Data Article

Data on the survival times of breast cancer patients in a Teaching Hospital, Osogbo

Phillip Oluwatobi Awodutire [a,*], Oladapo Adedayo Kolawole [b], Oluwatosin Ruth Ilori [b]

[a] University of Africa, Toru Orua, Nigeria
[b] Ladoke Akintola University of Technology Teaching Hospital Osogbo, Nigeria

ARTICLE INFO

ABSTRACT

In a bid to assess the contribution of prognostic factors to survival times of breast cancer patients from day of presentation in Nigeria, a data was collected from Ladoke Akintola University of Technology Teaching Hospital Osogbo. This is a retrospective data culled from the case note files of the breast cancer patients. The survival time of the patients was recorded as the difference between the day of presentation and the day of last contact. The data is censored at 1 year. The prognostic factors considered are years of breastfeeding(In years), Age at menarche, stage at presentation, neoadjuvant treatment offered and use of contraceptives. Four survival models were applied to the data to assess the contribution of the prognostic factors to survival times

---

* Corresponding author.
  E-mail address: Phillip.awodutire@uat.edu.ng (P.O. Awodutire).

Specifications Table

Subject	Statistics and Probability
Specific subject area	Survival Analysis
Type of data	Table
How data were acquired	The data was culled from the case notes of the breast cancer patients.
Data format	Raw/Analysed
Parameters for data collection	This is a retrospective study. To collect this data, we considered breast cancer patients that reported for treatment at the hospital between 2000 and 2014. Random Censoring was prefered since the patients didnt report for treatment at the same time. The censored time was taken to be 365 days.
Description of data collection	We traced the case note files of the patients, both at the Clinic wards and at the Biostatistic Records. The case files were randomly sampled. From the sampled files, we culled out the variables needed for the research.
Data source location	Institution: Ladoke Akintola University of Technology Teaching Hospital Osogbo City/Town/Region:Osogbo Country:Nigeria Latitude and longitude (and GPS coordinates, if possible) for collected samples/data:
Data accessibility	Repository name: Mendeley Direct URL to data: "https://data.mendeley.com/datasets/r3zhfg9gwv/draft?a=baf9c7ea-ee26-498e-9ab7-b1aed495369e" doi:10.17632/r3zhfg9gwv.1
Related research article	P.O. Awodutire, O.A. Kolawole, O.R. Ilori, Parametric Modeling of Survival Times Among Breast Cancer Patients in a Teaching Hospital, Osogbo, Journal of Cancer Treatment and Research. Vol. 5, No. 5, 2017, pp. 81-85. doi: 10.11648/j.jctr.20170505.12

Value of the Data

- The data are useful and important to serve as a reference for the creation of a data base on breast cancer patients in Nigeria, as such data base is currently not existent. Also based on the fact that these data are primarily obtained, directly from patients' case note files.
- These data can serve to benefit clinical researchers researching in the field of breast cancer, breast diseases, and oncology research.
- They can also be of benefit to statisticians in clinical fields as the data relate to a clinical condition, and can also be useful for epidemiologists in having a reference point when planning screening and interventional programmes.
- The data can be used or reused by any researcher who is interested in the field of breast cancer research, oncology, or medical statistics and especially for statisticians who are interested in developing new survival models. Also, they can be used to assess or evaluate the performance of their new models as compared to existing ones.
- With regards to impact, the data can positively impact society if accessed by epidemiologists and clinical researchers, and the data are studied, and are used to develop feasible plans for screening programmes in order to help reduce the disease burden and mortality (ie improve survival) in breast cancer patients.
- Thus, in the short term, use of the data can help in planning for public health programmes, and in the long term, to help reduce the disease burden of breast cancer and improve survival.
- The data can also have additional benefits in the sense that they can be used by clinicians and medical researchers in fields not related to breast cancer or oncology in general, in which survival prediction is of value.

1. Data description

In developing or low income countries, breast cancer was characterized by late clinical presentation and in advance stage of the disease, when only chemotherapy and palliative care could be offered, and therefore associated with high mortality [1]. The steady rise in breast cancer cases in Nigeria is an indication of inadequate or ineffective control measures to curtail the disease or due to diversion of global attention to HIV/AIDS and tuberculosis in the country [2]. The data is a secondary data culled from the case notes files of breast cancer patients that reported for treatment at Ladoke Akintola University of Technology, Osogbo. The patients under study are all female. The case files were picked at both the wards, Biostatistics and Record Units. The first date of report of patients to the hospital is taken with some prognostic factors to survival time of the patients were recorded. The survival time is censored at 1 year [3].

The data has 8 variables

1. Time(In Days): This measures the survival times of the patients. The survival time is the difference between the date of report for treatment at the hospital and the date of last contact. The time is measured in Days
2. AGE OF PATIENTS: This gives the age of the patients (in years) at the time of report to the hospital
3. CENSOR: This is the censor values. It is 1 when survival time is uncensored and 0 when survival time is censored.
4. AGE AT MENARCHE: This measures the age at which the patients first had a menstrual flow. This variable is measured in years
5. BREASTFEED: This measures the average years that the patient spent to breastfeed.
6. CONTRACEPT: Has the patient used contraceptive before? 1-Yes, 2-No
7. DETECTION: This variable gives the stage of tumor development as at the time of presentation. There are four stages viz: Stage I,II,III,IV.

We categorized this into two; Stage I/ Stage II (Early Detection), Stage III/Stage IV(Late Detection)

1-Early Detection, 2- Late Detection

1. NEOADJUVANT: Was Neoadjuvant administered during treatment? 1-Yes, 2-No

2. Experimental design, materials and methods

For the data analysis, four survival models were considered. They are Exponential, Weibull, Lognormal and Loglogistics. For model comparison to determine the best of the four, we used the Akaike Information Criterion(AIC).

$$Exponential: \quad f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \qquad x > 0, \ \theta > 0 \tag{1}$$

$$Log - Logistic: \quad f(x) = \frac{x^{\alpha-1}\mu\alpha}{(1+\mu x^{\alpha})^2} \qquad x > 0, \ \alpha > 0, \ \mu > 0 \tag{2}$$

$$Weibull: \quad f(x) = \alpha\mu x^{\alpha-1}e^{-\mu x^{\alpha}} \qquad x > 0, \ \alpha > 0, \ \mu > 0 \tag{3}$$

$$Log - Normal: \quad f(x) = \frac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \qquad x > 0, \ \mu > 0, \ \delta > 0 \tag{4}$$

The Akaike Information Criterion (AIC) was used for comparative studies of the model. The AIC is given as Eq. (5).

$$AIC = 2k - 2\ln(L) \tag{5}$$

3. The following gives the R-Code used to analyse the data

```
info<- read.csv("C:/Users/phillip4real/Desktop/DATA1.csv")
t=info[,1] #Survival Times of Patients#
c=info[,3] #Censor#
x1<-info[,2] #Age of Patients#
x2<-info[,4] # Age at Menarche of Patients#
x3=info[,5] #Average year of Breast Feeding#
x4<-info[,6]#Use of contraceptives#
x5<-info[,7]#Point of Detection#
x6<-info[,8]#Neoadjuvant Application#
library("fBasics")
library("MASS")
library("STAR")
library("ActuDistns")
library("predfinitepop")
library("GeneralizedHyperbolic")
library("ghyp")
library("glogis")
library(ggplot2)
plot1 <- ggplot(info, aes(x=TIME))
plot1 <- plot1+geom_histogram(aes(y = ..density..),binwidth=500)
plot1 <- plot1+ggtitle("Histogram of Survival time in
days")+labs(x="Survival time")
print(plot1)
#parameter estimation for the Exponential Distribution
expara <- fitdistr(t,"exponential")
expg<- rexp(n=nrow(info),rate=0.0025198901)
#parameter estimation for the Weibull Distribution
weipara<- fitdistr(t,"weibull")
weig<- rweibull(n=nrow(info),shape= 0.6406347 ,scale=294.3278921)
#parameter estimation for the lognormal
lognpara <- fitdistr(t,"lognormal")
logng<- rlnorm(n=nrow(info),meanlog=4.7529652,sdlog=1.9460617)
#parameter estimation for loglogistic distribution
loglpara <- glogisfit(t)
loglg<- rglogis(n=nrow(info),location=-1357.606,scale=288.4,shape=211.7
#plotting the real data and laying density functions of the fitted
distribution
hist(t,prob=T,xlab="Survival Time",main="Histogram of Survival
Time and the Fitted Distributions",ylim=c(0.00,0.0025))
lines(density(t),lwd=1,col=2)
lines(density(expg),lwd=1,col=3)
lines(density(weig),lwd=1,col=4)
lines(density(loglg),lwd=1,col=5)
lines(density(logng),lwd=1, col=6)
legend("topright",lwd=c(1,1,1,1,1),col=c(2,3,4,5,6),legend=c("Type
I GHL ",
"Fitted Exponential",
"Fitted Weibull",
"Fitted Loglogistic",
"Fitted Lognormal"))
library(survival)
```

```
  er=(survreg(Surv(t,c)~x1+x2+x3+x4+x5+x6,dist="exp"))#
Exponential Survival Model
  er
  AIC(er)
  qq=(survreg(Surv(t,c)~x1+x2+x3+x4+x5+x6,dist="logn"))#
Lognormal Survival Model
  qq
  AIC(qq)
  bb=(survreg(Surv(t,c)~x1+x2+x3+x4+x5+x6,dist="logl"))#
Log-logistic Survival Model
  bb
  AIC(bb)
  orr=(survreg(Surv(t,c)~x1+x2+x3+x4+x5+x6,dist="wei"))#Weibull
Survival Model
  rr
  AIC(rr)
```

Ethics statement

An informed consent was obtained for experimentation with human subjects from respective administrative officers for data collection and no harm was done to any of the participants.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgement

I acknowledge the efforts of Staff of the department of Biostatistics and Records, LAUTECH Teaching Hospital in Retrieving the case note files.

References

[1] F.A Adetifa, R.K. Ojikutu, Prevalence and trends in breast cancer in Lagos State, Nigeria, Afr. Res. Rev. 3 (5) (2009).
[2] A. Afolayan, Breast cancer trends in a Nigerian population: an analysis of cancer registry data, Int. J. Life Sci. Pharma Res. 2 (3) (2012).
[3] P.O. Awodutire, O.A. Kolawole, O.R. Ilori, Parametric modeling of survival times among breast cancer patients in a teaching hospital, Osogbo, J. Cancer Treat. Res. 5 (5) (2017) 81–85, doi:10.11648/j.jctr.20170505.12.