**THE UNIVERSITY OF ZAMBIA**

**SCHOOL OF PUBLIC HEALTH**

**ASSIGNMENT 1**

**`**

****

**ADVANCED BIOSTATISTICS**

**STUDENT’S ID:** PH2000067

**PROGRAMME:** MSc EPIDEMIOLOGY

**QUESTION: “**Write a report based on the analysis of cancer (prostate9.dta) of the prostate data. Your report should be similar to a published paper in a peer reviewed journal.”

**WORD COUNT:** 1,645 words

**PREDICTING THE SPREAD OF PROSTATE CANCER TO THE SURROUNDING LYMP NODES USING NON SURGICAL METHODS.**

**ABSTRACT**

*This study had an aim of determining whether a combination of five variables could be used to ascertain whether or not the prostate cancer had spread to the surrounding lymph nodes in patients presenting with cancer.*

*This cross sectional study collected data from 53 male patients presenting with prostate cancer and had gone through a laparotomy. The median age was 60 years and the ages ranged from about 56 years to 65 years.*

*the best predictors of whether or not prostate cancer has spread to the surrounding lymph nodes (nodal involvement), were level of serum acid phosphate, X-ray examination results, and the size of the tumour determined by rectal examination (P vale 0.033, 0.010, 0.018, respectively). All the predictors showed an increasing relation of the odds of nodal involvement for every unit increase in acid levels, positive X-ray examination results, and increase in the size of the tumour respectively.*

**Keywords:** Nodal involvement, model

**BACKGROUND**

Prostate cancer, a cancer of a man’s prostate gland, is a very common malign type of cancer in males that has a high rate of death (Buhmeida, Pyrhönen, Laato, & Collan, 2006). The treatment pathway taken for prostate cancer patients is highly dependent on if or not the prostate cancer has spread to the surrounding lymph nodes. Prognosis of prostate cancer can be done using x-ray examinations, rectal examination to observe the size of the tumour, biopsy to determine the pathological grade of the cancerous tumour and also by measuring the level of the serum acid phosphate of the patient (Byron, et al., 1976).

The most radical prostate cancer treatment is limited to the patients with organ-confined disease (Buhmeida, Pyrhönen, Laato, & Collan, 2006). Prostate cancer is predominant in males of advanced age, the risk of having prostate cancer increases as the man grows older than 40 years of age. The age of the male is important factor that should prompt ever male to undergo prostate cancer screening, especially for men above the age of 50 years (Freedland S, 2004).

A laparotomy, a surgical incision into the abdominal cavity, is a risky approach to investigate whether or not prostate cancer cells have spread to the surrounding lymph node, due to the risks associated with surgery (Lilja. et al, 2008). This study assessed whether the combination five predictors of prostate cancer severity would help to ascertain whether or not the cancer had spread to the surrounding lymph nodes (Presence or absence of prostate cancer nodal involvement), as the main aim and objective.

METHODOLOGY

**Participants**

This was a cross-sectional study. The study was conducted with a sample size of 53 male patients presenting with prostate cancer, who had also undergone a laparotomy. With the minimum age being 45 years and the maximum age 68 years old at the time of diagnosis. The subject in this study were first screened and ascertained whether or not the cancer has spread to the surrounding lymph nodes by means of a laparotomy. Patients that had gone through a laparotomy and were presenting with prostate cancer were considered for this study.

**Procedure**

During this study, five different independent variables that would likely help to ascertain the extent of nodal involvement of prostate cancer were measured on the participants to model their ability to measure whether or not prostate cancer had spread to the surrounding lymph nodes.

The outcome of interest was nodal involvement, a binary outcome which measured as to whether or not the cancer had spread to the surrounding lymph nodes. In determining the outcome, the following independent variables were measured; the age in years of the patient at diagnosis, the level of serum acid phosphates (abbreviated as acidlevo). The age of the patient and the level of serum acid phosphate were measured on a continuous scale. The other independent variables measured were data from the Xray examination recorded as either positive or negative, the size of the tumour (tumasize) was determined by a rectal examination and measured as either small or large. Then the summary of the pathological grade of the tumour was determined from a biopsy and measured as either less serious or more serious.

**Statistical analysis**

This study had five independent variables, three variables were categorical and two (age of patient and acid level) were continuous variables. The continuous variables were not uniformly distributed. qq-plots were done to test for normally assumption (note, the plots are not shown). Since the two continuous variables were not normally distributed, the descriptive statistics for this the continuous variables are reported with the median and inter-quartile range. This was done to minimise the possibility of inflating the type-1 error, residual confounding and bias (Burbos, et al., 2010).

The age of the patients and the level of serum acid phosphate, were log transformed because they showed skewness. This was done to make the variables approximately be normally distributed.

In the modelling process, univariate analyses were done to observe the relationship of the independent variables to the binary outcome variable individually to determine the level of correlation between the independent variable(s) to the dependent. Due to the nature of the outcome variable, a logistic regression was a more suitable path in the modelling process (Guzman, 2000). The generalised linear model (glm) of the family (binomial) and logit as link function approach was used. The regression results were transformed in the form of odds ratio (exponential form) for easy interpretation instead of log odds. After the univariate analysis, a multivariate model was run which had all the independent variables against the outcome to see if put together they all explain the outcome better.

The investigator led approach, as a more preferred direction, was used to select the best model which had good predictors of prostate cancer nodal involvement. Step by step, the variables which had the highest p values (taking account of prior knowledge about the variables) were removed. Following the principle of parsimony (Which stresses that when faced with two competing models, the model with the simplest and better/informative predictors is preferred), the parsimonious model was arrived at by using the Bayesian Information Criterion (BIC). Further, to estimate how well the model classified those with the presence or absence of nodal involvement correctly to the group they belonged to, a test of sensitivity and specificity was done. Then later the receiver operating characteristic (ROC) curve was done to graphically illustrate the diagnostic ability of the nodal involvement predictors. All these analyses were done with STATA statistical software, version 15.0 (StataCorp).

RESULTS

**Descriptives.**

In this study, the 53 males who had gone through a laparotomy and presenting with prostate cancer were screened based on the possible predictors of nodal involvement. The distribution of patients’ ages was such that, the median age was 60 years and the ages ranged from 56 years to 65 years. The measure of the levels of serum acid phosphate was distributed from about 0.50 – 0.78 King Armstrong units, with the median level of serum acid phosphate being 0.65 King Armstrong units.

Risk factors

At Simple logistic regression analysis, the age of the patient at diagnosis suggested to have a decreasing effect 0.95 on the odds of nodal involvement per year increase in the age with a p value (P=0.300) which was highly not statistically significant. And when log transformed, the p value P=0.273 still remained not significant the significant level of P=0.05 which was used in this study. The level of serum acid phosphate suggested to have an increasing effect of 8.77 on the odds of having nodal involvement (P=0.088) but would not allow us to rule out chance finding. A positive Xray examination significantly predicted an increase of 8.86 in the odds of nodal involvement (P=0.002). The odds of nodal involvement significantly increased as the size of the tumour (tumasize) increased (P=0.009). Then finally, the pathological grade of the tumour (tumagrad) fairly suggested an increase in the odds of nodal involvement the more serious it gets (P=0.047).

After log transforming the two continuous variables age and acidlevo, a multivariate analysis with all the risk factors was done and then using the BIC, a parsimonious model was arrived at. The results of the best fit, parsimonious multiple logistic regression model are show in the table 1 below.

**Table 1.** The adjusted predictors of nodal involvement (odds ration). Best fit model

|  |  |  |
| --- | --- | --- |
| **Nodal Involvement Predictors** | **Odds ratio**  **(95% Confidence interval)** | **P-value** |
| lacidlevo | 11.09 (1.21-101.84) | 0.033 |
| Xray | 7.86 (1.63-37.99) | 0.010 |
| tumasize | 5.92 (1.35-25.88) | 0.018 |

lacidlevo = log transformed, level of serum acid phosphate (measured in King Armstrong units). Xray = Xray examination (measured as positive or negative). tumasize = size of the tumour (measured as either small or large).

From the results of the analysis, the best predictors of whether or not prostate cancer has spread to the surrounding lymph nodes (nodal involvement), were level of serum acid phosphate, Xray examination results, and the size of the tumour determined by rectal examination (P vale 0.033, 0.010, 0.018, respectively). All the predictors showed an increasing relation of the odds of nodal involvement for every unit increase in acid levels, positive Xray examination results, and increase in the size of the tumour respectively.

**DISCUSSION**

The main aim of this study was to ascertain if a combination of five variables could be used to determine whether or not the prostate cancer had spread to the lymph nodes. Only three variables explain this association better, as noted in the explanation of the univariate analysis. At 95% Confidence level, the relationship of the main predictors to the dependent variable was such that; One unit increase in the serum acid phosphate of the patient increase the odds of having a presence of nodal involvement by a factor of 11.09 in King Armstrong units (P=0.033, CI; 1.21-101.84). Having a positive X-ray examination result also increased the odds of having a presence of prostate cancer in the surrounding lymph nodes by 7.86 times (P=0.010), this increase in the actually population can be as low as 1.63 and as high as 37.99 for every positive X-ray result at 95% confidence level. The odds of having a presence of nodal involvement are 5.92 times greater in the patients with large tumour sizes compared to those with small tumour sizes (P=0.018, CI; 1.35-25.88) at 95% confidence level. The sensitivity and specificity tests which measured how effective this model classifies the patients has having a presence of nodal involvement or absence of nodal involvement correctly, showed a sensitivity of 60% and specificity of up to 88%. This model has a higher specificity than sensitivity. The level at which in correctly classifies presence or absence of nodal involvement is 77% which is good enough.

Table 2. Below shows the analysis of well the model predicts positive results as truly positive (sensitivity) and negative results as truly negative (specificity).

--------------------------------------------------

Sensitivity Pr( +| D) 60.00%

Specificity Pr( -|~D) 87.88%

Positive predictive value Pr( D| +) 75.00%

Negative predictive value Pr(~D| -) 78.38%

--------------------------------------------------

False + rate for true ~D Pr( +|~D) 12.12%

False - rate for true D Pr( -| D) 40.00%

False + rate for classified + Pr(~D| +) 25.00%

False - rate for classified - Pr( D| -) 21.62%

--------------------------------------------------

Correctly classified 77.36%

The relation between the sensitivity and specificity of this model in clearly seen in the graphical representation below. The area under the curve shows that the model with the selected predictors is good enough



**Figure 1** ROC curve for the prediction the presence or absence of nodal involvement.

CONCLUSION / RECOMMENDATIONS

The sample size in this study was 53, which was small and there susceptibility of selection biases. The size of the sample is probably one of the reasons why out of the five predictor variables age and tumour grade were not statistically significant, and also led to the two continuous variables to exhibit skewness. The participants in the study were not well randomly selected to be part of the study, due to the inclusion criteria of that only the patients presenting with prostate cancer and had gone through a laparotomy were selected.

Generally from the statistical analysis, the level of serum acid phosphate, X-ray examination results, and the size of the tumour predict the level of prostate cancer spread to the surrounding lymph nodes (nodal involvement). The following process is advised in managing patients that presenting with prostate cancer.

Patients presenting with Prostate cancer

Perform more tests for verification.

Assume low nodal involvement.

Low acid levels, negative Xray, small tumour size

Considering to start the best treatment regime

Assume high nodal involvement.

High acid levels, positive Xray, large tumour size

Examination of risk factors, other clinical examination. (Serum phosphate acid levels, Xray examination, tumour size)

**Chart 1.** Proposed problem solving procedure for managing prostate cancer.

The design of this study was a limitation if the findings would be used to determine whether these predictors of nodal involvement would be as good as or more effective than laparotomy or another specific method. An experimental design, or Randomized clinical trial would be a more informative approach.

REFERENCES