Biomedical Statistics and Informatics
2019; 4(2): 15-21
http://www.sciencepublishinggroup.com/j/bsi
doi: 10.11648/j.bsi.20190402.11
ISSN: 2578-871X (Print); ISSN: 2578-8728 (Online)

Statistical Analysis of Age Reporting Errors by Insured and Uninsured Patients in Cape Coast Teaching Hospital of Ghana

Eyiah-Bediako Francis[1], Bosson-Amedenu Senyefia[2], Borbor Bridget Sena [3]

[1]Department of Statistics, University of Cape Coast, Cape Coast, Ghana

[2]Department of Mathematics and ICT, Holy Child College of Education, Takoradi, Ghana

[3]Department of Mathematical Sciences, University of Mines and Technology, Tarkwa, Ghana

Email address:
senyefia@yahoo.com (Bosson-Amedenu S.)

To cite this article:
Eyiah-Bediako Francis, Bosson-Amedenu Senyefia, Borbor Bridget Sena. Statistical Analysis of Age Reporting Errors by Insured and Uninsured Patients in Cape Coast Teaching Hospital of Ghana. *Biomedical Statistics and Informatics.* Vol. 4, No. 2, 2019, pp. 15-21.
doi: 10.11648/j.bsi.20190402.11

Received: July 12, 2019; **Accepted**: August 4, 2019; **Published**: August 15, 2019

Abstract: Age is a very important variable that guides clinicians to carryout diagnosis, treatment, as well as administering medical procedures to patients. Misreporting of age by patients to clinicians can have dire consequences on the patients' health. This retrospective study used a 10 year demographic data involving the ages reported by 906,383 patients. Demographic indexes such as Whipples, Myers Blended and Joint Score were employed to analyse reported ages among insured and uninsured patients at the Cape Coast Teaching Hospital. The computed joint score values of 76.88 and 85.60 respectively for uninsured and insured patients qualified the data as highly inaccurate by the standards of interpretation of UN index. The summary of the digit preference of the uninsured and insured patients by Myers blended index approach were 29.34 and 29.87 respectively. The blended sum at the digits 0, 1, 2 and 5 exceeded 10% of the total blended population, an indication of over selection of ages ending in those digits by the insured and uninsured patients. Whipple's index for uninsured and insured patients was 149.3 and 287.1 respectively. These values respectively show that the reliability of the ages reported were rough and very rough, by the Whipple's index interpretation standards. The insured were found to have higher tendency of concentrating on ages ending in 0 and 5 than the uninsured. The study concluded that age data in Cape Coast Teaching Hospital is misreported and inaccurate and if not adjusted may result in wrong age-dependent medical procedures undertaken by clinicians. It was recommended among others for hospitals to institute innovative ways of recording ages such as using calendar of historical events technique where the patients could not recall their correct age.

Keywords: Myers Index, Whipples Index, UN Index, Hospital, Insured, Uninsured, Demographic Indexes

1. Introduction

Demographic indexes such as Whipples index, Myers Blended index and UN index have been widely used in measuring the quality of age data in diverse disciplines. Many reasons account for misreporting of age, these include: (a) Ignorance of correct Age-most especially illiterate population (b) Carelessness in reporting and in recording age (c) There is tendency to record or state age ending in certain figures known as preferred digits e.g age ending in 0 or 5 (d) There is a possible sub-conscious aversion to certain numbers (e) Exaggeration of length of life at advance ages (f) Misplacement of age, from some motive such as economic, social, political or purely personal [1-6].

Bwalya et al. (2015), [7] applied demographic indexes to assess if there had been improvement in the quality of age data with respect to three consecutive population censuses (1969, 1990 and 2010) in Zambia. Data was collected in single years and the Whipples, Myers Blended and UN joint score indexes were applied to the data. The study found improvement in the quality of data collected in successive censuses. That notwithstanding, digit preference of 0 and 5

were still found to be persisting in Zambia.

Bekele (2006), [8] used demographic indexes to measure the quality of data in two consecutive censuses in Ethiopia. The results showed that the quality of the data was reduced in the in the second census which was contrary to expectation. Bosson-Amedenu et.al (2019), [9] in their study analysed the quality of age data reported by Health Insurance Scheme registrants for the month of June, 2018 in the Sekondi-Takoradi District of Ghana. The results showed among others that the quality of data reported was highly inaccurate with a UN index value of 138.14.

Pardeshi (2010), [10] applied demographic indexes in a cross-sectional study at Yavatmal district of India. The results showed among others that the Whipples index for terminal digits of "0" and "5" were determined to be 386.71 and 382.74 respectively. It was concluded that the ages reported were of poor quality. Bello (2012), [11] studied the ages of patients collected at the outpatients department of Dutsin-ma General Hospital, Katsina State of Nigeria, in January 2012 using demographic techniques. From the results, the accuracy of the ages reported was very rough in quality for both male and female ages reported. For the Myer's index, about 86 percent of male outpatients and 88 percent of female outpatients reported their ages with incorrect final digits.

Although a lot of studies have applied the concept of demographic indexes in other disciplines, not enough work has been done in the Health sector to find out the quality of age data reported to clinicians by their patients. Misreporting of ages by patients can have a great influence on the prescriptions by clinicians which inadvertently will have negative effect on the patients.

2. Method

A retrospective demographic data of 906,383 patients who visited the Cape Coast Teaching Hospital from the years January 2008 through to December 2017 were analysed for age reporting errors using demographic indexes. The data was in the form of single years of age and was also put in the form of five year age intervals. The research sought to determine the error of age reporting by the method of demographic indexes. The Whipples, Myers and Joint score indexes were used to determine the age heaping among patients. Microsoft Excel was employed in the analysis of the data.

2.1. Whipples Index

The Whipples index was used to measure the preference for or avoidance of a particular digit. This index operates on the assumption of rectangularity or of linearity of five year age range. Age heaping on terminal digits of 0 and 5 put together in the age range 23-62 may be measured by comparing the sum of the population at ages in this range ending in zero and five with one fifth of the total population in the age range.

$$Whipples\ Index = \frac{P_{25} + P_{30} + P_{35} + P_{40} + \cdots + P_{55} + P_{60}}{\frac{1}{5}\ (P_{23} + P_{24} + P_{25} + \cdots + P_{61} + P_{62}} \times 100$$

or

$$Whipples\ Index = \frac{\sum_{23}^{62} P_a ending\ in '0' or\ '5'}{\frac{1}{5}\sum_{23}^{62} P_a} \times 100$$

2.2. Myers Index

Myers index is one of the widely used methods of measuring age preference at each digit. It is useful in measuring the accuracy of age data. This index is used to measure the preference or dislike for each of the ten digits in age reporting. It is applicable when age data is in single years. It is referred to as blended index since it was developed to avoid the bias in the computed index. This bias is attributed to the Whipples index since for example, the terminal digit of numbers are more likely to be 0 than 1 and 9. The overall Myers summary index for all terminal digits is computed as one-half of the sums of absolute deviations from 10%. The range of the index are between 0 and 90.

The Myers blended index is given by:

$$\text{Myers Blended Index} = \left[\sum_{i=0}^{9} \frac{B_i}{\sum_{i=0}^{9} B_i} \times 100\right] - 10\%$$

B_i is the Blended population i ranging from $0 - 9$

$$\text{Grand Blended population} = \sum_{i=0}^{9} B_i$$

$$\text{Magnitude of preference} = \left[\sum_{i=0}^{9} \frac{B_i}{\sum_{i=0}^{9} B_i} \times 100\right]_{-10}$$

2.3. United Nations Index (Joint Score or Age Accuracy Index)

This index simultaneously uses age-sex ratios in its computation. Age and sex ratios are computed for five-year age groups up to about 70. With respect to the sex ratios; successive differences between consecutive age groups are found and the average taken after finding the absolute sum. With respect to the age ratios of either sex, deviations from 100 are computed; disregarding the sign, the sum is computed [12-15]. The mean of the deviations of the two age ratios ARS (M) and ARS (F) are then summed to 3 times of the mean of the sex ratio differences (SRS). Hence, the UN Joint score is computed as:

$$\text{Joint Score} = ARS\ (M) + ARS\ (F) + 3\ (SRS)$$

Table 1. *United Nations Index (Joint Score) Computation of Reporting errors by noninsured Patients.*

Terminal Digits	Male Population	Age Ratio	Deviation from 100	Female Population	Age Ratio	Deviation from 100	Sex Ratio	Successive Difference
0-4	10351			11462			90.30710173	
5-9	5544	72.4990	-27.5010	5668	66.6314	-33.3686	97.81227946	-7.50518
10-14	4943	69.9597	-30.0403	5551	71.1302	-28.8698	89.04701856	8.765261
15-19	8587	189.7470	89.7470	9940	179.0830	79.0830	86.38832998	2.658689
20-24	4108	67.2671	-32.7329	5550	74.7978	-25.2022	74.01801802	12.37031
25-29	3627	102.4721	2.4721	4900	102.4569	2.4569	74.02040816	-0.00239
30-34	2971	84.0096	-15.9904	4015	89.0540	-10.9460	73.99750934	0.022899
35-39	3446	118.3582	18.3582	4117	110.9404	10.9404	83.70172456	-9.70422
40-44	2852	100.0000	0.0000	3407	99.9853	-0.0147	83.71000881	-0.00828
45-49	2258	70.6729	-29.3271	2698	67.3995	-32.6005	83.69162342	0.018385
50-54	3538	159.2618	59.2618	4599	166.0589	66.0589	76.92976734	6.761856
55-59	2185	67.9627	-32.0373	2841	71.5978	-28.4022	76.90953889	0.020228
60-64	2892	144.1316	44.1316	3337	134.8828	34.8828	86.66466886	-9.75513
65-69	1828	49.2590	-50.7410	2107	46.6925	-53.3075	86.7584243	-0.09376
70+	4530			5688				
Absolute Total			432.3407			406.1335		57.68659
AVERAGE			33.25698			31.24104		4.43743
JOINT SCORE	77.81							

The joint score value of 76.88 qualifies the data as highly inaccurate by the standards of interpretation of UN index. The female noninsured patients reported more inaccurate ages than their male counterparts; evidenced from the magnitude of their absolute deviations (424.13 > 402.11). Male and female patients within the age group of 15-19 and 50-54 reported the most inaccurate ages.

Table 2. *United Nations Index (Joint Score) Computation of Reporting errors by Insured Patients.*

Terminal Digits	Male Population	Age Ratio	Deviation from 100	Female Population	Age Ratio	Deviation from 100	Sex Ratio	Successive Difference
0-4	44221			11462			-37.1355	
5-9	19185	62.7125	-37.2875	5668	62.8645	-37.1355	-36.8878	6.5977
10-14	16963	66.8796	-33.1204	5551	63.1122	-36.8878	18.8553	13.2674
15-19	31542	164.2086	64.2086	9940	118.8553	18.8553	17.7075	13.1635
20-24	21454	84.9899	-15.0101	5550	117.7075	17.7075	2.4685	33.3518
25-29	18944	102.4720	2.4720	4900	102.4685	2.4685	-8.6565	-0.0018
30-34	15520	82.5400	-17.4600	4015	91.3435	-8.6565	7.5483	0.0010
35-39	18662	120.5400	20.5400	4117	107.5483	7.5483	0.0000	-10.0295
40-44	15444	100.0000	0.0000	3407	100.0000	0.0000	-34.9545	0.0011
45-49	12226	59.1113	-40.8887	2698	65.0455	-34.9545	70.9934	0.0017
50-54	25922	183.6031	83.6031	4599	170.9934	70.9934	-27.6096	-8.7986
55-59	16011	64.3736	-35.6264	2841	72.3904	-27.6096	32.8087	-0.0023
60-64	23822	153.4132	53.4132	3337	132.8087	32.8087	-48.9311	-18.1872
65-69	15045	45.6020	-54.3980	2107	51.0689	-48.9311	-37.1355	0.0038
70+	42162			5688				
Absolute Total			458.0279			344.5567		103.4074
AVERAGE			35.23292			26.50436		7.954415
JOINT SCORE	85.60							

Again joint score value of 85.60 qualifies the data as highly inaccurate by the standards of interpretation of UN index. The male insured patients reported more inaccurate ages than their female counterparts; evidenced from the magnitude of their absolute deviations $(458.0279 > 344.5567)$. Male patients within the age group of 50-54 and 15-19 reported the most inaccurate ages. Also female patients within the age group of 50-54 and 65-69 reported the most inaccurate ages.

Discussions on Tables 1 and 2.

It has been established from the analysis by Joint Score approach of detection of age reporting error that the insured patients wrongly reported their ages more than the insured patients; evidenced by the comparing the sum of absolute deviations. Whereas the male and female patients within the age groups of 15-19 and 50-54 reported the most inaccurate ages in the case of the noninsured patients, male patients within the age group of 50-54 and 15-19 reported the most inaccurate ages among the insured. Again, female patients within the age group of 50-54 and 65-69 reported the most inaccurate ages among the insured.

Table 3. *Computation of reporting errors by Uninsured Patients by Myers Blended Index.*

Terminal Digit	Sum of ages 10-89	Coefficient	Ages 10-89 Coefficient Product	Sum of ages 20-89	Coefficient	Ages 20-89 Coefficient Product	Blended Sum	% Distribution	Deviation from 10
0	7147	1	7147	5209	9	46877	52086	32.23	22.23
1	3452	2	6904	2180	8	17443	19623	12.14	2.14
2	4603	3	13809	2847	7	19926	22772	14.09	4.09
3	3634	4	14536	2301	6	13809	16110	9.97	-0.03
4	3452	5	17261	2120	5	10599	12719	7.87	-2.13
5	4785	6	28708	3513	4	14051	17564	10.87	0.87
6	3271	7	22894	1999	3	5996	7995	4.95	-5.05
7	3210	8	25680	1999	2	3997	5996	3.71	-6.29
8	3937	9	35431	2483	1	2483	4966	3.07	-6.93
9	2847	10	28466	1756	0	0	1756	1.09	-8.91
							161587	100	58.68
			Summary Index of Age preference					Total Index	29.34

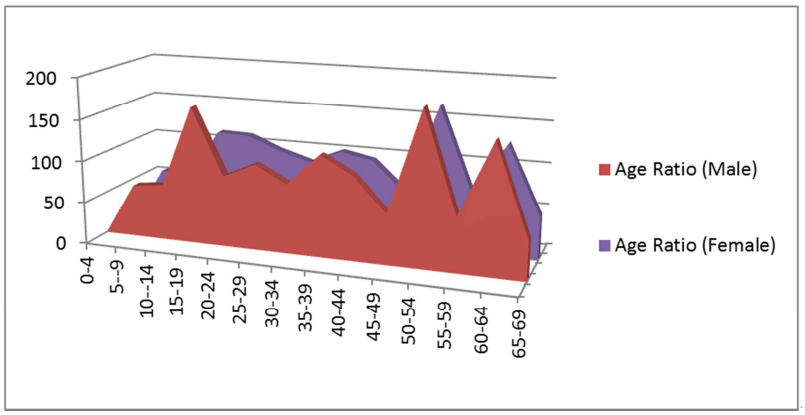

Figure 1. *Age Ratio Distribution of Insured Male and Female.*

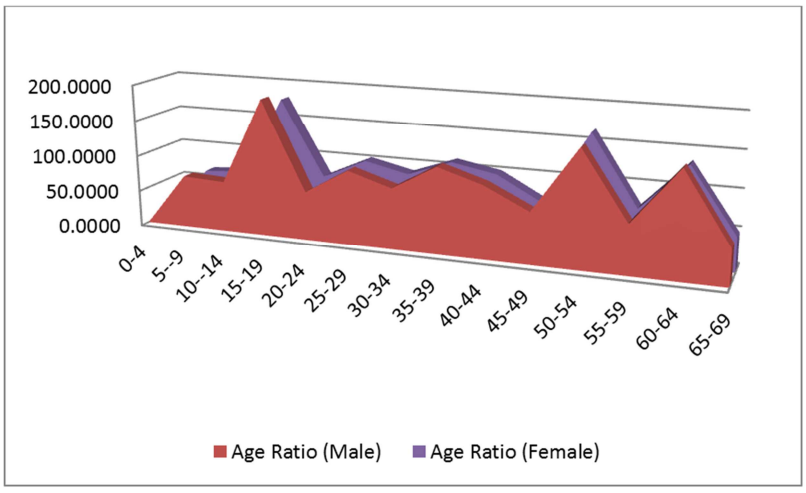

Figure 2. *Age Ratio Distribution of uninsured Male and Female.*

The use of the Myer's index was to facilitate in measuring the preference of the dislikes for each of the ten digits by the uninsured patients. The overall Myers summary index for all terminal digits is computed as one-half of the sum of absolute deviations from 10%. The results of the computation found the summary of the preference of the uninsured patients to be 29.34. About 59% of uninsured patients reported their ages with incorrect terminal digits. This is an indication that of a serious prevalence of age heaping among uninsured patients. The results further showed that there was more preference for the terminal digit of 0 (32%). The digit of 9 was the least preferred terminal digit. The blended sum at the digits 0, 1, 2 and 5 exceeded 10% (representing 16,159) of the total blended population, an indication of over selection of ages ending in those digits by both the uninsured patients. On the other hand there was under-selection of the remaining digits as their respective blended sums were less than 10% of the total blended population. The index was not approximately

zero, an indication of presence of age heaping.

Table 4. *Computation of reporting errors by insured Patients by Myers Blended Index.*

Terminal Digit	Sum of ages 10-89	Coefficient	Ages 10-89 Coefficient Product	Sum of ages 20-89	Coefficient	Ages 20-89 Coefficient Product	Blended Sum	% Distribution	Deviation from 10
0	99807	1	99807	59207	9	532865	592073	26.24	16.24
1	48212	2	96423	33833	8	270662	304494	13.49	3.49
2	64282	3	192847	49057	7	343402	392460	17.39	7.39
3	50749	4	202996	30449	6	182697	213146	9.45	-0.55
4	48212	5	241058	33833	5	169164	202996	9.00	-1.00
5	66820	6	400918	57516	4	230062	287578	12.74	2.74
6	45674	7	319719	25375	3	76124	101498	4.50	-5.50
7	44828	8	358627	22245	2	44490	66735	2.96	-7.04
8	54978	9	494804	35524	1	35524	71049	3.15	-6.85
9	39753	10	397534	24529	0	0	24529	1.09	-8.91
			58.92				2256558	100	59.73
			Summary Index of Age preference					Total Index	29.87

Again, the findings show that the summary of the preference of the insured patients to be 29.87. About 60% of insured patients reported their ages with incorrect terminal digits. This is further an indication of a serious prevalence of age heaping among insured patients. It could be inferred from the results that there was more preference for the terminal digit of 0 (26%). The digit of 9 was again the least preferred terminal digit. The blended sum at the digits 0, 1, 2 and 5 exceeded 10% (representing 225, 656) of the total blended population, an indication of over selection of ages ending in those digits by the insured patients. On the other hand there was under-selection of the remaining digits as their respective blended sums were less than 10% of the total blended population. The index was not approximately zero, an indication of presence of age heaping.

Table 5. *Computation of reporting errors by uninsured Patients by Whipple's Index.*

Age	Number of Uninsured Patients	Age	Number of Uninsured Patients
23	1017	25	1163
24	945	30	1199
25-29	3627	35	896
30-34	2971	40	969
35-39	3446	45	715
40-44	2852	50	703
45-49	2258	55	369
50-54	3538	60	618
55-59	2185	TOTAL	6632
60	618		
61	73	Digit Preference for '0' and '5'digits $= \frac{6632}{\frac{1}{5}(23639)} \times 100 = 140.3$	
62	109		
TOTAL	23639		

Table 6. *Computation of reporting errors by insured Patients by Whipple's Index.*

Age	Number of insured Patients	Age	Number of Uninsured Patients
23	14210	25	16240
24	13195	30	16747
25-29	18944	35	12518
30-34	15520	40	13533
35-39	18662	45	9981
40-44	15444	50	9811
45-49	12226	55	5159
50-54	25922	60	8627
55-59	16011	TOTAL	92617
60	8627		
61	1015	Digit Preference for '0' and '5'digits $= \frac{92617}{\frac{1}{5}(161298)} \times 100 = 287.1$	
62	1522		
TOTAL	161298		

The results from tables 5 and 6 show that the whipple's index for uninsured and insured patients are 149.3 and 287.1 respectively. These values respectively show that the reliability of the ages reported were rough and very rough, by the whipple's index interpretation standards. It can be inferred that the ages reported by the uninsured patients were of better quality compared to that of the insured. The insured have higher tendency of concentrating on ages ending in 0 and 5 than the uninsured. This may be due to high illiteracy rate among the insured.

3. Conclusion

Demographic indexes such as Whipples, Myers Blended and Joint Score were employed to analyse reporting errors among insured and uninsured patients at the Cape Coast Teaching Hospital. The computed joint score values of 76.88 and 85.60 respectively for noninsured and insured patients qualified the data as highly inaccurate by the standards of interpretation of UN index. The female noninsured patients reported more inaccurate ages than their male counterparts; evidenced from the magnitude of their absolute deviations (424.13 > 402.11). Male and female patients within the age group of 15-19 and 50-54 reported the most inaccurate ages. The male insured patients reported more inaccurate ages than their female counterparts; evidenced from the magnitude of their absolute deviations $(458.0279 > 344.5567)$. Male patients within the age group of 50-54 and 15-19 reported the most inaccurate ages. Also female patients within the age group of 50-54 and 65-69 reported the most inaccurate ages. The summary of the digit preference of the uninsured and insured patients by Myers blended index approach was 29.34 and 29.87 respectively. The blended sum at the digits 0, 1, 2 and 5 exceeded 10% of the total blended population, an indication of over selection of ages ending in those digits by the insured and uninsured patients. Whipple's index for uninsured and insured patients was 149.3 and 287.1 respectively. About 59% of uninsured patients reported their ages with incorrect terminal digits. This is an indication that of a serious prevalence of age heaping among uninsured patients. The results further showed that there was more preference for the terminal digit of 0 (32%). The digit of 9 was the least preferred terminal digit. The blended sum at the digits 0, 1, 2 and 5 exceeded 10% (representing 16,159) of the total blended population, an indication of over selection of ages ending in those digits by both the uninsured patients. On the other hand there was under-selection of the remaining digits as their respective blended sums were less than 10% of the total blended population. The index was not approximately zero, an indication of presence of age heaping. About 60% of insured patients reported their ages with incorrect terminal digits. This is further an indication of a serious prevalence of age heaping among insured patients. It could be inferred from the results that there was more preference for the terminal digit of 0 (26%). The digit of 9 was again the least preferred terminal digit. The blended sum at the digits 0, 1, 2 and 5 exceeded 10% (representing 225, 656) of the total blended population, an indication of over selection of ages ending in those digits by the insured patients. On the other hand there was under-selection of the remaining digits as their respective blended sums were less than 10% of the total blended population. The index was not approximately zero, an indication of presence of age heaping. These values respectively show that the reliability of the ages reported were rough and very rough, by the Whipple's index interpretation standards. The insured were found to have higher tendency of concentrating on ages ending in 0 and 5 than the uninsured. The study recommended among others for hospitals to institute innovative ways of recording ages such as using calendar of historical events technique where the patients could not recall correct their correct age.

4. Recommendations

The study recommended for health practitioners to find innovative ways of crosschecking reported ages before taking records of them.

References

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

[2] Shirley. L., V. Ravi and M. Margaret, (2004). An evaluation of the age and sex data from the census population of Canada, provinces and territories, 1971 to 2001, Canadian Population Society Annual Meeting, Winnipeg, Manitoba, (2004). http//web.uvic.ca/~canpop/2004/Loh-Verma-Michalowski-CPS04.ppt.

[3] S. Denic, F. Khatib and H. Saadi, Quality of age data in patients from developing countries, Journal of Public Health, 26 (2) (2003), 168–171.

[4] Bello Y., (2012). Error Detection in Outpatients' Age Data Using Demographic Techniques. International Journal of Pure and Applied Sciences and Technology. Int. J. Pure Appl. Sci. Technol., 10 (1) (2012), 27-36. ISSN 2229-6107. Available online at www.ijopaasat.in.

[5] Barua, R. K (2015). Detection of Digit Preference and Age Misreporting by using Demographic Techniques. A thesis submitted in partial fulfillment of the requirements for the degree of Master of Population, Reproductive Health, Gender and Development (MPRHGD) at East West University, Dhaka, Bangladesh.

[6] Kpedekpo G. M. K. (1982). Essentials of Demographic Analysis for Africa. Hernerman Educational Books Inc., ew Hemisphere.

[7] Bwalya B. B, Phiri M. and Mwansa C. (2015). International Journal of Current Advanced Research vol 4, Issue 5, pp 92-97, May 2015.

[8] Bekele S. (2006). Analysis on the quality of age and sex data collected in two population and housing censuses of Ethiopia. EthiopJournal of Science, 29 (2): 123-132, 2006. ISSN: 0379-2897.

[9] Bosson-Amedenu S., Oduro-Okyireh T., Osei-Asibey E. Detection of errors in age data of National Health Insurance Scheme registrants in Ghana: demographic indexes approach. International Journal of Mathematical Archive-10 (3), 2019, 6-13 Available online through www.ijma.info ISSN 2229-5046.

[10] 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): 391-5.

[11] Bello Y. (2012). Error Detection in Outpatients' Age Data Using Demographic Techniques. Int. J. Pure Appl. Sci. Technol., 10 (1) (2012), pp. 27-36.

[12] G. M. K. Kpedekpo, Essentials of Demographic Analysis for Africa, Hernerman Educational Books Inc., New Hemisphere, 1982.

[13] B. Kestenbaum, A description of the extreme age population based on improved Medicare enrollment data, Demography, 29 (1992), 565–580.

[14] H. S. Shryock, J. S. Siegel and Associates. The Methods and Materials of Demography, Condensed edition by Edward G. Stockewell, Academic Press, New York, 1976.

[15] I. Johannes and F. Polly, Bancroft's Introduction to Biostatistics, (2nd Edition), Harper and Row publishers, New-York, 1970.