Hindawi

Research Article

Estimation of Risk Factors Affecting Screening Outcomes of Prostate Cancer Using the Bayesian Ordinal Logistic Model

J. L. Sirengo ⓘ**, D. A. Alilah** ⓘ**, D. A. Mbete** ⓘ**, and R. Keli** ⓘ

Department of Mathematics Masinde Muliro University of Science and Technology, P.O Box 190-50100, Kakamega, Kenya

Correspondence should be addressed to J. L. Sirengo; jsirengo@mmust.ac.ke

Prostate cancer occurs when cells in the prostate gland grow out of control. Almost all prostate cancers are adenocarcinomas. The survival rate for prostate cancer patients depends on the screening outcome, which can be either no prostate cancer, early detection, and late detection or advanced stage detection. The main objective of this study was to estimate the risk factors affecting the screening outcome of prostate cancer. With ordinal outcomes, a generalized Bayesian ordinal logistic model was considered in the analysis. The generalized Bayesian ordinal logistic model helped in estimation of coefficient parameters of the risk factors affecting each level of prostate cancer-screening outcomes. In the study, positive coefficients, that is, $\beta_k > 0$, indicated that the higher values on the explanatory variable increased the chances of the respondent being in a higher category of the dependent variable than the current one, while the negative coefficients, that is, $\beta_k < 0$, signified that the higher values on the explanatory variable increased the likelihood of being in the current or lower category of prostate cancer. For instance, from the analysis, positive or negative outcomes of prostate cancer showed that an increase in weight lowered the chances of an individual having the disease.

1. Introduction

Cancer is a disease caused by abnormal cell growth, where the disease has the potential to invade or spread to other parts of the body [1]. Once infected, normal cells are transformed into tumor cells, which progress from a precancerous lesion to a malignant tumor [2]. There are several types of cancer, namely, prostate, skin, cervical, breast, colorectal, lung, brain, and bone [3].

Policies on cancer treatment and management in Kenya guided the development and implementation of legal frameworks for the delivery of cancer services at the national and county levels [4]. The key among them was to provide health insurance cover to allow the vulnerable, orphans, and elderly to access comprehensive health care in line with the Sustainable Development Goal (SDG) [5–7]. This intended to inspire prostate cancer patients to seek early screening devoid of the cost for treatment.

A report by the National Cancer Screening Guidelines in 2018 showed that most Kenyans did not turn up for early

screening and diagnosis [8]. This was based on data from Kenyatta National Hospital which indicated that, in 2014–2016, approximately 64% of cancer patients were diagnosed at stage III or IV, who were difficult to attain effective treatment [9].

According to Bratt et al. [10], different people had expressed different reasons, such as age, health, race, obesity status, and family history, which deterred them from screening for prostate cancer. It was also believed that majority of people failed to turn up for screening due to lack of awareness training, socioeconomic-related challenges, worry about examination discomfort, fear of finding positive cancer status due to associated myths and stigma, and also inability to establish effective follow-up treatment [11, 12]. Motivated by these reasons, the study estimated the risk factors affecting the screening outcome of prostate cancer screening.

2. Literature Review

A study by Kramer et al. [13] critically evaluated the evidence for recommending the screening of asymptomatic men for

prostate cancer with a blood test to detect a prostate-specific antigen (PSA). They found out that although screening for prostate cancer had the potential in saving lives, over diagnosis, screening, and subsequent therapy could actually have net unfavourable deaths or quality of life or both. Without controlled clinical trials and prospective ascertainment of costs at the individual level, it was not possible to determine the net cost for mass screening programs of prostate cancer. There was uncertainty because earlier disease stages resulted in lower treatment costs, but the potential cost increased from over diagnosis. Understanding the impact of risk factors on screening outcomes will help individuals seek screening services at the earliest time possible. When the disease is detected in early stages, there are high chances in the reduction of treatment costs and saving of life. However, Wolf et al. [14] sought to establish predictors of interest in PSA screening and the impact of informed consent from prostate cancer patients. A method of a randomized trial was used which indicated that informed consent decreased patient interest in prostate-specific antigen screening. The result showed that perceived screening efficacy, perceived seriousness of an abnormal prostate-specific antigen, and willingness to accept treatment risks were univariate predictors of PSA screening interest ($P < 0.05$) among patients who were not informed. Estimation of risk factors affecting screening outcomes in this study helped identify major risk factors that need to be considered to partake in this noble task.

Bozkurt et al. [15] compared Bayesian networks and binary logistic regression methods in prediction of prostate cancer. It was clear that serum PSA level, age, digital rectal examination, and clinical symptoms were helpful for early detection of the tumor. The aim of their study was to examine and compare the methods used for early detection of prostate carcinoma which was identified by both logistic regression and Bayesian networks. Karlsson [16] also performed microsimulation modeling of prostate cancer screening in Sweden. A method based on the approximate Bayesian computations and Markov chain Monte Carlo was developed. On predicting the effect of dynamically changing current opportunistic PSA testing patterns for regular screening, the result revealed a reduction in the prostate cancer incidence and an increase in mortality when 8 yearly screening in men aged 55–69 was introduced for over 20 years. This study focused on the effect of risk factors considered by individuals before screening for prostate cancer. The effect was based on different outcomes (measurement levels) by using a generalized Bayesian ordinal logistic regression model.

Brant et al. [17] carried out research on screening for prostate cancer by using random-effects models. Male participants in a long-term longitudinal study were screened by using posterior probabilities, where each male was classified into one of the four diagnostic states for prostate cancer, i.e., normal, benign prostatic hyperplasia, local cancer, and metastatic cancer. Repeated measures of PSA were collected when there was no clinical evidence of prostate cancer diseases which were used in a classification process. The result showed that, overall, 86.8% or 88.3% were correctly classified by using the longitudinally collected PSA measurement, depending on whether or not a distinction was made between local and metastatic cancer. However, predictors like diet and family history were not included in the model because of lack of availability or insufficient information in data. Their study focused on the screening measurement of PSA and the effect of the age and the groups that an individual belonged to in predicting the next measurement level (i.e., PSA level) over a period of time. This study focused on the effect of unmeasured risk factors on the screening outcome (measurement level) of an individual by using a generalized Bayesian ordinal logistic regression model without random effects.

On the other hand, Liu and Koirala [18] indicated that the proportional odds (PO) assumption for ordinal regression analysis is often violated because it is strongly affected by the sample size and the number of covariate patterns. To address this issue, the partial proportional odds (PPO) model and the generalized ordinal logistic model were developed. However, these models are not typically used in research. One likely reason for this is the restriction of current statistical software packages: SPSS cannot perform the generalized ordinal logistic model analysis, and SAS requires data restructuring. Their article illustrated the use of generalized ordinal logistic regression models to predict mathematics proficiency levels using Stata and compared the results from fitting PO models and generalized ordinal logistic regression models.

Kulkarni et al. [19] also indicated that generalized linear models (GLMs) such as logistic regression are among the most widely used arms in the data analyst's repertoire and often used on sensitive datasets. A large number of prior works that investigated GLMs under differential privacy (DP) constraints provide only private point estimates of regression coefficients and are not able to quantify parameter uncertainty. In their work, with logistic and Poisson regression as running examples, they introduced a generic noise-aware DP Bayesian inference method for a GLM at hand, given a noisy sum of summary statistics. Quantifying uncertainty allowed them to determine which regression coefficients were statistically significantly different from zero. They provided a tight privacy analysis and experimentally demonstrated that the posteriors obtained from their model, while adhering to strong privacy guarantees, were close to nonprivate posteriors. This study incorporated the Bayesian paradigm to estimate significant risk factors affecting each screening outcome of prostate cancer.

A Schorgendorfer et al. [20] indicated that logistic regression is a popular tool for risk analysis in medical and population health science. With continuous response data, it is common to create a dichotomous outcome for logistic regression analysis by specifying a threshold for positivity. Fitting a linear regression to the nondichotomized response variable assuming a logistic sampling model for the data has been empirically shown to yield more efficient estimates of odds ratios than ordinary logistic regression of the dichotomized endpoint. They illustrate that risk inference is not robust to departures from the parametric logistic distribution. Moreover, the model assumption of proportional

odds is generally not satisfied when the condition of a logistic distribution for the data is violated, leading to biased inference from a parametric logistic analysis. The novel Bayesian semiparametric methodology for testing the goodness of fit of parametric logistic regression with continuous measurement data was developed. The testing procedures hold for any cutoff threshold, and our approach simultaneously provides the ability to perform semiparametric risk estimation. Bayesian factors were calculated using the Savage–Dickey ratio for testing the null hypothesis of logistic regression versus semiparametric generalization. When parametric logistic regression fails, an empirical Bayesian approach that is fully Bayesian and computationally efficient is used to test the methods that have been proposed and presented for semiparametric risk estimation and odds ratio estimation. Theoretical results establish the consistency of the empirical Bayes test. Results from simulated data show that the proposed approach provides accurate inference irrespective of whether parametric assumptions hold or not. Evaluation of risk factors for obesity shows that different inferences are derived from an analysis of a real dataset when deviations from a logistic distribution are permissible in a flexible semiparametric framework.

In this study, Bayesian methods were used to estimate the parameters, where the coefficients β_j varied across predictor variables. This was useful in estimating the effects of the risk factors on each outcome of prostate cancer screening. Thus, the generalized Bayesian ordinal logistic model was used in estimating the parameters for each outcome of prostate cancer.

3. Methodology

Since McCullagh's 1980 research paper on regression for ordinal data, cumulative-type models have become a standard tool in ordinal regression [21]. The ordered logit model is a regression model for an ordinal response variable. It is based on the cumulative probabilities of the response variable. In particular, the logit of each cumulative probability is assumed to be a linear function of the covariates with constant or varying coefficients across response categories [22].

In this study, the outcomes of screening for prostate cancer are ordinal. The predictor variables were measured as continuous and/or categorical data. The risk factors considered by those who turned up for screening prostate cancer in this study are age (continuous), traces of prostate cancer in family/history (binary with no = 0 or yes = 1), weight control (ordinal with underweight = 0, normal weight = 1, overweight = 2, and obese = 3), and hereditary breast and ovarian cancer syndrome (binary with no = 0 or yes = 1). The screening outcome for the ith individual was therefore given by

$$
Y_i = \begin{cases} 0, & \text{no} - \text{prostate (no} - \text{prostate)}, \\ 1, & \text{early stage (stage I and early stage II)}, \\ 2, & \text{late stage (late stage II and early stage III)}, \\ 3, & \text{advanced stage (late stage III and stage IV)}. \end{cases}
$$
(1)

The effects of the risk factors on each outcome are estimated by a generalized Bayesian ordinal logistic model. This was based on the coefficient parameters, β_{kj}, where $k = 1, 2, \cdots K$ is the number of risk factors and $j = 1, 2, \cdots J - 1$ is the number of prostate cancer outcomes. This implies that there is the $K \times (J - 1)$ matrix of the coefficient parameters, where $J - 1 = 3$ and X_{ik} is the kth risk factor considered by the ith; $i = 1, 2, \cdots n$ individual for prostate cancer screening. The risk factors for each individual form an $n \times K$ matrix as follows:

$$
X_{ik} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{pmatrix} \beta_j^T = \begin{pmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{K1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{K2} \\ \beta_{13} & \beta_{23} & \cdots & \beta_{K3} \end{pmatrix}.
$$
(2)

The generalized ordinal logistic regression model for every outcome of prostate cancer is given by

$$
C_{(ij)} = \frac{P(Y_i \leq j | X_{ik})}{P(Y_i > j | X_{ik})} = \xi(\alpha_j + X_{ik}\beta_{kj}),
$$
(3)

where ξ is the antilogarithm transformation function and α_j is the $J - 1$ threshold for the prostate cancer outcome. From the above representation, j as indicated on the parameters represents the outcomes of prostate cancer. Since there are four outcomes, the generalized Bayesian ordinal logistic model consists of a series of binary regression models, and therefore,

$$
\text{logit}(C_{ij}) = \log\left(\frac{P(Y_i \leq 0 | X_{ik})}{P(Y_i > 0 | X_{ik})}\right) = \alpha_j + X_{ik}\beta_{kj},
$$
(4)

whereby the three sets of coefficient parameters represented by $j = 1, 2, 3$ are defined as follows:

(i) $\beta_{k1}, k = 1, 2, \ldots, K$ are the coefficient parameters for risk factors of the model that represented the chances of an individual having prostate cancer or not having the disease, that is, either no prostate or an individual being in one of the following groups: early, late, or advanced stages (0 versus 1, 2, and 3).

(ii) $\beta_{k2}, k = 1, 2, \ldots, K$ are the coefficient parameters for the risk factors of individuals who were in the early stage or no prostate versus other stages of prostate cancer (i.e., 0 and 1 versus 2 and 3).

(iii) $\beta_{k3}, k = 1, 2, \ldots, K$ are the coefficient parameters for the risk factors of the individuals who were in the advanced stage versus other stages (that is, 0, 1, and 2 versus 3).

The latent propensity of the generalized Bayesian ordinal regression model in equation (4) follows a series of logistic distributions with the conditional mean matrix $X_{ik}\beta_{kj}$. Therefore, there are a series of latent continuous random variables $Z_{ij} \sim N(X_{ik}\beta_{kj}, I)$, and thus, the variable Y_i is observed such that $Y_i = j$ if $\alpha_{j-1} < Z_{ij} \leq \alpha_j$ with $\alpha_1 = -\infty$ and $\alpha_j = \infty$. To be more specific, a latent propensity variable Z_{ij}

is used as a basis for modeling the ordered ranking of prostate cancer screening outcomes for the ith individual. It is assumed to be a linear function of the covariates X_{ik}, and therefore, in this study, a series of latent variables for each screening outcome are given by

$$Z_{ij} = X_{ik}\beta_{kj} + \epsilon_{ij}; \epsilon_{ij} \sim N\left(0, \sigma^2\right), \qquad (5)$$

where the screening outcome of the ith individual is defined as

$$Y_i = \begin{cases} 0, & Z_{ij} \leq \alpha_1, \\ 1, & \alpha_1 < Z_{ij} \leq \alpha_2, \\ \vdots \\ j, & \alpha_j < Z_{ij} \leq \alpha_{j+1}, \\ \vdots \\ J, & Z_{ij} > \alpha_{j+1}. \end{cases} \qquad (6)$$

$$\pi\left(\beta_j, \alpha_j | Y_i\right) \propto \pi\left(\beta_j, \alpha_j\right) \prod_{i=1}^{n} \sum_{j=1}^{J} I\left(Y_i = j\right) \times \left[\Phi\left(\alpha_j + X_{ik}\beta_{kj}\right) - \Phi\left(\alpha_{j-1} + X_{ik}\beta_{kj}\right)\right]. \qquad (7)$$

For the model parameters (β_j, α_j), with $\alpha_j = (\alpha_2, \ldots, \alpha_{J-1})$, the following priors are considered:

$$\beta_j \big| \mu_{\beta_j}, \sigma_\beta^2 \sim N\left(\mu_{\beta_j}, \sigma_\beta^2\right),$$

$$\pi\left(\mu_{\beta_j}\right) \propto 1, \qquad (8)$$

$$\sigma_{\beta_j}^2 \sim IG(\mathbf{a}, \mathbf{b}).$$

Considering the Albert and Chib [23] joint posterior model for ordered multinomial responses, in this study, the joint distribution model for the $J - 1$ set of parameters is given by

To get the conditional distribution of the parameter estimates β_j, the likelihood function of the model parameters β_j and the latent variables $Z_{ij}; j = 1, 2, 3$ is given by the following joint posterior distribution:

$$\pi\left(\beta_j, \mu_{\beta_j}, \sigma_\beta^2, \alpha_j, Z_j | Y_i\right) \propto \prod_{i=1}^{n}\left[\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(Z_{ij} - X_{ik}\beta_{kj}\right)^2\right)\right) \times \sum_{j=1}^{J}\left\{I\left(Y_i = j\right)I\left(\alpha_{j-1} < Z_{ij} < \alpha_j\right)\right\}\right], \qquad (9)$$

where, in the equation, $\prod_{i=1}^{n}\left((1/\sqrt{2\pi})\exp\ \left(-(1/2)\left(Z_{ij} - X_{ik}\beta_{kj}\right)^2\right)\right)$ is the likelihood function of the model parameters. Therefore, the fully conditional distribution of β_j is given by

$$\beta_j \big| \mu_{\beta_j}, \sigma_\beta^2, \alpha_j, Z_j, Y_i \sim N\left(\widehat{\beta}_j, \widehat{\Sigma}_{\beta_j}\right). \qquad (10)$$

To prove equation (10), the chosen priors of the parameters in equation (8) and that of α_j (assumed diffuse), $j = 1, 2, 3$ are used whereby

$$\pi\left(\beta_j \big| \mu_{\beta_j}, \sigma_\beta^2, \alpha_j, Z_j, Y_i\right) \propto \exp\left[-\frac{1}{2}\left\{\left(Z_j - X_k\beta_j\right)^2 + \frac{1}{\sigma_\beta^2}\left(\beta_j - \mu_{\beta_j}\right)^2\right\}\right]. \qquad (11)$$

Let $L_1 = \left(Z_j - X_k\beta_j\right)^T\left(Z_j - X_k\beta_j\right) + (1/\sigma_\beta^2)\left(\beta_j - \mu_{\beta_j}\right)^T\left(\beta_j - \mu_{\beta_j}\right)$ in the equation. This implies that

$$L_1 = Z_j^T Z_j - 2\beta_j^T X_k^T Z_j + \beta_j^T X_k^T X_k \beta_j + \frac{1}{\sigma_\beta^2}\beta_j^T \beta_j - 2\beta_j^T \frac{\mu_{\beta_j}}{\sigma_\beta^2} + \frac{\mu_{\beta_j}^2}{\sigma_\beta^2}$$

$$= \beta_j^T \left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)\beta_j - 2\beta_j^T \left(X_k^T Z_j + \omega_j \right)$$

$$+ \left[\left(X_k^T Z_j + \omega_j \right)\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(X_k^T Z_j + \omega_j \right) \right]$$

$$- \left[\left(X_k^T Z_j + \omega_j \right)\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(X_k^T Z_j + \omega_j \right) \right] \tag{12}$$

$$+ Z_j^T Z_j + \eta_j$$

$$= \left(\beta_j - \widetilde{\beta}_j \right)^T \left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)\left(\beta_j - \widetilde{\beta}_j \right)$$

$$- \left(X_k^T Z_j + \omega_j \right)\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(X_k^T Z_j + \omega_j \right) + Z_j^T Z_j + \eta_j,$$

where

$$\widehat{\Sigma}_{\beta_j} = \left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1} \text{ and } \widehat{\beta}_j = \left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(X_k^T Z_j + \omega_j \right). \tag{13}$$

Thus,

$$\beta_j | Y_i, Z_j, \mu_{\beta_j}, \sigma_\beta^2, \alpha_j \sim N_m \left(\left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1}\left(X_k^T Z_j + \omega_j \right), \left(\frac{I}{\sigma_\beta^2} + X_k^T X_k \right)^{-1} \right). \tag{14}$$

On the other hand, with the defined prior for $\sigma_{\beta_j}^2$, the fully conditional distribution shown in the following equation is derived:

$$\sigma_{\beta_j}^2 \left| \beta_j, \mu_{\beta_j}, \alpha_j, Z_j \sim IG\left(\widehat{\mathbf{a}}_{\sigma_{\beta_j}^2}, \widehat{\mathbf{b}}_{\sigma_{\beta_j}^2} \right). \right. \tag{15}$$

Since σ_β^2 follows inverse gamma with parameters $\mathbf{a}$ and $\mathbf{b}$, then its prior is given by

$$\pi\left(\sigma_{\beta_j}^2 | \mathbf{a}, \mathbf{b} \right) = \frac{\mathbf{b}^{\mathbf{a}}}{\Gamma(\mathbf{a})}\left(\sigma_{\beta_j}^2 \right)^{-\mathbf{a}-1} \exp\left(-\frac{\mathbf{b}}{\sigma_{\beta_j}^2} \right), \tag{16}$$

and the fully conditional distribution of $\sigma_{\beta_j}^2$ is given by

$$
\begin{aligned}
\sigma_{\beta_j}^2 \big| \beta_j, \mu_{\beta_j}, \alpha_j, Z_j &\propto \pi\left(\beta_j \big| \mu_{\beta_j}, \sigma_{\beta_j}^2\right) \times \pi\left(\sigma_{\beta_j}^2\right) \\
&\propto \prod_{j=1}^{n_j} \left\{ \frac{1}{\sqrt{2\pi\sigma_{\beta_j}^2}} \exp\left[-\frac{1}{2\sigma_{\beta_j}^2}\left(\beta_j - \mu_{\beta_j}\right)^2 \right] \right\} \\
&\quad \times \frac{\mathbf{b^a}}{\Gamma(\mathbf{a})} \left(\sigma_{\beta_j}^2\right)^{-\mathbf{a}-1} \exp\left(-\frac{\mathbf{b}}{\sigma_{\beta_j}^2} \right) \\
&\propto \left(\frac{1}{\sigma_{\beta_j}^2}\right)^{(J/2)+\mathbf{a}+1} \exp\left[-\frac{1}{\sigma_{\beta_j}^2}\left\{ \mathbf{b} + \frac{\sum_{j=1}^{J}\left(\beta_j - \mu_{\beta_j}\right)^2}{2} \right\} \right],
\end{aligned}
\tag{17}
$$

which is a kernel of the inverse gamma distribution with,

$$
\mathbf{a}_{\sigma_{\beta_j}^2} = \frac{J}{2} + \mathbf{a} \quad \text{and} \quad \mathbf{b}_{\sigma_{\beta_j}^2} = \mathbf{b} + \frac{\sum_{j=1}^{J}\left(\beta_j - \mu_{\beta_j}\right)^2}{2},
\tag{18}
$$

and hence,

$$
\sigma_{\beta_j}^2 \big| \beta_j, \mu_{\beta_j}, \alpha_j, Z_j \sim IG\left[\frac{J}{2} + \mathbf{a}, \mathbf{b} + \frac{\sum_{j=1}^{J}\left(\beta_j - \mu_{\beta_j}\right)^2}{2} \right].
\tag{19}
$$

The conditional distribution of other parameters in the model, that is, the fully conditional posterior distribution of the latent variables, is independent with

$$
Z_j \big| \beta_j, \mu_{\beta_j}, \sigma_{\beta_j}^2, \alpha_j, Y_i = j \sim N\left(X_{ik}\beta_{kj}, I\right)
$$

truncated at the left (right) by $\alpha_{j-1}(\alpha_j)$. $\qquad$ (20)

The fully conditional density of α given other parameters is given up to the proportionality constant by the equation

$$
\alpha_j \big| \beta_j, \mu_{\beta_j}, \sigma_{\beta_j}^2, Z_j, Y_i, \{\alpha_r, r \neq j\} \propto \prod_{i=1}^{N} \left[I(Y_i = j)I\left(\alpha_{j-1} < Z_{ij} < \alpha_j\right) + I(Y_i = j+1)I\left(\alpha_j < Z_{ij} < \alpha_{j+1}\right) \right].
\tag{21}
$$

The equation is seen to be uniform on the interval $[\max\{\max\{Z_{ij}; Y_i = j\}, \alpha_{j-1}\}$ and $\min\{\min\{Z_{ij}; Y_i = j+1\}, \alpha_{j+1}\}]$.

4. Results

In the study, Z_{i1}, Z_{i2}, and Z_{i3} are the latent variables for the model of individuals with negative or positive outcomes of prostate cancer, individuals who were in early or no prostate versus late or advanced stages, and individuals who were in advanced stages or lower stages of prostate cancer, respectively.

The solution of the coefficient parameters was then obtained, as shown in equation (13). In the solution, $\widehat{\beta_j}$ represent the coefficient parameter estimates of the risk factors in the generalized Bayesian ordinal logistic model where I is the $K \times K$ identity matrix,

$$
X_k = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{pmatrix} Z_j = \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & z_{n3} \end{pmatrix} \text{and } \omega_j = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \vdots & \vdots & \vdots \\ \omega_{K1} & \omega_{K2} & \omega_{K3} \end{pmatrix},
\tag{22}
$$

where $\omega_j = (\mu_{\beta_j}/\sigma_{\beta}^2)$; $j = 1, 2, 3$ is the coefficient parameter mean for each outcome divided by common variance σ_{β}^2.

The study assumed the normal prior with the mean μ and standard deviation σ on all of the fixed effects. For positive

Table 1: Parameter estimates for each outcome of prostate cancer.

| Variables	0 vs. 1, 2, 3		0, 1 vs. 2, 3		0, 1, 2 vs. 3	
	Estimate	Post.prob	Estimate	Post.prob	Estimate	Post.prob
Threshold	4.7141	1.0000	−1.9944	1.0000	0.6053	0.9895
Age (X_1)	0.8319	0.9452	2.4232	1.0000	0.6741	0.9928
Traces of family						
History (X_2)	2.8214	0.9862	2.8307	1.0000	1.0832	0.9992
Weight control (X_3)	−1.5621	1.0000	0.9845	1.0000	−0.0255	0.5720
Breast and ovarian						
Syndrome (X_4)	0.7685	0.9950	2.2911	1.0000	0.7251	1.0000
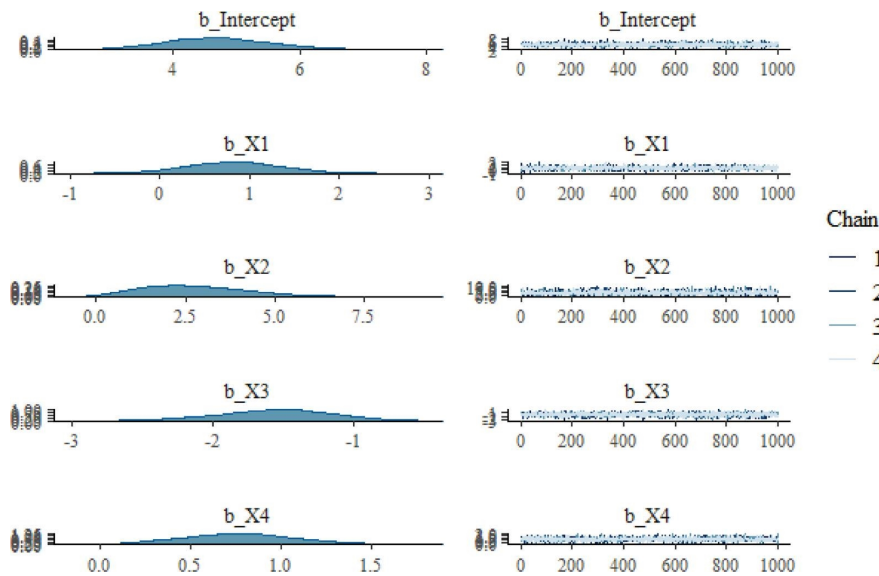
Figure 1: Graphical representation of coefficient parameters for negative or positive outcomes of prostate cancer.

coefficients of β_j, that is, $\beta_k > 0; k = 1, 2, \cdots, K$, the higher values on the explanatory variable increase the chance that the respondent will be in a higher category of the dependent variable than the current one, while the negative coefficients, that is, $\beta_k < 0$, signify that the higher values on the explanatory variable increase the likelihood of being in the current or lower category. Fitting the data in the model, the coefficient parameter estimates were obtained, and the results are presented in Table 1.

Posterior predictive probability values are used to test the hypothesis and estimation of significant risk factors affecting the outcome of prostate cancer. The one-sided hypotheses indicated that the posterior probability exceeded 95%; for two-sided hypotheses, the value tested against lied outside the 95% CI. In addition, the posterior probabilities of point hypotheses assumed equal prior probabilities. For instance, age (X_1) is the only variable that is not significant for the model that represented individuals who were in the category (0 vs. 1, 2, and 3). In the other categories, that is, 0 and 1 vs. 2 and 3, none of the variables are insignificant, and in the (0, 1, and 2 vs. 4) category, only weight control is not significant.

Using the coefficient estimates, age (X_1) had a coefficient of 0.83 which is positive, indicating that an increase in age increases the chances of an individual being in early, late, or advanced stages of prostate cancer. Similarly, traces of family history (X_2) and hereditary breast and ovarian cancer syndrome (X_4) had coefficients of 2.82 and 0.76, respectively, thus indicating that individuals who have family members with a history of prostate cancer or hereditary breast and ovarian cancer are likely to have the disease. However, the coefficient parameter of weight control was −1.56 which indicated that the higher values of individuals controlling weight increased their likelihood of not having prostate cancer. The results for other categories can be interpreted based on the signs of the coefficient parameters.

The graphical representation of the parameters in this category is shown in Figure 1. The figure shows how the intercept and coefficient parameters are distributed and their prediction from the first to the fourth chain. On the other hand, the marginal effects describe the average effect of change in the explanatory variable (i.e., risk factors) on the change in the probability of prostate cancer outcomes. They also provide a direct effect and easily interpreted answer to research questions. In this research, the marginal effect, which showed whether an individual was in advanced or lower stages of prostate cancer outcomes, i.e., 1, 2, 3, or 4, is illustrated in Figure 2.
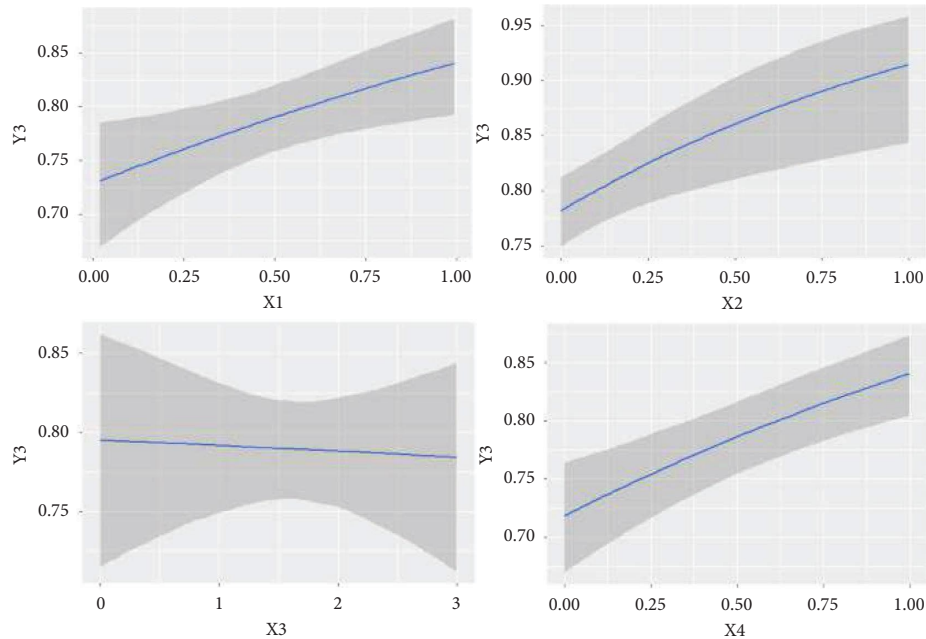
FIGURE 2: The marginal effects of the risk factors in the advanced stage of prostate cancer.

From Figure 2, as the values of age (X_1) increase, the values of prostate cancer outcomes also increase, indicating that aged individuals are most likely to be at an advanced stage of the disease. The same scenario is presented for individuals in X_2 and X_4. The other variable, that is, X_3, had negative coefficients/slope; thus, its graph had a decreasing function.

5. Discussion

The findings gave the specific effects of the risk factors on prostate cancer outcomes. Every outcome of prostate cancer patients was taken into account in the analysis, starting with estimation of the risk factors affecting individual outcomes, that is, no prostate or with prostate cancer. The analysis also focused on the risk factors that led individuals to have early stage detection or no prostate and those who were either in late or advanced stages of prostate cancer outcomes. The other category of prostate cancer outcomes considered in the analysis was individuals who were in advanced or lower stages of prostate cancer.

The effect of weight control indicated that individuals with more weight were likely to not have prostate cancer. Other risk factors like family history, hereditary breast and ovarian cancer syndrome, and age positively affected the outcome of individuals having prostate cancer. All risk factors had a positive impact on individuals who were in late or advanced stages versus lower stages of prostate cancer. It was also observed that age, traces of family history, and hereditary breast and ovarian cancer syndrome were the variables which positively affected individuals who were in the advanced stage.

Data Availability

The data used for analysis were generated using R software.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] M. A. J. Chaplain and B. D. Sleeman, "Modelling the growth of solid tumours and incorporating a method for their classification using nonlinear elasticity theory," *Journal of Mathematical Biology*, vol. 31, no. 5, pp. 431–473, 1993.

[2] A. Karagu, A. Ng'ang'a, J. Kibachio, and P. Gichangi, "Developing a national cancer control plan through effective partnerships: a case of Kenya national cancer control strategy 2017-2022," *Journal of Global Oncology*, vol. 4, no. 2, 2018.

[3] World Health Organization, *Early Detection*, Geneva, Switzerland, 2007.

[4] L. K. Makau-Barasa, S. Greene, N. A. Othieno-Abinya, S. B. Wheeler, A. Skinner, and A. V. Bennett, "A review of Kenya's cancer policies to improve access to cancer testing and treatment in the country," *Health Research Policy and Systems*, vol. 18, no. 1, pp. 2–10, 2020.

[5] E. W. Barasa, N. Mwaura, K. Rogo, and L. Andrawes, "Extending voluntary health insurance to the informal sector: experiencesand expectations of the informal sector in Kenya," *Wellcome open research*, vol. 2, 2017.

[6] D. N. Kimani, D. I. Muthaka, and D. K. Manda, *Healthcare Financing through Health Insurance in Kenya: The Shift to a National Social Health Insurance Fund*, Kenya Institute for Public Policy Research and Analysis, Nairobi, Kenya, 2004.

[7] G. Manyara and E. Jones, "Community-based tourism enterprises development in Kenya: an exploration of their potential as avenues of poverty reduction," *Journal of Sustainable Tourism*, vol. 15, no. 6, pp. 628–644, 2007.

[8] K. Nairobi, *Kenya National Cancer Screening Guidelines*, Ministry of Health, Nairobi, 2018.

[9] J. I. Rosser, B. Njoroge, and M. J. Huchko, "Knowledge about cervical cancer screening and perception of risk among women attending outpatient clinics in rural Kenya," *International Journal of Gynecology & Obstetrics*, vol. 128, no. 3, pp. 211–215, 2015.

[10] O. Bratt, L. Drevin, O. Akre, H. Garmo, and P. Stattin, "Family history and probability of prostate cancer, differentiated by risk category: a nationwide population-based study," *Journal of the National Cancer Institute*, vol. 108, no. 10, Article ID djw110, 2016.

[11] A. Muthoni and A. N. Miller, "An exploration of rural and urban kenyan women's knowledge and attitudes regarding breast cancer and breast cancer early detection measures," *Health Care for Women International*, vol. 31, no. 9, pp. 801–816, 2010.

[12] R. G. Mbugua, S. Oluchina, and S. Karanja, "Prostate cancer awareness and screening among men in a rural community in Kenya: a cross-sectional study," *African Journal of Urology*, vol. 27, no. 1, pp. 7–10, 2021.

[13] B. S. Kramer, M. L. Brown, P. C. Prorok, A. L. Potosky, and J. K. Gohagan, "Prostate cancer screening: what we know and what we need to know," *Annals of Internal Medicine*, vol. 119, no. 9, pp. 914–923, 1993.

[14] A. M. Wolf, J. T. Philbrick, and J. B. Schorling, "Predictors of interest in prostate-specific antigen screening and the impact of informed consent: what should we tell our patients?" *The American Journal of Medicine*, vol. 103, no. 4, pp. 308–314, 1997.

[15] S. Bozkurt, A. Uyar, and K. H. Gulkesen, "Comparison of bayesian network and binary logistic regression methods for prediction of prostate cancer," *International Conference on Biomedical Engineering and Informatics (BMEI)*, vol. 3, pp. 1689–1691, 2011.

[16] A. Karlsson, *Microsimulation Modelling of Prostate Cancer Screening in Sweden (Doctoral Dissertation, Karolinska Institutet (Sweden))*, Department of Medical Epidemiology and Biostatistics Karolinska Institutet, Stockholm, Sweden location, Solna, Sweden, 2019.

[17] L. J. Brant, S. L. Sheng, C. H. Morrell, G. N. Verbeke, E. Lesaffre, and H. B. Carter, "Screening for prostate cancer by using random-effects models," *Journal of the Royal Statistical Society: Series A*, vol. 166, no. 1, pp. 51–62, 2003.

[18] X. Liu and H. Koirala, "Ordinal regression analysis: using generalized ordinal logistic regression models to estimate educational data," *Journal of Modern Applied Statistical Methods*, vol. 11, no. 1, pp. 242–254, 2012.

[19] T. Kulkarni, J. Jälkö, A. Koskela, S. Kaski, and A. Honkela, "Differentially private bayesian inference for generalized linear models," in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, venue, city, July 2021.

[20] A. Schorgendorfer, A. J. Branscum, and T. E. Hanson, "A Bayesian goodness of fit test and semi parametric generalization of logistic regression with measurement data," *Biometrics*, vol. 69, no. 2, pp. 508–519, 2013.

[21] P. McCullagh, "Regression models for ordinal data," *Journal of the Royal Statistical Society: Series B*, vol. 42, no. 2, pp. 109–127, 1980.

[22] L. Grilli and C. Rampichini, *Ordered Logit Model*, 2014.

[23] J. H. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 669–679, 1993.