# Determination of Predictor Variables Contributing to the Survival of Prostate Cancer Patients in Komfo Anokye Teaching Hospital, Ghana

**Asiedu-Addo Samuel Kwes**
**Nwi-Mozu Isaac**
**Assuah CharlesKojo**
**AsumaduOppong Richard**
University of Education
Department of Mathematics Education
Winneba, Ghana

## Abstract

*In this study, we identify the variables associated with survival and hazard of prostate cancer patients. We also interpret the time until death of prostate cancer patients. A total of 199 prostate cancer cases diagnosed at the KomfoAnokye Teaching Hospital within a 9 year period were retrieved, representing the time at which the hospital commence her prostate cancer registry (2005 to 2013). The data were analyzed using descriptive statistics and Kaplan Meier Non-Parametric methods. The result indicated that, 13.6% of the patients diagnosed at the age of 68 died with the prostate cancer when the cancer has developed into stage three. Only the stage of cancer at which a patient was diagnosed contributed to the survival or hazard of a patient, p-value < 0.05. Age of patients and type of treatments given, do not statistically contribute to the survival of prostate cancer patient's p-values > 0.05.*

**Keywords**: survival analysis, Kaplan-Meier, censoring, events, prostate cancer

## *1. Introduction*

The human body is made up of trillions of living cells. Normal body cells grow, divide into new cells, and die in an orderly way. During the early years of a person's life, normal cells divide faster to allow the person to grow. After the person becomes an adult, most cells divide only to replace worn-out or dying cells or to repair injuries. Cancer begins when cells in a part of the body start to grow out of control. There are many kinds of cancer, but they all start because of out-of-control growth of abnormal cells. Cancer cell growth is different from normal cell growth. Instead of dying, cancer cells continue to grow and form new, abnormal cells. Cancer cells can also invade (grow into) other tissues, something that normal cells cannot do. Growing out of control and invading other tissues are what makes a cell a cancer cell. The prostate is an exocrine gland found only in males. It is located in front of the rectum and below the urinary bladder. The size of the prostate varies with age. In younger men, it is about the size of a pea, but it can be much larger in older men. The prostate's job is to manufacture some of the fluid that protects and nourishes sperm cells in semen, making the semen more liquid (www.ncbi.nlm.nih.gov/pubmed/21675880).

Prostate cancer starts with tiny alterations in the shape and size of the prostate gland cells- Prostatic intraepithelial neoplasia (PIN). Prostate cancer is the second most common cancer in males after hepatocellular carcinoma in Africa (Parkin, 2003), but in the United States and Northwestern Europe, it is the most common cancer in males and the second most common cause of cancer deaths (Jemal*et al.,* 2002). A similar study by Chu et al. (2011), indicated that prostate cancer is recorded as the second most common cancer in men. In contrast, its incidence is much lower in Asia, South America and Africa(Arthur et.al. 2005). In African countries where registers exist such as Nigeria, Uganda, South Africa and Zimbabwe, it has been observed that the incidence of prostate cancer is increasing between the ages of 40 to 70 years (Wabinga, 2003).Possible factors such as old age, genetics, diet and obesity are known to cause prostate cancer. The disease generally leads to stiffness in the pelvis, blood in urine or semen, painful during ejaculation, frequent urinating, and swelling of lower extremities. Prostate cancer has no national boundaries and vary greatly across the world.

Asian men have been known to have the least exposure to the peril of prostate cancer whether they live in Asia or outside Asia. African American men have among the highest reported prostate cancer rates in the world (Chu et. a1. 2011). The reported risk of the disease by countries is due to proper cancer registry by those countries. For instance, in African countries which have registered prostate cancer such as South Africa, Uganda, Nigeria and Zimbabwe, prostate cancer has been revealed to be increasing between the ages of 40 to 70 years (Wabinga, 2003).

Many Ghanaian men are unaware of the disease and its menace. As a result of this, many prostate cancer patients result into local herbs for treatment when early symptoms suggest prostate cancer. Preliminary hospital records indicate that the incidence of prostate cancer has been observed to be increasing, although there are no statistics to prove it.Whereas in the advanced countries, screening for Prostate Specific Antigen(PSA) has led to early detection and management of the disease, screening has been very low in Ghana, thus leading to low detection rate, poor management and increased mortality.(Arthur et al., 2005).A study conducted by Arthur et al., (2005), randomly sampled 196 men ranging from 40 to 95 years and screened them with PSA measurement ranging from 0 to 110ng/ml revealed that serum PSA increases with age. It is reported that as at 2004, Ghana has no prostate cancer registry making it difficult to be included in the world prostate cancer rate report. However report by WHO in 2011, revealed that the death rate of prostate cancer in Ghana was increasing at the rate of 9.5% per year and 45th in the world. If awareness is not raised among Ghanaian men, death rate will continue to increase yearly.

There hasnot been an attempt to statistically present the variables that influence the survival rate of a prostate cancer patient. Thus, knowing significant variables would be informative for governments, hospitals, Ministry of Health, non-governmental organizations and the country at large in the setting of priorities, formulation of policies and the allocation of funds in curbing the menace of prostate cancer. Insurance companies could use the findings in determining premiums and pricing products. Hence, this study investigated statistical variables that contribute to the survival rate of prostate cancer patients.

## *2. Materials and Methods*

The data consist of all prostate cancer patients diagnosed at Oncology Department at KomfoAnokye Teaching Hospital (KATH) in Kumasi, Ghana within a 9 year period (2005 to 2013). KATH is located at Bantama, in the Ashanti Region capital, Kumasi. KATH is the only tertiary health institution in the Ashanti Region, the second largest hospital in Ghana and serves as the main referral hospital for the Ashanti, Northern, Upper East and Upper West Regions in the country. Ashanti region has a total population of 4,725046 and is the region with the highest population in Ghana according to Ghana Statistical Service (2012). The population of male in Ashanti region is 50.2% as against 48.8% female and the region with the highest male population in Ghana. Hence, KATH is the most perfect hospital or medical facility to be used as a case study for this research. KATH started its prostate cancer registry in the year 2005. Thus, efficiency of this study would be achieved if all patients treated at the commencement of prostate cancer registry is taken at the time of the study. Patient's records were evaluated for the endpoints of either survival or death. In 11 cases out of 210, no information of patient's records were retrievable and hence excluded from the analysis. This approach is similar to the study conducted by Siddique et al (2001) in Pakistani breast cancer patients. Thus, in all 199 prostate cancer patients were used in this study.

## *3. Mathematical Formulation*

In order to have holistic findings on survival of a prostate cancer patient, predictor variables available at the time of the study were considered. These variables are age, treatment and stage of cancer. All these predictors were taken at the first time a patient was treated. The data were analyzed employing survival analysis method with R software version 2.15.3 (2013).Survival analysis is a set of statistical methods for analyzing data where the dependent variable is the time until occurrence of an event of interest. In this study, the dependent variable is continuous and is the time' until a patient experience the event death. Patients who experience the event are coded one on the bivariate variable relapse and those who have not experienced the event death at the time of study (censored) are coded zero on relapse. According to Allison (1995) survival analysis is best used when there are censored data. Observations are called censored when the information about their survival time is incomplete. In this study, the following category of patients were censored: patients who were alive and disease-free; patients who died from other causes other than prostate cancer; patients who were still alive with the disease as at their last follow up date; and, patients who were lost to follow-up. The survival and hazard functions are key concepts in survival analysis for describing the distribution of event times.

The survival function gives, for every time, the probability of surviving (or not experiencing the event) up to that time. The survival function denoted by $S$ is defined by

$$S(t) = \Pr(T > t)$$

where, $t$ is some time, $T$ is a random variable denoting the time of death, and "Pr" stands for probability. That is, the survival function is the probability that the time of death is later than some specified time $t$.The hazard function $\lambda(t)$ gives the potential that the event will occur, per time unit, given that an individual has survived up to the specified time. For continuous random variables, the hazard function is given by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\Pr(t \le T < t + \Delta t)}{\Pr(T \ge t)} = \frac{S'(t)}{S(t)}$$

A number of models are available to analyze the relationship of a set of predictor variables with the survival time. Methods include parametric, nonparametric and semiparametric approaches. A nonparametric estimator of the survival function,

$$S(t) = \Pr(T > t)$$

given byMachin et al. (2006) define Kaplan-Meier (KM) survivor function at time t as a product of the form

$$\hat{S}(t) = \prod_{j:a_j < t} \left(1 - \frac{d_j}{r_j}\right), a_k < t \le a_{k+1}.$$

where

$d_j$ is the number of deaths at $a_j$

$r_j$ is the number of risk at $a_j$

In this study, KM method was used to estimate and graph survival probabilities as a function of time. It was also used to obtain univariate descriptive statistics for survival data, including the median survival time, and compared the survival experience for two or more groups of subjects. Both the Log Rank and the Wilcoxon test given by Korosteleva (2008) for survival difference were employed in this study. The Log Rank test statistic is given by

$$Z = \frac{\sum_{i=1}^{k} d_{1i} - \left(\frac{n_{1i} d_i}{n_i}\right)}{\sqrt{\text{var}\left(\sum_{i=1}^{k}\left(d_{1i} - \frac{n_{1i} d_i}{n_i}\right)\right)}}$$

and the Wilcoxon statistics is given by

$$Z = \frac{\sum_{i=0}^{k} n_j (d_{1i} - E(d_{1i}))}{\sqrt{\text{var}\left(\sum_{i=0}^{k} n_j (d_{1i} - E(d_{1i}))\right)}}$$

where

$n_i$ is the population size

$n_{1i}$ is the size of the group of interest

$n_j$ is the total number at risk at each time point

$d_i$ is the sample size

## *4. Analysis of Result*

### 4.1Results of Descriptive Statistics

Table 1 provides the descriptive statistics for various events, prostate cancer stage and treatment. The stage of the cancer at the first time of diagnose is classified into stages: stage 1; stage 2; stage 3 and stage 4 according to the severity of the cancer. A patient is either treated with chemotherapy or a combination of chemotherapy and radiotherapy.

**Table 1: Variable Statistics**

| | Time | Event | Age | Treatment | Stage |
|---|---|---|---|---|---|
| Min | 10 | 0.0000 | 34 | 1.00 | 1.000 |
| Median | 1,684 | 0.0000 | 69 | 1.00 | 3.00 |
| Mean | 1,569 | 0.1357 | 68 | 1.352 | 3.261 |
| Max | 3,250 | 1.0000 | 95 | 2.000 | 4.000 |

The median and mean time period for a patient to experience the event death are 1,684 and 1,569 days respectively. The mean event is 0.13577. That is approximately 14% of the patients have experienced the event (death) at the end of the ninth year at age 68 years with a little above stage 3 cancer. The minimum and maximum ages of patients are 34 years and 95 years respectively.

From Table 2, thirteen (13) patients had a stage one cancer but none (100%) experienced the event at the end of the period of study. Out of 13 patients who had stage two cancer, only 1 (7.7%) experienced the event at the end of the study. Again, out of 82 patients who were found at stage three cancer11 (13.4%) experienced the event. Finally, stage 4 cancer was highest among patients who experienced the event at the end of the study. Out of 91 patients with stage four cancer 15 (16.5%) patients experienced the event.

**Table 2: Descriptive Statistics for Patient's Cancer Stage**

| | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
|---|---|---|---|---|
| Records( No. of patient) | 13 | 13 | 82 | 91 |
| No. of Event (death) | 0 | 1 | 11 | 15 |

Table 3 displays the descriptive statistics of patient's exposure to treatments. It is expected that if one treatment is better than the other, then patients using that treatment should have longer times to relapse and more censored observations (patient experience no death by the time the study ends) than those using the other.

**Table 3: Descriptive Statistics for Patient's Exposure to Treatments**

| | Total No. | No. of Events (death) | No. of Censored | Percentage of Censored |
|---|---|---|---|---|
| Treatment 1 | 129 | 16 | 113 | 87.65% |
| Treatment 2 | 70 | 11 | 59 | 84.3% |
| Overall | 199 | 27 | 172 | 86.4% |

Out of 199 patients under study, 129 patients were given treatment 1 out of which 113 (87.6%) were censored. Seventy (70) patients were given treatment 2 (that is second treatment) out of which 59 (84.3%) were censored. Overall 199 patients were included in the study out of which 172 (86.4%) patients were censored at the end of the study. It can be seen that the difference in percentage of censored for both groups is not wide.

### 4.2 Result of Kaplan-Meier Function

The Kaplan-Meier survival function is presented in graphical representation where the survival function for no predictor is produced. KM is easy to interpret when predictor is categorized. Thus, the survival function for each categorical covariate (treatments and cancer stages) is produced. The predictor age is a continuous function and will not be produced. However, its significant to survival of patients was examined using its corresponding p-value. Figure 1 provides the survival plot of a patient assuming that a patient survival does not depend on any of the predictors under consideration. It is observed that the survival curve decreases gradually until day 850 when there is a slight drop in survival and become stable till day 1380 when it dropped slightly again and become stable afterwards. The mean survival time is estimated to be 2829.6 days with standard error of 75.1 days. Though, the survival graph decreases gradually, patients tend to have a higher survival since the least survival probability is estimated as 0.89 (89 %)

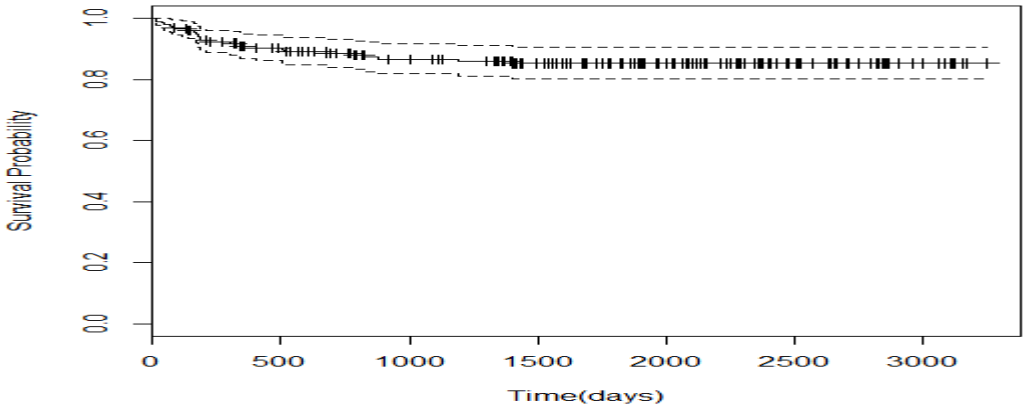


**Figure 1: Kaplan-Meier Function for no Predictors**

Figure 2 provides KM functions for treatment groups. It was observed from the results that treatment 1 continues to drop slightly until itbecame constant between 750 days and 1400 days and drop again at day 1400 and became constant afterwards. Treatment 2 also continued to fall in survivaluntil day 120 when it became constant in survival. Treatment 2 was working better up to some point and then stopped working after that point. This is the point at which the two curves cross each other. From the graph, it can be seen that treatment 2 was working better than treatment 1until day 180 when treatment 2 changes over for treatment 1 to work better.

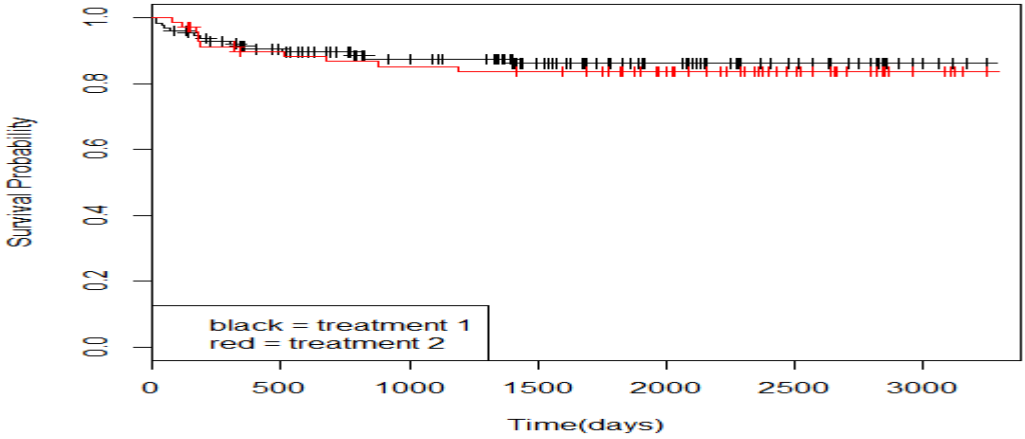


**Figure 2: Kaplan-Meier Survival Graph for Treatment Groups**

The crossing point of the two curves is the time at which patients resorted to treatment one. The gap between the two curves distinguishes between the survival distributions which indicates that the curve above the other has higher survival than the other; that is patients under treatment 1 slightly longer than patient receiving treatment 2. However, the two curves are very close to each other and may also lead to the conclusion that the two treatment effect are the same or similar. The mean survival time and the corresponding standard error for each treatment 1 and treatment 2 are recorded as (mean – 2,851days, standard error = 92.4) and (2,785days, standard error 128,6days) respectively.

Figure 3 presents the survival function for cancer stages. It is observed that patients with stage 1 cancer have perfect survival with no one experiencing death. Stage 2 cancer patientshave perfect survival (100% survivor) at the beginning of the study till 250 days when the survival probability dropped to 90%and maintain stable afterwards. Stage 2 cancer patients have moderately high survival than stage 3 cancer patients, whiles stage 4 cancer patients have the least survival. There is a wide gap between stage 1 and the other stages of cancer, indicating that they may not have close survival with stage 1 cancer. Stage 3 cancer patients decreases gradually in survival up to 85% until day 800 and became stable till day 1300 when there was a slightly drop in survival 84% and then became stable afterwards.

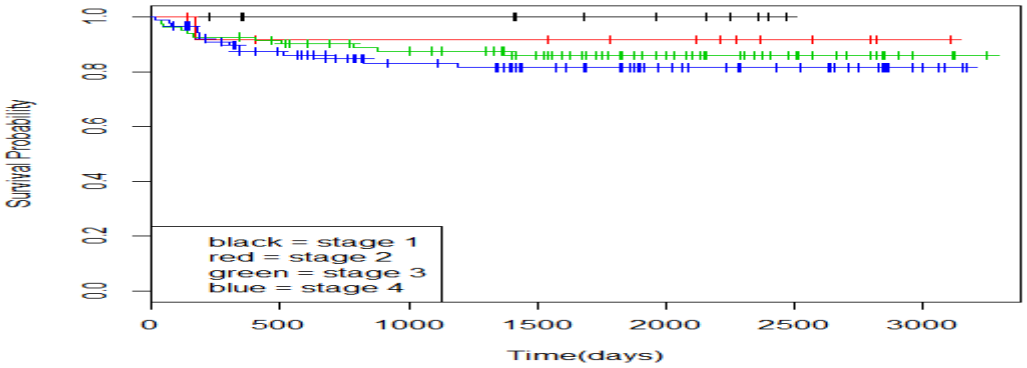


**Figure 3: Kaplan-Meier Survival Graph for Prostate Cancer Stages**

Stage 4 cancer patients continue to reduce in survival up to survival probability. 80% at day 1250 and became stable in survival afterwards. Stage three cancer and stage four cancer graph are close to each other indicating that they may have those survival (and thus similar risk). In all, patient cancer stage tends to have higher survival with the least survival probability 80% (0.8). The mean survival time and the corresponding standard error for each patient cancer stage 1, stage 2. Stage 3 and stage 4 are recorded as (mean = 3,138 , standard error = 0), (mean =2,890, standard error = 237), (mean=2,759, standard error =10) and (mean = 2629, standard error=119) respectively.

### 4.3 Assessing the Significant Effect of Predictors

The statistical significant effect of predictors on patient survival wasexamined. Testing the predictor that influence the survivor of a patient help in the best approach in improving the survival of the patients. Table 4 provides p-values for Wald test of' predictor. The Wald test give the p-value for each continues and categorical covariate. In performing Wald test, treatment 1 and prostate cancer stage 1 are used as the baseline.

**Table 4: P- Values for Full Predictors**

| Predictor | P-Value |
|---|---|
| Intercept | $1.500655e^{-28}$ |
| Age | $6.717104e^{-03}$ |
| Treatment | $2.647685e^{-01}$ |
| Stage 2 | $2.235348e^{-29}$ |
| Stage 3 | $6.523545e^{-198}$ |
| Stage 4 | $0.00e^{+00}$ |

The result of the Wald test indicated that stages of the prostate contributed significantly to the survival of a patient. Treatments and age at which a patient was detected prostate cancer do not statistically contributed to the survival of' thepatient.

### 4.4 Testing Survival Difference of Prostate Cancer Stages

Table 5 and 6 provide Wilcoxon and Log Rank test respectively for survival differences of prostate cancer stages. Both the Wilcoxon and Log rank test the alternative hypothesis that: patients do nothave equal survivor for prostate cancer stages.

**Table 5: Wilcoxon Test for Cancer Stages**

| | No. of Patients | Observed | Expected | $\frac{(O-E)^2}{E}$ | $\frac{(O-E)^2}{V}$ |
|---|---|---|---|---|---|
| Stage 1 | 13 | 0.000 | 1.63 | 1.6314 | 1.8663 |
| Stage 2 | 13 | 0.955 | 1.67 | 0.3039 | 0.3488 |
| Stage 3 | 82 | 10.295 | 10.88 | 0.0318 | 0.0602 |
| Stage 4 | 91 | 13.931 | 11.00 | 0.7813 | 1.4878 |

*Chisq = 2.9 on 3 degrees of freedom, p = 0.040*

**Table 6: Log Rank Test for Cancer Stages**

| | No. of Patients | Observed | Expected | $\frac{(O-E)^2}{E}$ | $\frac{(O-E)^2}{V}$ |
|---|---|---|---|---|---|
| Stage 1 | 13 | 0 | 1.63 | 1.7484 | 1.8699 |
| Stage 2 | 13 | 1 | 1.67 | 0.3481 | 0.3730 |
| Stage 3 | 82 | 11 | 10.88 | 0.0416 | 0.0736 |
| Stage 4 | 91 | 15 | 11.00 | 0.8897 | 1.5815 |

*Chisq = 3 on 3 degrees of freedom, p = 0.038*

The results indicate that both the Wilcox on and Log rank rejected the hypothesis that patients have equal survival for prostate cancer stage with p-values 4.01 % and 3.87% respectively.

## *5. Discussion and Conclusion*

The results obtained show that older men (from 74 years - 95 years) usually develop dreadful cancer (stage 3 or stage 4 cancer). The mean age of patient is 68 years with cancer stage recorded as 3.261 indicating that cancer starts to develop to stage 4 cancer at the age of 68 years. The minimum age for a patient is recorded as 34 years suggesting the age at which men develop stage 1 cancer. It could be deduced that most patients are detected of prostate cancer when the tumor has spread to the delicate part of the body. The reason could be inability for men to have early prostate cancer checkup. This is the reason of low number of records of patients from the year 2005to2013 where only 199 patients were recorded. The number of patients dying of the disease increases with an increase in the stage of the cancer. Results indicated that, till date no patient has ever died of prostate cancer with stage 1. Stage 4 had 15 people experiencing the event whiles stage three had 11 patients experiencing the event. Stage two had only one person experiencing the event. Stage 4had the least survival. The hypothesis that the survival cancer stages are equal is refuted (p-value $=$ 4%). Thus, stage 1 patient will not have the same survival with stage 2 and so on. The effect of age and treatment given to patients do not have any statistically significant effect on patient survival. The results of this findings is in parallel with the study conducted by Arthur et al (2006) which reported that prostate cancer is common with men at age from 56 to 85 years. All patients are given one kind of treatment at the initial diagnose. A combination of treatments is given to some patients perhaps due to the nature of the tumor or probably the age of the patient at the time of treatment.

## *References*

Allison, P. D. (1995). Survival Analysis Using SAS: A Practical Guide.

Arthur, F.K.N., Yeboah, F.A., Adu-Frimpong, M., Sedudzi, E. E., and Boateng, K. A. (2005). Prostate cancer screening in Ghana -a clinical benefit. Journal of Science and Technology, vol 25. No. 2.

Chu, L. W., Ritchey, J., Susan, S. D., Saba, M. Quraishi, H. Z., and Hsing, A. W. (2011). Prostate cancer incidence rates in africa. Division of Cancer Epidemiology and Genetics, No. 6.

Cleves, M., Gould, W., Gutierrez. R. G. and Marchenko, Y. V. (2010) .An Introduction to Survival Analysis Using Stata. Stata Press, third, edition

Ghana Statistical Service (2012). 20l0 Population and Housing Census (PHC) finalresults. www.statsghana.gov.gh/docfiles/2010phc/(October

Jemal, A., Thomas, A., Murray, T., & Thun, M. (2002). Cancer statistics. California Cancer Journal for Clinicians, 52:23-47.

Korostcleva, O. (2008). Clinical statistics: Introducing clinical trials, survival analysis, and longitudinal data analysis.

Machin, D., Cheung; Y. B., & Mahesh, K.B.(2006). Survival Analysis: A Practical Approach.

Parkin, D. M. (2003). Epidemiology of cancer in Africa. Proceedings of the 4th International Conference on Cancer in Africa.

Siddique, T., Sabih, M., Khan, S., & Salam, A. (2001). A survival analysis of metastatic breast cancer in Pakistani patients. Journal o] Pakistan Medical Association, 54(67).

Wabinga, H. (2003). Cancer registry development in Africa. Proceedings of the 4th International Conference on Cancer in Africa.

World Health Organization (2011). Prostate cancer death rate by country. World Life Expectancy. (World health rankings live longer live better). www.worldlifeexpectancy.com/cause-of-death/prostate-cancer/by-country www.ncbi.nlm.nih.gov/pubmed/21675880 (2013)