USING BAYES' THEOREM IN THE ESTIMATION OF PREVALENCE OF HIGH BLOOD PRESSURE IN CERTAIN AGE GROUPS

Supported by tetfund TERTIARY EDUCATION TRUST FUND

Rotimi K. Ogundeji[1]* and Ismaila A. Adeleke[2]
[1]Department of Mathematics, University of Lagos, Lagos State, Nigeria
[2]Department of Actuarial Science and Insurance, University of Lagos, State, Nigeria
*Corresponding author: rogundeji@unilag.edu.ng

Received: March 11, 2019 **Accepted:** June 03, 2019

Abstract: This is a community population-based study that considered the occurrence of high blood pressure in adults aged 18 years and older, classified into different age groups. The study collated data of patients' visits to a medical centre to obtain proportion of patients with high blood pressure in different age groups. The study employs Bayesian approach to estimate probability of having high blood pressure by different age groups as a measure of prevalence of high blood pressure. Results show the prevalence of having high blood pressure (measured in probability values) within different age groups. The correlation analysis shows a direct association between probabilities of the events values and age.

Keywords: Age groups, Bayesian methods, conditional probabilities, correlation analysis

Introduction

High blood pressure (hypertension) can damage the health of a person in many ways as it seriously hurts important organs like the heart and brain. High blood pressure among old aged groups is a known fact that has been well documented. However, the prevalence of high blood pressure among lower aged groups (i.e. below 30 years old) needs to be investigated. Age, race or ethnicity, obesity, gender, lifestyle habits and family history have been known to influence the risk of developing high blood pressure (HBP). This study defines high blood pressure (also called Hypertension) based on standard set reading ≥ 140/90 mmHg (Adeloyea *et al.*, 2014).

Many studies have been conducted on prevalence of high blood pressures at local and national levels, and at urban or rural areas; the study points to the fact that there has been increasing prevalence of hypertension across the globe (Adediran *et al.*, 2011; Ekwunife and Aguwa, 2011; Wang *et al.*, 2014). Studies also suggested that hypertension is high in Nigeria, but the overall awareness of this silent killer disease (hypertension) cases is low in Nigeria. The rate at which people are dying from this silent disease is quite alarming (Adeloyea *et al.* 2014; Rao *et al.*, 2014).

A number of researches have been carried out on hypertension at old age but only few of them considered having hypertension at early age (i.e. between 18 and 30 years old). The study considers age as a major factor as blood pressure tends to rise with age due to the fact that the structure of human heart and vasculature change with age. There is the need to investigate the prevalence of high blood pressures at different age groups and obtain the probabilities of the disease at the different age groups. The aim of this study is to estimate the probability of a person developing high blood pressure at different age groups using Bayes' theorem. Since hypertension is related to age, using Bayes' theorem, age can be used to more accurately assess the probability of a person having hypertension, than when age is not involved. Staff records at the University of Lagos medical centre were used in estimating the prior and posterior probability relative to the age groups. The correlation between the age groups and the probability of having high blood pressure at those age groups is also considered.

Bayesian inference has found application in a wide range of activities, including science, engineering, philosophy, medicine, sport and law (Davison, 2008; Jackman, 2009; Ogundeji and Okafor, 2010). Bayes' theorem tells how to update or revise beliefs in light of new evidence. As a formal theorem, Bayes' theorem is valid in all interpretations of probability. Bayesian inference derives the posterior probability as a consequence of two antecedents, a prior probability and a "likelihood function" derived from a statistical model for the observed data. Bayesian inference computes the posterior probability according to Bayes' theorem (Lee, 2004; Okafor and Ogundeji, 2012).

Materials and Methods

Data presentation

Data were collected from the University of Lagos Medical centre on the total number of patients, and the number of patients with high blood pressure (HBP) within different age groups in each month of the year 2015 and 2016, respectively as shown in Tables 1 and 2. The collated data is graphically displayed in Figs. 1 and 2.

The total number of patients within a particular age group h is represented by N_h while n_h represents the number of patients with hypertension within the age group. The following age groups are considered: below 30, 30-40, 41-50, 51-60, and above 60 years, respectively.

Table 1: Number of patients with hypertension within different age groups in 2015

Month	Total number of patients	Number of patients without hypertension	Number of patients with hypertension	Age group									
				Below30		30-40		41-50		51-60		Above60	
				N_h	n_h	N_h	n_h	N_h	n_h	N_h	n_h	N_h	n_h
Jan	322	192	130	29	8	31	13	120	38	128	62	14	9
Feb	435	303	132	6	2	59	13	163	16	177	83	30	18
Mar	617	392	225	13	1	70	15	249	79	239	99	46	31
April	426	276	150	3	1	56	9	201	69	138	60	28	11
May	448	284	164	14	2	98	17	167	63	150	71	19	11
Jun	432	321	111	28	4	287	56	68	23	35	21	14	7
July	75	53	22	8	0	27	7	22	7	13	4	5	4
Aug	125	67	58	5	2	25	7	58	20	32	25	5	4
Sept	220	121	99	10	5	58	25	97	45	52	23	3	1
Oct	103	68	35	3	0	48	15	20	10	22	5	10	5
Nov	291	139	152	18	7	92	35	118	83	57	25	6	2
Dec	347	230	117	12	3	202	45	71	36	42	18	20	15

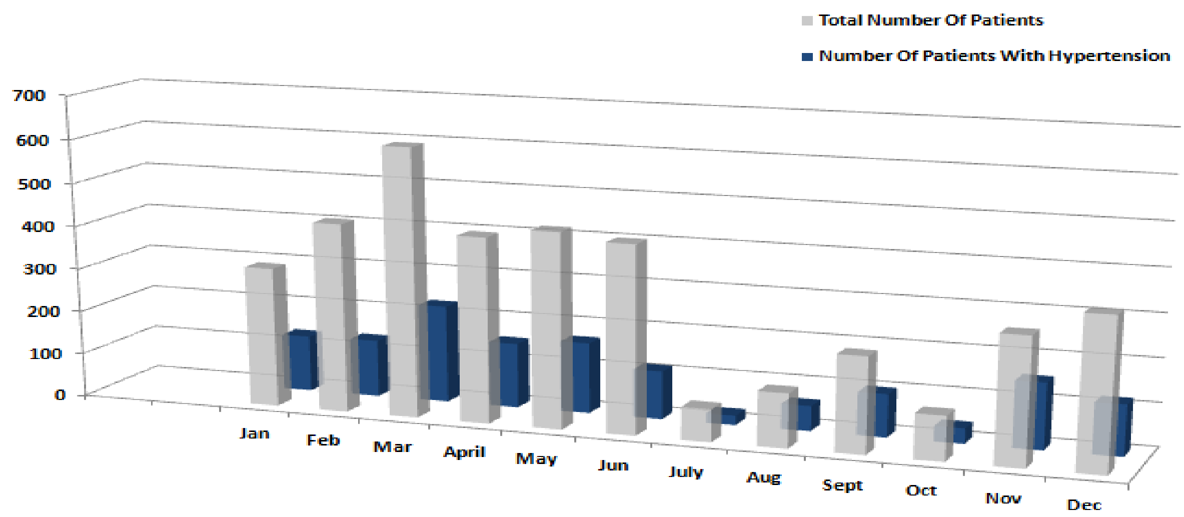

Fig. 1: Monthly distribution of number of patients and patients with hypertension, 2015

Table 2: Number of patients with hypertension within different age groups in 2016

Month	Total number of patients	Number of patients without hypertension	Number of patients with hypertension	Age group									
				Below 30		30-40		41-50		51 – 60		Above 60	
				N_h	n_h	N_h	n_h	N_h	n_h	N_h	n_h	N_h	n_h
Jan	65	47	18	3	0	15	1	27	9	15	5	5	3
Feb	56	33	23	2	0	7	2	18	7	26	11	3	3
Mar	60	45	15	4	3	12	1	17	2	23	7	4	2
April	154	113	41	5	0	29	2	57	11	51	19	12	9
May	249	172	77	9	2	62	11	88	29	73	26	17	9
Jun	230	166	64	10	2	59	11	76	22	68	23	17	6
July	259	177	82	15	1	57	10	85	24	86	40	16	7
Aug	226	163	63	14	2	60	7	74	22	64	24	14	8
Sept	244	175	69	28	3	23	7	96	23	82	27	15	9
Oct	197	141	56	6	0	61	8	60	18	63	26	7	4
Nov	148	105	43	5	1	26	4	54	13	54	19	9	6
Dec	234	164	70	22	0	49	14	80	20	67	29	16	7

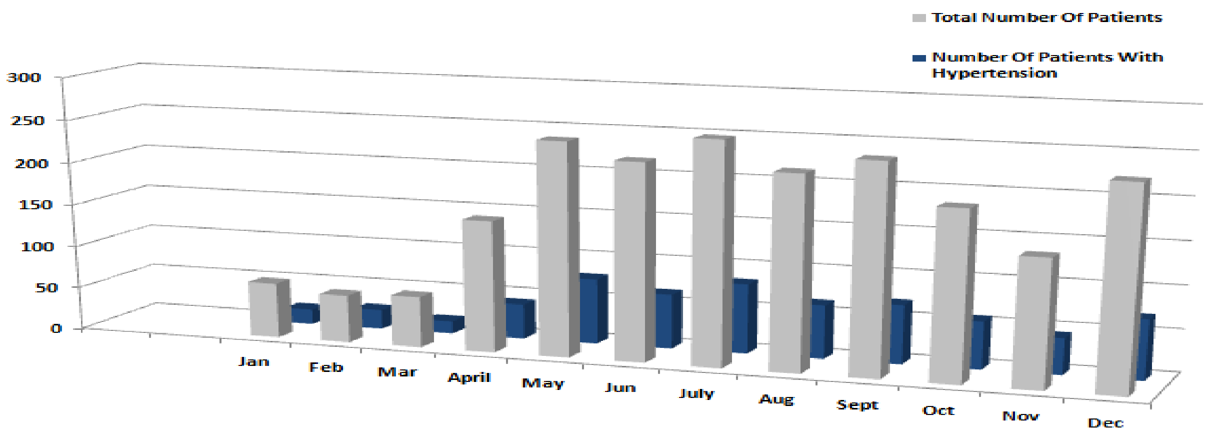

Fig. 2: Monthly distribution of number of patients and patients with hypertension, 2016

Bayes' theorem and conditional probabilities

Bayes' theorem is considered when the sample space is partitioned into a set of mutually exclusive events (A_1, A_2, . . . , A_n) and within the sample space, there exists an event B, for which P(B) > 0.

Bayes' theorem can be derived from the definition of conditional probability. The probability of event A given event B is;

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0. \qquad (1)$$

More generally,

$$P(A_i | B) = \frac{P(B|A_i)\ P(A\ )}{P(B|A_1)\ P(A_1) + P(B|A_2)\ P(A_2) + \ .\ .\ . + P(B|A_n)\ P(A_n)}. \qquad (2)$$

$$P(A_i | B) = \frac{P(A_i)\,P(B|A_i)}{\sum_i^n P(A_i)\,P(B|A_i)} \qquad (3)$$

is the Bayes' theorem.

An expression for expanded Bayes' theorem can be obtained as follows:

Let M_1, M_2, M_3, . . . be statistical models, D is the observed data and C represents all the conditions and knowledge surrounding the random experiment. The expanded Bayes' theorem for two models is given as:

$$P(M_1|D, C) = \frac{P(D|M_1,C).\ P(M_1|C)}{P(D|M_1,C).\ P(M_1|C) + P(D|M_2,C).\ P(M_2|C)} \qquad (4)$$

$$P(M_2|D, C) = \frac{P(D|M_2,C).\ P(M_2|C)}{P(D|M_1,C).\ P(M_1|C) + P(D|M_2,C).\ P(M_2|C)} \qquad (5)$$

The expanded Bayes' theorem for three models is also given as:

$$P(M_1|D, C) = \frac{P(D|M_1,C).\ P(M_1|C)}{P(D|M_1,C).\ P(M_1|C) + P(D|M_2,C).\ P(M_2|C) + P(D|M_3,C).\ P(M_3|C)} \qquad (6)$$

$$P(M_2|D, C) = \frac{P(D|M_2,C).\ P(M_2|C)}{P(D|M_1,C).\ P(M_1|C) + P(D|M_2,C).\ P(M_2|C) + P(D|M_3,C).\ P(M_3|C)} \qquad (7)$$

$$P(M_3|D, C) = \frac{P(D|M_3,C).\ P(M_3|C)}{P(D|M_1,C).\ P(M_1|C) + P(D|M_2,C).\ P(M_2|C) + P(D|M_3,C).\ P(M_3|C)} \qquad (8)$$

In general, the expanded Bayes' theorem for n models is given as:

$$P(M_i | D,C) = \frac{P(D|M_i,C).\ P(M_i|C)}{\sum_i^n P(D|M_i,C).\ P(M_i|C)} \qquad (9)$$

For the purpose of this study, M_i are the events of a patient developing high blood pressure whose posterior probability is computed based on the observed data D and prior information C. Thus, the posterior probabilities $P(M_i|D,C)$ is computed as given in equation (9), the prior probabilities $P(M_i|C)$ is computed from data (monthly) as the ratio of the number of patients with high blood pressures to total number of patients while the likelihood function $P(D|M_i, C)$ is computed from the observed data for each age group.

In Bayesian paradigm, the prior and likelihood are used to compute the conditional distribution (the posterior distribution) of the unknowns given the observed data, from which statistical inference arise (Adeleke and Ogundeji, 2009).

The three important ingredients of Bayesian analysis are the prior density $P(\theta)$; a likelihood function $P(y|\theta)$ and marginal data density $P(y) = \int P(y|\theta)P(\theta)d\theta$. These are used to derive the posterior distribution $P(\theta|y)$, using Bayes' theorem.

The basic relation in Bayesian statistics is:

Posterior ∞ prior x likelihood (10)

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int P(y|\theta)P(\theta)d\theta} \qquad (11)$$

which describes the shape of the posterior distribution up to a multiplicative constant and where θ is the set of free parameters (Okafor and Ogundeji, 2012).

Results and Discussion

The probability of patients having hypertension within different age groups is estimated using Bayes' theorem. Based on the data collated in Tables 1 and 2, the probabilities of patients' visits to the medical centre at different age groups in 2015 and 2016 were computed and presented in Tables 3 and 4, respectively. The mean probabilities of the patient's visits for 2015 and 2016 are graphically represented in Figs. 3 and 4, respectively.

Table 3: Probabilities of patients' visits to the medical centre under different age groups, 2015

Month	Age Groups				
	Below 30	30-40	41-50	51-60	Above 60
Jan	0.0901	0.0963	0.3727	0.3975	0.0435
Feb	0.0138	0.1356	0.3747	0.4069	0.069
Mar	0.0211	0.1135	0.4036	0.3874	0.0746
April	0.0070	0.1315	0.4718	0.3239	0.0657
May	0.0313	0.2188	0.3728	0.3348	0.0424
June	0.0648	0.6644	0.1574	0.081	0.0324
July	0.1067	0.3600	0.2933	0.1733	0.0667
Aug	0.0400	0.2000	0.4640	0.2560	0.0400
Sept	0.0455	0.2636	0.4409	0.2364	0.0136
Oct	0.0291	0.4660	0.1942	0.2136	0.0970
Nov	0.0619	0.3162	0.4055	0.1959	0.0206
Dec	0.0346	0.5821	0.2046	0.1210	0.0576

Table 4: Probabilities of patients' visits to the medical centre under different age groups, 2016

Month	Age group				
	Below 30	30-40	41-50	51-60	Above 60
Jan	0.0462	0.2308	0.4154	0.2308	0.0769
Feb	0.0357	0.1250	0.3214	0.4642	0.0536
Mar	0.0667	0.2000	0.2833	0.3833	0.0667
April	0.0325	0.1883	0.3701	0.3311	0.0779
May	0.0361	0.2490	0.3534	0.2932	0.0682
June	0.0435	0.2565	0.3304	0.2957	0.0739
July	0.0579	0.2200	0.3282	0.3320	0.0618
Aug	0.0619	0.2655	0.3274	0.2832	0.0619
Sept	0.1148	0.0943	0.3934	0.3361	0.0615
Oct	0.0305	0.3096	0.3046	0.3198	0.0355
Nov	0.0338	0.1757	0.3649	0.3649	0.0608
Dec	0.0940	0.2094	0.3419	0.2863	0.0684

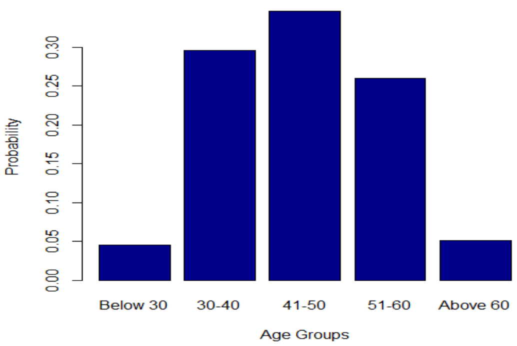

Fig. 3: Probability of patients' visits 2015

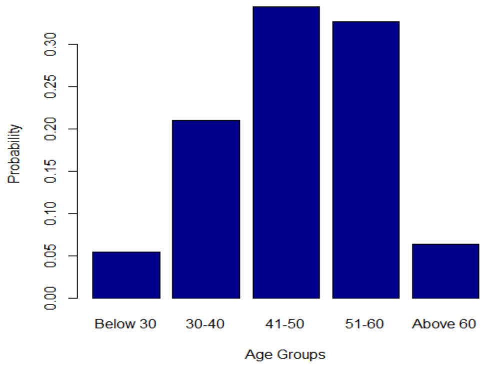

Fig. 4: Probability of patients' visits 2016

Using Bayes' theorem and based on the data collated in Tables 1 and 2, the posterior probabilities of having high blood pressures given a particular age group were computed for 2015 and 2016, respectively. These results are presented in Table 5 and graphically displayed in Figs. 5 and 6.

Table 5: Probabilities of high blood pressure (HBP) given an age group

Year	Age group				
	Below 30	30-40	41-50	51-60	Above 60
2015	0.2370	0.2724	0.3857	0.4584	0.5754
2016	0.1407	0.1696	0.2718	0.3745	0.5847

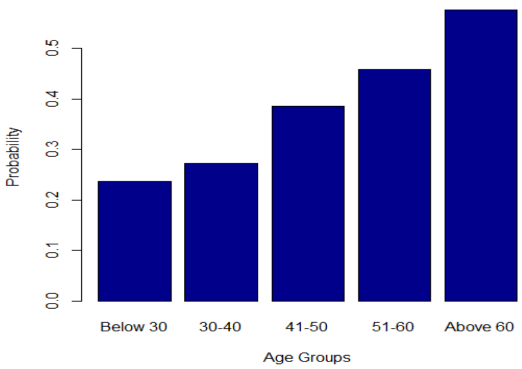

Fig. 5: Prob. HBP given age groups (2015)

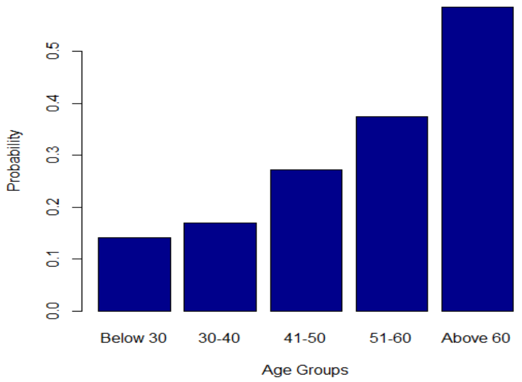

Fig. 6: Prob. HBP given age groups (2016)

For the purpose of comparing the relationship between the posterior probabilities and ages of patients with high blood pressures, both Pearson Moments and Spearman Rank Correlation Coefficients were computed for 2015 and 2016 data (Table 5). Furthermore, considering the number of observations (sample size n = 5), bootstrap samples of 50, 100, 1,000, 10,000 and 100,000 were obtained for correlation analysis. These results are shown in Table 7 and Figs. 9 – 10, respectively displayed the scatter plots for 2015 and 2016 results in Table 5.

Table 7: Results of correlation analysis with bootstrap samples

Year	Coefficients	N 5	Bootstrap Samples 50	100	1,000	10,000	100,000
2015	Spearman:	1.000	1.000	1.000	1.000	1.000	1.000
	Pearson:	0.976	0.976	0.976	0.976	0.976	0.976
2016	Spearman:	1.000	1.000	1.000	1.000	1.000	1.000
	Pearson:	0.967	0.967	0.967	0.967	0.967	0.967

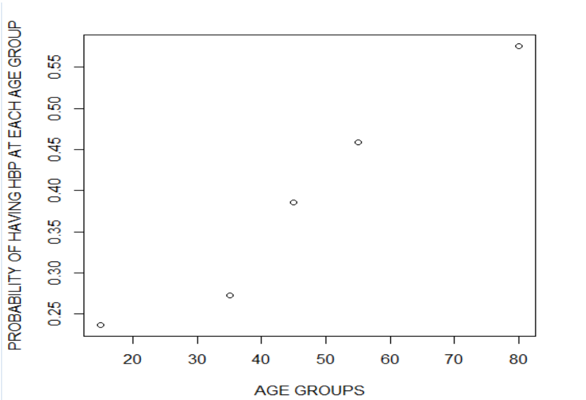

Fig. 9: Scatter plot of probabilities and age (2015)

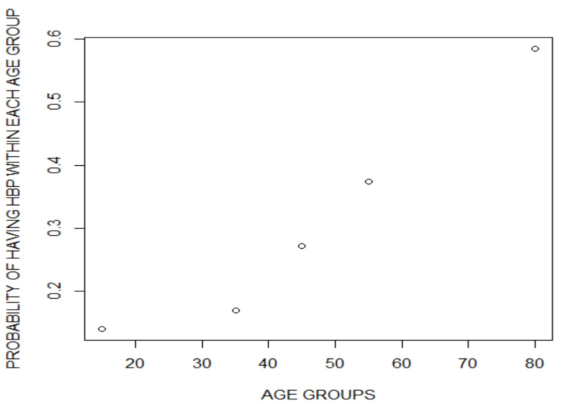

Fig. 10: Scatter plot of probabilities and age (2016)

Conclusion and Recommendations

The study shows that for the two years (2015 and 2016) considered, the prevalence of having high blood pressure (measured in probability values given an age group) is highest for patients above 60 years, followed by at age 51-60, 41-50, and 30-40 in descending order while the probability of having high blood pressure below age 30 is of the lowest. It means that even people below the age 30 do have the chances of developing high blood pressure as against presumable zero probabilities. Furthermore, it means that the probability of having high blood pressure below age 30 is low compared with other age groups and that young people (below 30years) are also likely to die of hypertension. Unlike in some health centres, the study suggests that blood pressure should be checked for all patients above the age of 18 years.

For 2015 and 2016, the correlation coefficients between the age groups and the probabilities of having high blood pressure are 0.976 and 0.967, respectively using Karl Pearson's Coefficient of Correlation. However, using a non-parametric method (Spearman's Rank Correlation method the correlation coefficient is 1.00. Even when computed with bootstrap samples to increase the sample sizes, the correlation coefficients remain unchanged. By whatever approach, there exist a very strong positive relationship between the age and the probability of developing high blood pressure. That simply means that an increase in age leads to a corresponding increase in the chances of having High Blood Pressure. Thus, people below 30 years of age can also die of high blood pressures. Both young and old people have to be conscious of their blood pressure status.

Conflict of Interest

Authors have declared that there is no conflict of interest reported on the work.

References

Adediran OS, Okpara IC, Adeniyi OS & Jimoh AK 2013. Prevalence of hypertension in Akwa Ibom State, South-South Nigeria: Rural versus urban communities study. *J. Med. and Med. Sci.*, 4(4): 149-154.

Adeleke IA & Ogundeji RK 2009. Bayesian estimation of the proportion of subscribers to Nigeria's national health insurance scheme. *J. Modern Math. and Stat.*, 3(3): 56 – 59.

Adeloyea D, Basquilla C, Aderemib AV, Jacqueline Y, Thompsonc JY & Obid FA 2014. An estimate of the prevalence of hypertension in Nigeria: A systematic review and meta-analysis. *Journal of Hypertension*, DOI: 10.1097/HJH.0000000000000413·Source: PubMed.

Davison AC 2008. Statistical Models. Cambridge University Press, New York.

Ekwunife OI & Aguwa CN 2011. A meta analysis of prevalence rate of hypertension in Nigerian populations. *J. Public Health and Epidemiol*., 3(13): 604 – 607.

Jackman S 2009. Bayesian Analysis for the Social Sciences. Wiley Series in Probability and Statistics.

Lee PM 2004. Bayesian Statistics: An Introduction. Oxford University Press, New York, Third Edition.

Okafor R & Ogundeji RK 2012. A Bayesian model for election forecasting: A case study of voting intentions of students of University of Lagos. *Joint Statistical Meeting, (JSM 2012)*: 1747 – 1755.

Ogundeji RK & Okafor R 2010. Empirical Bayes approach to estimate mean CGPA of University graduates. *Int. J. Appl. Math. and Stat.*, 17(10): 77-89.

Rao CR, Kamath VG, Shetty A & Kamath A 2013. *A Quantitative Analysis from Coastal Karnataka, India*. Hindawi Publishing Corporation. ISRN Preventive Medicine. Volume 2013, Article ID 574973, 6 pages .http://dx.doi.org/10.5402/2013/574973.

Wang Z, Qingyun D, Liang S, Nie K, Lin D, Chen Y & Li J 2014. Analysis of the spatial variation of hospitalization admissions for hypertension disease in Shenzhen, China. *Int. J. Environ. Res. Public Health*, 11: 713-733; doi:10.3390/ijerph110100713.