RELAP - Revista Latinoamericana de Población
Vol. 20: e202568
http://doi.org/10.31406/relap2026.v20.e202568
ISSN digital: 2393-6401

Revisiting the method for assessing the quality of centenarian census data*

Revisión de una propuesta de evaluación de la calidad de la recopilación de datos censales de las personas centenarias

Neir Antunes Paes
antunes@de.ufpb.br
ORCID: https://orcid.org/0000-0002-0927-9103
Universidade Federal da Paraíba, Brasil

Rozane Pereira de Sousa
enfermeirarozane@gmail.com
ORCID: https://orcid.org/0000-0002-2619-8161
Universidade Federal da Paraíba, Brasil

Vivian Conceição Alves Leite Pereira do Lago
vivianlago@hotmail.com
ORCID: https://orcid.org/0000-0002-3047-3605
Universidade Federal da Paraíba, Brasil

Abstract

Centenarians are one of the fastest-growing age groups globally. However, collecting data from this population poses various challenges, particularly regarding the quality of census data. In 2023, Medeiros and Paes proposed a method to evaluate the quality of centenarian counting worldwide based on national censuses (2010, 2011). The present study revisited this method, updating it for 87 countries with data from 2020 to 2023. The study calculated three indicators for each of these countries and then obtained a new range of variations for each indicator. Finally, the countries were classified according to this review. The study findings reveal heterogeneous

Keywords
Centenarians
Data Reliability
Census
Data Classification

* The authors thank the CNPq for the funding. A previous version of this paper was presented at the 21st Conference of the Applied Stochastic Models and Data Analysis International Society (ASMDA, 2025).

Esta obra está bajo una licencia de Creative Commons Reconocimiento-No Comercial 4.0 Internacional

patterns in the quality and composition of centenarian data across countries, categorized as: good, intermediate, and low quality. The reviewed and updated method uses a few indicators that are generally available in every country's population counting records.

Resumen

Los centenarios son uno de los grupos de edad de más rápido crecimiento a nivel mundial. Sin embargo, surgen varios desafíos al recopilar datos de esta población, particularmente en relación con la calidad de los datos censales. En 2023, Medeiros y Paes propusieron un método para evaluar la calidad global del recuento de centenarios basado en los censos de los países (2010, 2011). El presente estudio consiste en revisitar este método, actualizándolo para 87 países con datos comprendidos entre 2020 y 2023. Para cada uno de los países, se calcularon tres indicadores. Luego, se obtuvo un nuevo rango de variaciones para cada uno de estos indicadores. Por último, los países se clasificaron de acuerdo con esta revisión. Los hallazgos de este estudio revelan patrones heterogéneos en la calidad y la composición de los datos de centenarios entre los países, los cuales se representan en tres categorías: calidad buena, intermedia y baja. La propuesta revisada y actualizada utiliza pocos indicadores que, por lo general, están disponibles en los registros de recuento poblacional en todos los países.

Palabras clave

Centenarios
Fiabilidad de los datos
Censo
Clasificación de datos

Enviado: 02/12/25
Aceptado: 28/05/26

Introduction

Older adults form a very diverse group, and centenarians, as part of this group, are no exception. It is crucial to have specific and disaggregated data on them to fully understand the challenges affecting their health and well-being (WHO, 2023).

The relevance of such data to understanding longevity is undeniable, which is why the investigation of very old individuals has gained increasing attention. However, collecting information on this age group poses considerable challenges. Researchers face numerous barriers to obtaining reliable and complete data on this growing population group (Zaccaria et al., 2025).

National censuses, considered the largest generators of demographic information and conducted in virtually every country, are fundamental in this context. Despite the great logistical and technical complexity involved in their execution, there is a constant concern with minimizing errors in data collection (Paes, 2018).

The accuracy of these data is frequently questioned. Schmertmann, Queiroz, and Gonzaga (2024) demonstrate that errors from death under-registration or census undercounts cause much greater biases than age misreporting. This contributes to biases in life expectancy estimates, for instance, hindering demographic analyses. This problem is exacerbated in less developed countries, where the quality of age data is an even more critical challenge.

Discussions raised by Gavrilov and Gavrilova (2019) and Newman (2024) reveal that current information on the quality of age declaration at extremely advanced ages is rarely ideal, even in nations with good vital statistics. It should be noted that estimating the hazard rate at very advanced ages is difficult due to the very small number of survivors at these ages and the misreporting of age by older people, given that data errors tend to accumulate with age.

Thus, Medeiros and Paes (2022) proposed an innovative methodology to classify the quality of census data for centenarians. Their systematic approach is a key differentiator, as it uses indicators to generate a quality score, enabling comparisons of data reliability across different censuses or regions. This systematic approach to quality assessment is a differentiator. They recognize that censuses, despite their operational rigor, are susceptible to age misreporting, which directly affects the count of individuals at very advanced ages.

The census data on centenarians that served as the basis for developing the original method covered the period 2000-2009. Since population dynamics may change over time, particularly due to the ongoing aging process in most of the countries included in the database, it was deemed appropriate to update the proposal by incorporating the most recent census data available.

Therefore, this study aimed to revisit the method proposed by Medeiros and Paes (2022) to evaluate the quality of census data for centenarian counts.

Methods

This is an ecological study based on secondary data of individuals aged 100 or older from 87 countries. It used the method originally developed by Medeiros and Paes in 2022, which was reviewed and updated in 2023 – i.e., this study is the second review of that method.

This updated review prioritized data from each country's most recent available national census, conducted mostly between 2020 and 2023. In cases where the most recent census data was not accessible, the most up-to-date population estimate was considered to ensure the information was current.

The original method and first review included countries based on the first three predefined criteria below. This second review added a fourth criterion to improve data representativeness and comparability of the data. Hence, the country should have:

1. Specific information on the number of centenarians registered, as well as the population aged 85 or over, for both men and women.
2. At least one male and female centenarian registered in the last census.
3. At least 100 thousand inhabitants.
4. Political-administrative-economic autonomy – i.e., an independent country.

According to the original model's methodology, 87 out of 193 countries with data available from UNdata (2025) met all four inclusion criteria. The study calculated three indicators regarding the centenarian population of each selected country, which are essential for data quality assessment:

1. Centenarians' sex ratio (SR) (total number of centenarian men divided by total number of centenarian women).

2. Ratio of men aged 100 or over in relation to men aged 85 or over, multiplied by 100 [P_m(100+/85+)].
3. Ratio of women aged 100 or over in relation to women aged 85 or over, multiplied by 100 [P_f(100+/85+)].

Next, the relationship between the Human Development Index (HDI) and life expectancy at birth (e^0) was investigated using the three indicators related to centenarians. This association was tested by applying a simple linear regression model (n=87), which showed a significant association ($p< 0.001$; Table 1). That is, the HDI is a significant predictor of data indicators. The negative Beta coefficients and their respective 95 % CIs confirm that an increase in socioeconomic development is associated with the three indicators.

Table 1. Coefficient of Determination and simple linear regression indicators between the Human Development Index (HDI) and indicators derived from census data of centenarians across the 87 selected countries.

Centenarian Indicators	n	R^2	β	95 % CI	Standard Error	P-Value
Sex Ratio (SR)	87	0.3217	-1.0776	(-1.42; -0.74)	0.1697	<0.001
Ratio Cent./85+ Masc (Pm)	87	0.3646	-14.3837	(-18.47;-10,29)	2.0596	<0.001
Ratio Cent./85+ Fem (Pf)	87	0.4191	-16.0521	(-20.13;-11.98)	2.0499	<0.001

Note: R^2: Coefficient of determination; β: Regression coefficient; 95 % CI: 95 % Confidence Interval. Analysis conducted using RStudio software.

Source: The authors based on data from United Nations (UNSD, 2025).

The study also applied the multivariate classification technique of Ward's clustering method (1963) to deepen the analysis of data quality. The 87 countries were grouped into four clusters based on five indicators: the three centenarian indicators, the HDI, and e^0. The clusters were categorized as follows:

- Cluster 1: Countries with high HDI and e^0, representing a "high data quality standard".
- Clusters 2 & 3: Countries with intermediate HDI and e^0, characterizing an "intermediate data quality standard".
- Cluster 4: Countries with low HDI and e^0, indicating a "low data quality standard".

Weights were assigned to each type of Cluster: Cluster 1 (weight 3), Clusters 2 and 3 (weight 2), and Cluster 4 (weight 1). A range of variation was constructed for SR, $P_m(100^+/85^+)$, and $P_f(100^+/85^+)$, based on the values obtained per country belonging to the same cluster. The ranges for the three indicators in each cluster from the original proposal are shown in Table 3. For instance, if a country has an SR value ranging from 0.138 to 0.353, it will receive a weight of 3.

The final data quality for centenarians was then classified based on the total score (sum of the weights), using a scale that ranged from very good to very low, considering the three indicators: SR, $P_m(100^+/85^+)$, and $P_f(100^+/85^+)$. Table 2 summarizes the final classification of weights.

Table 2. Classification of scores to assess the quality of census counting of centenarians, according to the weights assigned to the selected indicators.

Classification	Sum of Indicator Weights [SR, Pm(100+/85+), Pf(100+/85+)]
Very Good (VG)	9
Good (G)	8 or 7
Moderate (M)	6
Low (L)	5 or 4
Very Low (VL)	3

Source: Medeiros, & Paes (2022).

The central hypothesis was that the quality of census data concerning centenarians and the population aged 85 and over tends to be better in countries with higher levels of development, represented by the HDI and e^0.

This hypothesis is grounded in the institutional robustness of Civil Registration and Vital Statistics (CRVS) systems. In high-development nations, demographic reliability is not a coincidental byproduct but the result of sustained investment in the digitalization of registry offices and the rigorous training of health professionals in certifying causes of death. Such institutional maturity significantly reduces classification errors and omissions (Institute for Health Metrics and Evaluation, 2025; Mikkelsen et al., 2015, 2023; Newman, 2024; Phillips et al., 2015).

Furthermore, to ensure the external validity of the identified clusters beyond institutional factors, a cross-validation analysis was performed

using Life Expectancy at Age 65 (e^{65}) as an independent biological benchmark (Table 3 and Figure 1). For the 87 countries included in the final selection, (e^{65}) data were obtained from the United Nations, Department of Economic and Social Affairs, Population Division, World Population Prospects 2024, ensuring temporal alignment with the analyzed census periods. The statistical robustness of this validation was tested through a one-way Analysis of Variance (ANOVA), using the F-test to determine whether statistically significant differences existed between the mean (e^{65}) values across the various data quality clusters. To complement this analysis, a graphical representation (Figure 1, Panel B) was included, showing the mean life expectancy at age 65 for each cluster along with 95 % confidence intervals (95 % CI). This visualization facilitates a direct comparison of the precision and distribution of e^{65} across the quality groups, further justifying the consistency of the classification.

Additionally, Tukey's Honest Significant Difference (HSD) test was applied for post-hoc multiple comparisons, allowing for the identification of which specific pairs of clusters exhibit significant distinctions. This step was essential to confirm that the classification of centenarian data quality remains consistent with internationally established mortality and longevity indicators (Figure 1, Panel A).

Various other technical mechanisms for evaluating the quality of age data in censuses and surveys can be found in the literature, most notably the Whipple and Myers indices. However, these were designed to detect age heaping or digital preference by measuring the accuracy of age reporting, which indicates flaws in data collection. Whipple focuses on digits 0 and 5, while Myers evaluates all digits. Both indicators essentially assess age. However, in most developing countries, there is a problem with the availability of census statistics for individuals over 100 years old, whereas the proposal presented in this article specifically assesses the centenarian population as a whole (100+) and not the quality of individual age groups above 100 years.

The database was developed using Microsoft Office Excel. Data analysis used the free statistical software R Studio version 4.3.3. Anonymized data were obtained from a secondary source and used in an aggregated manner, so review by a Research Ethics Committee was not required.

Results and discussion

Statistical validation of data quality clusters

Before examining the demographic implications, it was statistically evaluated whether the clusters differ regarding life expectancy at age 65 (e^{65}). A one-way ANOVA indicated substantial differences between groups ($F = 109.3$, $p < 0.001$). The effect size is large ($\eta^2 \approx 0.72$), suggesting that approximately 72 % of the variance in e^{65} across the 87 countries is explained by the cluster classification (Table 3).

Table 3. Analysis of variance for life expectancy at age 65 across clusters.

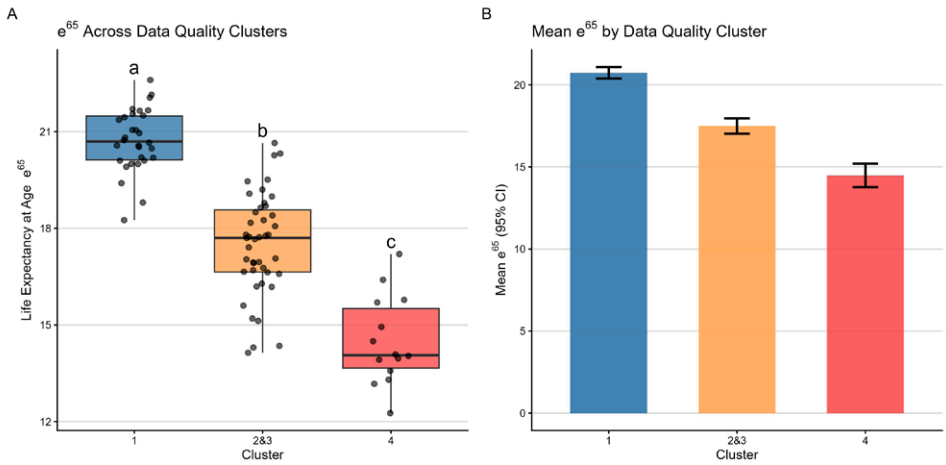
Source of variation	Df	Sum Sq	Mean Sq	F-Value	p-value	η2
Between clusters	2	406.7	203.35	109.3	<0.001	0.72
Within clusters (Residuals)	84	156.3	1.86	-	-	
Total	86	563.0	-	-	-	-

Source: The authors based on data from United Nations, Department of Economic and Social Affairs, Population Division (UNDESA-PD, 2024).

Post-hoc comparisons using the Tukey's Honestly Significant Difference Test confirm that all pairwise comparisons between clusters are statistically significant ($p < 0.001$). Figure 1 illustrates the distribution of life expectancy at age 65 across the identified clusters. Panel A presents the distribution through boxplots, where different letters above each group indicate statistically significant differences according to Tukey's HSD test ($p < 0.05$). Panel B displays the mean e^{65} values with their respective 95 % CI. Together, these visualizations provide both distributional and central tendency perspectives on cluster differences.

The results from Panel A reveal a pronounced gradient in life expectancy at age 65 across clusters. The approximately seven-year difference in e^{65} between Cluster 1 and Cluster 4 highlights a significant cross-national gap in late-life survival. Panel B reinforces the robustness of the results by demonstrating that the mean e^{65} values for each cluster are estimated with high precision. This is evidenced by the 95 % Confidence Intervals (95 % CI); while Clusters 1 and 2 & 3 exhibit extremely narrow intervals, Cluster 4 shows a slightly wider amplitude. This lower relative precision in Cluster 4 stems from its smaller sample size compared to the other groupings, which increases the standard error of the estimates.

Figure 1. Distribution and mean life expectancy at age 65 (e^{65}) by data quality clusters.



Note: R Studio was used.

Source: The authors based on data from United Nations, Department of Economic and Social Affairs, Population Division (UNDESA-PD, 2024).

Nevertheless, despite this greater dispersion, the mean e^{65} for Cluster 4 remains statistically lower and distinct from the others, aligning with the profile of regions that face greater challenges regarding the stability and quality of demographic data for the elderly and centenarians.

Comparison of quality indicators for counting centenarians in the three proposals (Original proposal, Review 1, and Review 2)

Review 2 selected 87 countries that met the four criteria – 29 in Cluster 1, 44 in Cluster 2 & 3, and 14 in Cluster 4, whose ranges of variation for the three indicators are detailed in Table 4.

These ranges were very close to those found in the two previous versions. Regarding Cluster 4, very high values were found for Vanuatu and India, and an even higher value was found for the Ivory Coast: 16.37 for P_m(100+/85+) and 16.13 for P_f(100+/85+). These values were well above the maximum limits of the ranges assigned for P_m and P_f in both the Original Proposal (7.607 and 9.514) and Review 1 (7.595 and 9.182). Since countries could have values close to or even higher than these, it was decided not to set a definitive upper limit for the ranges of these two indicators regarding the countries in Cluster 4.

Table 4. Reference intervals of selected quality indicators regarding centenarian census data of countries in the Original Proposal and Reviews 1 and 2.

Quality of Census Data on Centenarians	Sex Ratio	Pm (100+/85+)*	Pf (100+/85+)**	Weight
High (Cluster 1)				3
Original Proposal	0.138; 0.353	0.203; 0.762	0.536; 1.405	
Review 1	0.173; 0.322	0.214; 0.713	0.541; 1.347	
Review 2***	0.140; 0.385	0.278; 0.743	0.593; 1.691	
Low (Cluster 4)				1
Original Proposal	0.588; 1.102	2.410; 7.607	2.952; 9.514	
Review 1	0.591; 1.103	2.518; 7.595	2.967; 9.182	
Review 2***	0.524; 1.122	≥3.076	≥2.936	
Intermediate (Clusters 2&3)				2
Original Proposal	Outside other ranges	Outside other ranges	Outside other ranges	
Review 1	Outside other ranges	Outside other ranges	Outside other ranges	
Review 2***	Outside other ranges	Outside other ranges	Outside other ranges	

*Note: *P_m(100+/85+): Ratio of male population aged 100 years and older to population aged 85 years and older; **P_f(100+/85+): Ratio of female population aged 100 years and older to population aged 85 years and older. ***Excluded from the sample: Cluster 1 = (SR: Malta, Iceland, Singapore, Cyprus, and Israel), (P_m: Malta, Cyprus, Singapore, and Israel), and (P_f: Singapore and Israel); Cluster 4 = (SR: Kenya, Zimbabwe, Malawi, and Guatemala), (P_m: Bhutan, Bangladesh, Guatemala), and (P_f: Bhutan).*

Source: UNSD (2025); Original Proposal: Medeiros, & Paes (2022); Review 1: Medeiros, & Paes (2025). Review 2: Basic Data (UNSD, 2025).

The SR limits ranged from 0.140 to 0.385 to define reference intervals for the 29 countries in Cluster 1. However, the magnitudes for four of them (Israel, Iceland, Singapore, and Malta) deviated significantly from the others. Therefore, they were removed from Cluster 1. SR and P_m were very low in Malta compared to the other countries, with a significant number of centenarian women, but much higher in the other three countries than in the rest of Cluster 1.

Some studies, such as those by Bell and Miller (2005), Meyer (2012), and Lozada-Martínez et al. (2024) suggest that the survival rate among centenarians usually does not exceed 0.25. However, nine countries out of the 29 in Cluster 1 exceeded this rate: Germany, Australia, the United States, Ireland, Cyprus, Greece, Iceland, Singapore, and Israel, whose data are from 2020 to 2023 (Table 6). Overall, 55 of the 87 countries surveyed had a survival rate > 0.25 – i.e., about 63 %.

This behavior may be associated with a possible shift in the specific pattern of extreme longevity, where the proportion of women is traditionally more pronounced. The supposedly atypical SR values for these countries may indicate a trend of more men reaching very advanced ages than women, which suggests that there are elements to assume an increased proportion of male centenarians in various parts of the world, particularly in countries with an advanced demographic transition.

For example, Perls (2017) discusses the increasing proportion of centenarian men in the US in birth cohorts, growing from 14 % to 22 % in the 100-104 age range and from 10 % to 17 % among those aged 105 or older. Another study by Lozada-Martínez et al. (2024) with 2,362,436 people showed a prevalence of 0.12 % of centenarians, with slightly more centenarian women than men (49.3 %). However, the SR trend observed by Perls (2017) is not universal. In some specific regions, such as certain areas of Italy (in the Mediterranean Blue Zones of Sardinia) and Greece (Ikaria), the proportions between centenarian men and women approach one to one, diverging from the more common pattern.

Similarly, some countries with outliers in each indicator had to be excluded to construct the intervals for Cluster 4 (with 14 countries), namely: Kenya, Zimbabwe, Malawi, and Guatemala for SR; Bhutan, Bangladesh, and Guatemala for P_m; and Bhutan for P_f. These outliers made it difficult to define representative intervals for most of the other countries, justifying their removal and the need for a specific analysis of census data on their centenarians. Possible causes of these outliers include misreported data, problems with civil registration, overestimated age, failures in data entry into the systems, migration issues, and variations in survival rates.

Furthermore, the presence of these outliers may be related to age coding errors, a well-documented phenomenon in the literature, which tends to accumulate at advanced ages and artificially inflate the frequency at certain ages (Newman, 2024, 2018).

The ranges of the indicators in Clusters 2 & 3 are specified so that their values fall outside the ranges of Clusters 1 and 4. Table 5 details the descriptive statistics of the selected indicators for the 85 countries from the Original Proposal and Review 1, as well as the 87 countries from the Review 2 analysis.

Table 5. Mean age, standard deviation, minimum and maximum value of selected indicators, Human Development Index, and life expectancy at birth of selected countries in the Original Proposal (2010-2011), Review 1 (2021-2022), and Review 2 (2025).

Measure	SR			Pm(100+/85+)			Pf(100+/85+)			HDI	e^0
	Original Proposal	Review 1	Review 2	Original Proposal	Review 1	Review 2	Original Proposal	Review 1	Review 2	2023	
Mean age	0.372	0.316	0.403	1.371	1.285	1.908	1.905	1.787	2.469	0.825	77.1
SD	0.235	0.228	0.237	1.867	1.748	2.890	2.186	2.007	3.011	0.122	5.8
Minimum	0.029	0.047	0.089	0.044	0.037	0.078	0.123	0.197	0.006	0.416	55.1
Maximum	1.103	1.103	1.122	7.607	7.595	16.370	9.514	9.182	16.135	0.972	85.5

Note: SD: standard deviation; SR: sex ratio; P_m: male population; P_f: female population; HDI: Human Development Index; e^0: life expectancy at birth.

Source: Original Proposal: Medeiros, & Paes (2022). Review 1: Medeiros, & Paes (2025). Review 2: UNSD (2025).

The mean SR decreased from the Original Proposal (0.372) to Review 1 (0.316) and then increased in Review 2 (0.403). This suggests a possible change in the ratio of men to women in the analyzed country groups over time, but it may also be related to the change in country selection.

In the Original Proposal, the male population aged 100 or older represented approximately 1.37 % of the male population aged 85 or older, on average. There was a slight decrease in Review 1, and an increase in Review 2 (average: 1.90 %). This may indicate greater longevity within the group of men aged 85+ in the countries analyzed in this review, resulting in a higher percentage reaching 100+ years.

The minimum (0.416) and maximum (0.972) HDI in 2023 reveal the wide range of the sample, encompassing countries with very low to very high levels of development. This indicates a sharp disparity in human development levels around the world, which was further exacerbated by the effects of the COVID-19 pandemic, as pointed out in the 2025 Human Development Report. The latter details how the pandemic declined the global HDI in several countries, negatively impacting health, education, and the economy, which, in turn, hinders recovery and progress towards a global very high HDI scenario projected for 2030 (UNDP, 2025). As observed for the HDI, e^0 followed regional disparities, ranging from 55.1 to 85.1 years.

Descriptive statistics of the indicators in Review 2

Table 6 presents Review 2 data regarding the development characteristics at different extremes of this spectrum, the descriptive statistics of the indicators for the countries in Clusters 1 (with high HDI and e^0 values), 4 (with low HDI and e^0 values), and 2 & 3 (with intermediate values in these indicators).

Table 6. Mean age, standard deviation, minimum and maximum values of the selected indicators, Human Development Index (HDI), and life expectancy at birth (e^0) of the countries in Cluster 1, Cluster 2 & 3, and Cluster 4 in Review 2.

Measure	SR	Pm	Pf	HDI	e^0
			Cluster 1		
Mean age	0.206	0.459	1.105	0.939	82.6
SD	0.052	0.128	0.263	0.021	1.14
Minimum	0.140	0.278	0.593	0.890	79.3
Maximum	0.385	0.743	1.691	0.972	84.7
			Clusters 2&3		
Mean age	0.422	1.358	1.726	0.817	76.407
SD	0.234	1.646	1.779	0.061	2.971
Minimum	0.089	0.078	0.006	0.682	69.80
Maximum	1.004	6.326	8.630	0.915	85.50
			Cluster 4		
Mean age	0.698	7.590	7.892	0.605	67.2
SD	0.199	4.146	3.901	0.084	5.38
Minimum	0.524	3.076	2.936	0.416	55.1
Maximum	1.122	16.370	16.135	0.741	74.7

Note: Excluded: Cluster 1 (SR: Malta, Iceland, Singapore, Cyprus, and Israel); (P_m: Malta, Singapore, Cyprus, and Israel), and (P_f: Singapore and Israel); Cluster 4 (SR: Kenya, Zimbabwe, Malawi, and Guatemala); (P_m: Bhutan, Bangladesh, and Guatemala), and (P_f: Bhutan).

Sources: Original Proposal: Medeiros, & Paes (2022). Review 1: Medeiros, & Paes (2025). Review 2: UNSD (2025).

It is possible to observe a relationship between HDI and e^0, highlighting significant disparities between nations. Analysis of the descriptive data for Cluster 1 (countries with a high level of human development) shows an average e^0 of 82.6 years, reflected in an average HDI of 0.939.

In contrast, Cluster 4 (countries with low levels of development) exhibits an average e^0 of 67.2 years. The discrepancy of 15.4 years compared to Cluster 1 underscores the profound vulnerabilities and challenges to

living conditions these nations generally face. The average HDI of 0.605 suggests systemic deficiencies in crucial areas, such as limited access to health services, high rates of preventable diseases, poor sanitation conditions, and so forth. Low longevity, therefore, is not an isolated phenomenon, but rather a direct reflection of development below that found in countries with better overall living conditions.

Marcus et al. (2023), Aladejare (2023), Mackenbach et al. (2019), and Omokanmi et al. (2022) list some of the main factors that contribute to underdevelopment and low longevity, namely: higher prevalence of malnutrition, inadequate management of external debt, unequal income, and poor management of natural resources. When combined, these factors perpetuate the cycle of underdevelopment and low life expectancy.

Illustrative example

Census data from centenarians of some selected countries representative of the five continents were used to illustrate a practical application of the revised method (Review 2). Table 7 details the weights of the three indicators assigned to each country with the respective final classification of the quality of census data.

Table 7. Rating the quality of census counts of centenarians in selected countries, according to Revision 2.

Country	SR	W*	Pm	W	Pf	W	Sum W	C**
Austrália	0.266	3	0.554	3	1.321	3	9	VG
South Africa	0.528	1	7.136	1	5.668	1	3	VL
Brazil	0.455	2	1.411	2	1.705	2	6	R
Russian Federation	0.270	3	0.295	3	3.171	1	7	G
India	0.914	1	12.277	1	11.668	1	3	VL

Sources: UNDS, 2025.

*Note: *W = weight: assigned according to the reference range corresponding to the calculation value in Table 1. VG = Very Good; G = Good; R = Regular; L = Low; VL = Very Low.*

***C = Classification: presented according to the scale proposed by Medeiros and Paes (2022).*

Classification of the quality of centenarian data in selected countries

The revised final classification of the quality of census data on centenarians for the 87 selected countries is shown in Table 8, which is arranged by continent, from best (VG) to worst quality (VP).

Table 8. Classification of the quality of census data regarding centenarians of selected countries, 2010-2024.

Continent	Cluster	Country	Original	R1	R2 (Year, Data)
Africa	4	Chad	VP	VP	L (2019, E)
	4	Kenya	_	_	L (2019, E)
	4	Malawi	_	_	L (2018, E)
	4	Zimbabwe	_	_	L (2022, E)
	4	Côte d'Ivoire	VL	VL	VL (2021, C)
	4	Rwanda	VL	VL	VL (2012, C)
	4	Senegal	VL	VL	VL (2013, C)
	4	South Africa	VL	VL	VL (2011, C)
Americas	2 &3	Argentina**	VG	VG	VG (2022, C)
	1	Canada	VG	VG	VG (2021, C)
	1	United States	VG	VG	VG (2020, C)
	2 & 3	Uruguay*	VG	VG	VG (2023, C)
	2 & 3	Peru	M	M	G (2017, C)
	2 & 3	Brazil	M	M	M (2022, C)
	2 & 3	Chile	M	M	M (2017, C)
	2 & 3	Guatemala	_	_	M (2020, E)
	2 & 3	Mexico	L	M	L (2020, C)
	2 & 3	Panama*	M	M	M (2023, C)
	2 & 3	Colombia	L	M	L (2018, C)
	2 & 3	Ecuador	M	M	L (2022, C)
	2 & 3	Cuba	M	M	VL (2022, E)
	2 & 3	Dominican Republic	VL	VL	VL (2022, C)
	2 & 3	Venezuela	M	M	VL (2023, E)
Asia	2 & 3	Azerbaijan	M	M	VG (2019, C)
	2 & 3	Brunei Darussalam	_	_	VG (2021, E)
	4	India	VL	VL	VL (2011, C)
	1	Israel***	L	L	VG (2022, C)
	1	Japan	VG	VG	VG (2020, C)
	2 & 3	Mongolia	G	G	VG (2020, C)
	1	Singapore***	L	L	VG (2020, C)
	1	South Korea	VG	VG	VG (2020, C)

Table 8. Continued.

Continent	Cluster	Country	Original	R1	R2 (Year, Data)
	2 & 3	Tajikistan*	L	L	VG (2020, C)
	2 & 3	Thailand	M	M	VG (2022, E)
	2 & 3	Turkey	G	G	VG (2021, C)
	2 & 3	Uzbekistan	_	_	VG (2024, E)
	2 & 3	China**	G	G	G (2020, C)
	2 & 3	Kyrgyzstan	L	L	G (2022, C)
	4	Bhutan	VL	VL	M (2017, C)
	2 & 3	Iran	L	L	M (2016, C)
	2 & 3	Kazakhstan	G	G	M (2021, C)
	2 & 3	Philippines	M	M	M (2020, C)
	4	Bangladesh	_	_	P (2022, E)
	4	Nepal	_	_	VL (2021, E)
	2 & 3	Oman	VL	VL	VL (2020, C)
	2 & 3	Saudi Arabia	_	_	VL (2022, E)
Europe	2 & 3	Albania	G	G	VG (2023, C)
	1	Austria	VG	VG	VG (2021, C)
	1	Belgium	VG	VG	VG (2023, E)
	1	Cyprus***	_	_	VG (2022, E)
	1	Denmark	VG	VG	VG (2023, E)
	2 & 3	Estonia	VG	VG	VG (2021, C)
	1	Finland	VG	VG	VG (2021, C)
	1	France	VG	VG	VG (2015, C)
	1	Germany	VG	VG	VG (2023, E)
	1	Greece	VG	VG	VG (2023, E)
	1	Iceland***	VG	VG	VG (2023, E)
	1	Ireland	VG	VG	VG (2023, E)
	1	Italy	VG	VG	VG (2023, E)
	1	Luxembourg	VG	VG	VG (2023, E)
	1	Malta***	VG	VG	VG (2021, C)
	1	Netherlands	VG	VG	VG (2023, E)
	1	Norway	VG	VG	VG (2023, E)
	2 & 3	Poland	VG	VG	VG (2021, C)
	1	Portugal	VG	VG	VG (2021, C)
	1	Slovenia	VG	VG	VG (2021, C)
	1	Spain	VG	VG	VG (2021, C)
	1	Sweden	VG	VG	VG (2022, E)
	1	Switzerland	VG	VG	VG (2021, C)

Table 8. Continued.

Continent	Cluster	Country	Original	R1	R2 (Year, Data)
	1	United Kingdom	VG	VG	VG (2011, C)
	2 & 3	Belarus	VG	VG	G (2019, C)
	2 & 3	Bulgaria	G	G	G (2021, C)
	2 & 3	Croatia	G	G	G (2021, C)
	2 & 3	Czech Republic	G	G	G (2021, C)
	2 & 3	Hungary	G	G	G (2022, C)
	2 & 3	Latvia	G	G	G (2021, C)
	2 & 3	Romania	G	G	G (2021, C)
	2 & 3	Russian Federation	VG	VG	G (2021, C)
	2 & 3	Slovakia	G	G	G (2021, C)
	2 & 3	Georgia	G	G	M (2025, E)
	2 & 3	Lithuania	G	G	M (2021, C)
	2 & 3	Serbia	M	M	M (2022, C)
	2 & 3	Macedonia	G	G	P (2021, C)
	2 & 3	Montenegro	G	G	P (2023, C)
	2 & 3	Ukraine	G	G	P (2022, E)
Oceania	1	Australia	VG	VG	VG (2021, C)
	1	New Zealand	VG	VG	VG (2018, C)
	4	Vanuatu	_	_	VL (2021, E)

*Note 1: * Data obtained from the National Institute of Statistics, 2025; ** Classified according to the best adjacent level;*** Very good quality was attributed based on the assumption of an increased proportion of male centennials.*

Note 2: VL = very low, L = low, M = moderate, G = good, VG = very good, C = census, E = estimates.

Source: Original Proposal: Medeiros, & Paes (2022). Review 1: Medeiros, & Paes (2025); Review 2: Basic Data (UNSD, 2025).

Some countries had values very close to the limits of the predefined intervals for each category (very low, low, moderate, good, very good). In these borderline cases, the study chose the highest classification between the two adjacent categories to ensure that countries with indicators on the threshold of a higher category are not unduly classified at a lower level, reflecting a conservative approach in assessing data quality.

The assessment of the quality of census data on centenarians also considered the assumption of an increased proportion of male centenarians in some countries, such as Malta, Iceland, Singapore, Cyprus, and Israel. In other words, these countries' P_m and P_f values were classified as good quality, belonging to Cluster 1, with high HDI and e^0. In turn, SR values were above the upper limit of the interval [0.140; 0.385] to classify

countries in this cluster. Thus, considering the possibility of there being more males among centenarians, the final classification was assigned as "very good". This option recognizes this assumption as an additional indicator of data reliability, complementing the quantitative analysis based on the sum of weights.

Table 8 illustrates that the ability to accurately measure human longevity is a strong indicator of a country's level of development. High HDI countries (Cluster 1) lead statistical reliability, while low HDI countries (Cluster 4) still face significant barriers in obtaining high-quality data on their older populations.

This poses methodological challenges regarding the rigorous comparability of longevity statistics at extreme ages between clusters. The literature widely recognizes this concern, as the accuracy of centenarian data is notoriously vulnerable to age overreporting and underreporting.

As highlighted by Gavrilov and Gavrilova (2011) in their analysis of mortality measurement at advanced ages in the United States, reliability at extreme ages must be critically examined even in countries with great statistical infrastructure. This need is even more pressing when analyzing Cluster 4, where the fragility of the civil registration system accentuates the risk of systematic biases in longevity estimates. A regional analysis reveals a stark contrast in these trajectories:

Fragility in Africa (Cluster 4): Most African countries analyzed remain in Cluster 4, characterized by low data quality. This persistence underscores the structural fragility of vital registration systems in these regions, where limited resources hinder the accurate counting of the oldest-old.

Advances in Asia and Europe: In contrast, significant improvements were observed in several Asian (Singapore and Thailand) and European (Albania) nations. These countries moved from lower historical ratings to "Very Good" (VG) in this second review, reflecting successful national efforts to modernize demographic records.

Regional Heterogeneity in the Americas: While nations like Argentina, Canada, and Uruguay consistently maintain "Very Good" (VG) quality, the region faces notable heterogeneity. Countries such as Colombia, Ecuador,

Cuba, and Venezuela showed stagnation or even a decline in data quality. This downturn may be intrinsically linked to political instability, economic crises, or the discontinuation of funding for recent census operations.

The complexity of collecting data at extreme ages is substantial, even in well-structured surveys. Information quality can be compromised by operational factors, such as the high prevalence of interviews with third parties or non-response to specific items, which requires careful respondent selection to ensure reliability (Barbi et al., 2018; Medeiros, & Paes, 2022; Reis, & Turra, 2016).

This methodological vulnerability, inherent to specialized surveys, can be exponentially exacerbated in large-scale censuses conducted by nations with poor statistical infrastructure, culminating in low primary data quality ratings, as observed for the countries in Cluster 4.

However, the quality disparity acts as a catalyst for improving registration systems. In response to this widely recognized challenge, the comparison between Review 1 and Review 2 indicates a continuous global effort to improve the quality analysis of centenarian data. Table 8 shows marked improvements in Asian countries (e.g., Israel, Singapore, Tajikistan, and Thailand) and European ones (e.g., Albania), which moved from lower ratings (P, F, or G) to VG in Review 2.

Still, this progress is not universal. The quality seemingly worsened or stagnated at low levels in some cases, such as in Latin American countries (e.g., Colombia, Ecuador, Cuba, and Venezuela). This may be related to political instability, economic crises, or discontinued funding for recent censuses. This stagnation compromises the reliability of centenarian data, making it more difficult to validate demographic advances in these countries, a direct contrast to the reality observed in studies of developed countries.

The trend toward improved accuracy in age reporting and advances in registration systems are intrinsically linked to broader and better-documented demographic gains. As identified by Barbi, Caselli, and Vallin (2003), improving the reliability of primary data is crucial, as it lends greater validity to longevity measurements and corroborates observed increases in life expectancy.

Thus, the pursuit of high-quality data for future cohorts is essential to reduce bias in population and mortality estimates and to allow demographic models to more accurately reflect the true trajectories of human survival and the life potential yet to be achieved.

Conclusion

The study findings reveal heterogeneous patterns in the quality and composition of centenarian data across countries. Among the 87 countries analyzed, 55 had an SR above 0.25, indicating a growing proportion of male centenarians. It is plausible to consider that this trend may reflect recent shifts in the behavioral and social determinants of male survival, particularly the decline in risk factors historically associated with excess male mortality, such as smoking and heavy alcohol consumption, but this does not necessarily represent causality.

Moreover, 53 of the 87 countries studied with data from 2020 to 2023 were classified as good or very good quality, 12 as moderate, and 22 as low or very low quality. Compared to the Original Proposal (with data from 2010 and 2011), there was a global effort to improve the analysis of centenarian data quality.

These findings highlight the need for further research examining the relationship between health behavior transitions and the narrowing longevity gap between sexes, while accounting for regional and contextual differences in population data quality. Future studies should also focus on the demographic patterns of Cluster 1 countries, characterized by higher data quality and advanced stages of demographic transition, to assess whether the patterns observed in these contexts also apply or may eventually apply to countries that have not yet completed or are still undergoing demographic transition.

A relevant limitation of the study concerns the lack of recent census data in some developed countries. This absence, however, does not necessarily indicate a statistical weakness, but rather reflects an ongoing methodological transition in several nations that are replacing the traditional decennial census with systems based on administrative records and continuous population estimates. Although these methods provide

more frequent updates and better data integration, they may differ in coverage and variable definitions, affecting international comparability and the assessment of data quality, particularly at extreme ages.

Despite the limitations, as was well said in the original paper regarding this proposal, this reviewed, updated proposal had good internal consistency. Its results produced satisfactory levels of application, scalability, and reproducibility for different locations in the world. Furthermore, the reviewed proposal used a few indicators usually available in every country's population counting records. Therefore, using this reviewed proposal may favor the understanding of the true panorama in the distribution of long-lived older people worldwide.

In addition, these results underscore significant regional disparities; while some regions suggest robust statistical systems, others still face challenges in age reporting and census coverage.

References

Aladejare, S. (2023). Does external debt promote human longevity in developing countries? Evidence from West African countries. *Fudan Journal of the Humanities and Social Sciences*, *16*, 213-237. https://doi.org/10.1007/s40647-023-00365-1

Barbi, E., Caselli, G., & Vallin, J. (2003). Trajectories of extreme survival in heterogeneous populations. *Population*, *58*(1), 43-66. https://doi.org/10.3917/popu.301.0045

Barbi, E., Lagona, F., Marsili, M., Vaupel, J. W., & Wachter, K. W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science*, *360*(6396), 1459-1461. https://doi.org/10.1126/science.aat3119

Bell, F. C., & Miller, M. L. (2005). *Life tables for the United States Social Security Area 1900-2100* (Actuarial Study No. 120). Social Security Administration. http://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl_7_1900.html

Gavrilov, L. A., & Gavrilova, N. S. (2011). Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File. *North American Actuarial Journal*, *15*(3), 432-447. https://doi.org/10.1080/10920277.2011.10597629

Gavrilov, L. A., & Gavrilova, N. S. (2019). Late-life mortality is underestimated because of data errors. *PLoS Biology*, *17*(2), 1-7. https://doi.org/10.1371/journal.pbio.3000148

IHME. Institute for Health Metrics and Evaluation. (2025). *Global Burden of Disease 2023: Findings from the GBD 2023 Study*. Seattle, WA: IHME.

Lozada-Martínez, I., Marín, J., Castelblanco-Toro, S., Mazenett-Granados, E., Suárez, J., Sarmiento, M., & Anaya, J. (2024). Demographics and clinical characteristics of a new population of centenarians in Colombia: The COOLCEN cohort. *Archives of Gerontology and Geriatrics Plus*, *1*(1), 1-10. https://doi.org/10.1016/j.aggp.2024.100006

Mackenbach, J., Valverde, J., Bopp, M., Brønnum-Hansen, H., Deboosere, P., Kalediene, R., Kovács, K., Leinsalu, M., Martikainen, P., Menvielle, G., Regidor, E., & Nusselder, W. (2019). Determinants of inequalities in life expectancy: An international comparative study of eight risk factors. *The Lancet Public Health*, *4*(10), e529-e537. https://doi.org/10.1016/s2468-2667(19)30147-1

Marcus, S., Njoku, C., Ijieh, S., & Okon, I. (2023). Undernourishment on longevity in West Africa: A group outcome. *AKSU Journal of Administration and Corporate Governance*, *3*(3), 177-189. https://doi.org/10.61090/aksujacog.2023.027

Medeiros, J. B., & Paes, N. A. (2022). Classification proposal to evaluate the quality of census data of centenaries. *Revista Latinoamericana de Población*, *16*. https://doi.org/10.31406/relap2022.v16.e202207

Medeiros, J. B., & Paes, N. A. (2025). Updating the proposal for assessing the quality of census count of centenaries. In *Applied Stochastic Models and Data Analysis* (pp. 29-38). Springer. https://doi.org/10.1007/978-3-031-82275-9_3

Meyer, J. (2012). *Centenarians: 2010* (2010 Census Special Reports). United States Census Bureau. https://www.census.gov/library/publications/2012/dec/c2010sr-03.html

Mikkelsen, L., Hooper, J., Adair, T., Badr, A., & Lopez, A. (2023). Comparative performance of national civil registration and vital statistics systems: a global assessment. *Boletim da Organização Mundial da Saúde*, *101*, 758-767. https://doi.org/10.2471/blt.22.289033

Mikkelsen, L., Phillips, D., AbouZahr, C., Setel, P., Savigny, D., Lozano, R., & Lopez, A. (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, *386*(10001), 1395-1406. https://doi.org/10.1016/s0140-6736(15)60171-4

Newman, S. J. (2018). Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS Biology*, *16*(12), Article e2006776. https://doi.org/10.1371/journal.pbio.2006776

Newman, S. J. (2024). The global pattern of centenarians highlights deep problems in demography. *medRxiv*. https://doi.org/10.1101/2024.09.06.24313170

Omokanmi, O., Ibrahim, R., Ajide, K., & Al-Faryan, M. (2022). Exploring the dynamic impacts of natural resources and environmental pollution on longevity in resource-dependent African countries: Does income level matter? *Resources Policy*, *79*, Article 102959. https://doi.org/10.1016/j.resourpol.2022.102959

Paes, N. A. (2018). *Statistical Demography of Included Events: With Examples Based on Brazilian Experience*. (2ª ed.). CCTA Publisher.

Perls, T. T. (2017). Male centenarians: How and why are they different from their female counterparts? *Journal of the American Geriatrics Society*, *65*(9), 1904-1906. https://doi.org/10.1111/jgs.14978

Phillips, D., AbouZahr, C., Lopez, A., Mikkelsen, L., Savigny, D., Lozano, R., Wilmoth, J., & Setel, P. (2015). Are well functioning civil registration and vital statistics systems associated with better health outcomes? *The Lancet*, *386*(10001), 1386-1394. https://doi.org/10.1016/s0140-6736(15)60172-6

Reis, C. S. dos, & Turra, C. (2016). Spatial distribution of centenarians in Brazil: An exploratory analysis of the quality of data from the 2000 and 2010 censuses. *Espinhaço Journal*, *5*(1). https://doi.org/10.5281/zenodo.3958091

Schmertmann, C., Queiroz, B. L., & Gonzaga, M. (2024). Data errors in mortality estimation: Formal demographic analysis of under-registration, under-enumeration, and age misreporting. *Demographic Research*, *51*(9), 229-266. https://doi.org/10.4054/DemRes.2024.51.9

UNDESA-PD. United Nations, Department of Economic and Social Affairs, Population Division (2024). World Population Prospects 2024, Online Edition.

UNDP. United Nations Development Programme. (2025). *Human Development Report 2025*. https://hdr.undp.org/system/files/documents/global-report-document/hdr2025reporten.pdf

UNSD. United Nations Statistics Division. (2025). *UNdata: Demographic Statistics Database*. https://data.un.org/Data.aspx?d=POP&f=tableCode%3a22

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236-244. https://doi.org/10.1080/01621459.1963.10500845

WHO. World Health Organization. (2023). *Healthy ageing: A priority for delivering universal health coverage*. https://www.who.int/publications/m/item/healthy-ageing--a-priority-for-delivering-universal-health-coverage

Zaccaria, D., Falciola, J., Masotti, B., Von Gunten, A., Herrmann, F., Jopp, D. S., & Cavalli, S. (2025). Data quality in centenarian research: The proxy-centenarian relationship and item nonresponse in the SWISS100 study. *PLoS ONE*, *20*(1), Article e0311847. https://doi.org/10.1371/journal.pone.0311847