broader trends in African censuses, where systemic inaccuracies often necessitate post-collection adjustments (Fayehun & Ajayi, 2019; Malik & Malik, 2021). The findings underscore the need for ongoing efforts to standardise methodologies and address the socio-cultural factors influencing data quality.

Age ratio and sex ratio scores further highlighted gender-specific differences in reporting accuracy. Between 1992 and 2002, scores for both males and females improved significantly, reflecting better data capture practices. However, subsequent increases in male scores in 2012 and 2022 suggest inconsistencies, possibly linked to methodological shifts or socio-cultural factors. These findings emphasise the need for targeted interventions to address gender disparities in reporting and improve the overall reliability of census data.

Recommendations

To address the identified challenges, several recommendations are proposed. Firstly, comprehensive training for enumerators is essential. Training programmes should focus on accurate data collection techniques, cultural sensitivities, and resolving inconsistent age responses. Scenario-based training can help enumerators navigate real-world challenges effectively, ensuring consistency and reliability in data collection.

The integration of technology may be further expanded. Building on the success of CAPI, future censuses should incorporate advanced features,

such as real-time logic checks and electronic data verification systems. These enhancements can reduce errors and standardise reporting practices across diverse demographic groups. Additionally, continuous investment in digital infrastructure and technical support is critical to sustain the benefits of technological advancements.

Gender-specific interventions are vital to address persistent disparities. Awareness campaigns targeting women in rural and patriarchal households can empower them to report their ages accurately. Encouraging self-reporting by women, rather than relying on male household heads, can help mitigate inaccuracies. These efforts should be complemented by community engagement initiatives that challenge patriarchal norms and promote equitable participation in data collection processes.

Policymakers should incorporate these findings into national census frameworks, ensuring that methodologies for addressing age heaping and digit preference are standardised. Aligning these methodologies with broader socio-economic and development goals can enhance the utility of census data for planning and policymaking.

Future Research

Future research may explore the socio-economic determinants of age reporting discrepancies, particularly the role of literacy, education, and household dynamics. Comparative studies across African nations can also provide valuable insights into best practices and contextual factors affecting data quality. Additionally, longitudinal studies assessing the scalability and sustained impact of CAPI can inform its broader implementation in census practices. Examining urban versus rural disparities in age reporting can further refine interventions, tailoring them to the unique challenges of each demographic group.

Conclusion

This study provides a comprehensive assessment of the quality of age-sex data in Zimbabwe's censuses, highlighting significant improvements over time. The 2022 census represents a milestone in data accuracy, largely driven by the adoption of CAPI. However, persistent gender disparities and socio-cultural influences underline the need for targeted interventions and ongoing methodological enhancements. By addressing these challenges and implementing the proposed recommendations, Zimbabwe can continue to improve the reliability of its demographic data, ensuring a robust foundation

for evidence-based policy planning and socio-economic development. The findings also contribute to the broader discourse on demographic data quality in resource-constrained settings, offering valuable lessons for other nations seeking to enhance their census practices.

Competing Interests

The authors have no conflicts of interest to declare

Acknowledgments

The authors extend their sincere gratitude to ZIMSTAT for granting them permission to access the data.

References

Agrawal, Gopal, and Puneet Khanduja. 2015. "Influence of Literacy on India's Tendency of Age Misreporting: Evidence from Census 2011." *Journal of Population and Social Studies* 23. doi: 10.14456/jpss.2015.1.

A'Hearn, Brian, and Joerg Baten. 2009. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *The Journal Of Economic History* 69(3):783–808.

Bekele, Selome. 2006. "ANALYSIS ON THE QUALITY OF AGE AND SEX DATA COLLECTED IN THE TWO." 29(2):123–32.

Bello, Yusuf. 2012. "Error Detection in Outpatients' Age Data Using Demographic Techniques." 10(1):27–36.

Blum, Mathias, and Karl-Peter Krauss. 2018. "Age Heaping and Numeracy: Looking behind the Curtain." *Economic History Review* 71(2):464–79.

Bwalya, Bupe, Million Phiri, and Cynthia Mwansa. 2015. "Digit Preference And Its Implications On Population Projections In Zambia : Evidence From The Census Data Research Article Digit Preference And Its Implications On Population Projections In Zambia :" *International Journal of Current Advanced Research* 4(5):92–97.

Fayehun, Olufunke, and Anthony I. Ajayi. 2019. "Age Heaping among Adults in Nigeria : Evidence from the Nigeria Demographic and Health Surveys 2003 – 2013." 1–8. doi: 10.1017/S0021932019000348.

Helleringer, Stephane, Chong You, Laurence Fleury, Laetitia Douillot, Insa Diouf, Cheikh Tidiane Ndiaye, Valerie Delaunay, and Rene Vidal. 2019. "Improving Age Measurement in Low- and Middle-Income Countries through Computer Vision: A Test in Senegal." *Demographic Research* 40:219–60. doi: 10.4054/DemRes.2019.40.9.

Malik, Manzoor Ahmad, and Manzoor Ahmad Malik. 2021. "Communications in Statistics : Case Studies , Data Analysis and Applications Age Heaping Pattern and Data Quality : Evidence from Indian Household Survey Data (1991 – 2016) Household Survey Data (1991 – 2016)." *Multilevel Dimension Reduction for Credit Scoring Modelling and Prediction: Empirical Evidence for Greece* 0(0):1–20. doi: 10.1080/23737484.2021.1952492.

Mba, Chuks J. 2003. "Assessing the Reliability of the 1986 and 1996 Lesotho Census Data." *Journal of Social Development in Africa* 18(1):111–28. doi: 10.4314/jsda.v18i1.23821.

Myers, R. J. 2012. "Accuracy of Age Reporting in the 1950 United States Census." *Journal of the American Statistical Association* 49(268):826–31.

Nagi, M. H., E. G. Stockwell, and L. M. Snavley. 1973. "Digit Preference and Avoidance in the Age Statistics of Some Recent African Censuses: Some Patterns and Correlates." *International Statistical Review / Revue Internationale de Statistique* 41(2):165. doi: 10.2307/1402833.

Nations, United. 1982. "Provisional Guidelines on Standard International Age Classifications." *Department of International Economic and Social Affairs* 74(M):32.

Palamuleni, Martin E. 1995. "Age Misreporting in Malawian Censuses and Sample Surveys : An Application of the United Nations ' Joint Age and Sex Score Author (s): Martin E . Palamuleni Source : Southern African Journal of Demography , October 1995, Vol . 5, No . 1 (October Publi." 5(1):11–17.

Pullum, T., and S. Staveteig. 2017. *An Assessment of the Qulity Nd Consistency of Age Nd Te Reporting in Uveys, 2000-2015*. *DHS METHODOLOGICAL*. 19. Rockville, Maryland, USA: ICF 2 The DHS Program.

Randall, S., and E. Coast. 2016. "'The Quality of Demographic Data on Older Africans.' Demographic Research''." *Demographic Research* 34((1)):143-74.

Shipanga, Immanuel, and Tobias Shinyemba. 2023. "Evaluation of Age-Sex Data Collected in the 1991, 2001 and 2011 Population and Housing Censuses of Namibia." *Asian Journal of Population Sciences* 2(January):1–16. doi: 10.3126/ajps.v2i1.51087.

Singh, M., Kashyap, G., & Bango, M. , 54(4), 725-734. doi:10.1017/S0021932021000249. 2022. "Age Heaping among Individuals in Selected South Asian Countries: Evidence from Demographic and Health Surveys." *Journal of Biosocial Science* Age heapin(54(4)):725–34.

Singh, M., G., Chandra K., and Madh. Bango. 2022. "Age Heaping among Individuals in Selected South Asian Countries: Evidence from Demographic and Health Surveys." *Journal of Biosocial Science* 54(4):725–34. doi: 10.1017/S0021932021000249 .

Smeets, J. B. J. 2024. "The Bias and Precision of Reporting the Average Age of Human Participants." *Acta Psychologica* 249:104457. doi: 10.1016/j.actpsy.2024.104457.

Szoltysek, M., Radosław P., and Siegfried G. 2017. "Age Heaping Patterns in Mosaic Data." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51:1–26. doi: 10.1080/01615440.2017.1393359.

Tollneck, F. and Joerg, B. 2015. "Age-Heaping-Based Human Capital Estimates." Pp. 131–54 in *Handbook for Cliometrics*.

UNESCO. 2017. *Estimation of the Numbers and Rates of Out-of-School Children and Adolescents Using Administrative and Household Survey Data*.

Yadav, Akhilesh, Minakshi Vishwakarma, and Shekhar Chauhan. 2020. "The Quality of Age Data : Comparison between Two Recent Indian Censuses." *Clinical Epidemiology and Global Health* 8(2):371–76. doi: 10.1016/j.cegh.2019.09.005.