An assessment of the age reporting in Tanzania population census 2012

A Sathiya Susuman[1]

[1]Professor (Associate) in Demography, Department of Statistics and Population Studies, University of the Western Cape,
Cape Town, South Africa
sappunni@uwc.ac.za

Hamisi F Hamisi[2],

[2] Demographer / Research Associate, Hope Centre for Children, Girls and Women, Msasani - Dar es Salaam | Tanzania
fakhijr@gmail.com

Siaka Lougue[3]

[3] Lecturer, School of Mathematics, Statistics and Computer Science
University of Kwazulu Natal, Westville Campus
siaka@aiims.ac.za

Nigatu Regassa[4]

[4] Professor, School of Environment, Gender and Development, Hawassa University, Ethiopia
negyon@yahoo.com

Kanayo Ogujiuba[5]

[5] Senior Research Fellow,Research Division,National Institute for Legislative Studies, Abuja – Nigeria,
kannyog@gmail.com

Eric Mwambene[6]

[6] Professor, Department of Mathematics, University of the Western Cape Cape Town,
South Africa
emwambene@uwc.ac.za

A Sathiya Susuman[1]

Correspodence: sappunni@uwc.ac.za

Abstract

The objective of this paper is to provide data users with a worldwide assessment of the age reporting in the Tanzania Population Census 2012 data. Many demographic and socio-economic data are age-sex attributed. However, a variety of irregularities and misstatements are noted with respect to age-related data and sex ratio data because of its biological differences between the genders. Noting the misstatement / misreporting, inconsistence of age data regardless of its significant importance in demographic and epidemiological studies, this study assess the quality of the 2012 Tanzania Population and Housing Census data relative to age. Data were downloaded from Tanzania National Bureau of Statistics. Age heaping and digit preference were measured using summary indices viz., Whipple's index, Myers' blended index, and Age-Sex Accuracy index. The recorded Whipple's index for both sexes was 154.43, where males had the lower index of about 152.65 while females had the higher index of about 156.07. For Myers' blended index, the prefrences were at digits '0' and '5' while avoidance were at digits '1' and '3' for both sexes. Finally, the age-sex index stood at 59.8 where the sex ratio score was 5.82, and the age ratio scores were 20.89 and 21.4 for males and female respectively. The evaluation of the 2012 Population Housing Censes data using the demographic techniques has qualified the data as of poor quality as a result of systematic heaping and digit preferences/avoidances in recorded age. Thus, innovative methods in data collection along with measuring and minimizing errors using statistical techniques should be used to ensure accuracy of age data.

Keywords: Age heaping, Digit preference/avoidance, Summary indices, Whipple's Index, Myer's Index, Age-sex accuracy index.

INTRODUCTION

Age is an important demographic variable utilized for descriptive and statistical analyses of population structure and to forecast of population growth. Many demographic and socio-economic data are age and sex attributed[1]. The latter is understandable because of the biological differences between males and females. A variety of anomalies and misstatement are noted with respect to age-related data [1, 2] and thus, age tends to be more susceptible to such anomalies regardless of its crucial significance in demographic and epidemiological analysis [2, 3]. Likewise, (4,5) report that age data tend to be more misreported than sex, yet its very important variable in health related studies of demography and epidemiology; it's also considered as one of the constant components in community-based surveys. Thus, the misreporting/misstatement of age constitute one of the most demographic challenges [6].

_____

[1] _Statistical computing and graphic language for statistical analysis_

The accuracy of demographic data, especially that of age, varies from one country to another and depends on numerous factors [3, 7 and 8]. The deficiencies are mostly in developing countries due to lack of administrative machinery and problems in the collection and tabulation of data [9]. Other common errors arise from under-reporting of children less than one year, and over-statement to exact age at very advance age to qualify for political or socio-cultural affairs[1]. Factors like ignorance of one's true age, low numeracy skills, open hostility to some inquiry. Also, cultural preferences were pronounced as source of age heaping in [2, 3 and 10] reports.

Anomalies in demographic data comprise of mainly two types _viz._, (i) coverage error and (ii) content error. Coverage errors occur when a person is being counted twice or missed during enumeration. The scope of this paper will however consider the second type (i.e. content errors) that occur when persons' characteristic during survey/census/registration of vital event are incorrectly reported or tabulated. Leaving apart other subcategories under this type, the study's objective is on respondents' voluntary or involuntary failure to give the appropriate information as required. The classic example is the misstatement/misreporting of age. To detect it, demographers have developed various techniques to ascertain the quality of age data [1]. Age heaping summary indices (_viz.,_ Myer's blended index, Whipple's index, and age-sex accuracy index) are among the indices developed to detect preference or avoidance of certain end digits in reported age (i.e. they are used to measure quality and consistency of age data).

Age anomalies or heaping is a common phenomenon of misstatement or misreporting of age data. Due to this, age data frequently displays excess frequencies at 'round' or 'attractive' age digits. For example, attractions or repulsion is normally on even numbers or multiple of "5". Consequently, this study uses age heaping summary indices _viz.,_ Whipple's index, Myer's blended index, and Age-sex accuracy index to assess the quality and accuracy of age data reported from 2012 PHC.

Study Methods

This paper uses age and sex data from 2012 Population Housing Census from NBS made available at http://www.nbs.go.tz.

Age heaping summary indices generally assumes that individual ages are evenly distributed over specific age groups and by extension entire age spectrum [3, 6]. However, each index has its own technique of calculation, as presented below. For Whipple's Index, data are imported into R-language[1] for analysis, and detailed codes were used. The remaining indices _viz.,_ Myer's index and age-sex accuracy indices are calculated in their natural way using Microsoft Excel spreadsheet.

Whipple's index measures heaping on ages ending in "0" and "5" reported in single years of age returned between "23" and "62" years inclusive [6]. Whipple's index is calculated by summing the number of persons in

_____

[1] _Statistical computing and graphic language for statistical analysis_

the age bracket 23 and 62 inclusive, and calculating the ratio of reported ages ending at 0 and 5 to 1/5 of the total population or sample in the same age groups, multiplied by 100 [6], mathematically written as:

$$WI = \Sigma(P25,P30,P35,...,P60) \Big/ \frac{1}{5}\Sigma(P23,P24,P25,...,P62) \qquad (1)$$

In the absence of heaping or preferences at "0" or "5", the index will have a value of 100 while a value of 500 indicate a complete heaping or preference at "0" or "5". The inferences about the index are; <105 = highly accurate data; 105 – 109.9 fairly accurate data; 110 – 124.9 = approximate data; 125 – 174.9 = rough data; and ≥ 175 = very rough data [4, 12].

Myer's blended index calculates the preference or avoidance of ages reported in any of the 10 digits expressed as percentages [13]. In the absence of systematic irregularities in age reporting in any of the digits ending at "0" through "9", the blended sum at each terminal digit should be approximately equal to 10% of the total blended population [4, 6]. If the sum of the blended index exceeds 10% then this indicates over selection of age ending in that digit (i.e. digit preference). If it is less than 10% it indicates under selection of age ending in that digit (i.e. digit avoidance).

The following are steps in calculating Myer's blended index as referred from [4, 12]:

1. Sum of populations ending in each digit over the whole range starting with the lower limit of range (i.e. 10, 20, 30, 40; 11, 21, 31,…; 12, 22, 32…).
2. Ascertain sum excluding the first populations combined in step 1 (i.e. 20, 30, 40,…;21, 31, 41,…).
3. Weigh the sum in step 1 and 2 and add the results to obtain a blended population (i.e. weights 1 and 9 for 0 digit, weights 2 and 8 for 1 digit etc.).
4. Convert the distributions in step 3 into percentages.
5. Take the deviations of each percentage in step 4 from 10 that are the expected value for each percentage.
6. A summary index of preference or avoidance for all terminal digits is derived as one half of the sum of the deviation from 10% each without regards to signs.

Age-Sex Accuracy Index (ASAI) measures the level of the quality of age and sex data in five-year age groups [2, 11]. The index is calculated as three times the average of sex ratio differences plus the average of the deviation from 100 of male and female age ratios, algebraically written as:

$$ASAI = 3 * SRS + \left(ARS_m + ARS_f\right) \qquad (2)$$

(where, $ASAI$ = age sex accuracy index, SRS = sex ratio scores, ARS_m = age ratio scores for males, and ARS_f = age ratio scores for females).

United Nations recommendations for scaling the estimate of reliability of the data is categorized as 0 – 19.9 for accurate data, 20 – 39.9 for inaccurate data, and above 40 for highly inaccurate data [12]. When working on the ASAI, two important parts *viz.,* (i) sex ratio scores and (ii) age ratio scores as shown in equation 2 above are calculated separately as follows:

Age ratio score (ARS): ratio of the population in the given age group to one-half of the population in the two adjacent groups multiplied by 100 [2, 3]. The ratio score is calculated separate for each male and female age group. Mathematically, let $_nARS_x$ be the age ratio score for age group x to $x+n$, $_nP_x$ be the age group from age x to $x+n$, $_nP_{x-n}$ and $_nP_{x+n}$ be the proceeding and succeeding age groups respectively, then:

$$_nARS_x = \left[2*_np_x \Big/ \left(_np_{x-n} + _np_{x+n}\right)\right]*100 \qquad (3)$$

The computed age ratio is compared with the expected value of 100. The discrepancy at each age group is a measure of net age misreporting [2, 3]. Taking the average deviation (regardless of the sign) from 100 of the age ratios and summing over all the age groups derives the measure of the accuracy of an age distribution [4]. An

age ratio of under 100 implies that members of the age group are either selectively under enumerated or that errors in age reporting resulted in misclassifying persons who belong to the age group. A ratio of more than 100 suggests the opposite of either one or the other or both of these conditions [2].

Sex ratio score (SRS): ratio of the population of males in the given age group to the population of females in that age group multiplied by 100 [1]. Mathematically, let $_nSRS_x$ be the sex ratio score of age group x to $x+n$, $_n p_x^m$ be population of males aged x to $x+n$, and $_n p_x^f$ for female population in similar age group x to $x+n$. Then:

$$_nSRS_x = \frac{_n p_x^m}{_n p_x^f} \times 100$$

(4)

The accuracy sex ratio index is obtained through summing over the successive differences between one age group and the next one (irrespective of the sign) and then taking the average of the summation [14].

R-Codes for Whipple's Index Calculation

```
> getwd ()  #returns a working directory
> TP <- read.csv ("Tanzania.csv", header=TRUE)  #returns raw data of Tanzania population in 2012 in single years
(TP = Tanzania Population).
> numerator <- sum(TP[26+0:7*5, "BS"])   #returns sum of population of ages 25,30,… to 60 for both sexes (BS =
Both Sexes).
> numerator  #sums of the numerator
[1] 4945941
> denominator <- sum(TP[24:63, "BS"])    #returns sum of population of ages from 23 to 62 for both sexes
> denominator  #sums of denominator
[1] 16013058
> WIbs = (numerator/denominator)*5*100  #returns the Whipple's Index for both sexes (WIbs = Whipple's Index for
both sexes).
> WIbs  #Whipple's Index for both sexes
[1] 154.4346
> NM <- sum (TP[26+0:7*5, "M"]) #returns sum of population of ages 25,30,… to 60 for males (NM = Numerator for
Males)
> NM  #sums numerator for male populations
[1] 2333121
> DM <- sum (TP[24:63, "M"]) #returns sum of population of ages from 23 to 62 for males (DM = Denominator for
Males).
> DM  #sums denominator for male populations
[1] 7642140
> WIm = (NM/DM)*5*100 #returns the summary of the Whipple's Index for male populations.
> WIm  # Whipple's Index for male populations
[1] 152.6484
> NF <- sum (TP[26+0:7*5, "F"])  #returns sum of population of ages 25, 30, …, to 60 for females (NF = Numerator
for Female).
> NF  # sums numerator for female populations
[1] 2612820
> DF <- sum (TP[24:63, "F"]) #returns sum of  population of ages from 23 to 62 for females (DF = Denominator
Female).
> DF # sums denominator female populations
[1] 8370918
> WIf = (NF/DF)*5*100   #returns the summary of the Whipple's Index for female populations.
> WIf  # Whipple's Index for female populations
[1] 156.0653
```

Ethical Considerations

This study used secondary data from the *2012 Tanzania Population and Housing Census data*. Prior to using these data, permission was obtained from *Tanzania Population Census*, which has allowed us to download the data from their Web Site available at http://www.nbs.go.tz.

Results

The Whipple's index of age data from PHC 2012 is presented in **Table 1** below. The table indicates Whipple's index for both sex is 154.43. Whipple's index for male is 152.65, and for female is 156.07. All of these values fall under the category of 125 – 174.9 that indicate the quality of age reporting in Tanzania's 2012 PHC was rough.

Though both data are rough, the quality of male is slightly better compared to that of female. Three indices *viz.,* Tables 1, 2 and 3 presented separately here.

Table 1 Myers Blended Index for both male and female in Tanzania 2012 PHC

1	2	3	4	5	6	7	8	9	10	11
0	5407161	1	5407161	4066889	9	36602001	42009162	16.50	6.50	Preference
	2407370	2	4814740	1455843	8	11646744	16461484	6.46	-3.54	Avoidance
2	3975916	3	11927748	2532193	7	17725351	29653099	11.65	1.65	Preference
3	2561955	4	10247820	1539119	6	9234714	19482534	7.65	-2.35	Avoidance
4	2662590	5	13312950	1628361	5	8141805	21454755	8.43	-1.57	Avoidance
5	3547724	6	21286344	2576009	4	10304036	31590380	12.41	2.41	Preference
6	2517393	7	17621751	1582296	3	4746888	22368639	8.78	-1.22	Avoidance
7	2296767	8	18374136	1424509	2	2849018	21223154	8.33	-1.67	Avoidance
8	3115597	9	28040373	2057635	1	2057635	30098008	11.82	1.82	Preference
9	2029528	10	20295280	1257724	0	0	20295280	7.97	-2.03	Avoidance
All							**254636495**	100.00	24.76	
Both Male and Female Whipple's index for Tanzania = 154.43										

Note: Terminal digit(1), Sum (10 - 79) (2), Coefficients(3), Products(4)=(2)*(3), Sum (20 - 79) (5), Coefficients(6),Products(7)=(5)*(6),Blended sum(8)=(4)+(7),Percentage distribution (%)(9),Deviation from 10%(10),Remarks(11)

Table 2 Myers Blended Index for male population in Tanzania 2012

1	2	3	4	5	6	7	8	9	10	11
0	2534431	1	2534431	1862754	9	16764786	19299217	15.86	5.86	Preference
1	1178829	2	2357658	710062	8	5680496	8038154	6.61	-3.39	Avoidance
2	1957275	3	5871825	1227617	7	8593319	14465144	11.89	1.89	Preference
3	1250171	4	5000684	740099	6	4440594	9441278	7.76	-2.24	Avoidance
4	1303099	5	6515495	790750	5	3953750	10469245	8.61	-1.39	Avoidance
5	1720088	6	10320528	1229483	4	4917932	15238460	12.53	2.53	Preference
6	1218886	7	8532202	762250	3	2286750	10818952	8.89	-1.11	Avoidance
7	1127617	8	9020936	698623	2	1397246	10418182	8.56	-1.44	Avoidance
8	1477678	9	13299102	969665	1	969665	14268767	11.73	1.73	Preference
9	920211	10	9202110	565483	0	0	9202110	7.56	-2.44	Avoidance
All							**121659509**	100.00	24.02	
Male Whipple's index for Tanzania = 152.65										

Note: Terminal digit(1), Sum (10 - 79) (2), Coefficients(3), Products(4)=(2)*(3), Sum (20 - 79) (5), Coefficients(6),Products(7)=(5)*(6),Blended sum(8)=(4)+(7),Percentage distribution (%)(9),Deviation from 10%(10),Remarks(11)

Table 3 Myers Blended Index for female population in Tanzania 2012

1	2	3	4	5	6	7	8	9	10	11
0	2872730	1	2872730	2204135	9	19837215	22709945	17.08	7.08	Preference
1	1228541	2	2457082	745841	8	5966728	8423810	6.33	-3.67	Avoidance
2	2018641	3	6055923	1304576	7	9132032	15187955	11.42	1.42	Preference
3	1311784	4	5247136	799020	6	4794120	10041256	7.55	-2.45	Avoidance
4	1359491	5	6797455	837611	5	4188055	10985510	8.26	-1.74	Avoidance
5	1827636	6	10965816	1346526	4	5386104	16351920	12.30	2.30	Preference
6	1298507	7	9089549	820046	3	2460138	11549687	8.69	-1.31	Avoidance
7	1169150	8	9353200	725886	2	1451772	10804972	8.13	-1.87	Avoidance
8	1637919	9	14741271	1087970	1	1087970	15829241	11.90	1.90	Preference
9	1109317	10	11093170	692241	0	0	11093170	8.34	-1.66	Avoidance
Total							**132977466**	100.00	25.4	
Female Whipple's index for Tanzania = 156.07										

Note: Terminal digit(1), Sum (10 - 79) (2), Coefficients(3), Products(4)=(2)*(3), Sum (20 - 79) (5), Coefficients(6),Products(7)=(5)*(6),Blended sum(8)=(4)+(7),Percentage distribution (%)(9),Deviation from 10%(10),Remarks(11)

Considering the three indices in Table 1, 2 and 3, our results indicate some anomalies for the age and sex data reported in 2012 PHC. There was age heaping at ages ending with certain terminal digits, indicating preferences or avoidances in reporting such ages. The approximation of age manifests itself in the phenomenon of age heaping in self-reported or proxy age data. Individuals lacking knowledge of their age rarely state this openly. Instead they choose a figure they think is plausible [4, 12]. Their choice is not random but has a systematic tendency to prefer attractive numbers such as those ending with '0' and '5' or even numbers or in some societies, numbers with other specific terminal digits due to cultural beliefs [4]. Consequently, age heaping in the country

indicates ignorance of one's own age or a tendency to round ages [2, 3]. It should however be noted that, high heaping observed in female to male might indicate that women in Tanzania are more illiterate and might lack numeracy skills [9, 15]. High anomalies in age reported data for female might be the result of proxy reporting of age data by male who mostly tend to be the head of household in developing countries. Thus, the evaluation of the 2012 PHC data using the demographic techniques above has finally qualified age related information of poor quality as the results of systematic heaping and digit preferences/avoidances. This means that, age awareness in the country is quite low and many have only a vague idea about their age. The impact of such misreporting can lead to misclassification bias and wrong assessment of demographic rates and interfere with planning effective interventions [9, 15]. It should however be noted that, these demographic techniques are very quick and inexpensive on general quality of data and provide some evidence of errors on specific segments of the population but not the magnitude of those errors.

Therefore, this study recommends the need to work with other assessment methods [12] to ensure the accuracy of the age data. For example, methods highlighted in [3, 4 and 17] of using local calendar of event were proved to be remarkably useful in elderly populations for generating high quality age related information. Thus, other innovative methods of data collection that ensure the accuracy of age data in all age categories are recommended [17]. Other methods, like age smoothing are recommended to correct age related information. But the most important precaution should be to collect data using methodologies that will ensure quality of age data. Notable example is the use date of birth (in the form of date, month, year) to overcome the challenges of using completed year (year at last birthday) that proved to be a source of age heaping since individual might round to the nearest age [2]. Alternatively, vital registration system should be adequately improved so as to get reliable demographic information to supplements census and survey data.

Figure 1 and 2 below indicate age in singe years against population and age grouped in 5-years age group respectively. Both figure show there is anomalies in age reported even if after the population being grouped into 5-years age groups.

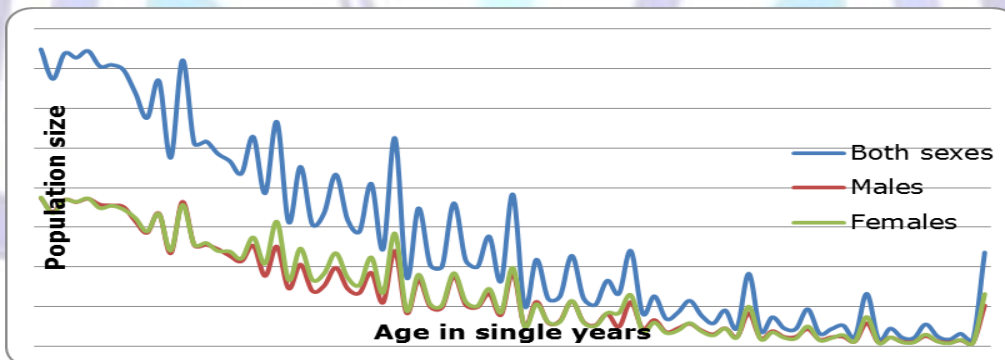Figure1 Age in single years against population size

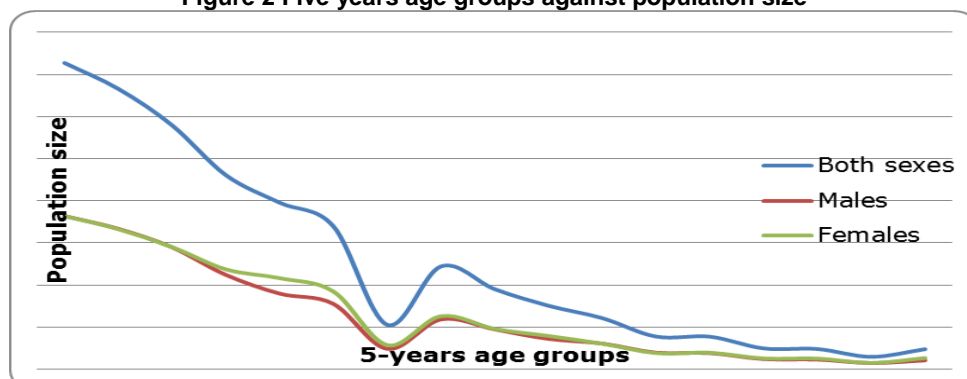**Figure 2 Five years age groups against population size**

June 03, 2015

Figure 3 presents the results of Myer's Blended Index for Tanzania 2012 PHC age data. Minus signs indicate avoidance of certain terminal digits while positive signs are for preference to those terminal digits. It can be clearly seen from the table and figure that, preference is at terminal digits '0', '2', '5', and '8'. Avoidance is at terminal digits '1', '3', '4', '6', '7', and '9'. Also indicates that high preference is at terminal digit '0' and much anomalies are observed more on females than males, while avoidance is at terminal digit '1'. Again much deviation is observed on females than males.

Figure 3 Myer's blended index for Tanzania 2012 Population Census age data
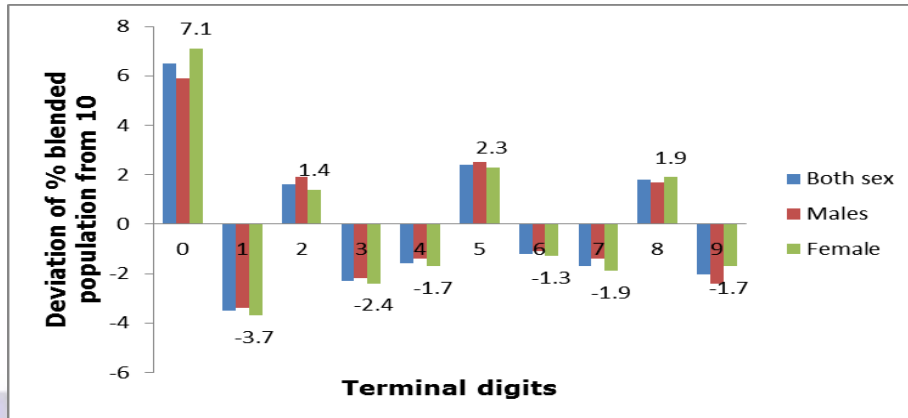
Detailed calculations and tabulations of the procedure for Myer's Blended Index for Tanzania 2012 PHC are presented in the **table 4.** Calculation of Age Sex Accuracy Index presented in the **table 5**.

Table 4 Myer's Blended Index for Tanzania 2012 PHC age data

Terminal digits	Deviation of % blended population from 10		
	Both sex	**Male**	**Female**
0	6.5	5.9	7.1
1	-3.5	-3.4	-3.7
2	1.6	1.9	1.4
3	-2.3	-2.2	-2.4
4	-1.6	-1.4	-1.7
5	2.4	2.5	2.3
6	-1.2	-1.1	-1.3
7	-1.7	-1.4	-1.9
8	1.8	1.7	1.9
9	-2.03	-2.4	-1.7

Table 5 Calculation of Age Sex Accuracy Index in Tanzania 2012

Age (1)	Population		Age ratios		Sex ratios (6)	Age ratio deviation		Differences of sex ratio (9) abs[cell(6,i)-cell(6,i-1))]
	Male (2)	Female (3)	Male (4)	Female (5)		Male (7) abs[col(4)-100]	Female (8) abs[col(5)-100]	
0 - 4	3,637,982	3,635,850			100.06			-0.20
5 - 9	3,333,835	3,325,202	102.1	101.75	100.26	2.1	1.75	0.52
10 - 14	2,892,583	2,900,004	103.81	101.84	99.74	3.81	1.84	5.27
15 - 19	2,238,976	2,369,860	95.57	93.65	94.48	-4.43	-6.35	11.50
20 - 24	1,793,053	2,160,986	94.68	102.57	82.97	-5.32	2.57	-1.03
25 - 29	1,548,762	1,843,732	136.15	135.03	84.00	36.15	35.03	-0.57
30 - 34	481,986	569,920	35.29	36.7449	84.57	-64.71	-63.2551	-9.42
35 - 39	1,182,651	1,258,301	165.87	165.02	93.99	65.87	65.02	-4.86
40 - 44	944,033	955,081	99.32	93.37	98.84	-0.68	-6.63	7.63
45 - 49	718,372	787,547	92.61	101.05	91.22	-7.39	1.05	-9.40
50 - 54	607,361	603,666	109.39	103.25	100.61	9.39	3.25	-2.10
55 - 59	392,104	381,736	79.41	76.8	102.72	-20.59	-23.2	5.35
60 - 64	380,223	390,494	120.5	122.87	97.37	20.5	22.87	3.24
65 - 69	238,972	253,864	78.78	79.11	94.13	-21.22	-20.89	4.03

|							
70 - 74	226,484	251,346	118.08	125.25	90.11	18.08	25.25	-7.96
75 - 79	144,643	147,489	66.89	57.22	98.07	-33.11	-42.78	20.06
80+	206,073	264,165			78.01			
Average Accuracy Index					59.8	20.89	21.45	5.82

*Average Accuracy Index (AAI); ASAI = 3 * SRS + (ARSM + ARSF), SRS = 5.82, ARSM = 20.89, ARSF = 21.45, ASAI = 3*5.82 + (20.89 + 21.45), **ASAI = 59.8***

Table 5 shows an average sex ratio score of approximately 5.82, and male and female age ratio scores of about 20.89 and 21.45 respectively. We get an age-sex accuracy index of about 59.8. According to the scale that UN proposes, the calculated age sex index is extremely very high. This rates the age data as of poor quality.

Table 6 Age Sex Accuracy Index for Tanzania 2012 Population Census

Age ratio deviation		Difference of sex ratios
Male	**Female**	
		-0.20
2.1	1.75	0.52
3.81	1.84	5.27
-4.43	-6.35	11.50
-5.32	2.57	-1.03
36.15	35.03	-0.57
-64.71	-63.25	-9.42
65.87	65.02	-4.86
-0.68	-6.63	7.63
-7.39	1.05	-9.40
9.39	3.25	-2.10
-20.59	-23.2	5.35
20.5	22.87	3.24
-21.22	-20.89	4.03
18.08	25.25	-7.96
-33.11	-42.78	20.06
20.89	**21.45**	**5.82**

Figure 4 Age ratio deviation by sex for Tanzania 2012 Population Census
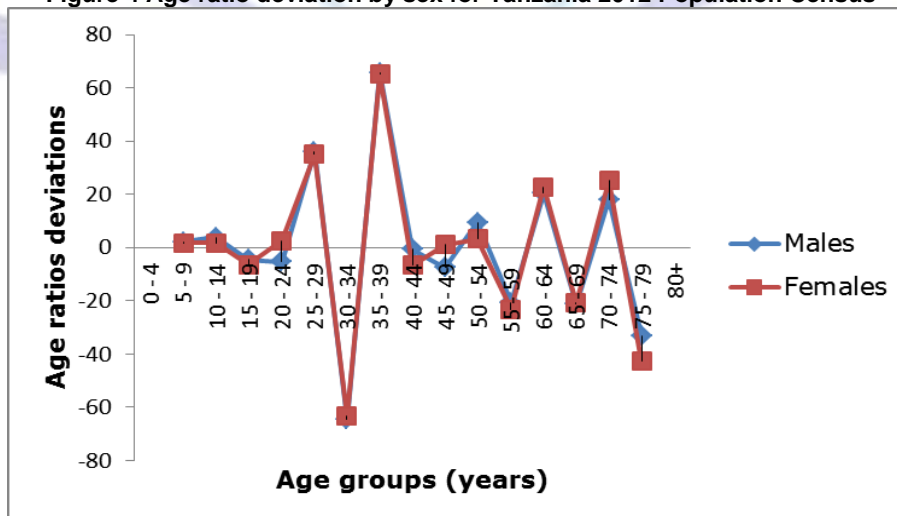
In addition to that the difference of age ratio deviation between male and female is slightly comparable as indicated in Table 6 and, Figure 4 indicates that, maximum positive deviations in both sexes were observed at age group 35 – 39 (65.87% for male and 65.02% for females). Likewise, maximum negative deviations were noted in the preceding age group 30 – 34 (-64.71% for male and -63.26% for female). It should also be noted that there is high deviation at older age groups compared to earlier age groups **(Figure 4).**

Conclusion

In conclusion, it is urged that the quality of age data is important because age-sex distribution is not only an invariable part of a survey report [3, 4 and 18], but the bias introduced can lead to wrong inferences. Innovative methods in data collection along with measuring and minimizing errors using statistical techniques should be used to ensure accuracy of age data. In the case of the Tanzania Population Census 2012, all statistical indices support that age data is of poor quality.

_____

Reference

[1] Bello, Y. (2012). Error Detection in Outpatients' Age Data using Demographic Techniques. *International Journal of Pure and Applied Science and Technology*. 10(1): 27 – 36.

[2] Denic, S., Khatib, F. & Saadi, H. (2004). Quality of age data in patients from developing Countries. *Journal of Public Health*, 26(2):168–71.

[3] Obonyo, J. B. (2011). *Kenya 2009 Population and Housing Census Analytical Report on Population Projections*. Kenya National Bureau of Statistics, Ministry of Planning, National Development and Vision 2030, Nairobi, 6.

[4] Pardeshi, G.S. (2010). Age heaping and accuracy of age data collected during a community survey in the Yavatmal district, Maharashtra. *Indian Journal of Community Medicine,* 35(3): 391-395.

[5] Smaller, S.W. (2004). *Biostatistics and Epidemiology, (3rd Ed.).* New York: Sponger-Verlay.

[6] Spoorenberg, T. (2009). *Is the Whipple's index really a fair and reliable measure of the quality of age reporting? An analysis of 234 censuses from 145 countries*. New York: United Nations Department of Economic and Social Affairs.

[7] Unisa, S., Dwivedi, L.K., Reshmi, R.S., & Kumar, K (2009). Age reporting in Indian census: An insight. Paper presented at the 26th *IUSSP International Population Conference*. Morocco, 27 September-October 2009.

[8] Talib, A.L., Ali, M.S., Hamid, M.S., & Zin, K.M. (2001). *Age Reporting Behavior: A Case Study of 1991 and 2000 Population and Housing Censuses,* Malaysia. Department of Statistics, 1:61–84.

[9] Gonzalez, J. F., Attanasio, L., & Trang Ha, J. (2014). An Assessment of the Age Reporting in the IPUMS-I Microdata. Paper submitted for presentation at the 2014 Annual Meeting of the *Population Association of America.*

[10] Nagi, M.H., Stockwell, E.G., & Snavley, L.M. (1973). Digit Preference and Avoidance in the Age Statistics of Some Recent African Censuses: Some Patterns and Correlates. *International Statistical Review,* 41(2): 165 – 174.

[11] United Nations. (1956). *Manuals on Methods of Estimating Population. Manual II: Methods of Appraisal of Quality of Basic Data for Population Estimates*. Sales No. E.56. XIII.2: Online: Retrieved July 1st 2014 from http://www.un.org/esa/population/pubsarchive/migration_publications/UN_1955_Manual2.pdf.

[12] United Nations. (2013). *International Workshop on Population Projections using Census Data*. Beijing, China.

[13] Hobbs, F. B. (2004). Age and sex composition. In Siegel, J.S., & Swanson, D.A., (Eds). The methods and materials of demography (2nd Ed., pp 125 – 173). San Diego, California: Elsevier Academic Press.

[14] Kpedekpo, G.M.K. (1982). *Essentials of Demographic Analysis for Africa*. New Hemisphere: Hernerman Educational Books Inc.

[15] West, K.K., Robinson, J.G., & Bentley, M. (2005). Did Proxy Respondents Cause Age Heaping in the Census 2000? ASA Section on Survey Research Methods: 3658 – 65: Online: Retrieved June 20 2014 from http://www.amstat.org/sections/srms/Proceedings/y2005/Files/JSM2005-000443.pdf

[16] Tanzania National Bureau of Statistics (2013). *Population Distribution by Age and* Sex. Dar es Salaam: Ministry of Planning, Economic and Empowerment.

[17] Haandrikman, K. (2004). Using a local time-path calendar to reduce heaping in durations of postpartum Amenorrhea, breastfeeding, postpartum Abstinence and contraceptive use. *Time Soc,* 13: 339-62.

[18] Spoorenberg, T., & Duttreuilh, C. (2007). Quality of age reporting: Extension and application of the modified Whipple's index. *Population-E,* 62(04): 729-741.

Authors Biography

Prof A Sathiya Susuman

Sathiya Susuman APPUNNI completed his Ph D in Demography from University of Kerala, India. His research efforts are concentrated in the field of population components, such as fertility, mortality, and migration. Keeping abreast of his research aspirations in the field of technical demography, he focus mainly on the 'demography of health,' with attention to a life course analysis of health risks, interventions, and outcomes. He worked on several research projects in India, Ethiopia and South Africa. He has published several articles in reputed journals. He is a member of Population Association of America, Union for African Population Studies, International Union for Scientific Study of Population, Asian Population Association, Statistics South Africa and Population Association of South Africa. At present he is working as a full time Faculty member of Natural Sciences, University of the Western Cape, Cape Town, South Africa.

Mr Hamisi F Hamisi

Mr. Hamisi's field level is in social sciences and health research, and statistical analysis. Mr. Hamisi acquired intensive and extensive experiences on Population health, Demographic analysis, Population projections and Official statistics, Applied and Business statistics, and Reproductive health within a span of 5 years. Mr. Hamisi graduated a Master of Philosophy (M Phil) with Cum Laude in Population Studies/demography with specialization on Childhood Mortality, B Sc. (Honours) with Cum Laude in Statistics with specialization in population study and demography, and BA in Geography and Management Science. His areas of research are maternal and childhood mortality, gender and reproductive health issues, and population projection for developments and official statistics.

Dr Siaka Lougue

Dr. Siaka LOUGUE is lecturer at the school of Mathematics, Statistics and Computer Sciences at the University of Kwazulu Natal, Durban, South Africa. He is specialized

in applied statistics/demography with focus on health and migration.

Dr Ogujiuba Kanayo

Ogujiuba Kanayo is a member of several research and management development networks. He has published several articles in reputed journals. His consulting and research interests are in Economic Demography, Public Policy and Economic Theory. He has worked on several projects and has been a consultant/senior policy adviser on World Bank, USAID, DFID, ECA, UNDP, EU, WAMI projects and to government agencies in Nigeria in the last 15 years. He holds a PhD in Development Economics, which was funded through a research grant from the African Economic Research Consortium (AERC). At present he is a Research Associate at the Statistics and Population Studies Department, University of Western Cape.

Prof Nigatu Regassa

Nigatu Regassa Geda is full Professor of demography (Population Studies), and has been teaching at one of the federal universities of Ethiopia, Hawassa University. He completed his doctoral degree in 2006 at the International Institute for Population Studies, Mumbai; and got Post-Doctoral research Fellowship at Oklahoma State University, U.S.A. in July 2011 where he involved in research works related to household food security, nutrition and environment. Nigatu has long years of experience in teaching and leadership with various rankings: During the last 18 years of his employment at Hawassa University, Nigatu has been active in various research projects having quite significant monetary and academic values to Ethiopia. He won several international and local grants, and represented the university and the education sector at many international forums. Since May 2014, Nigatu has moved to Canada as a visiting research scholar at the University of Saskatchewan.

Prof Eric Mwambene

Eric Mwambene completed doctoral studies at University of Vienna in 2001. Prior to that, he had done an MSc in discrete mathematics at University of Zimbabwe and an MSc in Artificial Intelligence at Imperial College, University of London. Prof Mwambene has recently been analyzing the statistical properties of biological networks. He also looks at symmetries of discrete structures and explores codes of algebraic systems. A member of International Mathematical Union and South African Mathematical Society, Prof Mwambene is a full time employed in the Department of Mathematics at University of Western Cape.