Advances in Computer Science Research, volume 96
Proceedings of the International Conference on Mathematics, Geometry, Statistics, and Computation (IC-MaGeStiC 2021)

Bayesian Accelerated Failure Time Model and Its Application to Preeclampsia

Dennis Alexander*, Sarini Abdullah

Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok, 16424, Indonesia
**Corresponding author. Email: dennis.alexander@ui.ac.id*

ABSTRACT
Preeclampsia (PE) often described as new-onset hypertension and proteinuria during the third trimester of pregnancy. PE, is one of the most feared complications of pregnancy because it can progress rapidly to serious complications, including death of both mother and fetus. It is important to get a better understanding about the factors that might affect the PE condition in pregnant women. Therefore, in this study, we tried to model the relationship between several factors and the time until deliveries under the PE condition. Data on 924 patients at obstetric and gynecology department in a hospital in Jakarta were used in the analysis. A survival regression model, Accelerated Failure Time (AFT) model, was proposed to model the delivery time under PE condition and important factors that influenced the time. Model parameters were estimated using Bayesian method. The results revealed some important factors in explaining the time of deliveries and we also produced the formulation for calculating the estimated probability of delivery given a specific gestational time and patient's characteristics.

Keywords: *Delivery time, Gestational time, Survival regression model.*

1. INTRODUCTION

Hypertension disorder can be experienced by more or less 10% of pregnant women when the gestational more than 20 weeks, with category of ≥ 140 mmHg systolic dan ≥ 90 mmHg diastolic, based on International Society for the Study of Hypertension in Pregnancy (ISSHP) [1]. World Health Organization (WHO) also categorize hypertension disorders in pregnancy as a disrect obstetric/maternal deaths [2]. The hypertension for a pregnant women is one of the indicator of a disease called Preeclampsia (PE). PE can become a worse effect in the future either for the mother and the fetus. National Institute for Health and Care Excellence (NICE) said that some of the factors that made a higher possibility of PE is when the mother has experienced PE at the last pregnancy, diabetes, pregnant when she is 40 years old, have body mass index (BMI) ≥ 35 kg/m, or has a family who had experienced PE also [3]. Until now the doctor have not found the cure of PE yet and so far the best practice to manage PE is by giving aspirin to the patient, because it may decrease the risk of PE if it is identified before 16 weeks of gestational [4].

Actually PE condition can be predicted by using a risk scoring method based on the clinical record of the patient. Unfortunately, risk scoring method has some weakness, not only the uneffectively on predicting the time when PE will occur but also the method cannot predict the patient's risk personally [5]. The next popular method is logistic regression which can be used to measure the risk personally but it is not flexible enough to choose a different gestational age to make a category of the severity level of PE [6].

Other than those two methods, there is another method that can predict the time until one possible event happened which is called survival analysis [7]. Survival analysis techniques are usually used to study the amount of time between entry into observation and a subsequent event [8]. Survival analysis can be called as a failure time analysis will be applied analyze this condition, where only one conditon can be happened from two possible events. One of the popular method in this analysis is Accelerated Failure Time (AFT) model where the method can predict or regress one value of the failure time with conditons that the event can be censored or uncensored [9].

In this study we are discussing about pregnancy cases, which there is only one possibility of the event to be happened (delivering with or without PE condition). If

the first event (delivering with PE condition) happened, then there is impossible to have the second event (delivering without PE condition). It also applies to the second possibility, if the second event happens, then there is impossible to expect that the first event will happen also [10].

Speaking of censoring, there are only two kinds of censoring events which are applied in this study, the right censored and uncensored events. The right censored event is applied when the patient delivered the baby normally (without PE condition) and the uncensored event is applied when the patient delivered the baby with PE condition [11].

Some previous studies about Bayesian AFT model were used to make a model about predicting the median of survival time and also to get the posterior distribution about the occurness of the event. Alvares, et al combined AFT model with Bayesian method to make a model using larynx cancer dataset from Kmsurv package in R programming language. The analysis predicted the median survival time and compared the median of survival time between two indivuals [10].

Basharat, et al developed Bayesian joint AFT model with Log-Logistic and Weibull distribution by using joiner package from R programming language. The analysis performed not only the effect of time independet and time dependent covariates, but also the occurance of death after the surgery [12].

Hu, et al combined AFT with Bayesian approach to demonstrate the spatially varying effects on survival rate from prostate cancer in Louisiana. Hu used three different prior distributions in Bayesian estimation for the AFT model to produce highly accurate parameter estimation [13].

Based on the explanations above, this study will focus on building a model using Bayesian Accelerated Failure time to predict preeclampsia condition of the patient and also the time of the event. R programming language will be used to build the model and the data itself consist of 924 patients with 860 of them are censored patients and 64 uncensored patients are gathered from a hospital in Jakarta, Indonesia.

2. RESEARCH AND METHODOLOGY

2.1 Accelerated Failure Time Model

Accelerated Failure Time (AFT) is used in this study for predicting the failure time when the patient got the PE condition. AFT models can be expressed as a survival time T in logarithmic scale in terms of a linear combination of covariates $\boldsymbol{x}$ with regression coefficients β and a measurement error ε as follows:

$$\log(T) = x^T \beta + \sigma\varepsilon \quad (1)$$

Where σ is a scale parameter, T is a shape parameter (when the T value is changed, then the shape of the curve will be changing also), and ε is the error term which usually expressed via a normal, logistic, or any other probabilistic distributions [14].

2.2 Bayesian

The Bayesian approach is an appealing alternative to the frequentist approach since its conception allows to measure the uncertainty associated with covariates, models, hypotheses, latent variabels, and missing data in probabilistic terms. Bayesian also can be used to incorporate prior knowledge in a natural way, such the historical clinical record (PreviousPE) [15].

3. ANALYSIS AND RESULT

3.1 Dataset

Table 1. Preeclampsia dataset

No	Feature	Data Type	Description
1	First_Pregnant	Categorical	{'No'=0, 'Yes'=1}
2	Conception_ IVF	Categorical	
3	PreviousPE	Categorical	{'No'=0, 'Yes'=1}
4	DiabetesMellitus Type2	Categorical	{'No'=0, 'Yes'=1}
5	ChronicHT	Categorical	{'No'=0, 'Yes'=1}
6	AnyFamily HistoryofPE	Categorical	{'No'=0, 'Yes'=1}
7	Smoking	Categorical	{'No'=0, 'Yes'=1}
8	UseofAspirin	Categorical	{'No'=0, 'Yes'=1}
9	UseofAntiHT Drug	Categorical	{'No'=0, 'Yes'=1}
10	Age	Numerical	Age
11	BMI	Numerical	Body Mass Index
12	CRL_mm	Numerical	Crown Rump Length
13	MAP	Numerical	Mean Arteial Pressure
14	MeanUtAPI	Numerical	Uterine Artery Pulsatility Index
15	Opthalmica	Numerical	Opthalmica value
16	PLGF Concentration	Numerical	Placental Growth Factor
17	Days_GA_ Delivery	Numerical	Delivery time
18	PE	Categorical	{'No'=0, 'Yes'=1}

Dataset was retrieved from an obstetrics and gynecology department of a hospital in Jakarta,

Indonesia. There are 924 patients in total where 860 of them were delivered with normal conditon (censored patients) and 64 were delivered with PE condition (uncensored patients). There are 16 features in the dataset that consist of 9 categorical data and 7 numerical data. These features will be used to predict preeclampsia. The features of the dataset are shown at Table 1.

3.2. Bayesian Accelerated Failure Time Model

Survival times for this study can be expressed through the following AFT model:

$$\log(T) = \beta_1 + \beta_2 FirstPregnant + \beta_3 Age + \cdots + \beta_{17} Opthalmica + \sigma\varepsilon \quad (2)$$

Where $\beta_k, k = 1, \ldots, 17$ are the coefficient for each predictor variables to be regressed.

As given in Table 1, in this study there were 16 covariates with 924 patient's medical records. AFT model together with Bayesian approach was applied by considering the distribution of the data. Right censored and uncensored were applied to differentiate the data into two graups and Weibull distribution were used to fit the model. By using 10000 iterations on Bayesian approach with 2000 first iterations as a burn in stage to make a better posterior distribution. The result of regression for each predictor variables are shown at Table 2.

Table 2. Posterior estimates of the parameter observed

Covariates	Coef	Coef Lower	Coef Upper
FirstPregnant	-0.0754	-0.146	-0.009
Age	-0.0057	-0.039	0.030
BMI	0.0002	-0.027	0.028
Conception_IVF	0.0285	-0.097	0.202
PreviousPE	-0.0915	-0.182	-0.002
DiabetesMellitusType2	-0.0596	-0.229	0.156
ChronicHT	0.0924	-0.013	0.212
AnyFamilyHistoryofPE	-0.0309	-0.121	0.077
Smoking	-0.0963	-0.312	0.227
UseofAspirin	-0.1593	-0.290	-0.003
UseofAntiHTDrug	-0.1250	-0.259	0.031
CRL_mm	-0.0211	-0.050	0.008
MAP	-0.0890	-0.118	-0.061
MeanUtAPI	-0.0249	-0.057	0.005
PLGFConcentration	0.0451	0.004	0.089
Opthalmica	-0.0247	-0.055	0.007

Table 2 showed that every covariates had their own coefficient mean score, coefficient lower bound score, and coefficient upper bound score. However, there were some covariates variables that did not have sufficient evidence to explain the incidence of PE, since the 95% confidence intervals (Coef Lower and Coef Upper) were between a negative value and positive value.

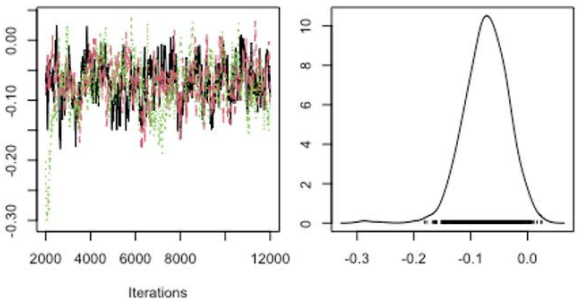

Figure 1 Trace and density plot FirstPregnant.

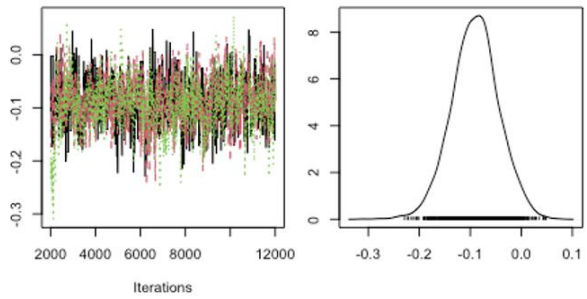

Figure 2 Trace and density plot PreviousPE.

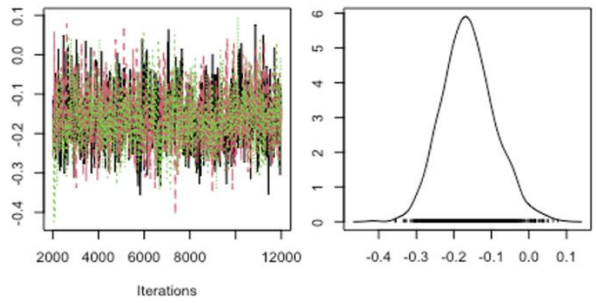

Figure 3 Trace and density plot UseofAspirin.

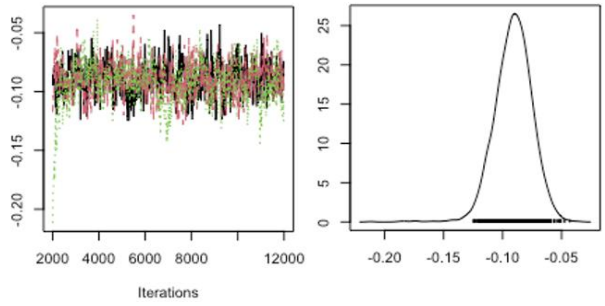

Figure 4 Trace and density plot MAP.

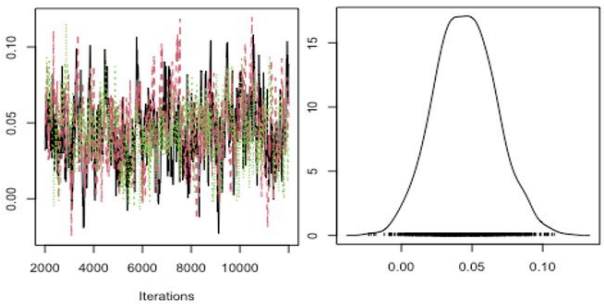

Figure 5 Trace and density plot PLGFConcentration.

Figures 1-5 above are the trace and density plots from predictor variables which give a significant effect to the incident of PE (Coef Lower and Coef Upper score are not between positive and negative number). The three colors in trace plot are blurry, it means that each iterations while building the model give the same score or convergence. For the density plot, it can be seen that the predictor variables are approaching normal distribution.

4. CONCLUSION

From **Table 2**, we could see some predictor variables which gave significant effect to the model such that FirstPregnant, PreviousPE, UseofAspirin, MAP, and PLGFConcentration. The other predictor variables were not having a sufficient evidence to explain the incidence of preeclampsia since the 95% credible intervals are between a negative and positive values.

AUTHORS' CONTRIBUTIONS

Conceptualization, D.A., S.A.; methodology, D.A., S.A.; writing—original draft preparation, D.A.; writing—review and editing, D.A., S.A.; funding acquisition, D.A. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

This research has been supported by Universitas Indonesia.

REFERENCES

[1] M.A. Brown, L.A Magee, L.C. Kenny, S.A. Karumanchi, F.P. McCarthy, S. Saito, D.R. Hall, C.E. Warren, G. Adoyi, S. Ishaku, Hypertensive disorders of pregnancy: isshp classification, diagnosis, and management recommendations for international practice, Hypertension, 2018, 72, 24–43. [CrossRef]

[2] WHO, Trends in maternal mortality: 2000 to 2017 – estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division, World Health Organization, 2019.

[3] National Guideline Alliance (UK), Hypertension in pregnancy: diagnosis and management (NG133), 2019, Available online: https://www.nice.org.uk/guidance/ng133 (accessed on 3 October 2021).

[4] D. L. Rolnik, K. Nicolaides, L Poon, Prevention of preeclampsia with aspirin, American Journal of Obstetrics & Gynecology, 2020, pp. 1-12.

[5] R. Fox, J. Kitt, P. Leeson, C.Y.L. Aye, A.J. Lewandowski, Preeclampsia: risk factors, diagnosis, management, and the cardiovascular impact on the offspring, MDPI Journal of Clinical Medicine, 2019.

[6] D.M. Syaharutsa, Y.A. Purwosunu, Scoring system for preeclampsia screening based on maternal and biophysical factors: result from a 3-month cohort study in Jakarta, Indonesia, Advanced Science Letters, 24(9), 2018, 6361-6365.

[7] T. Badriyah, M. Tahrir, I. Syarif, Predicting the risk of preeclampsia with history of hypertension using logistic regression and naive bayes, International Conference on Applied Science and Technology (ICAST), 2018, pp. 399-403.

[8] N.H.E. Hasanthika, L.A.L.W. Jayasekara, Analyzing the customer attrition using survival techniques, in: International Journal of Statistics and Probability, 2017, DOI: https://doi.org/10.5539/ijsp.v6n6p85

[9] D. Alvares, E. Lazaro, V. Gomez-Rubio, C. Armerp, Bayesian survival analysis with BUGS, Statistics in Medicine, 2021. [Wiley]

[10] D. Wright, A. Wright, K. Nicolaides, The competing risk approach for prediction of preeclampsia, American Journal of Obstetrics & Gynecology, 2020, pp. 12-23.

[11] A.F Majeed. Accelerated failure time models: an application in insurance attrition. The Journal of Risk Management and Insurance, Bangkok, Thailand : The University, 2020, pp:02953269.

[12] N. Basharat, Joint analysis of longitudinal and time to event data using accelerated failure time models: a bayesian approach, University of Saskatchewan, Canada, 2019.

[13] G. Hu, Y. Xue, F. Huffer, A comparison of bayesian accelerated failure time models with spatially varying coefficients, University of Connecticut, Department of Statistics, 2020.

[14] Y. Xue, E.D. Schifano, G. Hu, Geographically weighted Cox regression for prostate cancer survival data in Louisiana, Geographical Analysis, Forthcoming, 2019.

[15] Z. Ma, Y. Xue, G. Hu, Geographically weighted regression analysis for spatial economics data: a bayesian recourse, 2019, Technical Report 19-10, University of Connecticut, Department of Statistics.