INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

A Bayesian Logistic Regression on Sickle Cell Anaemia: Variation in Survival by Age and Sex

Okewole Dorcas Modupe
Lecturer, Redeemer's University, Ede, Osun state, Nigeria
Ojo Adedayo Edwin
Intern/Volunteer, National Bureau of Statistics, F.C.T, Nigeria
Ogunlaja Aemere
Lecturer, Redeemer's University, Ede, Osun, Nigeria

Abstract:
A study on variation in survival of sickle cell anaemia by age and sex was carried out. The dichotomous nature of the response variable (survival) suggested modelling with the logistic function. There is a dearth of Studies on the specific way in which survival of sickle cell anaemia varies across sex and different age groups of patients from the South-western part of Nigeria. Adequate information on the pattern of response of patients in the presence of significant covariates will yield more effective medical intervention. This study therefore was focused on the survival of sickle cell anaemia across sex and different age groups. The goal was to find out which of these two variables contribute significantly to the survival of sickle cell anaemia and the specific pattern across the different age groups. Data on Sickle cell anaemia patients obtained from the University College Hospital, University of Ibadan, in the South-western part of Nigeria was analysed in the Bayesian logistic Regression approach. Survival varied significantly across different ages, the estimate of the coefficient was -0.04215 with a Bayesian confidence interval (-0.04215, -0.02338). For the female, the percentage alive reduced at the child bearing ages compared with other ages. Sex on the other hand did not significantly contribute to the variation in survival, having the estimate of its coefficient as -0.1298 with a Bayesian confidence interval (-0.13 and 0.452). Age-specific medical attention might increase the life expectancy of these people.

Keywords*: Logistic regression, bayesian approach, sickle cell, variation*

1. Introduction

Logistic regression deals with the regression involving categorical or non-numerical dependent or response variable being predicted by numerical and/or non-numerical variables. Apart from the conventional classical method, the Bayesian approach, which uses probability distributions to quantify uncertainty about any parameter of interest and is able to change those distributions by conditioning on new data as it arrives, is worth applying to all fields of research. Some applications of the Bayesian methods to logistic models are contained in Genkin et al (2007), Chaloner and Larntz (1989), Genkin et al (2007) applied Bayesian logistic regression to language text categorization, where the result showed that the Bayesian approach competes effectively with other standard approaches. The application of logistic regression, just like other statistical methods, cuts across a wide range of research area as long as the dependent variable is categorical with only two levels. In this paper, the Bayesian approach was used to analyse alogistic regression of sickle cell anaemia survival rate. Extensive research has been carried out on sickle cell anaemia using the classical approach. Chijioke et al (2009) presented the result of a study carried out in a city in Nigeria. It was a study on the longevity and clinical pattern of adult sickle cell anaemia based on a 10-year data obtained in a teaching hospital. The conclusion of the Chijioke et al research was that life expectancy of sickle cell anaemia patients was low, while some survive beyond the 4th decade given optimal management. Another population study on mortality of children with sickle cell anaemia was carried out by Fernandes et al (2010) where mortality rate of children with sickle cell disease was also shown to be high. Chaloner and Larntz (1989)derived a general Bayesian design criterion for non-linear models and applied it to logistic regression. Their theoretical results showed the designs to be optimal. These studies showed several factors involved in the pattern, trend or distribution of sickle cell anaemia survival. None of the studies found in literature has shown the specific way in which the survival of sickle cell anaemia varies with different age groups and sex. We focused this paper on how survival varies with age and sex in order to make possible suggestions on areas requiring more

medical as well as research attention. Age and sex are important covariates in some studies and might have a lot of interference with other covariates if not properly adjusted for.

2. Methodology

2.1. About the Data

The study is based on a secondary data obtained from the University College Hospital, Ibadan, the teaching hospital of the University of Ibadan, Oyo State. Ibadan is the capital of Oyo state, Nigeria and the most populous city in the state with population of over 3 million people. In the study, all reported patients with sickle cell disease over a period of 10 years were the study population. There were 395 females and 381 males. It includes the status of the patient (Dead/Alive) as the response variable, then age and sex as independent variables. The following is a summary of the data.

Age groups	Male		Female	
	Alive	Dead	Alive	Dead
0 - 5	43 (95.6%)	2(4.4%)	28(96.6%)	1(3.4%)
6 - 10	33 (91.7%)	3(8.3%)	25(96.2%)	1(3.8%)
11 - 15	25(89.3%)	3(10.7%)	30(96.8%)	1(3.2%)
16 - 20	63(96.9%)	2(3.1%)	71(98.6%)	1(1.3%)
21 - 25	73(94.8%)	4(5.2%)	66(93.0%)	5(7.0%)
26 - 30	49(96.1%)	2(3.9%)	68(91.9%)	6(8.1%)
31 - 35	27(93.1%)	2(6.9%	31(91.2%)	3(8.8%)
36 - 40	18(100%)	0(0%)	19(82.6%)	4(17.4%)
41 - 45	7(87.5%)	1(12.5%)	12(100%)	0(0%)
46 - 50	4(80%)	1(20%)	10(83.3%)	2(16.7%)
51 - 55	5(100%)	0(0%)	5(100%)	0(0%)
56 - 60	4(80%)	1(20%)	3(100%)	0(0%)
61 - 70	1(20%)	4(80%)	1(50%)	1(50%)
71	2(66.7%)	1(33.3%)	0(0%)	0(0%)
Total	354(93.2%)	26(6.8%)	369(93.7%)	25(6.3%)

Table 1: Frequency distribution of Sickle Cell Survival Status by Age and Sex

2.2. Model and Theoretical Background

The response variable is survival, which provides observations on those that were alive and those already dead at the time of data collection. This was regressed over age and sex using the logistic regression model;

The regression model is given as $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ (1)

Since the response variable is dichotomous, a logistic function is applicable and is given as

$$\log \text{ odds} = \text{logit} = \ln\left(\frac{P_X}{1-P_X}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (2)$$

Where,

$$\hat{\text{odds}} = \exp(\hat{\text{logit}})$$

The logistic function is therefore

$$P_X = \frac{\exp(\log it)}{1+\exp(\log it)} \quad (3)$$

$$\text{and } \hat{P}_X == \frac{\hat{\text{odds}}}{1+\hat{\text{odds}}} = \frac{\exp(\log \hat{it})}{1+\exp(\log \hat{it})}$$

2.3. Framework for the Bayesian Analysis

2.3.1. Likelihood Function

Since the survival status is a dichotomous variable (dead or alive), the function for each individual is a bernoulli function given as:

$$f(y_i) = P^{y_i}(1-P)^{(1-y_i)}$$

$$= \left(\frac{\exp(\text{logït})}{1+\exp(\text{logït})}\right)^{y_i}\left(1-\frac{\exp(\text{logït})}{1+\exp(\text{logït})}\right)^{(1-y_i)} \quad (4)$$

Since individual subjects are assumed to be independent, the likelihood function over a data set of *n* subjects is then given as:

$$L(\beta_1,\beta_2/Y,X) = \prod_{i=1}^{n}\left[\left(\frac{\exp(\text{logït})}{1+\exp(\text{logït})}\right)^{y_i}\left(1-\frac{\exp(\text{logït})}{1+\exp(\text{logït})}\right)^{(1-y_i)}\right] \quad (5)$$

2.3.2. Prior Distribution

We shall use the most commonly used non-informative prior which is given as

$$\beta_j \square N(\mu_j,\sigma_j^2), \quad j=0,1,2.$$

In practice, one of the ways to present a non-informative prior is that μ is stated as zero while σ is taken between 10 and 1000, large enough to be considered non-informative

2.3.3. Posterior Distribution

This is the product of the likelihood function and the prior distribution

$$= \prod_{i=1}^{n}\left[\left(\frac{e^{\beta_0+\beta_1X_1+\beta_2X_2}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1X_1+\beta_2X_2}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2}}\right)^{1-y_i}\right] X \prod_{j=0}^{2}\frac{1}{\sigma_j\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\beta_j-\mu_j}{\sigma_j}\right)^2\right\} \quad (6)$$

2.4. Analysis Method

The analysis was carried out with the use of WinBUGS, which is the windows version of the BUGS (Bayesian analysis Using Gibbs Sampling) project. This software uses Markov chain Monte Carlo method through the Gibbs sampling, to draw samples from the posterior distribution. Summary measures such as measures of central tendencies, partition and dispersion would then be computed from the samples to represent the posterior distribution. We carried out 100,000 iterations after observing signs of convergence as seen in Fig.1for one of the parameters of the model. For more details on checking convergence, see Gilks et al (1996).

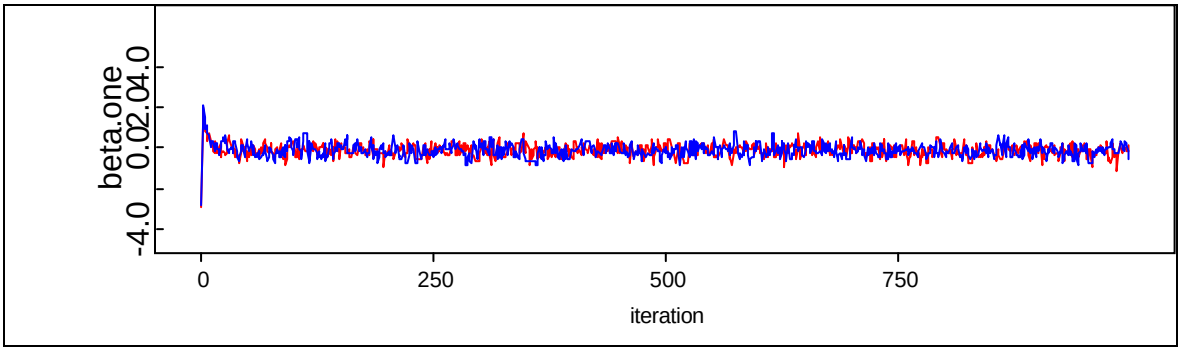

Figure 1: History plot for parameter β_1

3. Results Discussion and Interpretation

Parameter	Mean	Std Dev	MC Error	2.5 Percentile	Median	97.5 Percentile
β_0	3.84	0.3517	0.001251	3.168	3.834	4.551
β_1	-0.1298	0.2956	$9.82\text{x}10^{-4}$	-0.7091	-0.13	0.452
β_2	-0.04215	0.009538	$3.213\text{x}10^{-5}$	-0.0609	-0.04215	-0.02338

Table 2: Summary report of the parameters

There is an indication that β_1 which is the coefficient of sex in the model, is not significantly different from zero since the Bayesian confidence interval which is the 2.5percentile and 97.5percentile, contains zero (-0.13 and 0.452). The DIC (366.7) for the model without sex was smaller than the DIC (368.5) for the model with Sex. Age on the other hand appears to contribute significantly to the variation in the sickle cell survival since β_2 is significantly different from zero (Bayesian confidence interval -0.04215 and -0.02338 does not contain zero).

The 5-year odd ratio of surviving as against not surviving is below one, indicating that the chance of surviving till the next five years is lower than not surviving. Although the regression results implied that survival of sickle cell disease does not vary by sex, there is a little sign of variation in the odd ratio for male and female as can be seen in Table 3 and Fig.2. Further studies with larger samples might be able to give more information on this.

Ages	Odd Ratio	
	Male	Female
5	0.7437	0.8109
10	0.6042	0.6591
15	0.492	0.5369
20	0.4015	0.4384
25	0.3284	0.3587
30	0.2693	0.2942
35	0.2213	0.2419
40	0.1823	0.1993
45	0.1505	0.1646
50	0.1245	0.1363
55	0.1033	0.1130
60	0.0858	0.0940

Table 3: Sex Odds Ratio for survival in Sickle Cell Disease

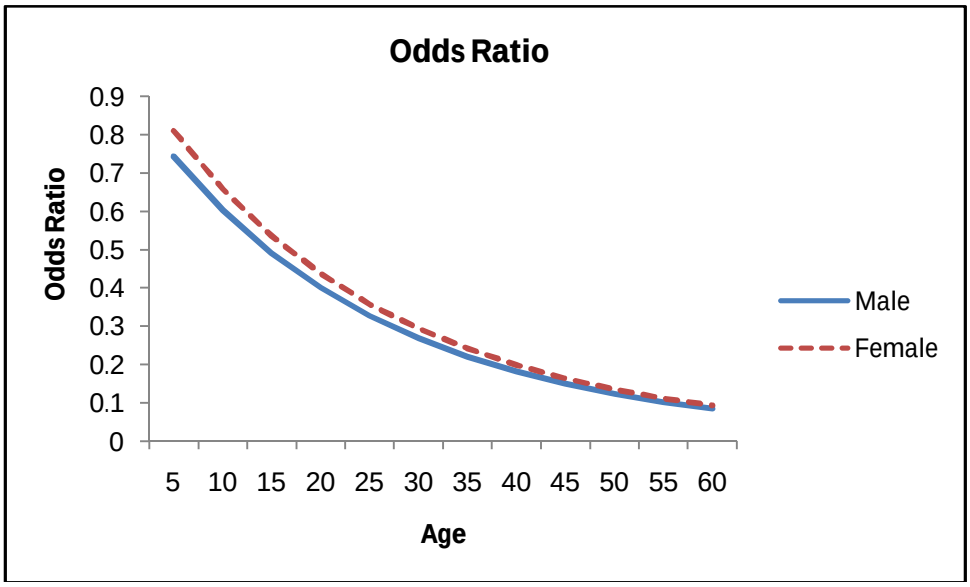

Figure 2: Odds Ratio for Survival in Sickle Cell anaemia

Table 1 also indicate that there is a little difference in the percentage that survived for male and female. When age is not considered, the percentage that survived is a little more for female (93.7%) than male (93.2%) and the median age at death showed this as well having 24years for male and 30years for female. This median value implies that half of the patients die after 30years while half die before 30years while the age for male is lower, 24years. Also, past research gave an indication of significant difference between mean age at death for males (33.4years) and females (36.9years) in Lanzkron *et al* (2013). However, a possible drift in survival by sex reflect in ages 21 – 40 where the percentage alive was smaller for females than males, as can be seen in Fig.3. This might be as a result of increased risk for the women due to child birth since this age range is the child-bearing period for women.

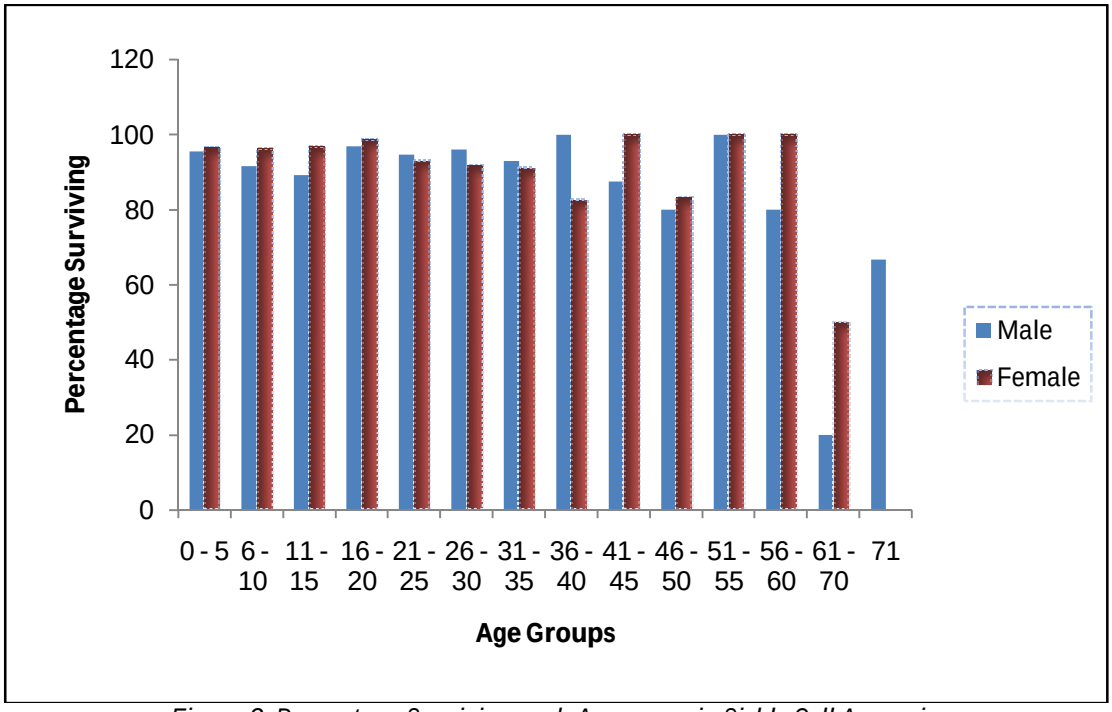

Figure 3: Percentage Surviving each Age group in Sickle Cell Anaemia

The result implies that, although some variation in survival by sex was observed, this variation does not significantly contribute to the total variation in survival of the sickle cell anaemia. The survival however varies significantly by age, with more deaths (occurring between age 21 and 40 for females

4. Conclusions

The study focused on age and sex as factors that could interfere with other variables and should therefore be adjusted for in studies like this. Studies on sickle cell ought to factor in socio-economic variables such as income, education, occupation, access to quality health services, age and sex. This study, based on the data obtained from the teaching hospital of the University of Ibadan, Nigeria, has shown age as an important factor to be considered in modelling survival of sickle cell anaemia. Sex was not significant in the model, suggesting that survival does not vary across sex. It is however important to have further studies on this because of the little variation noticed in the data, the odds ratios and previous results. Another important result was the fact that at the child-bearing age, the percentage surviving is less for females than males unlike other ages. This requires more medical attention which might bring improvement on the health of such groups of people. The variation in survival across different age groups with other factors and wider coverage of health institutions would be considered in further studies.

5. References

i. "Large-Scale Bayesian Logistic Regression for Text Categorization" Alexander GENKIN, David D. LEWIS and David MADIGAN. Technometrics, Vol.49, no. 3, (2007) pages 291- 304.

ii. "The Longevity and Clinical pattern of Adult Sickle Cell Anaemia in Ilorin". Chijioke A. And Kolo P. M. European Journal of Scientific research. Vol. 32 No. 4. (2009). pages 528-532.

iii. "Mortality of Children with Sickle Cell Disease: A Population Study". Fernandes Ana Paula Pinheiro Chagas, Januario Jose Nelio, Cangussu Camila Blanco, Macedo Daniella and Viana Marcos Borato. Journal de Peadiatria. Vol. 86 No. 4, (2010) pages 279 – 284.

iv. "Optimal Bayesian Design Applied to Logistic Regression Experiments". Kathryn Chaloner and Kinley Larntz. Journal of Statistical Planning and Inference. Volume 21. (1989) pages 191 – 208.

v. "Mortality Rates and Age at Death from Sickle Cell Disease: U.S., 1979-2005"Lanzkron Sophie, Carroll C. Patrick, and Haywood, Jr. Public Health Rep. Volume 128, No. 2. (2013) pages 110 – 116

vi. "Mortality in Sickle cell disease: life expectancy and risk factor for early death".Patt O. S., Bambrilla D. J., Rosse W. F., and Klug P. P.,N. Engl. J. Med. Volume 330. (1994)pages 1639 – 1644.