

RESEARCH ARTICLE

Bayesian Inference on Malignant Breast Cancer in Nigeria: A Diagnosis of MCMC Convergence

Ropo Ebenezer Ogunsakin*, Lougue Siaka

Abstract

Background: There has been no previous study to classify malignant breast tumor in details based on Markov Chain Monte Carlo (MCMC) convergence in Western, Nigeria. This study therefore aims to profile patients living with benign and malignant breast tumor in two different hospitals among women of Western Nigeria, with a focus on prognostic factors and MCMC convergence. **Materials and Methods:** A hospital-based record was used to identify prognostic factors for malignant breast cancer among women of Western Nigeria. This paper describes Bayesian inference and demonstrates its usage to estimation of parameters of the logistic regression via Markov Chain Monte Carlo (MCMC) algorithm. The result of the Bayesian approach is compared with the classical statistics. **Results:** The mean age of the respondents was 42.2 ± 16.6 years with 52% of the women aged between 35-49 years. The results of both techniques suggest that age and women with at least high school education have a significantly higher risk of being diagnosed with malignant breast tumors than benign breast tumors. The results also indicate a reduction of standard errors is associated with the coefficients obtained from the Bayesian approach. In addition, simulation result reveal that women with at least high school are 1.3 times more at risk of having malignant breast lesion in western Nigeria compared to benign breast lesion. **Conclusion:** We concluded that more efforts are required towards creating awareness and advocacy campaigns on how the prevalence of malignant breast lesions can be reduced, especially among women. The application of Bayesian produces precise estimates for modeling malignant breast cancer.

Keywords: Bayesian - malignant breast cancer – MCMC

Asian Pac J Cancer Prev, **18** (10), 2709-2716

Introduction

Breast cancer is the commonest cause of mortality and morbidity among women worldwide, and currently the most common cancer among Nigerian women (Adebamowo and Ajayi, 1999; Ebughe et al., 2013; Ojewusi and Arulogun, 2016; Oladimeji et al., 2015; Banjo, 2004). The human breast is a pair of mammary glands composed of specialized epithelium and stroma in which both benign and malignant lesions can occur (Dauda et al., 2011). Benign breast constitutes the larger of the breast lesions but much concern is given to malignant lesions of the breast since breast cancer is the most frequent malignancy in the majority of the women (Uwaezuoke and Udoye, 2014). Globally, breast cancer accounts for 18.4% of cancers associated with women. In 2012, (Jedy-Agba et al., 2012) reported that the incidence of breast cancer in Nigeria has risen significantly with the incidence in 2009-2010 reported to be at 54.3 per 100 000, thereby representing a 100% increase within the last decade. The report about patients diagnosed with breast cancer in eastern Nigeria suggested that every 1 out of 5, representing 23%, are malignant in nature (Yusufu et al., 2003). From literature, we found that previous studies

only focused on benign breast cancer (Abudu et al., 2007; Adesunkanmi and Agbakwuru, 2000; Forae et al., 2014; Guray and Sahin, 2006; Kumar et al., 2014; Anyikam et al., 2008; Godwins et al., 2011; Ochicha et al., 2002).

The application of the Bayesian technique and its usage to analyze cancer data has proliferated in recent years. Several researchers such as (Acquah, 2013) studied the comparison of Bayesian and classical and found that Bayesian gave a better result than the classical statistics. Other studies have also shown similar result (Yu and Wang, 2011; Mila and Michailides, 2006; Albert, 1996; Congdon, 2014; Marrelec et al., 2003; Daíz and Batanero, 2016; Lozano-Fernández, 2008; Gordóvil-Merino et al., 2010). In general, studies comparing both methods find that Bayesian technique proffers a better solution compared to classical statistics. The Bayesian technique assumes model parameters as random variables and not as constants, while the probability of the unascertained parameters can be obtained via Bayes theorem (Congdon, 2005; O'Neill, 2002; O'Neill et al., 2000; Wong and Ismail, 2016). This provides information regarding parameter uncertainty that might be very difficult to obtain using the classical technique. Classical technique fits the logistic regression by means of an iterative approach

and in some cases, as a result of this iterative approach, convergence may be difficult to achieve. The robustness and accuracy of the results produced by Bayesian approach makes its gain popularity in data analysis. As such, this paper investigates the significant predictors as well as characterizing patients diagnosed of benign and malignant breast cancer lesion using both classical approach and Bayesian approach.

Materials and Methods

Data Collection

Ethical approval was obtained from the ethics committee of the Federal Medical Teaching Hospital, Ekiti State, Nigeria. This data was extracted from cancer registry of the Federal Medical Teaching Hospital. We accessed 237 records and 20 variables of breast cancer data. Some of these variables describe socio-demographic and cancer-specific information of an incidence of breast cancer. Extensive variable selection procedures were performed on the 20 variables. The records of patients aged 20 years and above were sorted out for this analysis. Information collected includes age, marital status, educational level, religion, race, type of breast cancer, occupation, Lab number, case number, site of the female breast cancer, type of diagnosis and histological type. Other information recorded was the modality of treatment received: surgery, chemotherapy, hormonal therapy, radiotherapy. The software R was used for the classical statistical analysis and the software WinBUGS14 for the Bayesian analysis. As a requirement of the Bayesian approach, several diagnostics tests were performed to answer convergence of the Markov chain Monte Carlo (MCMC) algorithm and the true reflection of the posterior distribution.

Bayesian Binary Logistic Regression

Bayesian logistic regression, which applies Bayesian inference, has the formulation of a logistic equation and includes both continuous and categorical explanatory variables. Binary regression model is used to describe the probability of a binary response variable as function of some covariates. The logistic regression model belongs to the class of Generalized Linear Models. Generalized linear models generalize the standard linear model:

$$\Phi(\eta) = \tau_i^T \lambda \tag{1}$$

Binary logistic regression model is represented as:

$$\pi_i | \eta_i \sim \text{Bin}(\eta_i) \tag{2}$$

If the response under consideration is observed, we have $\pi_i = 1$ for the i th individual and zero otherwise. And π_i is the probability that the i th individual presents the response under consideration, λ is the j vector of unknown parameters, $\tau_i^T = (\tau_{i1}, \dots, \tau_{ij})$ represent vector of known covariates associated to the i th individual. Therefore, G is now define as any transformation assuming values between 1 and 0. Since function G can be any arbitrary

cumulative distribution, this study consider only the logistic part as earlier mentioned. Hence,

$$G(\tau_i^T \lambda) = \frac{\exp(\tau_i^T \lambda)}{(1 + \exp(\tau_i^T \lambda))}$$

The link function defines the linear predictor as expressed below

$$\varphi_i = G^{-1}(\eta_i) = \lambda_{-1} \tau_{i1} + \dots + \lambda_j \tau_{ij} \tag{3}$$

Suppose η_i denotes the probability of having malignant or benign breast lesion, the logit transformation is expressed as

$$\text{logit}(\eta_i) = \log\left(\frac{\eta_i}{1 - \eta_i}\right) \tag{4}$$

The likelihood function for data $\pi = (\pi_1, \dots, \pi_k)^T$ is expressed as

$$P(\pi_i | \lambda) = \prod_{i=1}^n [G(\tau_i^T \lambda)]^{\pi_i} [1 - G(\tau_i^T \lambda)]^{1 - \pi_i} \tag{5}$$

For the Bayesian analysis, it is important to provide a joint prior distribution over the parameter space. The preferred prior for logistic regression parameters is a multivariate normal distribution and is given by (Ojo et al., 2017; Ntzoufras, 2011):

$$\lambda_k \sim N(b_0, \Sigma_0^2) \tag{6}$$

where $\Sigma_1^{-1} = \Sigma_0^{-1} + \Sigma_{ML}^{-1}$ and

$b_1 = \Sigma_1 | \Sigma_{ML}^{-1} | \lambda_{ML} + \Sigma_1 | \Sigma_0^{-1} | b_0$ Note ML is the maximum likelihood estimate. Hence we have

$$p(\lambda_k) \propto \prod_{i=1}^k \exp\left[-\frac{1}{2}(\lambda_k - b_0)' \Sigma_k^{-1} (\lambda_k - b_0)\right] \tag{7}$$

Therefore, the posterior distribution is represented as follows

$$P(\lambda | \pi) \propto p(\lambda_k) \prod_{i=1}^n [G(\tau_i^T \lambda)]^{\pi_i} [1 - G(\tau_i^T \lambda)]^{1 - \pi_i} \tag{8}$$

The latter part of expression (8) can be regarded as normal distribution for parameters λ and it has no closed form. Posterior distribution is usually of high dimension and analytically intractable which sometimes required knowledge of powerful integration. In order to overcome these difficulties, Markov chain Monte Carlo (MCMC) algorithm is needed (Ngesa et al., 2014). MCMC technique is among of the technique employed to generates the estimates of unknown parameters θ and corrects the values generated in order to have a better estimate of the desired posterior distribution, $p(\theta | \eta)$ (Ojo et al., 2017; Ntzoufras, 2011). When MCMC is employed to generate a sample of $p(\theta | \eta)$, there is need to check that the MCMC

Table 1. Heidelberger and Welch Stationarity and Half-Width Tests for the Bayesian Chains Used in the Diagnosis of MCMC

Param.	Stationarity Test	P-Value			Half-width Test	Mean			Half width		
		C1	C2	C3		C1	C2	C3	C1	C2	C3
λ_0	passed	0.927	0.888	0.308	passed	0.117	0.132	0.129	0.054	0.054	0.055
λ_1	passed	0.591	-0.219	0.364	passed	-1.148	-1.144	-1.148	0.01	0.009	0.01
λ_2	passed	0.0572	0.818	0.204	passed	1.348	1.343	1.35	0.01	0.011	0.011
λ_3	passed	0.394	0.51	0.994	passed	0.836	0.838	0.839	0.013	0.012	0.012
λ_4	passed	0.915	0.987	0.112	passed	1.22	1.216	1.22	0.016	0.016	0.016
λ_5	passed	0.893	0.815	0.313	passed	-0.216	-0.226	-0.228	0.044	0.044	0.044
λ_6	passed	0.808	0.914	0.237	passed	-0.977	-0.977	-0.983	0.038	0.037	0.038
λ_7	passed	0.824	0.94	0.507	passed	-1.536	-1.55	-1.551	0.044	0.044	0.044
λ_8	passed	0.64	0.954	0.163	passed	0.6168	0.613	0.612	0.02	0.042	0.019
λ_9	passed	0.4	0.896	0.092	passed	1.054	1.058	1.053	0.009	0.009	0.009

algorithm converges to the desired posterior distribution (Ojo et al., 2017).

Assessing Bayesian Markov Chain Monte Carlo and Convergence

In this study, non-informative prior were assumed in order not to influence the posterior distribution and it was assumed that $\lambda_k \sim N(1, 0.0001)$. All the Bayesian analysis was carried out using WinBUGS 14 (Ntzoufras, 2011). We ran 1,500,000 Markov chain Monte Carlo (MCMC) iterations, with the initial 200,000 discarded to cater for the burn-in period. The 5,000 iterations left were used for assessing convergence of the MCMC. We assessed MCMC convergence of our model parameters by checking Heidelberger-Welch diagnostic, autocorrelation plot, Gelman-Rubin plots (Gelman et al., 2014a), and running quantiles of the MCMC output.

Gelman-Rubin

The diagnostic of Gelman and Rubin requires two or more chains from over-dispersed starting points by computing the within and between chains variability respectively. Large deviation between two variances implies non-convergence of the chain. If all the chains have converged as expected, the posterior marginal variance estimate is expected to be very close within the chain variance. The test statistics for the Gelman-Rubin diagnostic test can be estimated as follows (Lesaffre and Lawson, 2012):

$$W = \frac{1}{M} \sum_{m=1}^M S_m^2$$

$$B = \frac{n}{M-1} \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta})^2$$

$$\hat{v} = \frac{n-1}{n} W + \frac{1}{n} B$$

where k is the number of iterations of the chains.
 $\hat{R} = \frac{\hat{v}}{W}$

Convergence is monitored when $\hat{R} \rightarrow 1$. \hat{R} is called the estimated potential scale reduction factor (PSRF). Brooks and Gelman (Gelman et al., 2014a) proposed an alternative approach that generalizes the initial method to consider more than one parameter concurrently. The estimate of the posterior variance covariance is now computed as:

$$\hat{v} = \frac{n-1}{n} W + \left(1 + \frac{1}{m}\right) B/n$$

where

$$W = \frac{1}{m(n-1)} \sum_{m=1}^M \sum_{t=n+1}^{2n} (\theta'_m - \bar{\theta}_m)(\theta'_m - \bar{\theta}_m)'$$

and

$$B/n = \frac{1}{M-1} \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta})^2$$

denote the q -dimensional within and between covariance matrix estimates of the q -variate. It then imply, if λ_m is the highest eigen value of, hence

$$\hat{R}^c = \frac{n-1}{n} + \left\{\frac{m+1}{m}\right\} \lambda_m \quad (9)$$

Where \hat{R}^c is the multivariate potential scale reduction factor (MPSRF). Convergence is attained when multivariate shrink factor converges to 1.

Heidelberger and Welch's: In order to test the hypothesis of stationarity, we first propose that we have a sequence $\{X^t : t = 1, 2, \dots, k\}$ from a covariance stationary process with unknown spectral density, $S(\omega)$. Therefore, for, $k \geq 1$

$$R_0 = 0, \quad R_k = \sum_{t=1}^k \theta^t$$

and

$$\bar{\theta} = \frac{1}{k} R_k$$

$$\hat{B}_k(r) = \frac{R_k(r) - [kr]\bar{\theta}}{\{k\hat{S}(0)\}^{1/2}}, \quad 0 \leq r \leq 1 \quad (10)$$

Table 2. WinBUGS Posterior Summaries for Breast Cancer Patients

	Mean	SD	MC error	2.50%	Median	97.50%	start	Sample
λ_0	0.126	1.973	0.01568	-3.514	0.049	4.251	5,000	49,749
λ_1	-1.147	0.669	0.003012	-2.453	-1.146	0.176	5,000	49,749
λ_2	1.347	0.637	0.002987	0.18	1.316	2.695	5,000	49,749
λ_3	0.838	0.815	0.003789	-0.637	0.796	2.561	5,000	49,749
λ_4	1.219	0.92	0.004805	-0.633	1.224	3.012	5,000	49,749
λ_5	-0.223	1.672	0.001283	3.91	-0.102	4.685	5,000	49,749
λ_6	-0.979	1.523	0.01116	-4.418	-0.836	1.603	5,000	49,749
λ_7	-1.545	1.686	0.01269	-5.25	-1.415	1.371	5,000	49,749
λ_8	0.614	1.133	0.005761	-1.454	0.554	3.002	5,000	49,749
λ_9	1.055	0.59	0.00283	-0.093	1.047	2.241	5,000	49,749

$\vec{B}_k(r)$ is approximately distributed as a Brownian bridge for large k , where

$$\vec{B}_k(r) = \{\vec{B}_k(r) : 0 \leq r \leq 1\}$$

Hence, the null hypothesis for stationarity is now tested using Cramer-von Mises statistic.

Results

Socio-demographic profile of participants

The main goal of this paper is to investigate the significant predictors as well as characterizing patients diagnosed of benign and malignant breast cancer lesion and presents diagnosis of MCMC convergence in western Nigeria, comparing the classical approach and Bayesian approach. Various prognostic factors are considered which include: intercept (λ_0), marital status: separated (λ_1), level of education: at least high school (λ_2), religion: Christian (λ_3), tribe: yoruba (λ_4), age: 35-49 (λ_5), 50-69 (λ_6), 70+ (λ_7), occupation: retired (λ_8), self employed (λ_9). A total of 237 breast cancer patients' data was extracted for analysis in the current study. Of these, 192 cases accounting for (81.01%) were malignant breast lesions, while 45 cases (18.99%) were benign giving a ratio of 4.3:1 for malignant to benign breast lesion. The mean age of the respondents was 42.2 ± 16.6 years with 52% of the women aged between 35-49 years. Table 1 shows the Heidelberg and Welch stationarity tests for the Bayesian Markov chain Monte Carlo. It shows the stationarity and convergence during the burn-in period.

Table 2 present the result of MCMC diagnostics for the patients diagnosed with benign and malignant breast cancer. The posterior means were obtained after a burn-in period of 5,000 with Monte Carlo error less than 2%. The posterior means and medians of the coefficient and indicate significance. The results of the posterior provide some evidence about the important variable to be selected while profiling patients diagnosed with malignant breast cancer. For , Table 2 shows that those with at least high school education are 1.3 times more likely than others to have benign breast cancer. The results indicate that women with age ≥ 35 years were at a higher risk of been diagnosed with malignant breast cancer than those with

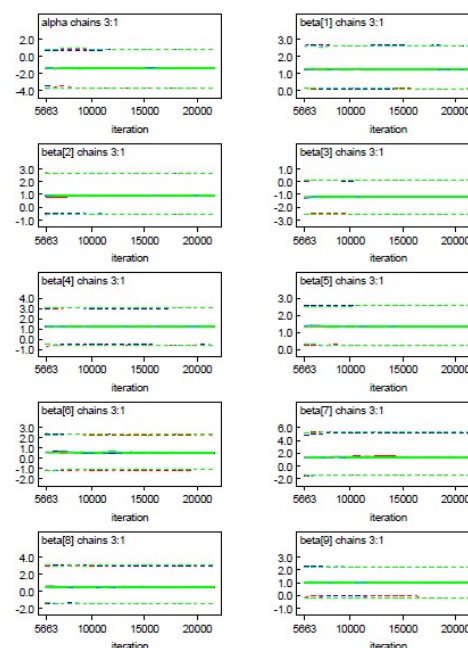


Figure 1. Running Quantiles for the Posterior Parameters in the Case of Female Benign and Malignant Breast Cancer Patients

age < 35 years.

Table 3 shows the result of a classical logistic analysis of the malignant breast cancer. The results indicate that malignant was observed to be strongly associated with age and educational status. This indicates that women with

Table 3. Result of Classical Logistic Regression for Patients Diagnosed of Benign and Malignant

	Est	Std Error	z value	Pr(> z)
λ_0	-2.421	1.2308	-1.967	0.0492
λ_1	1.2479	0.5976	2.088	0.04459
λ_2	0.5926	0.7774	0.762	0.0368
λ_3	1.0782	0.6392	1.687	0.0916
λ_4	1.2048	0.8515	1.415	0.1571
λ_5	1.1952	0.5732	2.085	0.0371
λ_6	0.5034	0.8188	0.615	0.5387
λ_7	1.0534	1.4439	0.73	0.4656
λ_8	0.4823	1.0432	0.462	0.6439
λ_9	0.9898	0.5658	1.749	0.0802

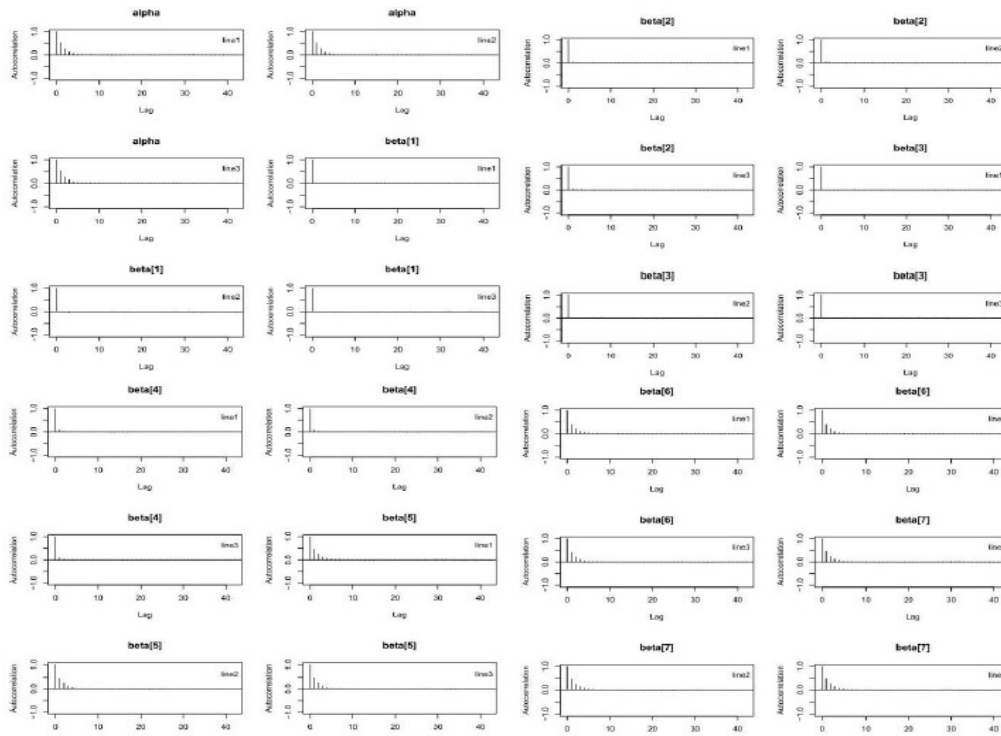


Figure 2. Auto-Correlation Plots for the Female Benign and Malignant Breast Cancer Patients

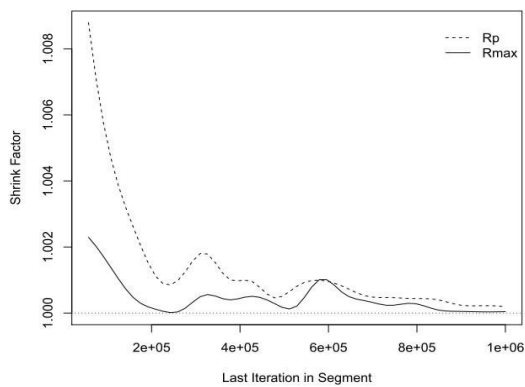


Figure 3. The Plot of the Brooks-Gelman MPSRF for Three Chains of 49,749 Iterations

at least high school education have a significantly higher risk of being diagnosed with malignant breast tumors. *Assessing the performance of Markov Chain Monte Carlo (MCMC) chains in WinBUGS*

When the results of the model are computed, it is necessary to check for the stationarity of Markov chain Monte Carlo algorithm. Both Figure 1 and Figure 2 were presented to demonstrate that there is no problem of autocorrelation among the MCMC chain. The blue and red lines in Figure 4 denote the variance within and between chains. To support that the chain is converged, the ratio must converge to one and the blue and red lines must converge to a stable value. It also displays the red lines representing the potential scale reduction factor

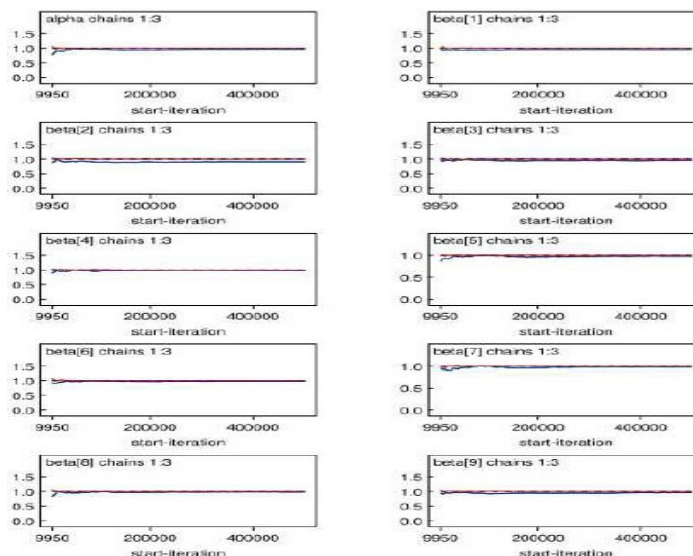


Figure 4. Gelman Rubin Convergence Diagnosis for Independent Variables

denoted by $\hat{\theta}$. Hence, Figure 4 indicates that all the $\hat{\theta}_i \rightarrow 1$ which suggests that the algorithm converges. Both Figure 3 and Figure 4 explain the same thing but one is obtained through CODA/BOA and the other through WinBUGS.

Discussion

The present confirm findings from studies conducted in Nigeria over the past years, on Breast cancer among women in the western Nigeria (Olugbenga et al., 2012; Abudu et al., 2007). All these studies showed that age could be a risk factor for malignant breast lesion. Similar studies have been documented in other parts of Africa and the rest of the world (Arora and Simmons, 2009). From the results of analysis, patient's age 35-49 years constituted the majority of patients (52%) in our study, indicating that women age 35-49 have a higher risk of developing malignant breast cancer than their other counterpart in the group. Therefore, more attention on breast cancer treatment are necessary for this age group. This agrees with breast cancer facts and figures released between 2011-2012. However, this corresponds to the working class population and it is also the child bearing age for many women. This may be as a result of the use of contraceptive and hormonal imbalance which common among the women (Onyeanus, 2015; Olugbenga et al., 2012). From the study, malignant breast lesions appeared to have higher distribution among those who had at least high school education, an observation which supports previous studies (Yuksel et al., 2017; Ibrahim et al., 2015, Yusufu et al., 2003; Ntekim et al., 2009). This result was supported by the descriptive statistics which shows that 52.3% of those diagnosed had at least high school education meaning that those who are educated were more interested in presenting their health problems to rather than consulting fake medical doctors. The high proportion of malignant breast lesion might also be attributed with lifestyle changes among those educated. In addition, this may also be as a result of their exposure to advancement in life like the nature of occupation, diet, without observing caution to health management. We found that the mean age of breast cancer patients in western is 42 years; this is similar to several Nigerian institution based studies, Adebamowo reported 43 years (Adebamowo and Adekunle, 1999), Ikpat et al. 42.7 years (Ikpat et al., 2002) and 44.9 years by Ebughe et al (Ebughe et al., 2013). Although our variables' interactions did not categorize age and educational status as components of socioeconomic status (SES), our findings are similar to those of some studies which showed that higher socioeconomic status (SES) is associated with higher breast cancer incidence (Pudrovska and Anikputa, 2012; Krieger et al., 2010; Vainshtein, 2008). Additional studies have provided a possible explanation for these findings that women with high SES are more likely to obtain routine breast cancer screening due to better access to preventive healthcare based on their level of education and increasing age, hence, increasing the detection of breast cancer (Akinyemiju et al., 2015).

Although our study did not investigate the risk factors for breast cancer in association with its sub molecular

types, a recent study conducted by (Akinyemiju et al., 2015) evaluated the association between SES and breast cancer subtypes using a valid measure of SES and the Surveillance, Epidemiology and End Results (SEER) database. Socioeconomic status based on measures of income, occupational class, education and house value, were categorized into quintiles and explored. Their findings showed that a positive association between SES and breast cancer incidence is primarily driven by hormone receptor positive lesion. Malignant breast lesions which can be subdivided into non-invasive and invasive tumors are documented to be more commonly diagnosed in postmenopausal women (LehmannChe et al., 2013). A molecular classification of breast cancer, with more than five reproducible subtypes (basal-like, ERBB2, normal-like, luminal A, and luminal B) has been defined through gene expression profiling and microarray analysis (Lønning et al., 2007). In addition, performing the gene set enrichment analysis (GSEA), a gene set linked to the growth factor (GF) signaling was observed to be significantly enriched in the luminal B tumors (Loi et al., 2009). Another study states that multiple pathways were identified by mapping gene sets defined in Gene Ontology Biological Process (GOBP) for estrogen receptor positive (ER+) or estrogen receptor negative (ER-); and among them, in a separate set, pathways related to apoptosis and cell division or G-protein coupled receptor signal transduction were associated with the metastatic capability of ER+ or ER- tumours, respectively (Jack et al., 2007).

The plot of Gelman Rubin convergence in Fig 4 suggesting that the MCMC sequence has converged on the posterior density as red line fall towards one for all parameters. Our findings are similar to the result obtained by Salameh et al (Salameh et al., 2014), Jackman et al (Jackman, 2000). Figure 3 shows a plot of the Brooks-Gelman MPSRF (denoted $\hat{\rho}$) along with the maximum PSRF (denoted $\hat{\rho}_{max}$) for successively larger segments of the chains. This plot suggests that although the chains differs significantly for the first few thousand iterations, they mix together after that and three chains of 1,500,000 iterations each is probably sufficient to ensure convergence of the chains. It also suggests using a burn-in of about 200,000 each. The result in Table 2 shows that each parameter passes the stationarity and half-width test respectively. This suggests that for the current study, the stationarity of the Markov chain and the sample size obtained is adequate for the estimation of mean values of the three iterations.

Findings from Bayesian and classical inference are not significantly different which could be due to the non-informative prior utilized in the Bayesian model. When both techniques produced similar results, findings from Bayesian are given more attention because it is more robust compared to the classical. The model used in this paper updates quickly and adding complexity will also improve the required time for updating. This diagnostic are necessary to ensure that we are actually sampling from a chain that has converged after a desirable burn-in. Using the posterior mean as a point estimate, Table3 compares the classical statistics estimates with the simulation (MCMC) result. The estimated means and standard errors appear quite close with minimum results

show a reduction of standard errors associated with the coefficients obtained from the Bayesian approach, hence resulting in higher stability to the coefficients. Other studies have also shown similar result (Gordóvil-Merino et al., 2010; Acquah, 2013).

Findings of this study shows that age of the patients and those with at least high school education are at higher risk of being diagnosed with malignant breast lesion than benign breast lesion in Western Nigeria. The higher proportion of those affected by malignant breast lesion is found among the educated and younger women. Therefore, this shows that non-educated women do not patronize these services based on our findings. More efforts are required towards creating awareness and advocacy campaigns on how the prevalence of malignant breast lesions can be reduced, especially among women.

We recommend that governments, non-governmental organizations and other sectors involved in policy making to put in place policies, strategies and sensitization that target non-educated women to enhance their patronization of breast cancer screening in the health facilities, so as to access the appropriate management health assessment as well as providing financially supported treatments for breast cancer patients.

Statement conflict of Interest

The authors have declared no conflict of interest.

References

- Abudu E, Banjo A, Izegbu M, et al (2007). Malignant breast lesions at olabisi onabanjo university teaching hospital (oouth), sagamu-a histopathological review. *Niger Postgrad Med J*, **14**, 57–9.
- Acquah HDG (2013). Bayesian logistic regression modelling via markov chain montecarlo algorithm. *J Soc Dev Sci*, **4**, 193–7.
- Adebamowo C, Adekunle O (1999). Case-controlled study of the epidemiological risk factors for breast cancer in nigeria. *Br J Surg*, **86**, 665–8.
- Adebamowo CA, Ajayi O (1999). Breast cancer in Nigeria. *West Afr J Med*, **19**, 179–91.
- Adesunkanmi A, Agbakwuru E (2000). Benign breast disease at wesley guild hospital, ilesha, nigeria. *West Afr J Med*, **20**, 146–51.
- Akinyemiju TF, Pisu M, Waterbor JW, Altekruze SF (2015). Socioeconomic status and incidence of breast cancer by hormone receptor subtype. *Springerplus*, **4**, 508.
- Albert JH (1996). Bayesian computation using Minitab. Duxbury Press, pp 117-24.
- Anyikam A, Nzegwu MA, Ozumba BC, Okoye I, Olusina DB (2008). Benign breast lesions in eastern Nigeria. *Saudi Med J*, **29**, 241–4.
- Arora N, Simmons RM (2009). Malignant breast disease: Diagnosis and assessment. *General Surgery*, pp 1481–94.
- Banjo A (2004). Overview of breast and cervical cancers in Nigeria: are there regional variations. In Paper presentation at the International workshop on new trends in Management of breast and cervical cancers, Lagos, Nigeria, pp 1-12.
- Congdon P (2005). Bayesian models for categorical data. John Wiley and Sons, pp 1-10.
- Congdon P (2014). Applied bayesian modelling, *Comput Stat Data Anal*, volume 595. John Wiley and Sons, pp 34-97.
- Dauda A, Misauno M, Ojo E (2011). Histopathological types of breast cancer in gombe, north eastern Nigeria: A seven-year review. *Afr J Reprod Health*, **15**, 107-9.
- Díaz C, Batanero C (2016). ¿cómo puede el método bayesiano contribuir a la investigación en psicología y educación?. *Paradigma*, **27**, 35–53.
- Ebughe G, Ekanem I, Omoronyia O, et al (2013). Age specific incidence of breast cancer in calabar, nigeria. *Int J Trop Dis*, **16**, 1-12
- Forae GD, Nwachokor FN, Igbe AP, et al (2014). Benign breast diseases in warri southern Nigeria: A spectrum of histopathological analysis. *Ann Nigerian Med*, **8**, 28.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2014a). Bayesian data analysis, volume 2. Chapman and Hall/CRC Boca Raton, FL, USA, pp 261-78.
- Gelman A, Hwang J, Vehtari A (2014b). Understanding predictive information criteria for bayesian models. *Comput Stat*, **24**, 997–1016.
- Godwins E, David D, Akeem J (2011). Histopathologic analysis of benign breast diseases in makurdi, north central nigeria. *Int J Med Sci*, **3**, 125–8.
- Gordóvil-Merino A, Guàrdia-Olmos J, Però-Cebollero M, de la Fuente-Solanas EI (2010). Classical and bayesian estimation in the logistic regression model applied to diagnosis of child attention deficit hyperactivity disorder. *Psychol Rep*, **106**, 519–33.
- Guray M, Sahin AA (2006). Benign breast diseases: classification, diagnosis, and management. *Oncologist*, **11**, 435–9.
- Ibrahim IM, Iliyasu Y, Mohammed AZ, et al (2015). Histopathological review of breast tumors in kano, northern nigeria. *Sub-Saharan Afr J Med*, **2**, 47.
- Ikpat O, Ndoma-Egba R, Collan Y (2002). Influence of age and prognosis of breast cancer in Nigeria. *East Afr Med J*, **79**, 651–7.
- Jack XY, Sieuwerts AM, Zhang Y, et al (2007). Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer*, **7**, 182.
- Jackman S (2000). Estimation and inference via bayesian simulation: An introduction to markov chain montecarlo. *Am J Political Sci*, **44**, 375–404.
- Jedy-Agba E, Curado MP, Ogunbiyi, et al (2012). Cancer incidence in Nigeria: a report from population-based cancer registries. *Cancer Epidemiol*, **36**, 271–8.
- Krieger N, Chen JT, Waterman PD (2010). Decline in us breast cancer rates after the women's health initiative: socioeconomic and racial/ethnic differentials. *Am J Public Health*, **100**, 132–9.
- Kumar V, Abbas AK, Fausto N, Aster JC (2014). Robbins and cotran pathologic basis of disease. Elsevier Health Sciences, pp 1043-52.
- Lehmann-Che J, Hamy AS, Porcher R, et al (2013). Molecular apocrine breast cancers are aggressive estrogen receptor negative tumors overexpressing either her2 or gdfp15. *Breast Cancer Res*, **15**, R37.
- Lesaffre E, Lawson AB (2012). Bayesian biostatistics. John Wiley and Sons, pp 176-86.
- Loi S, Sotiriou C, Haibe-Kains B, et al (2009). Gene expression profiling identifies activated growth factor signaling in poor prognosis (luminal-b) estrogen receptor positive breast cancer. *BMC Med Genomics*, **2**, 37.
- Lønning PE, Chrisanthar R, Staalesen V, Knappskog S, Lillehaug J (2007). Adjuvant treatment: the contribution of expression microarrays. *Breast Cancer Res*, **9**, S14.
- Lozano-Fernández LM (2008). Bayesian inference for binomial populations. bayesian estimation for child depression prevalence. *Adv Appl Stat*, **9**, 13–35.
- Marrelec G, Benali H, Ciuciu P, Pel'egrini-Issac, M, Poline JB (2003). Robust bayesian estimation of the hemodynamic response function in event-related bold fmri using basic

- physiological information. *Hum Brain Mapp*, **19**, 1–17.
- Mila AL, Michailides TJ (2006). Use of bayesian methods to improve prediction of panicle and shoot blight severity of pistachio in california. *Phytopathology*, **96**, 1142–7.
- Ngesa O, Mwambi H, Achia T (2014). Bayesian spatial semi-parametric modeling of hiv variation in kenya. *PLoS One*, **9**, e103299.
- Ntekim A, Nufu F, Campbell O (2009). Breast cancer in young women in ibadan, nigeria. *Afr Health Sci*, **9**, 242-6.
- Ntzoufras I (2009). Bayesian modeling in winbugs. hoboken. **698**, pp 447-60.
- Ntzoufras I (2011). Bayesian modeling using WinBUGS, volume 698. John Wiley and Sons.
- Ochicha O, Edino S, Mohammed A, Amin S (2002). Benign breast lesions in kano. *Niger J Surg Sci*, **4**, 1–5.
- Ojewusi AA, Arulogun OS (2016). Breast cancer knowledge and screening practices among female secondary schools teachers in an urban local government area, ibadan, nigeria. *J Public Health Epidemiol*, **8**, 72–81.
- Ojo OB, Lougue S, Woldegerima WA (2017). Bayesian generalized linear mixed modeling of tuberculosis using informative priors. *PLoS One*, **12**, e0172580.
- Oladimeji KE, Tsoka-Gwegweni JM, Igbodekwe FC, et al (2015). Knowledge and beliefs of breast self-examination and breast cancer among market women in ibadan, south west, nigeria. *PLoS One*, **10**, e0140904.
- Olugbenga AM, Olanrewaju MJ, Kayode OM (2012). Profile of cancer patients attending tertiary health institutions in southwestern nigeria. *Asian J Pharm Clin Res*, **5**, 34–7.
- O'Neill PD, Balding DJ, Becker NG, Eerola M, Mollison D (2000). Analyses of infectious disease data from household outbreaks by markov chain monte carlo methods. *J R Stat Soc Ser C Appl Stat*, **49**, 517–42.
- Onyeanusi CG (2015). Effect of family support on medication adherence and glycemic control of type 2 diabetes outpatients in a tertiary hospital in south-eastern Nigeria. PhD thesis, pp 51-67.
- O'Neill PD (2002). A tutorial introduction to bayesian inference for stochastic epidemic models using markov chain monte carlo methods. *Math. Biosci*, **180**, 103–14.
- Pudrovskaya T, Anikputa B (2012). The role of early-life socioeconomic status in breast cancer incidence and mortality: unraveling life course mechanisms. *J Aging Health*, **24**, 323–44.
- Salameh P, Waked M, Khayat G, Dramaix M (2014). Bayesian and frequentist comparison for epidemiologists: A non mathematical application on logistic regressions. *Open Epidemiol J*, **7**, 17-26.
- Uwaezuoke SC, Udoye EP (2014). Benign breast lesions in bayelsa state, niger delta nigeria: a 5 year multicentre histopathological audit. *Pan Afr Med J*, **19**, 395-5717
- Vainshtein J (2008). Disparities in breast cancer incidence across racial/ethnic strata and socioeconomic status: a systematic review. *J Natl Med Assoc*, **100**, 833–9.
- Wilhelmsen M, Dimakos XK, Husebø T, Fiskaen M (2009). Bayesian modelling of credit risk using integrated nested laplace approximations, NR publication, pp 1-25.
- Wong RSY, Ismail NA (2016). An application of bayesian approach in modeling risk of death in an intensive care unit. *PLoS One*, **11**, e0151949.
- Yu L, Wang L (2011). Bayesian and classical estimation of mixed logit model for simulated experimental data. *IEEE J Biomed Health Inform*, **28**, 255–8.
- Yüksel S, Ugras GA, Çavdar, I, et al (2017). A risk assessment comparison of breast cancer and factors affected to risk perception of women in turkey: A cross-sectional study. *Iran J Public Health*, **46**, 308.
- Yusufu L, Odigie V, Mohammed A (2003). Breast masses in zaria, nigeria. *Ann Afr Med*, **2**, 13–6.