



RESEARCH ARTICLE

Apply Parametric Shared Frailty Models to Colorectal Cancer Patients

Hevi J. Hameed, Mohammad M. Faqe

College of Administration and Economics, University of Sulaimani, Kurdistan Region, Iraq

ABSTRACT

Colorectal cancer is a combination of colon and rectal cancer that indicates an abnormal growth of cells in either the colon or rectum and is named according to its original location. After treatment, cancer may return to the primary site of the original tumor or to a different location in the body once or more, which is called recurrent. This paper aimed to model this type of data from 128 colorectal cancer patients collected at Hiwa hospital in Sulaimani considering the gamma shared and inverse Gaussian shared frailty models for analyzing the patient's survival times with colorectal cancer recurrence and estimate the prognostic factor's impact on their survival. The results of these frailty models compared to those without frailty models using Weibull, log-logistic, and lognormal as a baseline distribution for both models. To identify the best model for the data, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were also used. Results showed that the cancer stage was the only significant factor affecting survival in recurrent events, as well as evidence of existing heterogeneity in colorectal patients. According to AIC and BIC, the Weibull as baseline distribution with shared Gamma frailty model proved the most efficient model for the colorectal recurrent data. In conclusion, the shared frailty model is better than no frailty when analyzing this type of data.

Keywords: Survival analysis, recurrent events, inverse Gaussian shared frailty, failure time model, gamma shared frailty model

INTRODUCTION

Survival analysis is a branch of statistics that contains survival data analysis methods with the response variable being the time required for occurring an event of interest.^[1] The event may be death, the onset of a disease, and the required time the patient takes to therapy response or the interval between response and disease recurrence. The event itself is a state transition, and the time to an event is a variable with a positive real value and a continuous distribution.^[2]

Regression models with the response distribution set to time-to-failure are known as parametric survival models. The capacity to take into consideration the censoring and truncation is another feature that distinguishes survival models from normal regression models. Specifically, the response distribution must have positive support, Weibull, log-normal, and log-logistic are some examples of such distributions.^[3]

A frailty model is the random effects time-to-event model, in which the frailty effects the baseline hazard multiplicatively. It can be applied for independent lifetime data (univariate), or for multivariate (dependent) duration times. This multivariate method mostly used for related individuals survival times, such as family members or twins, when the assumption of independence cannot stand, or for recurring occurrences in the same individual, or for times of various events in the same individual, for instance disease start, and relapse recurrence.^[4]

Shared frailty models are the time-to-event data analog to random effects model. It is a model wherever frailties shared by groups of observations or multiple records in the same observation instead of specific observation, thus causing the same group's observations to be correlated. In these models, the main idea is that each individual has different frailty, and the frailest will experience the event of interest earlier than the less frail subjects in the data.^[5]

According to the World Health Organization (WHO), the second-leading cause of cancer-related mortality globally is colorectal cancer. Patients getting curative resection for localized illness had a 5-year survival rate of 70–90%. In contrast, when there are metastases in the regional lymph node, this probability lowers to 40–80%. The most effective

Corresponding Author:

Hevi J. Hameed, College of Administration and Economics, University of Sulaimani, Kurdistan Region, Iraq.
E-mail: hevi.hameed@univsul.edu.iq

Received: July 18, 2022

Accepted: August 11, 2022

Published: November 01, 2022

DOI: 10.24086/cuesj.v6n2y2022.pp119-124

Copyright © 2022 Hevi J. Hameed, Mohammad M. Faqe. This is an open-access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0).

technique is a surgical treatment to colorectal cancers' cure; however, recurrences happen at a steady rate depending on the disease stage.^[6]

Colorectal cancer frequently returns after at least a year. When it is simply an advanced form of the original cancer, or when it occurs within a few months. In this case, the cancer frequently returns, because the first round of treatment did not eliminate all of the cancer cells, the patient may be more tired than usual, and depending on whether and where the cancer has spread, the doctor performs a physical examination and an endoscopy of the colon at regular intervals during these examinations, which usually take place every 3–6 month.^[6]

Ignoring the presence of heterogeneity leads to inaccurate parameter estimates and standard errors in survival analysis. It may also overstate life expectancy. Underestimation of covariate effects when heterogeneity is disregarded. Discounting frailty causes the estimate of regression coefficients to be biased toward zero by an amount that relies on the distribution and variability of the frailty factors.^[7]

While there are several studies that have considered frailty and shared frailty models, to the best of our knowledge, none have considered the recurring events of colorectal cancer, and this is the first examination of the factors affecting patient survival without colorectal recurrence using shared frailty model in accelerated failure time (AFT) metric. This article demonstrated the application of frailty models to recurrent tumor events in the colorectal cancer dataset. Frailty models are the extension of common survival models such as (AFT models Cox proportional hazard and parametric proportional hazard). This article focused only on AFT models and their extension. Various extensions from univariate frailty models are possible to multivariate frailty models and one of these is considered in this study, namely, shared frailty models, which will share frailty among groups or clusters, where the cluster can be an individual or a group. Here, groups recurring events in the same person in patients with colorectal cancer.

MATERIALS AND METHODS

Constructing random effects models can be useful in recurrent occurrences studies over time or other sorts of multivariate survival data; it is a common phenomenon to tackle such problems utilizing frailty models. Many statistical strategies for adjusting and evaluating unobserved heterogeneity have been developed. Modeling unobserved variability in survival data is a topic that has received a lot of attention in recent decades. Unobserved heterogeneity, often known as hidden variability, is a significant source of variation in medical and biological applications. Variability is classified into two types: that produced by observable risk variables and that affected by unknown variables that are ideally unpredictable.^[8]

In the present paper, we considered the no frailty model and the shared frailty model. The no frailty model AFT models are discussed without taking frailty into account, while, in shared frailty, model survival models are analyzed using frailty effects. Our study interest is about the colorectal cancer recurrent disease. Frailty models are best models can be used to describe such unobserved heterogeneity in time-to-event analysis. The effect of heterogeneity, or frailty, has been recognized for a long time, there are several methods

to access the frailty effect. We choose Gamma distribution and inverse Gaussian distribution for evaluating the frailty effect, because they have a simple density function, for which parameters are easily obtained through likelihood estimation.

Survival Analysis Functions

We start with a time-to-failure random variable (T) and specify the density function for T as $f(t)$, the survivor function $S(t) = P(T > t)$ the cumulative distribution function can be described as $F(t) = P(T \leq t) = 1 - S(t)$. In this context, one more function that stands out is the hazard function $h(t) = \frac{f(t)}{S(t)}$ it is the instantaneous failure rate up till time t given survival.^[3]

Cox Proportional Hazards

The Cox proportional hazards model is a method of multiple regressions for determining the impact on survival time from several covariates.

$$h(t | z) = h_0(t) \exp(\beta'Z) \quad (1)$$

Where $h_0(t)$ denotes a legitimate hazard function (failure rate) for some unspecified life distribution model, β , is a vector of parameters, and Z is a covariate vector.^[9]

AFT Model

When examining survival time data, an alternative to the Proportional Hazard (PH) model is the AFT Model. We investigated the direct effect of explanatory variables on survival time rather than risk using AFT models. For the reason that the parameters assess the corresponding covariates effect on survival time, this feature enables the result's interpretation be more easier than other models.^[9]

According to the AFT model, an individual at time t with covariate Z has the survival function as same as an individual at time $\{t \times \exp(\beta'Z)\}$ with a baseline survival function, where $\beta^t = (\beta_1, \beta_2, \dots, \beta_p)$ is a regression coefficients vector; also, the AFT model can be defined by this relationship:

$$S(t | Z) = S_0\{t \times e^{(\beta^t z)}\} \text{ for all } Z \quad (2)$$

The AFT model is a log-scale in relation to time that it gives the analogous to the ordinary linear regression method. In this method, we model the $\log(T) = Y$, in which it is the survival time's natural logarithm. This is how positive variables are naturally transformed into observations along the full real line in linear models. For Y , a linear model is assumed:

$$\log(T) = Y = \mu + \beta^t z + \sigma \epsilon \quad (3)$$

Where $\beta^t = (\beta_1, \beta_2, \dots, \beta_p)$ is regression coefficients vector

μ = Intercept

σ = Scale parameter

The error distribution = ϵ which assumed to have a specific parametric distribution.

When we denote by S_0 the survival function when $Z = 0$.

We adopt the following *baseline hazard* functions $h(t)$ and survival functions $S(t)$ the parameter λ, p :^[9]

1. Weibull
 $h(t) = \lambda p t^{p-1}$ (4)

$S(t) = e^{-\lambda t^p}$ (5)

2. Lognormal

$$h(t) = \frac{\varnothing\left(\frac{\log t - \mu}{\sigma}\right)}{\left[1 - \varnothing\left(\frac{\log t - \mu}{\sigma}\right)\right] \sigma}$$
 (6)

$S(t) = 1 - \varnothing\left(\frac{\log t - \mu}{\sigma}\right)$ (7)

3. Log-logistic

$h(t) = \frac{\lambda p t^{p-1}}{1 + \lambda t^p}$ (8)

$S(t) = \frac{1}{1 + \lambda t^p}$ (9)^[9]

Frailty Models

A generalization of the survival regression model is a frailty model. In addition to the observed covariates, a frailty model also allows for the latent multiplicative effect to be presence on the risk function. This impact, or frailty, is inferred rather than directly determined from the data. It is supposed having limited variance and its mean equal to one. When the frailty is larger than one, people have a higher risk of failing and are considered to be frailer than others. Thus, frailty models can be a valuable alternative to ordinary time –to-event models, while the standard models are unable to fully account for the variation in failure observed times.^[3]

AFT Shared Frailty Model

The applicable choice for multivariate time-to-event data is the shared AFT frailty model. Supposing $\log T_{ij}$ is the natural logarithm of the j^{th} recurrent event’s survival time in the i^{th} subject for colorectal cancer patient.

Where $i = 1, 2, 3, \dots, n$, Patients number $j = 1, 2, 3, 4$, recurrent events and Z_{ij} be the covariates vector related to each individual. Now, the shared frailty model in AFT form can be given by,

$\log T_{ij} = \mu + z_{ij} \beta + U_i + \sigma \epsilon_{ij}$ (10)

Where μ is the parameter of intercept β , the unknown regression coefficients vector, the scale parameter is σ the ϵ_{ij} are the random errors, in which distributed identically independent and the U_i are the random effects for specific individual (each patient) which is supposed to be a random variable with density function that is independent identically distributed.

$f(U_i)$ In this paper, we supposed that the random effect (shared frailty) had inverse Gaussian and gamma distribution with variance θ and unit mean, as described in the density functions bellow.^[10]

Gamma distribution

$$f_u(U) = \frac{u^{(1/\theta)-1} \exp(-\frac{u}{\theta})}{\Gamma(\frac{1}{\theta}) \theta^{1/\theta}}$$
 (11)^[10]

Inverse Gaussian Distribution

$$f_u(U) = \left(\frac{1}{2\lambda\theta}\right)^{1/2} U^{-3/2} \exp\left(\frac{-(u-1)^2}{2\theta u}\right) \quad \theta > 0, u > 0$$
 (12)

The conditional hazard function and survival function could be written as follows for the j^{th} observation in the i^{th} patient:

$h_{ij}(t | u_i) = \frac{1}{\sigma t_{ij}} h_0(\epsilon_{ij} | U_i)$ (13)

$S_{ij}(t | u_i) = S_0(\epsilon_{ij} | U_i)$ (14)

Where

$\epsilon_{ij} = \log t_{ij} - \mu - z_{ij} \beta - U_i / \sigma$ (15)

$H_0(\cdot)$, $S_0(\cdot)$ are and survival function of ϵ_{ij} respectively, and β refers to the covariate coefficient’s vector associated with the covariates vector z_{ij} measured in the j^{th} event of i^{th} subject.^[10]

Model Assessment

We employed Akaike Information Criterion (AIC) in addition to Bayesian Information Criterion (BIC) to compare and select the efficient model among parametric AFT models without frailty with three distributions as baseline and those AFT models with two different shared frailty model (Gamma and Inverse Gaussian).

Akaike’s (1974) information criterion defined by Akaike is:

$AIC = -2 \times \ln L + 2(h + m)$ (16)

Where h indicates the covariate numbers in the model, for instance $m = 2$ in Weibull, because it has two parameters. The smaller AIC value is considered as better model.

Another fit measure is Bayesian information criterion defined by Schwarz in (1978),

$BIC = -2 \times \ln L + 2 \ln(n)$ (17)

The number of data points is represented by n , where n is the sample size. The BIC estimate has the main advantage of including the BIC penalty for the number of parameters being evaluated. The best model is determined by having the lowest BIC values.^[11]

DATA DESCRIPTION AND ANALYSIS

Description of the Data

In this study, 128 cases of colorectal cancer were recorded at Hiwa Hospital in Sulaimani city. Cases were collected over a period of (59) months; beginning from January 7, 2017, to July 20, 2021, for the colorectal cancer patients who were alive at the time that they took part in the study. Data from colorectal

Table 1: The first and last 8 observations from colorectal cancer patient’s data

ID	Follow-up time	Time 1	Time 2	Time 3	Time 4	Treatment	Initial size	Stage	Age
1	31	-	-	-	-	0	2.6	4	51
2	41	26	36	41	-	0	4	4	47
3	52	7	12	41	45	0	5	3	42
4	19	-	-	-	-	0	9	3	62
5	16	-	-	-	-	0	3.5	2	62
6	24	9	-	-	-	0	4.5	2	60
7	21	-	-	-	-	0	1.5	2	78
8	39	33	-	-	-	0	4	4	35
121	12	-	-	-	-	1	5.2	3	79
122	11	-	-	-	-	1	4.5	2	36
123	10	1	2	-	-	1	2.5	3	48
124	9	-	-	-	-	1	5	3	69
125	11	-	-	-	-	1	3.5	3	44
126	10	-	-	-	-	1	4.6	3	66
127	6	-	-	-	-	1	5	3	66
128	32	8	13	-	-	1	5	2	38

Table 2: Explain frequencies and percentages of explanatory variables

Variables	Categories code	Frequency (n=128), n (%)
Treatment group	Chemotherapy (0)	68 (53.1)
	Chemotherapy with radiotherapy (1)	60 (46.9)
Cancer stage	Low stage (0)	48 (37.5)
	High stage (1)	80(62.5)
Initial size	≤2.5 cm (1)	18 (14.06)
	2.6–5 cm (2)	52 (40.63)
	≥5 (3)	58 (45.31)
Age	<50 (0)	47 (36.7)
	≥50 (1)	81 (63.3)
Response	The response is a time to recurrent event or follow-up	

cancer patients who were diagnosed at the time the patients entered the study. The event of interest is recurrences of tumors. Of those patients, 68 were randomized to the group that had just received chemotherapy, and 60 were randomized to the group that received both chemotherapy and radiation therapy. Multiple tumor recurrences occurred in many patients during the study, and at each visit, the new tumors were removed. The dataset contains the first four tumor recurrences of each patient. Furthermore, each time of recurrence was measured from the patient’s time of entry into the study. About 28.9% had at least the first recurrence and 11% of the patients had the second recurrence, while 5.5% had the third recurrent event, 2.3% had all four recurrences. The data consist of the following variables as demonstrated in Table 1 ID, patient identification (this is the sequence number of the subject), follow-up time: the time from the (diagnosis) or the start of the study to the end of it. Time 1,

Table 3: The values of Bayesian Information Criterion and Akaike Information Criterion for the parametric accelerated failure time shared frailty models

Distributions of baseline	The distribution of frailty	AIC	BIC
Weibull	Without frailty	302.4751	325.0553
	Gamma	294.9285	320.7345
	Inverse Gaussian	296.3048	322.1108
Lognormal	Without frailty	305.4182	327.9984
	Gamma	297.7116	323.5176
Log logistic	Inverse Gaussian	299.0783	324.8843
	Without frailty	304.4661	327.0464
	Gamma	297.0295	322.8355
	Inverse Gaussian	298.4006	324.2066

BIC: Bayesian Information Criterion, AIC: Akaike Information Criterion

Time 2, Time 3, and Time 4, are the times of the four potential recurrences of colorectal cancer in months. A patient with only two recurrences has missing values in Time 3 and Time 4. Four observations are made for each patient, one for each of the four possible tumor recurrences. The treatment type that each patient received is demonstrated as (0 = chemotherapy and 1 = chemotherapy with radiotherapy).

Initial size and initial tumor size measured in centimeters (cm). Stage (the cancer stage at which the patient was first diagnosed), and age (the age of the patient). For the AFT without frailty model, the data considered have only the first event (recurrent) or the follow-up time if no relapse is observed.

The frequencies and percentages of explanatory variables are given in Table 2 that 53.1% of the patients received chemotherapy, while (46.9%) received chemotherapy with radiation therapy. In respect to our data, the patients were

Table 4: Weibull accelerated failure time with and without shared frailty model for colorectal cancer patients

Parameter	Weibull (without frailty)			Weibull (Gamma frailty)			Weibull (inverse Gaussian frailty)		
	β	TR	P	β	TR	P	β	TR	P
Treatment	0.093	1.098	0.705	0.0791	0.804	1.0823	0.0612	1.063	0.846
Age	0.2351	1.264	0.356	0.243	0.461	1.275	0.280	1.323	0.388
Stage	-1.288	0.275	0.001*	-1.217	0.005*	0.295	-1.222	0.294	0.005*
Initial size (cm)									
≤ 2.5	Ref	1		Ref	1		Ref	1	
2.6–5	0.489	1.63	0.236	0.359	1.43	0.487	0.384	1.468	0.459
≥ 5	0.258	1.29	0.491	0.105	1.11	0.823	0.157	1.17	0.740
Intercept	4.549	-	0.0000	4.5792	-	0.0000	4.538	-	0.0000
P	1.078	-	-	1.1008	-	-	1.097	-	-
Frailty (θ)	---			$\theta=(1.100818)$			$\theta=(1.068641)$		
				Prob \geq chibar2=0.001*			Prob \geq chibar2=0.002*		

*Significance at 99% level or (0.01). Ref: Reference level which its time ratio always equals to one. TR: Time ratio

Table 5: The survival rate at lower stage (0) and higher stage (1) for Weibull with Gamma shared

ID	Time	Stage	S (t)
1	31	1	0.6443853
2	26	1	0.627423
2	36	1	0.5416252
2	41	1	0.5063942
126	10	1	0.8720858
127	6	1	0.9003416
128	8	0	0.9505355
128	13	0	0.9184979
128	32	0	0.8074583

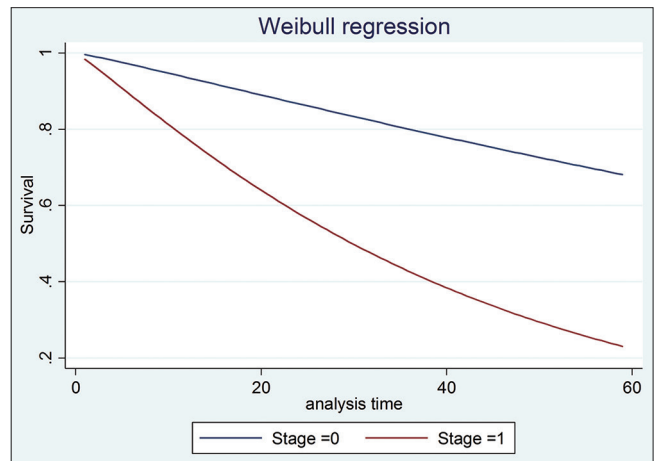


Figure 1: The survival rate at the levels of stage

diagnosed according to their initial cancer stages with 37.5% which were diagnosed in lower stage (stage one and stage two), while patients of higher stages (stage three and four) made up the top proportion, progressing to (62.5%). Regarding patients' initial tumor sizes (14.06%), their tumors were less than or equal to 2.5 cm, whereas 40.63% their tumor sizes were between 2.6 and 5 cm as well as sizes larger or equal to 5 cm correlated to 45.31% of cases, finally in regards to the patients age groups, 36.7% were less than 50 years, while those more than or equal to 50 years were (63.3%). The data analyzed with Stata 14 program.

RESULTS

According to the results in Table 3, the values of AIC and BIC for several parametric AFT models employing inverse Gaussian and gamma shared frailty models, as well as models without frailty. The Gamma shared frailty model with Weibull baseline distribution recorded the lowest value of AIC and BIC among other models tested, demonstrating the efficiency of this model for describing the colorectal cancer recurrence dataset using some parametric AFT models.

Table 4 demonstrates the results of the Weibull AFT model (without frailty) and with inverse Gaussian as well

as Gamma shared frailty model using Weibull distribution as baseline which was the best model for data from patients with colorectal cancer. The table presented estimated values (β , time ratio (TR), p-value, estimated baseline distribution parameters (P), and frailty variance θ) with their likelihood-ratio test. In all three models, the only significant prognostic factor was the patient's initial cancer stage with time ratio values (0.275, 0.295, and 0.294) for Weibull without frailty, Weibull with gamma shared model, and Weibull with inverse Gaussian shared models, respectively, all of them are less than one indicates that patients with a higher initial stage are accelerating the time to relapse for the patient. In addition, the both results of frailty models indicated the significant of heterogeneity (θ) between patients according to the significance-likelihood-ratio test of theta ($\theta = 0$) with values 1.100818 for the gamma shared frailty model and 1.068641 for the shared inverse Gaussian model indicating that there is a frailty effect affects the resulting of the data which tell us that the data is not homogeneous, but heterogeneous some patients are more frails than others. Although each patient has the same covariate value, the heterogeneity impacts the patient's recurring event. The estimates from the inverse

Gaussian shared frailty model are quite close to the results from the gamma shared frailty models, showing the analysis's robustness with respect to the baseline function choice.

From Table 5, it is clear that with regular repetition of recurrent events the patient's survival decreases steadily.

Figure 1 represents the survival rate at lower stage (0) and higher stage (1) for Weibull with Gamma shared frailty which was the best fitting model; furthermore, this figure demonstrates the same results as the numerical does that the patients with lower cancer stages at most survive as the higher cancer stage.

CONCLUSION AND RECOMMENDATION

Conclusion

From performing the recurrent event or more commonly multivariate survival data, according to the results the paper obtained, the following conclusions were drawn:

1. The shared frailty model, as an extension of the Cox model for time-to event data is an excellent choice of recurrent events, particularly when observations of the same subject share an unobservable common frailty. It obviously considers the possibility of dependence between downtimes.
2. Based on AIC as well as BIC to compare the shared frailty AFT models, it is concluded that the shared frailty AFT model using Weibull distribution as the baseline was the most appropriate model to the data set used in this paper.
3. Among (treatment group, age, cancer stage, and initial size) as prognostic factors the only variable that affecting survival time to recurrent event was initial cancer stage, it is noted that the higher stage will accelerate the time to event, meaning that the diagnoses with the early stages of cancer reduce the chances of cancer recurrence.
4. Ignoring un observed heterogeneity may lead to overestimate or underestimate the estimated coefficients.

Recommendation

1. Performing univariate frailty models and shared frailty models by comparing both results to get more information about the use of frailty models in each circumstance and differences among them.

2. More studies should be done in this field because such studies are important and relate to people's lives.
3. Providing qualified endoscopy machines in several hospitals and care units for the early detection of colon cancer hence lowering the incidence as well as recurring rates of it.
4. Data should be registered in health authorities so that researchers can conduct the research in detail and get interesting results.

REFERENCES

1. M. R. Karim and M. A. Islam. *Reliability and Survival Analysis*. Singapore: Springer Nature Pvt Ltd., 2019.
2. D. D. Hanagal. *Modeling Survival Data Using Frailty Models*. Boca Raton, FL: Chapman & Hall/CRC, 2011.
3. R. G. Gutierrez. Parametric frailty and shared frailty survival models. *The Stata Journal*, vol. 2, no. 1, pp. 22-44, 2002.
4. A. Wienke. *Frailty Models in Survival Analysis*. London, United Kingdom: Chapman and Hall/CRC, 2010.
5. G. Grover and D. Seth. Application of frailty models on advance liver disease using gamma as frailty distribution. *SRL*, vol. 3, pp. 42-50, 2014.
6. H. Kobayashi, H. Mochizuki, K. Sugihara, T. Morita, K. Kotake, T. Teramoto, S. Kameoka, Y. Saito, K. Takahashi, K. Hase, M. Oya, K. Maeda, T. Hirai, M. Kameyama, K. Shirouzu and T. Muto. Characteristics of recurrence and surveillance tools after curative resection for colorectal cancer: A multicenter study. *Surgery*, vol. 141, no. 1, pp. 67-75, 2007.
7. U. Abdulkarimova. *Frailty Models for Modelling Heterogeneity*. Canada: McMasters University, 2013.
8. K. Adeleke and G. Grover. Parametric frailty models for clustered survival data: Application to recurrent asthma attack in infants. *Journal of Statistics Applications and Probability Letters*, vol. 6, pp. 89-99, 2019.
9. A. Gebeyehu. *Survival Analysis of Time-to-first Birth after Marriage among Women in Ethiopia: Application of Parametric Shared Frailty Model*. Jimma: Department of Statistics, College of Natural Sciences, Jimma University, 2015.
10. K. D. Fentaw, S. M. Fenta, H. B. Biresaw and S. S. Mulugeta, Time to first antenatal care visit among pregnant women in Ethiopia: Secondary analysis of EDHS 2016; Application of AFT shared frailty models. *Archives of Public Health*, vol. 79, no. 1, pp. 1-14, 2021.
11. P. K. Swain and G. Grover. Determination of predictors associated with HIV/AIDS patients on ART using accelerated failure time model for interval censored survival data. *American Journal of Biostatistics*, vol. 6, pp. 12-19, 2016.