A reliability assessment of the age-sex data from 1991 and 2006 Nigeria population censuses

Emmanuel U. Ohaegbulem *

Department of Statistics, Imo State University, Owerri, 460222, Nigeria
*Corresponding author E-mail: emmanx2002@yahoo.com

Abstract

This study assessed the quality of the age-sex data from 1991 and 2006 Nigeria population censuses using some conventional techniques of evaluating demographic data quality. Whipple and Myers indices were used to determine the extent of digit preference for the age-sex data presented in single years. There were very obvious preference for ages with end-digits 0 and 5 while other end-digits were avoided in the two censuses; and this was more pronounced with the females than males. From 1991 census to 2006 census, the Whipple index for both sexes declined from 294 to 251, while the Myers index declined from 60.8 to 49.9; indications of deficiencies in the Nigeria censuses data. The Joint scores computed for the five-year abridged distribution of the age-sex data yielded 54.83 and 38.52 for the 1991 and 2006 censuses, respectively. This showed that the data are poor in quality and not completely reliable as a result of misreporting. The 29.7% decline in Joint score from 1991 census to 2006 census was an indication of a modest improvement in quality over than fifteen years interval, even as the 2006 age-sex data are usable with adjustment.

Keywords: *Age-Sex Data; Joint Score; Myers Index; Nigeria Population Censuses; Whipple Index.*

1. Introduction

The role and impact of a formidable planning, policy formulation, and monitoring in national development can never be over-emphasized. The high levels of development attained by most developed countries of the world are clearly testimony. It is inconceivable to visualize a modern society, where meaningful development activities can be carried out without first considering the population; its size, distribution, growth and characteristics [7]. Thus, demographic data collection, evaluation and analysis have assumed great importance in the day-to-day administration and functioning of societies.

Census is the only source of comprehensive demographic data on all persons in a country required for planning, policy formulation, and monitoring of development goals [2]. According to Gunasekera [5], the basic inputs for national policy formulations and most of the demographic research studies are the data obtained from population censuses. Intuitively, the formulations of good policies as well as the validity of the findings of demographic research studies depend upon the accuracy of census data. However, the collection of census data is inevitably not free from errors; irrespective of the country.

One major challenge to meaningful development in some African countries is the scarcity and quality of demographic data. Censuses and surveys are not regular; for example, after the officially-endorsed national census in 1963, Nigeria conducted a successful and acceptable national census only in 1991 (28 years after). More often than not, census data collected in some countries in Africa have content and coverage errors which mar their quality and reliability for planning purposes (see, for example, [6]).

According to Mba [1], the usefulness and reliability of demographic parameters derived from censuses depend on the quality of the data collected. In other words, one of the fundamental precautions that must be taken before embarking on the analysis and interpretation of demographic data is that the quality of the observed data should be ascertained.

Among the demographic variables captured in a census, age and sex data play a vital role in population studies. The age-sex structure is one of the most fundamental characteristics of population composition; as it affects fertility behaviour, mortality and morbidity levels, migratory movements, labour force participation, and a host of other factors (see, for example, [5]). In addition, the age-sex data are the basic inputs for making population projections. Age-sex data are, therefore, almost always essential for analysis of population dynamics. Hence, evaluation of age-sex data is of the most important steps in a census evaluation programme.

Age and sex variables have been emphasized as critical to most demographic investigations. Consequently, the age-sex data of 1991 and 2006 censuses in Nigeria are put up for evaluation in this study. This is in a view to ascertaining the qualities, hence the reliability levels, of the two censuses' data; and also comparing them in terms of quality. In addition, the data evaluation will provide basis for improved data quality in future censuses in Nigeria.

2.   Methods

In this assessment for reliability of the age-sex data from 1991 and 2006 Nigeria censuses, both the reported ages in single years and five-years abridged distributions (presented in Table1) will be considered. In appraising the single year age data, two conventional indices will be employed; and they are Whipple Index (WI) and the Myers Index (MI). Also, the United Nations Age-Sex Accuracy Index will be used to appraise the grouped age-sex data.

Table 1: Age and Sex Distributions of 1991 and 2006 Nigeria Population Censuses

Age Group	1991			2006		
(Years)	Males	Females	Both Sexes	Males	Females	Both Sexes
0-4	7,344,454	6,999,435	14,343,889	11,569,218	11,025,749	22,594,967
5-9	7,374,314	7,126,144	14,500,458	10,388,611	9,616,769	20,005,380
10-14	5,812,538	5,336,143	11,148,681	8,504,319	7,631,631	16,135,950
15-19	4,528,811	4,806,977	9,335,788	7,536,532	7,362,887	14,899,419
20-24	3,314,303	4,357,267	7,671,570	6,237,549	7,197,530	13,435,079
25-29	3,304,739	4,006,932	7,311,671	5,534,458	6,676,968	12,211,426
30-34	2,808,629	3,105,298	5,913,927	4,505,186	4,962,352	9,467,538
35-39	2,206,871	2,008,062	4,214,933	3,661,133	3,670,622	7,331,755
40-44	1,971,197	1,874,721	3,845,918	3,395,489	3,060,981	6,456,470
45-49	1,355,101	1,061,602	2,416,703	2,561,526	2,029,767	4,591,293
50-54	1,388,650	1,182,149	2,570,799	2,363,937	1,885,282	4,249,219
55-59	638,375	481,394	1,119,769	1,189,770	876,477	2,066,247
60-64	898,801	791,573	1,690,374	1,363,219	1,087,067	2,450,286
65-69	406,540	357,400	763,940	628,436	522,612	1,151,048
70-74	492,186	394,116	886,302	765,988	564,609	1,330,597
75-79	195,455	156,368	351,823	327,416	252,422	579,838
80-84	258,059	222,627	480,686	408,680	351,373	760,053
85+	230,585	194,404	424,989	404,021	311,204	715,225
TOTAL	**44,529,608**	**44,462,612**	**88,992,220**	**71,345,488**	**69,086,302**	**140,431,790**

2.1. Whipple index (WI)

The Whipple Index is a summary measure of age-heaping for ages ending in 0 or 5 (see [3]). It is an index of preference for or avoidance of the end-digits 0 and 5 in the age range, $23 - 62$ years [2]. The index of preference for each end-digit is the percentage of the one-tenth of the total population aged $(23 - 62)$ years reported at the ages ending with the digits. The Whipple Index for the end-digit 0 is given by,

$$WI(0) = \frac{(P_{30}+P_{40}+P_{50}+P_{60})}{\frac{1}{10}\left({}_{40}P_{23}\right)} \times 100 \qquad (1)$$

And for end-digit 5;

$$WI(5) = \frac{(P_{25}+P_{35}+P_{45}+P_{55})}{\frac{1}{10}\left({}_{40}P_{23}\right)} \times 100 \qquad (2)$$

The Whipple Index for the end-digits 0 and 5 combined is given by

$$WI(0,5) = \frac{(P_{25}+P_{30}+P_{35}+P_{40}+P_{45}+P_{50}+P_{55}+P_{60})}{\frac{1}{5}\left(_{40}P_{23}\right)} \times 100 \qquad (3)$$

Where P_x and $_nP_x$, respectively, are the population sizes at age x years and age-group (x to $x + n$) years.

This measure assumes that the population is equally distributed among the ten end-digits. The index varies between a minimum of 100 (indicating no preference at all) and a maximum of 500 (if only 0 and/or 5 were reported). The close the index for each end-digit (or both) is to 100, the lower the degree of preference for it (or them) and the better the quality of age reporting. The more the index deviates from 100, the higher the degree of preference for the end-digit(s). Table 2 shows the United Nations recommended standard for measuring age heaping on digits 0 and 5 using Whipple Index; which classifies the quality of data from very accurate to very bad (see, for example, [1]; [8]).

Table 2: UN Recommendation for Age-Heaping on the Basis of Whipple Index

Whipple Index (WI)	<105	105 − 109.9	110 − 124.9	125 − 174.9	>175
Quality of Data	Very/Highly Accurate	Relatively/Fairly Accurate	Approximately Accurate	Bad/Rough	Very Bad/Rough
% Deviation from Perfect	<5%	(5.9 − 9.9)%	(10.24 − 24.99)%	(25 − 74.99)%	>75%

2.2. Myers index

The Myers Index was developed to detect preference for or avoidance of all the end-digits; from 0 to 9 for the population in age range $10 - 69$ years [2]. It is defined as,

$$MI = \sum_{i=1}^{9}|\%B_i - 10| \qquad (4)$$

Where B_i is the blended population reported at ages with the end-digit i; and is expressed as,

$$B_i = (i + 1)P_{1i} + [10 - (i + 1)]P_{2i} \qquad (5)$$

and

$$\%B_i = \frac{B_i}{\sum_{i=0}^{9} B_i} \times 100 \qquad (6)$$

Where, P_{1i} and P_{2i} are, respectively, the total populations reported at ages with end-digit i; in the ranges $10 - 59$ years and $20 - 69$ years.

The range of Myers Index is 0 to 180; but when it is to be calculated as half of the total absolute deviation of $\%B_i$ from 10, the range of MI is 0 to 90. Furthermore, when all age end-digits are equally chosen the index will be 0; and the index will be 180 or 90 (depending on the style of calculation) where a single end-digit is chosen.

2.3. United nations age-sex accuracy index

The United Nations Age-Sex Accuracy Index (also called Joint Score (JS)) is a single index for the joint evaluation of data on age and sex, proposed by the United Nations based on empirical evidence [2].

This index is computed as a weighted sum of the United Nations Age Ratio Score (ARS) and the United Nations Sex Ratio Score (SRS) of a study population. It is expressed symbolically as,

$$JS = ARS_{males} + ARS_{females} + 3(SRS) \qquad (7)$$

Where, ARS which is defined for each sex, is given as

$$ARS = \frac{1}{k-3}\sum_{i=2}^{k-2}\left|\,_nAR_x - 100\right| \qquad (8)$$

Such that $\left|\,_nAR_x - 100\right|$ denotes the absolute deviation of $_nAR_x$ (the age ratio for the age (x to $x + n$) years) from 100; and k is the total number of age groups including the last open-ended age group. Age ratios provide no figures for the first and the last closed-ended age groups. However, if computation of ARS is restricted to closed-ended age groups, the denominator of (8) becomes $k - 2$. The computation of ARS is restricted to the age range (0-69) years, and this is because of the wide range fluctuation beyond this age range.

To obtain the age ratio for each sex, especially when digit preference/avoidance is highly pronounced, use is made of the Ramachandran modification given as,

$$_nAR_x = \frac{_nP_x}{\frac{1}{4}\left(_nP_{x-n} + 2\,_nP_x + _nP_{x+n}\right)} \times 100 \qquad (9)$$

Where $_nP_x$ is the population aged $(x \text{ to } x+n)$ years.
SRS is commonly used for sex data evaluation, and is given by

$$SRS = \frac{1}{k-1}\sum_{i=1}^{k}\left|SR_{(i)} - SR_{(i-1)}\right| \qquad (10)$$

Where, k is the number of ages or age groups involved in the computation, and $SR_{(i)}$ denotes the sex ratio for the ith age or age group, defined for specified area and time as the number of males per 100 females. It is expressed mathematically as,

$$SR = \frac{_nN_x^{(m)}}{_nN_x^{(f)}} \times 100 \qquad (11)$$

Where $_nN_x^{(m)}$ and $_nN_x^{(f)}$ denote, respectively, the number of males and females reported in the age group $(x \text{ to } x+n)$ years.
One major assumption considered in computing this index is that if a population is changing only through fertility and mortality, the corresponding sex ratios should not fluctuate much from one age group to another. Consequently, a low value of SRS suggests that data quality may be good, while a high value suggests that data quality may not be good.
For the purpose of evaluation of age and sex data in demographic study, the quality of any age and sex data is based on the place of the computed value of the United Nations Age–Sex Accuracy Index for such data with reference to the United Nation's criteria for demographic data quality standard presented in Table 3.

Table 3: United Nations Criteria for Data Quality on the Basis of 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 utilize for any inference.

3. Results

The Whipple indices computed for the age-sex data of 1991 and 2006 Nigeria censuses, presented in Table 4, indicate patterns of digit preferences in the two censuses. The indices are well above 175, which implies age-heaping for ages ending in 0 and 5; and this was more pronounced with the females than the males in both censuses. On the overall, however, there is about 14.6% decline in age-heaping for ages ending in 0 and 5 from 1991 to 2006 censuses.

Table 4: Whipple Indices for 1991 and 2006 Nigeria Population Censuses

Sex	Census Year		% Decline
	1991	2006	
Males	280	247	11.8
Females	307	255	16.9
Both Sexes	294	251	14.6

The Myers indices (for both sexes) computed from1991 and 2006 Nigeria censuses, as well as the mean indices (presented in Table 5) also show that there were general preferences forages ending in 0 and 5(with ages ending in 0 being more preferred than 5) while there was avoidance of other age end-digits especially with ages ending in 1, 4, and 9 being most avoided. Myers index dropped from 60.8 in 1991 census to 49.9 in 2006 census; a decline of about 17.9%.

Table 5: Myers Indices for 1991 and 2006 Nigeria Population Censuses

Census Year	End-Digit										MI	% Decline
	0	1	2	3	4	5	6	7	8	9		
1991	19.6	-5.5	-2.3	-4.7	-5.1	10.8	-4.0	-3.6	0.0	-5.2	60.8	
2006	16.1	-5.1	-1.5	-4.0	-4.2	8.5	-3.2	-3.0	0.3	-3.8	49.9	17.9
Mean Index	**17.9**	**-5.3**	**-1.9**	**-4.4**	**-4.7**	**9.7**	**-3.6**	**-3.3**	**0.2**	**-4.5**	**55.5**	

Table 6 shows cases the age ratios and sex ratios derived from the 1991 and 2006 Nigeria population censuses data. The Ramachandran modification was adopted in determining the age ratios since digit preference/avoidance were highly pronounced in the two censuses. The overall sex ratios derived from the 1991 and 2006 censuses data are 100.15 and 103.27, respectively, and the sex ratio at birth was 104.93 for the 1991 census data; this remained the same in 2006 census data. The sex ratios for the two censuses are above 100, except for the ages 15-34 years in 1991 and 20-39 years in 2006; while the age ratios exhibited no defined pattern.

Table 7 features the age ratio scores (for males and females), sex ratio scores, and the age-sex accuracy indices computed for the 1991 and 2006 Nigeria population censuses data. The age ratio scores for the males are less than those of the females in the two censuses, and the age ratio scores dropped from 1991 census to 2006 census with a decline of about 30.1% and 35.0% for males and females, respectively. In the same vein, the sex ratio scores dropped from 11.19 in 1991 census to 8.08 in 2006 census (about 27.8% decline), while the joint score dropped from 54.83 in 1991census to 38.52 in 2006 census (about 29.7% decline).

Table 6: Age Ratios and Sex Ratios for 1991 and 2006 Nigeria Population Censuses

Age Group (Years)	$_nAR_x$				$SR_{(i)}$	
	1991		2006		1991	2006
	Males	Females	Males	Females		
0-4	-	-	-	-	104.93	104.93
5-9	105.70	107.21	101.72	101.52	103.48	108.03
10-14	98.82	94.42	97.38	94.68	108.93	111.44
15-19	99.62	99.59	101.11	99.65	94.21	102.36
20-24	91.67	99.43	97.67	101.25	76.06	86.66
25-29	103.82	103.56	101.50	104.68	82.48	82.89
30-34	100.95	101.60	98.98	97.91	90.45	90.79
35-39	96.02	89.29	96.20	95.56	109.90	99.74
40-44	105.07	109.97	104.37	103.57	105.15	110.93
45-49	89.30	81.98	94.15	90.15	127.65	126.20
50-54	116.43	121.02	111.52	112.95	117.47	125.39
55-59	71.64	65.57	77.93	74.19	132.61	135.74
60-64	126.48	130.73	119.98	121.69	113.55	125.40
65-69	73.78	75.22	74.24	77.51	113.75	120.25
70-74	124.10	121.08	123.16	118.60	124.88	135.67
75-79	68.51	67.29	71.59	71.06	125.00	129.71
80-84	109.56	111.87	105.55	110.99	115.92	116.31
85+	-	-	-	-	118.61	129.83
Overall					**100.15**	**103.27**

Table 7: Age Ratio Scores, Sex Ratio Scores, and Joint Scores from 1991 and 2006 Nigeria Population Censuses

Census Year	ARS		SRS	JS
	Males	Females		
1991	9.28	11.98	11.19	54.83
2006	6.49	7.79	8.08	38.52
% Decline	**30.1**	**35.0**	**27.8**	**29.7**

4. Discussion

This study evaluated the age-sex data from the 1991 and 2006 Nigeria population censuses, and has revealed the extent of the qualities of the two censuses' data. The indices derived from the data have demonstrated that the age-sex data collected from the two censuses are very deficient from the point of view of age-reporting; and hence, the data are of poor quality.

The Whipple indices for the two censuses indicate very high preferences for ages with end-digits 0 and 5; although the indices are slightly better for males than females. Similarly, the Myers indices for the two censuses show that end-digit 0 was more preferred than end-digit 5; while the other end-digits were very much avoided. By the assessment criteria of Myers index, the indices for 1991 and 2006 censuses are well above 0; and these are indicators of deficiencies in the two censuses data. However, there was general improvement in the quality of age reporting as preference for certain age end-digits declined and avoidance of certain age end-digits also declined. Also, the Whipple and Myers indices recorded about 14.6% and 17.9% declines, respectively, from the 1991 to 2006 Nigeria population censuses; an indication of significant improvement in the 2006 Nigeria population census data over that of 1991.

The overall sex ratios for 1991 and 2006 censuses' data (100.15 and 103.27, respectively) fall within the model sex ratio range of 100 – 107 for Africa [10]. The sex ratio at birth (104.93) remained the same in the two censuses. Sex ratio at birth is usually 104 because of the biological fact that male births generally exceed female births [5]. The overall sex ratios and the age-specific sex ratios reveal that there were more males than females in the 1991 and 2006 censuses data, except for the ages (15–34) years for the 1991 census and (20–39) years for the 2006 census. However, majority of the age-specific sex ratios fall very much outside the range, 95 – 105 males per 100 females and are somehow different from general expectation; an indication of the extent of age misreporting by five-year age groups in the two population censuses. Under normal circumstance, it is a valid expectation in all population to have a slightly higher number of males at young ages and a consistently greater number of females at older ages since females live longer than males (see, for example, [9]; [4]).

The age ratio scores obtained for males and females in the two censuses suggest that the quality of age data may be better for males than for females. The about 30.1% and 35.0% drops in age ratio scores from 1991 census to 2006 census for males and females, as well as the about 27.8% decline in sex ratio scores from 1991 census to 2006 census are indications that 2006 census data were better than 1991 census data in terms of age-reporting.

The Joint Score for the 1991 Nigeria census data (= 54.83) which lies between 40 and 60 suggests that the age and sex data are deficient and require massive adjustment before they could be meaningfully used, and any interpretation with the data should be with care and caution. In a similar vein, the Joint Score for the 2006 Nigeria census data (= 38.52) lies between 20 and 39; and this suggests that the age and sex data are usable with adjustment.

5.   Conclusion

The age-sex data from the 1991 and 2006 Nigeria population censuses are poor in quality as a result of age misreporting. However, the quality of the age-sex data had improved somewhat with the passage of time between 1991 and 2006 censuses; as the data moved from being deficient and require massive adjustment before any meaningful usage in 1991 census to being usable with little adjustment in 2006 census. Much more significant improvements are expected in subsequent population censuses in Nigeria should adequate measures be put in place to, at least, substantially reduce the incidence of age misreporting; such as improving the literacy levels of Nigerians and people consequently becoming more aware of their ages, as well as the National Population Commission (NPC) using good and diligent interviewers/enumerators and the use of calendar of local events to estimate the ages of a respondents in cases of unknown ages.

References

[1]   C.J. Mba, Assessing the reliability of the 1986 and 1996 Lesotho census data, Journal of Social Development in Africa. Vol. 18 (1) (2003). http://dx.doi.org/10.4314/jsda.v18i1.23821.

[2]   E.C. Nwogu, I.S. Iwueze, Introduction to Demography, Supreme Publishers, Owerri, Imo State, Nigeria, 2009.

[3]   F.B. Hobbs, Age and sex composition, In Jacob A. Siegel and David A. Swanson (Eds), The methods and materials of demography, Elsevier Academic Press, USA, 2004, pp 125 – 173. http://dx.doi.org/10.1016/b978-012641955-9/50041-2.

[4]   G. Bicego, O.B. Ahmad, Infant and child mortality, DHS comparative studies, No. 20, Institute for Resource Development/Macro Systems, Inc., Maryland, Columbia, 1996.

[5]   H.R. Gunasekera An evaluation of age-sex data of census of population and Housing 2001 – Sri Lanka. A paper presented to the 24th Population Census Conference, Hong Kong, China 25 – 27 March, 2009.

[6]   I. Sembajwe, The evaluation of the vital registration system in Lesotho: The case of Mantsebo. Working papers in demography, Department of Statistics, Demography Unit, Maseru National University of Lesotho, 1990.

[7]   P.K.A. Makinwa, Population Data: The importance of census, sample surveys and vital registration systems, Unpublished Monograph, No. 15, Nigerian Educational Research Council, Lagos State, Nigeria, 1985, page 53.

[8]   T. Dahiru, H.G. Dikko, Digit preference in Nigerian censuses data of 1991 and 2006, Epidemiology Biostatistics and Public Health. Vol. 10, No. 2 (2013).

[9]   United Nations, The determinants and consequences of population trends: New sample of findings on interaction of demographic, economic and social factors, Vol. 1, New York, 1973.

[10]   United Nations, Manual X: Indirect techniques for demographic estimation. Population Division, ST/ESA/SER.A/81,New York, 1983.