ECRONICON OPEN ACCESS

EC NURSING AND HEALTHCARE
Research Article

Application of Bayesian Nonparametric Estimation on Cervical Cancer: A Case Study of University College Hospital, Nigeria

Onatunji Adewale P[1], Folorunso Serifat A[2]*, Oluwasola Timothy AO[3], Adesina Oluwaseun A[4] and Uthman Kafayat T[5]

[1]*LAUTECH Int'l College, Ogbomoso, Oyo State, Nigeria*

[2]*University of Ibadan Laboratory for Interdisciplinary Statistical Analysis (UI-LISA), Nigeria*

[3]*Department of Obstetrics and Gynecology, College of Medicine, University of Ibadan, Nigeria*

[4]*Department of Mathematics and Statistics, The Polytechnic, Ibadan, Oyo State, Nigeria*

[5]*National Centre for Genetic Resources and Biotechnology, Ibadan, Oyo State, Nigeria*

***Corresponding Author:** Folorunso Serifat A,Department of Statistics, University of Ibadan, Oyo State, Nigeria.

Received: October 28, 2020; **Published:** February 12, 2021

Abstract

Cervical cancer is the third most common malignancy in women worldwide, and in developing countries, it remains a leading cause of cancer-related death for women. The disease is a significant illness in women in Sub Sahara Africa Countries like Nigeria. Cervical cancer ranks as the 2nd most prevalent cancer among women in Nigeria and the most prevalent cancer among ages 15 - 44 years. Some of the identified risk factors of this disease have been wrongly captured in some statistical models. This consequentially makes the result of the analysis wrongly interpreted and concluded when the risk factors are continuous. This study is aimed at examining the Bayesian estimation of cervical cancer on a woman's age and survival period of the disease using the application of non-parametric techniques. The study is a means to apply Bayesian Nonparametric estimation on cervical cancer. The study findings reveal that women between ages 40 - 60 years have a significant increase in the probability of mortality from cervical cancer when the survival period is between stage I-IV. Also, there is an increase in woman's age and survival period of living with the disease which is significantly likely to decrease the mortality from cervical cancer in the study area.

Keywords: *Bayesian Estimation; Mortality; Risk Factors; Probability; Survival Time*

Introduction

Globally, approximately 10 million people are diagnosed with cancer annually. The increase of cancer incidence is occurring at a faster rate of 48% in developed countries and 52% in developing countries confirmed by Kanavos [1]. According to the World Health Organization, 100,000 new cancer cases have been reported in the country each year; and projected that by 2020, cancer incidence for Nigerian males and females would rise to 90.7/100,000 and 100.9/100,000, respectively. In Nigeria, with a population of nearly 200 million people, complex diseases such as cancer are currently emerging as important health care priorities for the future. A report by Stewart., *et al*. [2] mentioned that many of the new cancer cases are now occurring among women from low and middle-income countries such as Nigeria. Cervical cancer is one of the most common diseases which leads to mortality in African countries. WHO [3] reported that cervical cancer is the second killer disease in Africa. As developing countries succeed in achieving lifestyles similar to those in advanced economies, they will encounter much higher cancer rates, particularly gynecological cancers such as cervical cancer. This disease has continued

to be a threat to women's health in developing countries. Marc., *et al*. [4] revealed that about 570,000 cases of cervical cancer and 311000 mortality from it occurred in 2018. In Malawi, 132 (44%) patients in stage 1 cervical cancer and 168 (56%) patients in stage 1 - 4 were reported for the study periods of 310 patients (Pandaora., *et al.* 2017). Cervical cancer grows in the cervix (the entrance from the vagina to the uterus) of a woman. Almost all cases of cervical cancer (99%) are associated with high-risk human papillomavirus (HPV) infection, an extremely contagious virus that is transmitted through sexual contact. Once diagnosed, as long as it is treated early and controlled efficiently, cervical cancer is one of the most successfully treatable types of cancer. With adequate treatment and palliative care, cancers diagnosed in late stages can also be monitored [5].

According to Human Papillomavirus and Related Cancers, Fact Sheet 2018, it was reported that there are 50.33 million women aged 15 years and older in Nigeria who are at risk of developing cervical cancer. The latest figures show that 14943 women are diagnosed with cervical cancer every year and 10403 dies from the illness. It is estimated that approximately 3.5% of women in the general population in Nigeria are projected at a given time to have harbor cervical HPV-16/18 infection and 66.9% of invasive cervical cancers are due to HPVs 16 or 18. Yahya and Mande [6] concluded that awareness of cervical cancer and cervical screening did not reflect adequate knowledge about cervical cancer and screening methods. There is a need for healthcare providers to offer adequate health education about the disease and screening methods.

Materials and Methods

University College Hospital (UCH) of the University of Ibadan, Nigeria is a leading tertiary Centre for cancer care in Nigeria and as well as West Africa. The Department of Obstetrics and Gynecology, University College Hospital admits patients from all states within the country and beyond. This marks it a hospital of the best choice for a large percentage of cancer patients. The data used in this study are extracted from medical records of cervical cancer in this hospital. This study examined 310 patients with cervical cancer from reproductive-age women and other data with considered variables such as age and survival time which is the period used in admission and this is measured in a month. The months are categorized as follows: stage I = 12 months, stage II = 24 months, stage III = 36 months, stage IV = 48 months, and stage V = 60 months and above. The ages of the patients are also grouped into 4 as shown in table 1. The mortality by the disease is a binary event indicator (response) measured alive = 1 and dead = 0 with continuous explanatory variables that are captured nonparametrically.

The Bayesian nonparametric regression model was applied to estimate cervical cancer women with their age and the survival period with cervical cancer measured in several months. Age of the woman and this period are modeled by basic functions (f_i and f_j) approximated by a linear combination (K = m + l) of basis functions introduced by Eilers and Marx [7] where f_i is the seasonal effect and f_j the penalized spline effect of woman's age. The unknown regression coefficients are estimated through unknown functions approximated by a polynomial spline of degree l and with equally spaced knots $x_{\min} = \zeta_0 < \zeta_1 < \ldots < \zeta_{m-1} < \zeta_m = x_{\max}$. To ensure enough flexibility and give a roughness penalty based on the second-order random difference of adjacent B-spline coefficients for sufficient smoothness of the fitted curve, 20 equally spaced knots were chosen. Also, to overcome the problem of curse of dimensionality and that assumption of a strictly linear effect on the continuous variables (woman's age and length of stay) is not appropriate, non-parametric modeling of cervical cancer is then developed and compactly written as

$$mortality = f_i\left(time \text{ period}\right) + f_j\left(woman's \text{ age}\right) \qquad (1)$$

The estimated nonlinear functions f_i and f_j depend considerably on the hyperprior assumptions with the choice of hyperparameters $\alpha_{ij} = 0.001$ and $\beta_{ij} = 0.001$. Binomial logit presented in Holmes and Held [8] was used to capture binary response variable, the patient status regresses nonparametrically on woman's age and survival period of the disease in Bayesian approach.

Statistical analysis

The cases of cervical cancer are denoted as y_i is a binary variable (alive=1 and dead=0) associated with the study area conditional on the probability of the considered variables. The mortality from the disease defined by conditional Binomial model assumes the binary outcome (p_i), the probability of mortality caused by cervical cancer given the health Care Centre in the study area.

$$p_i : y \sim \text{Bin}(p_i, n_i) \text{ where } \text{p}_\text{i} = \frac{\exp(\beta_1 woman's \text{ age} + \beta_2 \text{time period})}{1+\exp(\beta_1 woman's \text{ age} + \beta_2 \text{time period})} \quad (2)$$

The likelihood function is given by $L\left(y_i|\text{p}_\text{i}, n_i\right) = \binom{n_\text{i}}{y_i} p_i^{y_i} \left(1-p_i\right)^{1-y_i}$ (3)

Here, the study area is UCH situated in the Ibadan area of Nigeria and n_i is the total population of women diagnosed with cervical cancer in this study. The probability will have either a prior distribution with or link to other parameters and covariate through a linear predictor replaced by structured additive predictors such as $\log it\left(p_i\right) = x_i\beta_i, i = 1,2$ where x are the i^{th} design matrix (not including an intercept) of women's age and period of cervical cancer. The parameters distribution for the continuous variables considered have variance with a gamma hyperprior fixed and common parameter at the status of an individual woman living with cervical cancer is included in p_i. The hyperprior distribution for the variance parameter on the positive real line follows a gamma distribution as against Jeffrey's prior distribution (e.g. beta distribution for binomial response likelihood). Jeffrey's prior distribution is not a complete non-informative prior because it has asymptotes close to 0 and 1 [9]. Before the unknown functions f_i and f_j depend on the woman's age and length of stay of the disease and prior beliefs about the smoothness f_i and f_j .The vector of function evaluation $f = \left(f_i \ , \ f_\text{j}\right)'$ as the matrix product of a design matrix $X = \left(\text{wowan's age} \ , \ \text{time period of living with the disease}\right)$.

The non-informative prior for a function f is defined suitable for design matrix(X) and its prior distribution for the vector of unknown parameters of the considered covariates takes general form as follows:

$$p\left(\beta/\tau^2\right) \propto \frac{1}{\left(\tau^2\right)^{rank(K)/2}} \exp\left(-\frac{1}{2\tau^2}\beta' K \beta\right) \quad (4)$$

Where K is the penalty matrix, for the β is partially improper Gaussian prior that follows $\beta/\tau^2 \sim \text{N}\left(0, \ \tau^2 K\right)$ and variance parameter τ^2 to be inverse smoothing parameter in a generalized nonparametric regression model which controls the tradeoff between flexibility and smoothness. The problems of analytically impossible of posterior simulation in the study consequentially lead to the posterior marginal distribution of cervical cancer is then given as

$$p\left(\beta, \tau^2/y_i\right) = L\left(y_i|\beta\right) \text{x } p\left(\beta/\tau^2\right) \text{ x } p\left(\tau^2\right) \quad (5)$$

The choice of before estimating the smoothing parameter τ^2 simultaneously with β, a proper hyperprior is assigned to it by choosing $\tau^2 \sim \text{IG(a=0.001, b=0.001)}$. The posterior marginal distribution is obtained using efficient sampling from the full conditionals for the regression parameters β of woman age and length of stay of the disease using MCMC by BayesX software.

Bayesian logistics regression model

The likelihood of the available data given in equation 3 is then multiplied by Jefffrey prior for the model parameter to give posterior distribution. Thus, the posterior distribution is estimated by random-walk Metropolis-Hastings sampling.

Results and Discussion

Variables/ Covariates	Alive	Dead	Total
Patients' age			
20 - 40	31 (10)	16 (5.16)	47 (15.16)
41 - 60	136 (43.87)	36 (11.61)	172 (55.48)
61 - 80	55 (17.74)	30 (9.68)	85 (27.42)
81 - 100	4 (1.29)	2 (0.65)	6 (1.94)
Total	226 (72.90)	84 (27.10)	310 (100)
Period (Months)			
Stage I	144 (46.45)	50 (16.13)	194 (62.58)
Stage II	49 (15.81)	13 (4.19)	62 (20)
Stage III	15 (4.84)	9 (2.90)	24 (7.74)
Stage IV	8 (2.58)	7 (2.26)	15 (4.84)
Stage V	10 (3.23)	5 (1.61)	15 (4.84)
Total	226 (72.90)	84 (27.10)	310 (100)

Table 1: *Distribution of characteristics of risk factors of mortality.*

This table shows the total number of women that died or survived cervical cancer at different age groups and the period of the disease measured in months. The period is categorized into five stages, I to V. Out of the 310 women examined, 226 women survived cervical cancer while 84 of them could not survive it. In this study patients between 41 and 60 years of age have the highest number, 172 (55.48 percent) women that suffered from cervical cancer from which 36 (11.61 percent) of them could not survive the disease. However, those between 81 and 100 years have the smallest number, 6 (1.94 percent) women from which 2 (0.65 percent) suffering the disease died of it. From the length of stay of the disease, stage I contains the highest number, 194 (62.58 percent) women that stayed in the disease measured in months, but 50 (16.13 percent) women died of the disease in stage I. Stage IV and V have the same total number, 15 (4.84) women that stayed in the disease but stage 4 has 7 (2.26 percent) and stage 5 has 5 (1.61 percent) women that died.

Covariates	Mean	Std. Dev.	Odds Ratio	NSE	95%Cred.Interval
Woman' s age	-0.0054687	0.0041589	0.9945462	0.000772	-.0198049 .0004026
Time period of Disease	-0.0072252	0.0040295	0.9928008	0.000634	-.0140168 .0062294

Table 2: *Bayesian estimation of the logistic regression model.*

Table 2 contains the posterior mean and standard deviation of women and period of living with the disease in the study. It is important to know that women's age and period of the disease is kept in this study because none of the estimated coefficients of the considered explanatory covariates contains zero in 95% credible interval. The means of women's age and period are insignificant to the mortality. Also, a unit change in a woman's age and period causes an insignificant decrease by 0.55% and 0.72% respectively in the probability of mortality from cervical cancer. This implies that age and period of living with the disease are not risk factors of the disease.

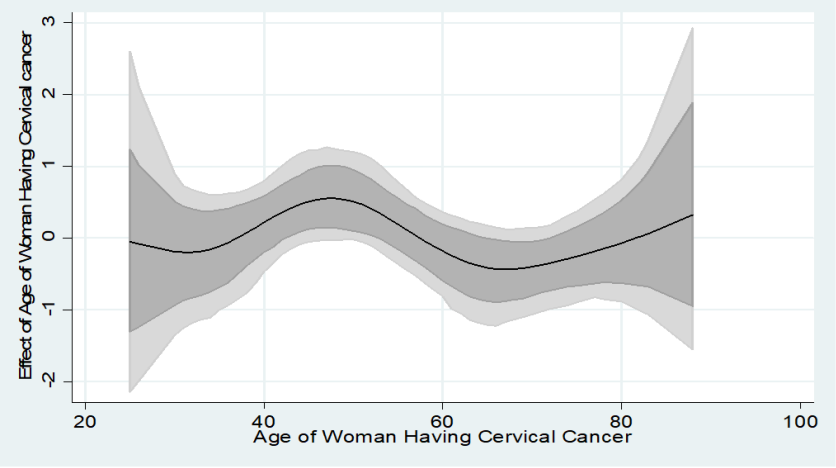


Figure 1: *Result of the nonparametric effect of woman's age with hyperparameters a = b = 0.001.*

This figure shows that women between 20 - 40 and 60 - 80 years of age are significantly less likely to die of cervical cancer. However, patients between 40 - 60 years of age are significantly likely to die of it. The woman's age has significant effect on the disease.

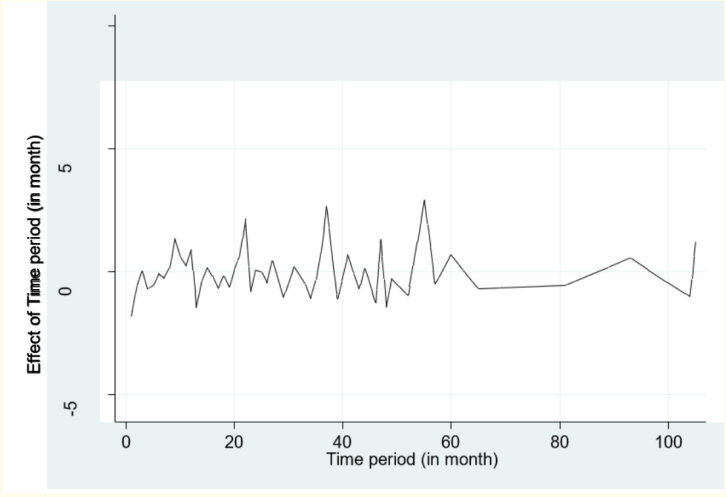


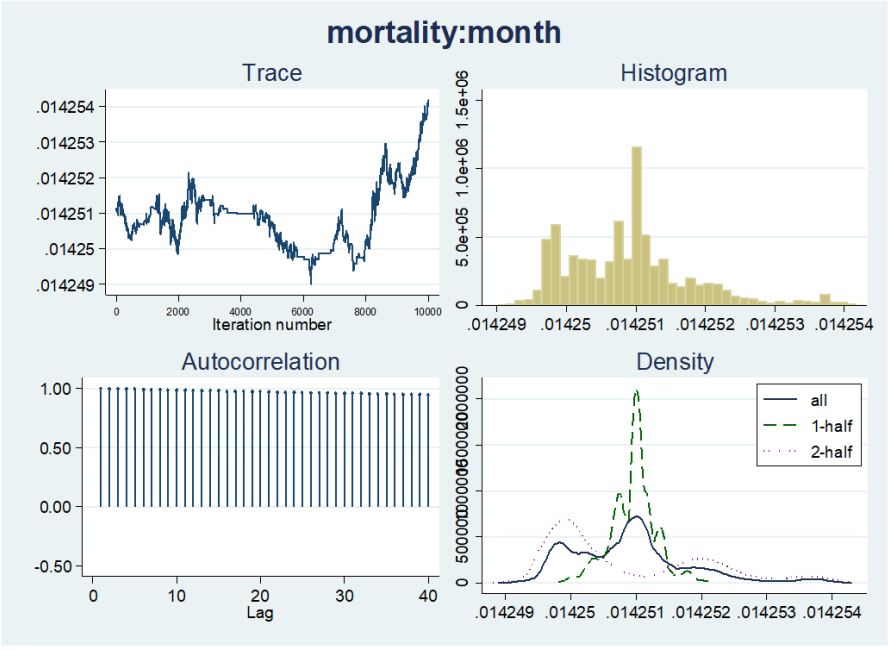
Figure 2: *Result of nonparametric effect of survival time period (months) with cervical cancer with hyperparameters a = b = 0.001.*

Figure 2 shows the seasonal variations in the period of living with the disease by adopting Fahrmeir and Lang's [10] study of seasonal effect. The time spanned 108 months. Finding reveals that the estimated seasonal effect significantly varied from stage I to V. This implies that the patients are significantly likely to die of the disease from stage I to IV. However, patients are not likely to die of it in stage V.

Conclusion

This study aimed at examining spline and seasonal effects of considered risk factors of cervical cancer. The application of Bayesian nonparametric techniques was used for analysis. The result of the age of the cancer women and their survival period was given. From table 1, it is revealed that more economically active women (41 - 60) as against economically inactive ones (e.g. 61 - 80) are involved in this study. Table 2 shows that a unit change in woman's age and period of living with the disease has insignificant likelihoods of reducing the mortality caused by cervical cancer by 0.55% and 0.72% respectively. Figure 1 and 2 reveal that age and period of living with the disease increase the probability of women surviving cervical cancer as the age and period increase. As the westernized lifestyle (with its associated reproductive patterns and dietary patterns) is continuously adapted, cancer incidents will also rise. Unless medical care and screening practices are improved, the risk factors responsible for the disease may not be identified and adequately analyzed. It is pertinent to agree with WHO and Yahya and Mande [6] that awareness of this disease should be encouraged and cervical cancer screening should be sponsored by NGOs and the Ministry of health. Early detection will assist in reducing the prevalence of cervical cancer in Nigeria [11-18].

Appendix



Figure

Bibliography

1. Kanavos P. "The rising burden of cancer in the developing world". *Annals of Oncology* 17 (2006): 15-23.

2. Stewart BW and Kleigues P. "World Cancer Report". IARC Press; Lyon France (2003).

3. World Health Organization (WHO). Human papillomavirus infection and cervical cancer (2010).

4. Marc A., *et al.* "Estimates of incidence and mortality of cervical cancer in 2018: A Worldwide Analysis (2020).

5. World Health Organization (WHO). Cervical cancer (2019).

6. Yahya A and Mande AT. "Awareness and knowledge of cervical cancer and its screening methods among women attending primary healthcare centers in Zaria, North-Western, Nigeria" (2019).

7. Eilers PHC and Marx BD. "Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder)". *Statistical Science* 11 (1996): 89-121.

8. Holmes C and Held L. "Bayesian auxiliary variable models for binary and multinomial regression". *Bayesian Analysis* 1 (2006): 145-168.

9. Lawson A. "Bayesian disease mapping: Hierarchical modeling in spatial Epidemiology". Andrew B. Lawson. P; Cm. -- (Chapman and Hall/CRC Interdisciplinary Statistics Series) includes Bibliographical References and Index (2008).

10. Fahrmeir L and Lang S. "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors". *Journal of the Royal Statistical Society C* 50 (2001): 201-220.

11. Bray F., *et al.* "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". *CA: A Cancer Journal for Clinicians* 68 (2018): 394-424.

12. Brezger A and Lang S. "Generalized Structured Additive Regression Based on Bayesian P-Splines". *Computational Statistics and Data Analysis* 50 (2006): 947-991.

13. Parkin DM., *et al.* "Cancer incidence in Five Continents Lyon: IARC Scientific Publications 7 (1997): 13-14.

14. Siegel RL., *et al.* "Cancer statistics, 2015". *CA: A Cancer Journal for Clinicians* 65 (2015): 5-29.

15. WHO. "Cancer" (2007a).

16. WHO. WHO calls for prevention of cancer through healthy workplaces (2007b).

17. https://emedicine.medscape.com/article/253513-overview

18. https://www.hpvcentre.net/statistics/reports/NGA_FS.pdf

Volume 3 Issue 3 March 2021