Title: REFINED Age Distribution and Demographic Parameters Estimation in Nigeria: An Indirect Approach

Adebowale, S.A: Fagbamigbe, F.A; Bamgboye, E.A

***Corresponding Author** (Email:Adesteve2006@yahoo.com)

Brief authors' profile:

Dr. Ayo Stephen Adebowale and Adeniyi Francis Fagbamigbe are researchers and lecturers in Department of Epidemiology, Medical Statistics and Environmental health, Faculty of Public Health, college of medicine, University of Ibadan, Ibadan, Nigeria.

ABSTRACT

Poor census results in terms of age misreporting are common in developing countries. This poses serious challenge to the true estimate of basic demographic parameters. The study uses 2006 census data. Analysis was done using indirect demographic method. After data adjustment, the results showed that male birth (49.4) and death (18.3) rates were higher than female birth (45.3) and death (16.0) rates. The expectation of life at birth was lower for males (46.7) than females (50.0). Data adjustment using logit transformation reveals gross age misstatement across all ages, but age misreporting are more pronounced among females than males. Also, there is tendency to under-report ages at ages 15 to 17 years and above 55 years, whereas, gross over-reporting of age are seen in ages between 18 and 55 years. Priority attention should be accorded through public campaign on the need to report true age during any demographic survey.

Keywords: Age misreporting, adjustment, life table, demographic parameter, Nigeria

Background

There has been increase in awareness of the implications of the unprecedented population trends for economic, health and social development. Across nations, different policies have been designed to curtail the upsurge in population growth. However, adequate demographic parameters and data, especially in developing countries are lacking to monitor the pace (Raymer and Rogers, 2007). The developing countries who strived to have one through either census or survey; data classified by age are often full of errors. In Nigeria, demographic information on the nation's population is poor and inadequate for planning of socioeconomic development and projection. Thus the need for reliable data in Nigeria is critical.

Censuses in developing countries are far from acquiescent perfect data, because they suffer from poor age-reporting on the part of the population (respondents) which have tendencies to bias the estimate of basic demographic parameters. Differential in coverage of the population by age and sex is also very often present among young adults and adult population and therefore its effects are at times inseparable from misstatement of age (Chevan and SurtherLand, 2009).

A review of contemporary literature on census data in Nigeria reveals a striking lack of consistency and great discrepancies in the age distribution and estimates of basic demographic parameters. The inconsistencies arise in part from the degree of defectiveness of the data and in part from the sophistication of the analysis method.

One way of refining age data obtained from most developing countries involves comparing the age distribution with standard age distribution obtained from Coale-Demeny model life tables. Then, the adjustment of this standard age distribution reflects the features of the population under study. This is achievable by fitting a stable population to the studied population which is adopted as a true representation of the age distribution. Using a standard population model will enhance internal consistency of the study population. This methodology has been assessed by internal and external comparisons and yielded the expected results (Davern et al, 2009). The consistency of the results, its reliability and robustness make the method suitable for any settings. However, the underlying assumptions of the method must be justified and understood before use. These assumptions often take into account of some flaws that are likely to present in the raw data.

The task of this paper is to provide better measure of basic demographic parameters such as; differential by sex in death and birth rates, and expectation of life at birth. These parameters are indications of the way in which a population will evolve in terms of age structure and size over time. It also provided a refined age structure for Nigeria through re-analysis of the 2006 census data by method which eliminated bias and problems that make the reported age-distribution lack any credibility. This is with the view to fulfilling the yearning for quality data by the planners and policy makers.

Materials and method

The data came from two sources namely; ICF Macro Calverton, Maryland, USA and National Population Commission (NPC), Nigeria which provided the NDHS 2008 and census figure for 2006. A brief description of the methodology involved during the NDHS data collection is discussed below.

The sample was designed to provide population and health indicators at the national, zonal, and state levels. The primary sampling unit (PSU), referred to as a cluster for the 2008 NDHS, was defined on the basis of Enumeration Areas (EAs) from the 2006 EA census frame. The 2008 NDHS sample was selected using a stratified two-stage cluster design consisting of 888 clusters, 286 in the urban and 602 in the rural areas. A representative sample of 36 800 households was selected, with a minimum target of 950 completed interviews per state. In each state, the number of households was distributed proportionately among its urban and rural areas.

All women age 15-49 who were either permanent residents of the households in the 2008 NDHS sample or visitors present in the households on the night before the survey were eligible to be interviewed. Three questionnaires were used. These are; the Household Questionnaire, the Women's Questionnaire, and the Men's Questionnaire. These questionnaires were adapted to reflect the population and health issues relevant to Nigeria. However, this study used only women questionnaire for its analysis. The NDHS data was used to estimate mortality level while adjusted were made on the age distribution reported by 2006 census.

The choice of the method employed in this study was dictated by the nature of the data. In the 2006 Nigeria Demographic and Health Survey, data on current mortality $l(x)$ were collected and growth rate r(x) was estimated using 1991 and 2006 census figures. It was therefore decided to approach the analysis through a method referred to as $l(x)$, r(x) method of stable population. Essentially, the procedure followed is as contained in the United Nations Manual X. According to the Manual, a model stable population can be defined by at least two parameters: the growth rate of the population and mortality level. The appropriate mortality level was selected with the aid of $l(2)$ estimated using Coale and Trussel model . Given the distortion-ridden age-distribution characteristic of African data, it is impossible to find a reported age distribution whose r(x) value agrees in totality with the model ones. Therefore, for value of r(x) and the given level of mortality, two stable populations, each having the "right" age distribution was selected. The right age distribution is those which almost certainly bracket the Nigeria age distribution provisionally assumed to be unknown.

In order to estimate the level of mortality, indirect method of estimating infant and childhood mortality which was first developed by William Brass in late 1960's was used. Brass formulated a procedure for converting proportions dead of children ever-born reported by women of childbearing age (15-49) into estimates of the probability of dying before attaining certain exact childhood ages. In an attempt to increase the flexibility of Brass original method, Sullivan reviewed the method by using a set of multipliers which are derived from least square regression to fit the expression $q(x) = K(i)D(i)$ to data generated from observed fertility schedule and Coale-Demeny model life-tables. This method was adjusted latter by Coale and Trussel in 1975.

The assumption of this method is that, fertility and childhood mortality have remained constant in the recent past. In Nigeria, going by the total fertility rates estimated for 2003 and 2008 by Nigeria Demographic and Health Survey which put the figure at 5.7 for the two periods. One can infer that fertility has remained constant during the period. Also the childhood mortality rate in Nigeria is reducing, but the reduction rate is steadily slow. Therefore, an assumption of constant childhood mortality is applicable to Nigeria.

The data required are; data classified by age of the mother, data on fertility experience of true cohorts. That is, data on children ever born and children surviving. This is used to estimate the number of children dead.

Results:

Procedures for estimating mortality levels

Step1: Calculation of average parity per woman P(i) i.e $\mathrm{P}(i) = \frac{\mathrm{CEB}(i)}{\mathrm{FP}(i)}$

Where; CEB(i) is the children ever born to women of age group i and FP(i) is the women population in that age group.

Step2: Calculation of proportion of children dead D(i) for each age group of mother

i.e $D(i) = \frac{CD(i)}{CEB(i)}$

Where; CD(*i*) is the number of children dead reported by women in age group *i*

Step 3: Calculation of multiplier K(*i*) i.e $K(i) = a(i) + b(i)[P(1)/P(2)] + c(i)[P(2)/P(3)]$

Where, a(*i*), b(*i*) and c(*i*) are constant coefficients and were chosen from the North model life table. See Manual X (UN, 1983)

$\frac{P(1)}{P(2)} = \frac{0.2352}{1.1919} = 0.1973$ and $\frac{P(2)}{P(3)} = \frac{1.1919}{2.5145} = 0.4740$

$$K(1) = 1.1119 - 2.9287(0.1973) + 0.8507(0.4740) = 0.9373$$
$$K(2) = 1.2390 - 0.6865(0.1973) - 0.2745(0.4740) = 0.9734$$
$$K(3) = 1.1884 + 0.0421(0.1973) - 0.5156(0.4740) = 0.9523$$
$$K(4) = 1.2046 + 0.3037(0.1973) - 0.5656(0.4740) = 0.9964$$
$$K(5) = 1.2586 + 0.4236(0.1973) - 0.5898(0.4740) = 1.0626$$
$$K(6) = 1.2240 + 0.4222(0.1973) - 0.5456(0.4740) = 1.0487$$
$$K(7) = 1.1772 + 0.3486(0.1973) - 0.4624(0.4740) = 1.0268$$

Step 4: Calculation of probabilities of dying q(x) and probabilities of surviving *l*(x) i.e.

$q(x) = K(i)D(i)$

Where, q(x) is the probability of dying. $\therefore l(x) = 1 - q(x)$ i.e *l*(x) is the probability of surviving.

Step5: Calculation of reference date

$$t(i) = A(i) + B(i)[P(1)/P(2)] + C(i)[P(2)/P(3)]$$

Where, A(*i*), B(*i*) and C(*i*) are constant coefficients and were chosen from the North model life table. The values are shown in Appendix (1)

$\frac{P(1)}{P(2)} = \frac{0.2352}{1.1919} = 0.1973$ and $\frac{P(2)}{P(3)} = \frac{1.1919}{2.5145} = 0.4740$

$$t(1) = 1.0921 + 5.4732(0.1973) - 1.9672(0.4740) = 1.24$$
$$t(2) = 1.3207 + 5.3751(0.1973) + 0.2133(0.4740) = 2.48$$
$$t(3) = 1.5996 + 2.6268(0.1973) + 4.3701(0.4740) = 4.19$$
$$t(4) = 2.0779 - 1.7908(0.1973) + 9.4126(0.4740) = 6.19$$
$$t(5) = 2.7705 - 7.3403(0.1973) + 14.9352(0.4740) = 8.40$$
$$t(6) = 4.1520 - 12.2448(0.1973) + 19.2349(0.4740) = 10.85$$
$$t(7) = 6.9650 - 13.9160(0.1973) + 19.9542(0.4740) = 13.66$$

Since the NDHS survey was conducted between June and October 2008, the period can be put at 2008.8. The t(i) values were subtracted from this date to give the reference date.

Step6: Estimation of mortality level:

Mortality level for each age is estimated by converting mortality estimates (column (8) of table 1) using the Coale-Demeny system. For instance, the level consistent with the estimated q(2) is equal to 0.1490 and the corresponding *l*(2) is 0.8510. Since only the *l*(x) values are tabulated in table 229 page 262 in Manual X (Both sexes, model North), 0.8510 is the used for Interpolation purposes to obtain column (12) of table 1.

The above procedures were also used to compute all the parameters in tables 2 and 3.

TABLE 1: Derivation of Mortality Levels, Both Sexes, Nigeria, NDHS 2008

1	2	3	4	5	6	7	8	9	10	11	12
Age Group	**FP(i)**	**CEB(i)**	**CD(i)**	**P(i)**	**D(i)**	**K(i)**	**q(x)**	***l*(x)**	**t(x)**	**Ref. Date**	**Mortality level**
15-19	6493	1527	195	0.2352	0.1277	0.9373	0.1197	0.8803	1.24	2007.6	12.63
20-24	6133	7310	1119	1.1919	0.1531	0.9734	0.1490	0.8510	2.48	2006.3	12.88
25-29	6309	15864	2392	2.5145	0.1508	0.9523	0.1436	0.8564	4.19	2004.6	14.29
30-34	4634	18256	3132	3.9396	0.1716	0.9964	0.1710	0.8290	6.19	2002.6	14.01
35-39	3912	20578	3922	5.2602	0.1906	1.0626	0.2025	0.7975	8.40	2000.4	13.82
40-44	3032	18727	3852	6.1765	0.2057	1.0487	0.2157	0.7843	10.85	1998.0	13.88
45-49	2872	19651	4649	6.8423	0.2366	1.0268	0.2429	0.7571	13.66	1995.1	13.50

TABLE 2: Derivation of Mortality Levels, Females, Nigeria, NDHS 2008

1	2	3	4	5	6	7	8	9	10	11	12
Age Group	**FP(i)**	**CEB(i)**	**CD(i)**	**P(i)**	**D(i)**	**K(i)**	**q(x)**	***l*(x)**	**t(x)**	**Ref. Date**	**Mortality level**
15-19	6493	720	84	0.1109	0.1167	0.9484	0.1107	0.8893	1.22	2007.6	12.55
20-24	6133	3553	500	0.5793	0.1407	0.9795	0.1378	0.8622	2.45	2006.4	12.94
25-29	6309	7831	1118	1.2412	0.1428	0.9558	0.1365	0.8635	4.14	2004.7	14.13
30-34	4634	8900	1476	1.9206	0.1658	0.9988	0.1656	0.8344	6.13	2002.7	13.78
35-39	3912	9956	1816	2.5450	0.1824	1.0644	0.1942	0.8058	8.34	2000.5	13.72
40-44	3032	9215	1847	3.0392	0.2004	1.0502	0.2105	0.7895	10.79	1998.0	13.67
45-49	2872	9503	2153	3.3088	0.2266	1.0281	0.2330	0.7670	13.61	1995.2	13.45

TABLE 3: Derivation of Mortality Levels, Males, Nigeria, NDHS 2008

1	2	3	4	5	6	7	8	9	10	11	12
Age Group	**FP(i)**	**CEB(i)**	**CD(i)**	**P(i)**	**D(i)**	**K(i)**	**q(x)**	***l*(x)**	**t(x)**	**Ref. Date**	**Mortality level**
15-19	6493	807	111	0.1243	0.1375	0.9269	0.1274	0.8726	1.26	2007.5	12.75
20-24	6133	3757	619	0.6126	0.1648	0.9677	0.1595	0.8405	2.51	2006.3	12.84
25-29	6309	8033	1274	1.2733	0.1586	0.9489	0.1505	0.8495	4.24	2004.6	14.43
30-34	4634	9356	1656	2.0190	0.1770	0.9941	0.1760	0.8240	6.24	2002.6	14.23
35-39	3912	10622	2106	2.7152	0.1983	1.0608	0.2104	0.7896	8.47	2000.3	14.00
40-44	3032	9512	2005	3.1372	0.2108	1.0472	0.2207	0.7793	10.92	1997.9	14.00
45-49	2872	10148	2496	3.5334	0.2460	1.0255	0.2523	0.7477	13.74	1995.1	13.56

The choice of stable population:

The approach in this study used the NDHS, 2008 data to obtain the levels of mortality as shown above. The 1991 and 2006 were used to obtain the growth rates for males and females separately with the assumption that Nigeria population grows exponentially. The growth rates and mortality levels were used to select the appropriate standard population from the Coale-Demeny model life tables as shown in tables 4 and 5. The age distribution for the standard was obtained through linear interpolation to match the calculated growth rates.

For Females; $l_f(2) = 0.8622$, the mortality level that corresponds to this value is 12.94. In 1991 census, the number of females = 44,462,612 and in 2006 the value = 69,086,302 thus giving a growth rate $= \frac{\ln\left(\frac{69,086,302}{44,462,612}\right)}{15} = 0.02938$.

For Males; $l_m(2) = 0.8405$, the mortality level that corresponds to this value is 12.84. In 1991 census, the number of males = 71,345,488 and in 2006 the value = 44,529,608 thus giving a growth rate $= \frac{\ln\left(\frac{71,345,488}{44,529,608}\right)}{15} = 0.0314$.

From the selected standard population, the following demographic parameters were obtained through linear interpolation to match with the levels of mortality and the growth rates computed above.

TABLE 4: Estimated of Basic Demographic Parameters

Parameter	Males	Females	Both Sexes
Birth Rate	49.43	45.34	44.6
Death Rate	18.03	15.96	14.2
Population 15-44	41.33	41.73	-
Dependency ratio	1.042	0.991	-
Expectation of life at birth	46.717	50.00	-

Then, the birth rate for both sexes (total population)

$$= \text{female birth rate} \times \frac{\text{Number of females}}{\text{Total population}} \times (1 + \text{sex ratio at birth}) \qquad (1)$$

$$= 0.04534 \times \frac{69086302}{140431790} \times (1 + 1.002 = 0.0446 \text{ i.e } 44.6 \text{ per } 1,000 \text{ population}$$

For the total population, the growth rate is $= \frac{\ln\left(\frac{140431790}{88992220}\right)}{15} = 0.0304$

Then, the death rate for both sexes (total population) = birth rate − rate of increase

$$= 0.0446 - 0.0304 = 0.0142 \text{ i.e } 14.2 \text{ per } 1,000 \text{ population}$$

For more understanding of (1) see Siegel and Swanson, 2004. The summary of the estimated demographic parameters are shown in table 4.

The following procedures were used for estimating entries in columns 1 to 12 of tables 5 and 6. The model standard population was selected using the mortality levels and growth rates of the study population. The selection was done separately for males and females. The standard age distributions $P_{MS}(x)$ (cumulated to $C_{MS}(x)$) which is the proportion of males in standard population at age x, x+5 were compared with the study population $P_{MR}(x)$ (cumulated to $C_{MR}(x)$) which is the proportion of males reported population at age x, x+5 using a logit transformation which linearizes the relationship between age and the cumulated proportion of the population under each age. This was also done for $P_{FS}(x)$ (cumulated to $C_{FS}(x)$) and $P_{FR}(x)$ (cumulated to $C_{FR}(x)$) for females. The transformations are;

$$Y_{MR}(x) = \ln\left[\frac{1.0+C_{MR}(x)}{1.0-C_{MR}(x)}\right] \text{ and } Y_{MS}(x) = \ln\left[\frac{1.0+C_{MS}(x)}{1.0-C_{MS}(x)}\right].$$

Where; $C_{MS}(x)$ is the proportion of males reported under age x, $Y_{MR}(x)$ is the logit transformation of the proportion of males reported under age x and $Y_{MS}(x)$ is the transformation of males for standard population. A second degree polynomial was fitted for the $Y_{MR}(x)$ and $Y_{MS}(x)$ based on the assumption that the expected parabola passes through the origin.

The equation is of the form $Y_{MR}(x) = \propto Y_{MS}^2(x) + \beta Y_{MS}(x)$ (1)

The values of the parameters of the equation (1) were estimated from the solution of the matrix

$$X = A^{-1}B \qquad \text{Where; } X = \begin{bmatrix}\hat{\propto}\\ \hat{\beta}\end{bmatrix}; \qquad A = \begin{bmatrix}\overline{Y_{MS}^2(x)} & \overline{Y_{MS}(x)}\\ \overline{Y_{MS}^3(x)} & \overline{Y_{MS}^2(x)}\end{bmatrix}; \qquad B = \begin{bmatrix}\overline{Y_{MR}(x)}\\ \overline{Y_{MS}(x)Y_{MR}(x)}\end{bmatrix}$$

$$\overline{Y_{MS}^2(x)} = \sum Y_{MS}^2(x)\,;\ \overline{Y_{MS}(x)} = \sum Y_{MS}(x);\ \overline{Y_{MS}^3(x)} = \sum Y_{MS}^3(x)$$

$$\overline{Y_{MR}(x)} = \sum Y_{MR}(x)\,;\ \overline{Y_{MS}(x)Y_{MR}(x)} = \sum Y_{MS}(x)Y_{MR}(x)$$

This transformation was also applied to females by simply replacing the subscript M with F, thereby, producing two polynomial equations.

For males; $Y_{ME}(x) = -0.02848Y_{MS}^2(x) + 0.9817Y_{MS}(x)$

For females; $Y_{FE}(x) = -0.02848Y_{FS}^2(x) + 0.9817Y_{FS}(x)$

Where; $Y_{ME}(x)$ and $Y_{FE}(x)$ are the estimated proportion of males and females population under age x respectively. $Y_{MS}(x)$ is the logit transformation of the proportion of males standard population under age x. $Y_{FS}(x)$ is the logit transformation of the proportion of females standard population under age x. These equations were used to obtain expected Y-transformation for the total distribution by sex and thereafter, the corresponding expected proportions under age x were estimated using the equation:

$$\text{For males; } C_{ME}(x) = \frac{\exp(Y_{ME}(x)) - 1.0}{\exp(Y_{ME}(x)) + 1.0}$$

$$\text{For females } C_{FE}(x) = \frac{\exp(Y_{FE}(x)) - 1.0}{\exp(Y_{FE}(x)) + 1.0}$$

Where; $C_{ME}(x)$ and $C_{FE}(x)$ are the estimated proportion of males and females population under age x respectively. The $C_{ME}(x)$ and $C_{FE}(x)$ were later used to obtain proportions in each age x, x+5 for males ($C_{ME}(x, x+5)$) and ($C_{FE}(x, x+5)$) for females. The proportions in each five year age group for males and females were multiplied by the total population of males and females separately. This paved way for the estimated population in each five-year age groups.

TABLE 5: Derivation of Logit transformation and Adjusted Age Distribution, Males, Nigeria 2006 Census

Age x	**Males**	**$P_{MR}(x)$**	**$P_{MS}(x)$**	**PROPORTION under age x**		**LOGIT Transformation**		**$Y_{ME}(x)$**	**$C_{ME}(x)$**	**C_{ME} (x,x+5)**	**Adjusted Male Age Distribution**
				$C_{MR}(x)$	**$C_{MS}(x)$**	**$Y_{MR}(x)$**	**$Y_{MS}(x)$**				
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
5	11569218	.1622	.1935	.1622	.1935	.3273	.3919	.3804	.1879	.1879	13405817
10	10388611	.1456	.1526	.3078	.3461	.6362	.7220	.6940	.3337	.1458	10402172
15	8504319	.1192	.1266	.4270	.4727	.9124	1.0271	.9782	.4535	.1198	8547189
20	7536532	.1056	.1057	.5326	.5784	1.1875	1.3201	1.2463	.5533	.0998	7120280
25	6237549	.0874	.0877	.6200	.6661	1.4500	1.6074	1.5044	.6365	.0832	5935945
30	5534458	.0776	.0722	.6976	.7383	1.7252	1.8935	1.7567	.7056	.0691	4929973
35	4505186	.0631	.0594	.7607	.7977	1.9957	2.1845	2.0086	.7634	.0578	4123769
40	3661133	.0513	.0487	.8120	.8464	2.2657	2.4866	2.2650	.8119	.0485	3460256
45	3395489	.0476	.0396	.8596	.8860	2.5836	2.8060	2.5304	.8525	.0406	2896627
50	2561526	.0359	.0319	.8955	.9179	2.8981	3.1510	2.8106	.8865	.0340	2425747
55	2363937	.0331	.0253	.9286	.9432	3.2963	3.5326	3.1125	.9148	.0283	2019077
60	1189770	.0167	.0196	.9453	.9628	3.5713	3.9658	3.4453	.9382	.0234	1669484
65	1363219	.0191	.0147	.9644	.9775	4.0106	4.4761	3.8236	.9572	.0190	1355564
70	628436	.0088	.0104	.9732	.9879	4.2990	5.1016	4.2670	.9723	.0151	1077317
75	765988	.0107	.0066	.9839	.9945	4.8140	5.8934	4.7964	.9836	.0113	806204
80	327416	.0046	.0036	.9885	.9981	5.1528	6.9581	5.4519	.9915	.0079	563629

TABLE 5: Derivation of Logit transformation and Adjusted Age Distribution, Females, Nigeria 2006 Census

Age x	Females	$P_{FR}(x)$	$P_{FS}(x)$	PROPORTION under age x		LOGIT Transformation		$Y_{FE}(x)$	$C_{FE}(x)$	C_{FE} (x,x+5)	Adjusted Female Age Distribution
				$C_{FR}(x)$	$C_{FS}(x)$	$Y_{FR}(x)$	$Y_{FS}(x)$				
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
5	11025749	.1596	.1825	.1596	.1825	.3220	.3691	.3957	.1953	.1953	13492555
10	9616769	.1392	.1461	.2988	.3286	.6164	.6825	.7248	.3473	.1520	10501118
15	7631631	.1105	.1224	.4093	.4510	.8695	.9719	1.0232	.4712	.1239	8559793
20	7362887	.1066	.1035	.5159	.5545	1.1415	1.2497	1.3046	.5732	.1020	7046803
25	7197530	.1042	.0871	.6201	.6416	1.4503	1.5218	1.5754	.6571	.0839	5796341
30	6676968	.0966	.0731	.7167	.7147	1.8017	1.7935	1.8411	.7262	.0693	4787681
35	4962352	.0718	.0611	.7885	.7758	2.1349	2.0695	2.1062	.7830	.0566	3910285
40	3670622	.0531	.0507	.8416	.8265	2.4533	2.3540	2.3744	.8297	.0467	3226330
45	3060981	.0443	.0419	.8859	.8684	2.8051	2.6531	2.6508	.8681	.0384	2652914
50	2029767	.0294	.0345	.9153	.9029	3.1185	2.9754	2.9422	.8998	.0298	2058772
55	1885282	.0273	.0281	.9426	.9310	3.5217	3.3317	3.2567	.9258	.0279	1927508
60	876477	.0127	.0225	.9553	.9535	3.7783	3.7379	3.6053	.9471	.0213	1471538
65	1087067	.0157	.0175	.9710	.9710	4.2190	4.2190	4.0046	.9642	.0171	1181376
70	522612	.0076	.0128	.9786	.9838	4.5268	4.8078	4.4731	.9774	.0132	911939
75	564609	.0082	.0086	.9868	.9924	5.0141	5.5689	5.0460	.9872	.0098	677046
80	252422	.0037	.0050	.9905	.9974	5.3448	6.6441	5.7922	.9939	.0067	462878

In order to see the level of errors in each five-year age group, the estimated population in each age group was subtracted from the reported values. Negative and positive values of error are indication of the level of under-reporting and over-reporting of people in that age group. For comparison of the level of errors for males and females, percentage errors were computed for each of the five year age groups as shown in columns (3) and (5) of table 6. This was done for the total population. To provide single index for comparing the level of errors between the two sexes, mean absolute deviation (MAD) was also computed for males and females. The result shows that MAD was higher in ages reported by females (14.5) than those reported by males (9.7). This means overall, females ages are more poorly reported than males reported ages.

$$\text{Mean absolute deviation}_{\text{Male}} = \frac{\sum|D_x^M|}{n-2}$$

$$= \frac{\sum|\text{column3}|}{n-2} = \frac{135.26}{16-2} = 9.7$$

$$\text{Mean absolute deviation}_{\text{Female}} = \frac{\sum|D_x^F|}{n-2}$$

$$= \frac{\sum|\text{column5}|}{n-2} = \frac{203.06}{16-2} = 14.5$$

TABLE 6: Derivation of Errors and Adjusted Age Distribution, Both sexes, Nigeria 2006 Census

AGE (X)	$ERROR_M$	$\%ERROR_M$ (D_x^M)	$ERROR_F$	$\%ERROR_F$ (D_x^F)	TOTAL POPULATION				
					Reported Age Distribution	Adjusted Age Distribution	%ERROR	$P_T(x)$	Adj. $P_T(x)$
1	2	3	4	5	6	7	8	9	10
5	-1836599	-15.87	-2466806	-22.37	22594967	26898372	-19.05	.1609	.1915
10	-13561	-.13	-884349	-9.20	20005380	20903290	-4.49	.1425	.1489
15	-42870	-.50	-928162	-12.16	16135950	17106982	-6.02	.1149	.1218
20	416252	5.52	316084	4.29	14899419	14167083	4.92	.1061	.1009
25	301604	4.84	1401189	19.47	13435079	11732285	12.67	.0957	.0835
30	604485	10.92	1889287	28.30	12211426	9717654	20.42	.0870	.0692
35	381417	8.47	1052067	21.20	9467538	8034054	15.14	.0674	.0572
40	200877	5.49	444292	12.10	7331755	6686586	8.80	.0522	.0476
45	498862	14.69	408067	13.33	6456470	5549541	14.05	.0460	.0395
50	135779	5.30	-29005	-1.43	4591293	4484518	2.33	.0327	.0319
55	344860	14.59	-42226	-2.24	4249219	3946585	7.12	.0303	.0281
60	-479714	-40.32	-595061	-67.89	2066247	3141023	-52.02	.0147	.0224
65	7655	.56	-94309	-8.68	2450286	2536940	-3.54	.0174	.0181
70	-448881	-71.43	-389327	-74.50	1151048	1989256	-72.82	.0082	.0142
75	-40216	-5.25	-112437	-19.91	1330597	1483250	-11.47	.0095	.0106
80	-236213	-72.14	-210456	-83.37	579838	1026508	-77.03	.0041	.0073
Total		**-135.26**		**-203.06**					

$P_T(x)$ is the proportion of total population at age x, x+5 and Adj. $P_T(x)$ is the adjusted proportion of total population at age x, x+5.

Figure 1 shows the comparison of the percentage errors between ages reported during census by males and females population in Nigeria. The data shows gross age misstatement across almost all the reported ages. In actual fact, if ages had been appropriately reported, the two curves should be straight lines passing through the horizontal axis (Age axis) as seen in the figure between the age 10-15 for males and 50-55 females. In the figure, errors in age reporting are more pronounced among females between ages 0 and 45 years than males. This is because the percentage error curve deviates slightly from horizontal axis than the females. However, between 45 and 55 years, females are more likely to have reported their true ages during this interval than males. At ages 55 years and above, ages reported by males are consistently less prone to errors than those reported by females.

The patterns of either underreporting or over-reporting of ages are similar for both sexes. Figure 2 shows that for the population of Nigeria as a whole, there is tendency to under-report ages at ages 15 to 17 years and above 55 years, whereas, gross over-reporting of age are seen in ages between 18 and 55 years. Similar pattern existed when the reported age distribution was compared with the standard as shown in figure 3.

FIGURE 1: Percentage Error in Age Reporting in Nigeria (Males and Females), 2006 Census

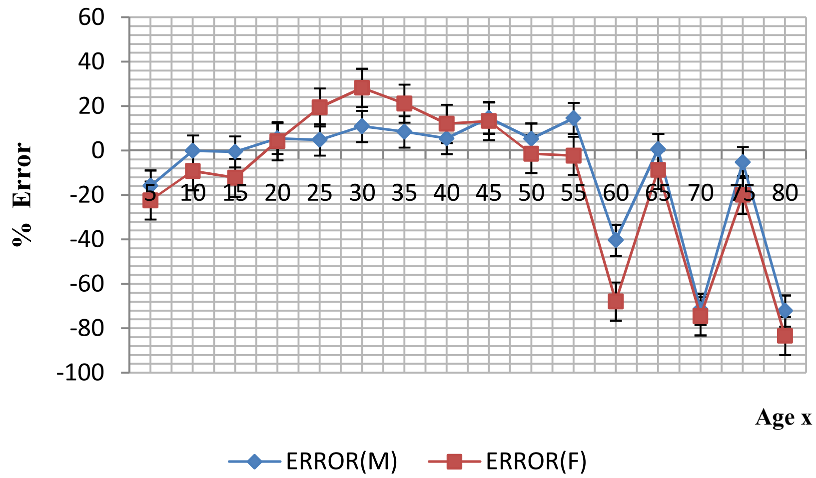


FIGURE 2: Percentage Error in Age Reporting in Nigeria, 2006 Census

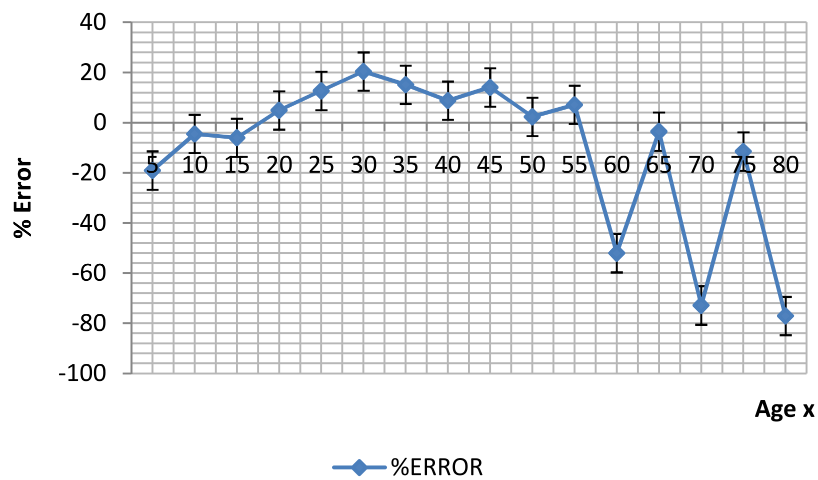
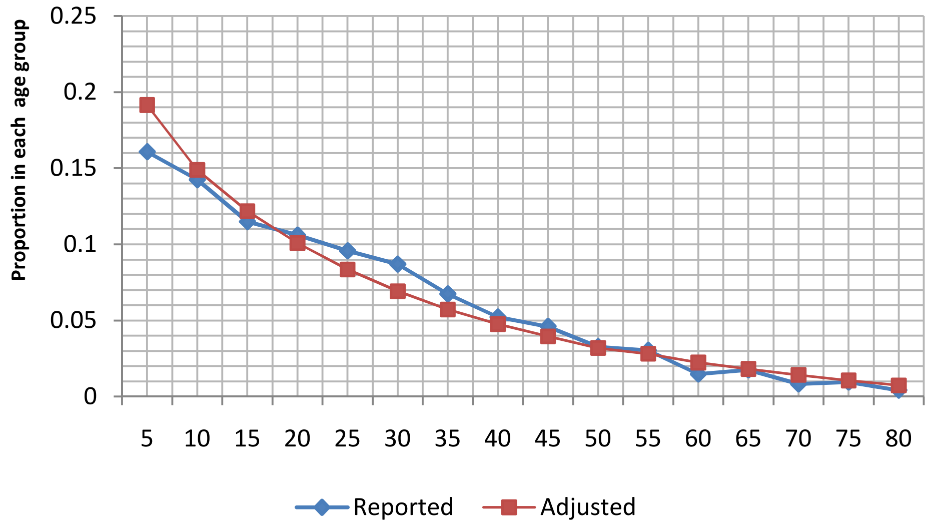


FIGURE 3:A comparison of the reported and adjusted proportion in each age group, Nigeria, 2006



Discussion:

Precise data on age, though essential to any demographic enquiry, are difficult to gather in many developing countries. This adverse situation is a product of a long-established way of life that makes little demand on the people to be aware of their ages and a polity that puts little attempt in registering vital events. Under these conditions, stated ages in a census or a survey tend to be predisposed by individual's physiological features (height, size and hair color), personal characteristics (marital status, education attainment and number of children ever born), penchant for digits such as 0 and 5, subjective biases to either over or under-report age. These introduce both random and systematic errors in age data. Ages can be grouped or smoothed to reduce the effect of digital preference and random errors in the data, but these are poor solutions if the aim is to minimize distortions that result from systematic over or under-reporting of age. Such systematic misstatements alter the true slopes of the age distribution and create intractable problems for the indirect assessment of basic demographic parameters. However, the presence of such errors can be detected by comparing the recorded age distribution with stable age distribution and through simulation will provide a refined age distribution as clearly shown in this study.

The study uses data on census which was conducted in 2006. Data analysis was done using Coale and Trussel method of estimating childhood mortality and Coale Demeny model life table based on specific assumptions. However, before using the approach the researchers ensured that all the models assumptions were satisfied as pointed out in the methodology section of this report. The results showed that male birth (49.4) and death rates (18.0) were higher than female birth (45.3) and death rates (16.0). This is in agreement with recent report of Nigeria Demographic and Health Survey, 2008 (NDHS, 2008) which puts the sex ratio at birth (ratio of males to females at birth) as 1.002. The expectation of life at birth was also lower for males (46.7 years) than females (50 years).

Data adjustment reveals that there is gross age misstatement across all ages. Poor vital registration system and high level of illiteracy in Nigeria (39.9% and 18.9% of the population had no and some primary education respectively, NDHS, 2006) can be a possible explanation in this regard. Errors in age reporting are more pronounced among females between ages 0 and 45 years than males. It is a common thing among African women to over-report their ages when they are younger in order to fit in to specific social group in their community and under-report ages at older ages to make people particularly men believe they are still younger. The net effect of both age under-reporting and over-reporting within this age interval make misstatement of age more rampant among females than males. However, between 45 and 55 years, females are more likely to have reported their true ages than males. At ages 55 years and above ages reported by males are consistently less prone to errors than those reported by females. In general, MAD was higher in ages reported by females (14.5) than those reported by males (9.7). This means females reported ages are more poorly reported than males reported ages across all ages. The patterns of either underreporting or over-reporting of ages are similar for both sexes. In Nigeria as a whole (both sexes combined), there is tendency to under-report ages at ages 15 to 17 years and above 55 years, whereas, gross over-reporting of age are seen in ages between 18 and 55 years.

There is one of the numerous reasons to suspect that in Nigeria, the tendency to exaggerate age becomes particularly strong at older ages among both men and women. Again, the results presented here-in do not contradict the common observation that the population reported at older ages is composed of persons younger than the stated age, but rather question its frequent interpretation that there is greater exaggeration of age among older persons.

Conclusion

Despite the care taken to ensure the quality of the data collected by census enumeration and registration, the final tabulations sometimes give obvious indications of errors in the basic information, more often; the errors can only be inferred. Although age misreporting and selective under-enumeration will continue to plague demographic studies (Ewbank, 1981), the recent evidence suggests that much can be done in terms of adjusting data for age errors.

Evaluations of these errors furnish those who use census data with a measure of their accuracy and are also important to census office as a guide to improvements in their procedures. More generally, the findings in such evaluations should be helpful to other offices that conduct surveys. Apart from providing some valuable insights into the pattern of age misreporting in developing countries, the approach used in this study can be useful in simulating the probable pattern of errors in estimates of demographic parameters and age distribution. Priority attention should be accorded to all preparatory census activities and population surveys on a regular basis so as to avoid problems of age misreporting.

References

Chevan, A and SurtherLand M. 2009. Revisiting Das Gupta: Refinement and Extension of Standardization and Decomposition. Demography. Volume 46-Number 3, Aug. 2009: 429-449.

Davern, M, Ruggles, S, Swenson, T et al. 2009. Drawing statistical inferences from historical census data, 1850-1950. Demography. Volume 46-Number 3, Aug. 2009: 589-603.

Ewbank, D.C. 1981. Age Misreporting and Age Structure Under-enumeration: Sources, Patterns and Consequences for Demographic Analysis. Washington, DC; National Academy Press.

National Population Commission. 2008. Nigeria Demographic and Health Survey. Macro-International Maryland, U.S.A

Raymer, J and Rogers, A. 2007. Using age and spatial flow structures in the indirect estimation of migration streams. Demography. Volume 44-Number 2, May. 2007: 199-223.

Siegel, J.S and Swanson, D.A. 2004. The methods and materials of demography. Amsterdam: Elsevier Academic Press.

United Nations Publication. 1983. Manual X Indirect Technique for demographic Estimation.