Determinant of flexible Parametric Estimation of Mixture Cure Fraction Model: An Application of Gastric cancer Data.

Chukwu, A. U. & Folorunso, S. A *
Department of Statistics, University of Ibadan, Nigeria.
***Corresponding Author:** serifatf005@gmail.com

Abstract

Cure models are survival models basically developed to estimate the proportion of patients cured in a clinical trial. These models estimate the cured proportion and also the probability of survival. Cure models are a special type of survival analysis model where it is assumed that there are a proportion of subjects who will never experience the event and thus the survival curve will eventually reach a plateau. Cure has become an important measure of long term survival benefit derived from therapy. This study was intended at determining the flexible Parametric Cure Fraction Model for Gastric cancer Data. Suitability of four parametric mixture cure models were considered namely; Log Normal (LN) cure fraction model, Log Logistic (LL) cure fraction model, Weibull (W) cure fraction model and Generalized Gamma (GG) cure fraction model. AIC, e^{μ} (mean time to cure), variance and cure fraction (c) were used to determine the flexible Parametric Cure Fraction Model among the considered models. Gastric Cancer data from 76 patients received adjuvant CRT and 125 receiving resection (surgery) alone were used to confirm the suitability of the models. The data was from a retrospective study in patients with gastric adenocarcinoma who underwent curative resection with D2 lymphadenectomy in the Barretos Cancer Hospital between January 2002 and December 2007. The survival for this data refers to the times until death in months since surgery.The Log-Logistic (LL) gave the minimum value for AIC, minimum means, minimum mean time to cure (e^{μ}) as well as cure fraction with values (525.865, 272.671); (0.529, 0.583); (1.697, 1.791) and (0.122, 0.123) using GCS and GCC, respectively. Also, GG gave the highest cure fraction with C = 0.374 and 0.599 for both GCS and GCC. Cure fraction obtained using GG presented the model as being best in terms of proportion to cure in GCS and GCC. The Log-Logistic (LL) showed a promising result being the best flexible model in terms of AIC, minimum variance, means and mean time to cure, while Generalized Gamma performed best in terms the value of cure fraction which gives the proportion of cured.

Keywords: *Cure Fraction Model, Gastric Cancer, flexible Model, AIC, Mean time to cure.*

1.0 Introduction

Survival analysis is commonly applied in many fields such as biology, medicine, public health, and epidemiology. A distinctive analysis of survival data comprises the modelling of time-to-event data, such as the time until death. The time to the event of interest is called either survival time or failure time [22]. In survival analysis, a set of data can be exact or censored, and it may also be truncated. Exact data set, also known as uncensored data, occurs when the particular time until the event of interest is known. Censored data arises when a subject's time until the event of interest is known only to occur in a certain period of time [22]. Recently, the development of new therapy has resulted in patients living longer with diseases such as cancer. The motivation of this work came from the fact that the achievement of cure is paramount in any epidemiological research. An effective Medical intervention is when a significant number of patients can be cured from a particular disease but due to some factors and different body chemistry,

not all patients will benefits from medical intervention. As medical treatments progress, one would like to distinguish between a change in the probability of cure and an increase in the expected survival time for uncured patients. The study covered four parametric mixture cure models and gastric cancer data were used to validate the most flexible one among the considered models.This paper is aimed to determine the flexible Parametric Cure Fraction Model for Survival Data, (b) to review cure fraction model theoretically and analytically, and (c) to compare the efficiency of the parametric cure fraction model that were considered, (d) to recommend for the flexible models within the parametric cure fraction models that best explain in practical situation, (d) to validate with hypothetical cancer data. The rest of the paper is organized as follows: the next sections briefly review some literature on cure models. Section 3 presents Method and model specification. Section 4 discusses the results. While the last section concludes the paper.

2.0 **Review of Past literatures**

Cure model was first modeled in the survival function of the uncured group as a product of the survival functions of a log-normal distribution and some background distribution for the normal population [4]. Berkson and Gage developed Boag model and the model was later studied extensively by several authors [2].

Several researchers started using a special case of the weighted Poisson distribution and formulated the long-term survival function [6], [18], [19] [20]. Reviewed on existing cure model and consideration of a fraction of long-term survivors of time-changing risk related to the relapse of leukaemia observed in patients treated for Hodgkin's disease was done in 2003 [16].

Development about Bayesian methods for right censored multivariate failure time data for populations with a cure fraction[9] as well as study that found out the most commonly used statistical methods for evaluating treatment or prognostic effects on cancer outcome are the logrank test and Cox regression analysis which relies on the proportional hazards (PH) assumption[15].

Employing a competing risks model which is equivalent to a mixture model that is the prevalent technique for cure data was also studied [3]. Yin *et al.* reviewed and worked on unified cure rate model [21]. Lambert *et al*. improved the parametric non-mixture cure fraction model to incorporate background mortality, by promoting estimates of the cure fraction in population-based cancer investigations [10] [11] [12]. Yi Li *et al*. derived new cure models from the perspective of competing risks and model the dependence between the censoring time and the survival time by means of a class of Archimedean copula models and consider the parameter estimation. They considered large sample outcomes by applying martingale principle though the simulation [23].

Abu Bakar *et al*. evaluated the cure fraction models and used it to monitor time trends in cancer patient survival which give valuable information to solve some problems with the cure fractions models [1]. Ortega, et al defined a negative binomial-generalized gamma distribution with a flexible cure rate survival model by assuming that the number of competing causes of the event of interest follows the negative binomial distribution and the time to event follows a generalized gamma distribution [13]. Fauzia *et al* investigate a survival model with cure fraction and change-point effect based on the bounded cumulative hazard model (BCH) [8].

3.0 **Method and Model Specification**

3.1 Method

Cure models are survival models which allow for a cured fraction in the study population. These models extend the understanding of time-to-event data by allowing for the formulation of more accurate and informative conclusions. These conclusions would otherwise be unobtainable from an analysis that fails to account for a cured fraction in the population.

If a cured component is not present, the model reduces to standard approaches of survival model. Most cure models assume that the susceptible individuals are homogeneous

in risk [17]. The model is divided into two categories namely:

Mixture cure model: can estimate the proportion of patients cured and the survival function of the uncured patients [2][4].
Non-Mixture Cure Model or the Bounded Cumulative Hazard Model (BCH) [18][19][20].

3.2 Model Specification (Mixture Cure Model)

This model was first developed by Boag (1949) and was modified by Berkson & Gage (1952) which can be defined as:

$$S(t) = c + (1 - c)S_u(t) \quad (1)$$

Where

- $S(t)$ = *The survival functions of the entire population*
- $S_u(t)$ = *The survival functions of the uncured patients*
- c = *The proportion of cure patients that is the cure fraction rate*

In cure models, the population is divided into two sub-populations so that an individual is either cured with probability 1- c, or has a proper survival function $S(t)$, with Probability c.

3.2.1 Estimations of Cure Fraction Model (Mixture Cure Model)

The parameter of mixture cure fraction model was employed using Maximum Likelihood (ML) approach. Parameters in cure models can be estimated parametrically and non-parametrically. Thus; the estimation employed shall be parametric in nature. Given the cure model in equation (1), the estimate of parameter c *is given as*:

$$c = \frac{S(t) - S_u(t)}{1 - S_u(t)} \quad (2)$$

3.2.2 Likelihood Estimation of Cure Fraction Model

The likelihood estimation of cure fraction model is:

$$L = [f(t_i)]^{d_i} [S(t_i)]^{1-d_i} \quad (3)$$

$$where\ d_i = \begin{cases} 0, if\ patients\ is\ cured \\ 1, \quad otherwise \end{cases}$$

d_i is zero if the patients is cured and one otherwise.

$f(t_i) = pdf of\ parametric\ cure\ model$

$S(t_i) = survival\ function$

put (3) *in* (1), we obtained

$$1 - F(t) = c + (1 - c)(1 - F_u(t)) \quad (4)$$

differenciating (4), we have;

$$f(t) = ((1 - c)\, f_u(t)) \quad (5)$$

put 5 *in equation* 3 *to obtain*

$$L = [(1 - c)(f_u(t)]^{d_i} [c + (1 - c)\, S_u(t)]^{1-d_i} \quad (6)$$

Taking the logarithms of equation (6) to have the log likelihood function of cure model *as*:

$$logL = \sum_{i=1}^{n} d_i\, log[(1 - c) + \sum_{i=1}^{n} d_i\, log(f_u(t_i)] + \sum_{i=1}^{n} (1 - d_i) log[c + (1 - c)\, S_u(t_i)] \quad (7)$$

3.3 Distributions of Cure Models

The Four different univariate distributions were examined for the mixture cure model to estimate the corresponding
(Proportion of cure patients), their mean and variance.

3.3.1 The Log Logistic Distribution

A random variable t is said to be distributed log-logistic if it pdf is given as;

$$f(t) = \frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{\left[1 + \left(\frac{t}{\beta}\right)^{\alpha}\right]^2}, \quad t > 0 \quad (8)$$

With $\beta = \frac{1}{\lambda}$, we have

$$f(t)=\frac{\alpha\lambda(\lambda t)^{\alpha-1}}{\left[1+(\lambda t)^{\alpha}\right]^{2}} \qquad (9)$$

Using the transformation $\mu=-\log\lambda, \sigma=\frac{1}{\alpha}$, which imply $\lambda=e^{-\mu}$ and $y=\log t \Rightarrow t = e^{y}$ then

$$f(t;\mu,\sigma)=\frac{\frac{1}{\sigma}\left(e^{\frac{\log t-\mu}{\sigma}}\right)^{1-\sigma}}{\left[1+\left(e^{\frac{\log t-\mu}{\sigma}}\right)\right]^{2}} \qquad (10)$$

The cdf of loglogistic cure model is given as

$$F(t)=1-\left[1+(\lambda t)^{\alpha}\right]^{-1} \qquad (11)$$

And the survival function is given as

$$S(t;\mu,\sigma)=\left[1+\left(e^{-\mu}e^{y}\right)^{\frac{1}{\sigma}}\right]^{-1} \qquad (12)$$

3.3.2 Log-normal Distribution

A random variable *t* is said to be distributed *log-normal* if it pdf is given as:

$$f(t;\mu,\sigma)=\frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log t-\mu}{\sigma}\right)^{2}} \qquad t>0,\quad \mu=0,\ \sigma=0 \qquad (13)$$

The cdf of lognormal cure model is given as

$$=\frac{1}{\sigma^{2}\sqrt{2\pi}}\left[ze^{-\frac{1}{2}z^{2}}+\frac{\mu}{\sigma}e^{-\frac{1}{2}(\frac{\mu}{\sigma})^{2}}\right] \qquad (14)$$

With survival function given as

$$S(t;\mu,\sigma)=1-\Phi\left(\frac{\log t-\mu}{\sigma}\right) \qquad (15)$$

3.3.3 Weibull Distribution

A random variable *t* is said to follow a Weibull distribution if it satisfies the density function:

$$f(t)=\frac{\alpha}{\beta^{\alpha}}t^{\alpha-1}e^{-\left(\frac{t}{\beta}\right)^{\alpha}}, \qquad t>0,\quad \alpha=0,\ \beta=0 \qquad (16)$$

The cdf of weibull cure model is given as

$$F(t)=1-e^{-\left(\frac{t}{\beta}\right)^{\alpha}} \qquad (17)$$

And the survival function is given as

$$S(t)=e^{-\left(\frac{t}{\beta}\right)^{\alpha}} \qquad (18)$$

3.3.4 The Generalized Gamma Distribution

The density function of the generalized gamma distribution is given by

$$f(t)=\frac{\beta}{\theta\Gamma(\alpha)}\left(\frac{t}{\theta}\right)^{\alpha\beta-1}e^{-\left(\frac{t}{\theta}\right)^{\beta}} \quad t>0,\quad \theta>0,\ \beta>0 \qquad (19)$$

The cdf of generalized gamma cure model is given as

$$F(t)=\frac{\gamma\left[\alpha,\left(\frac{t}{\theta}\right)^{\beta}\right]}{\Gamma(\alpha)} \tag{20}$$

And the survival function is given as

$$S\left(t^{1},\mu,\sigma\right)=1-\frac{\gamma\left[\alpha,e^{\frac{\log t-\mu}{\sigma}}\right]}{\Gamma(\alpha)} \tag{21}$$

3.4 Estimations and Inferences

3.4.1 Estimations

Using log likelihood functions of cure models in equation 7 at a baseline distribution, the parameters μ,σ and c can be found by differentiating it with respect to the parameters. The MLE of μ, σ and C are obtained using numerical approach

.

$$\frac{\partial \log L}{\partial \mu}=\sum_{i=1}^{n} d_i(\log t_i-\mu)$$

$$\frac{\partial \log L}{\partial \sigma}=-\frac{1}{\sigma}\sum_{i=1}^{n} d_i+\frac{1}{2\sigma^{3}}\sum_{i=1}^{n} d_i(\log t_i-\mu)^{2}$$

$$\frac{\partial \log L}{\partial c}=-\frac{1}{1-c}\sum_{i=1}^{n} d_i+\sum_{i=1}^{n}\phi\left(\frac{\log t_i-\mu}{\sigma}\right)$$

This is repeated for all the cure models distribution that we considered.

4.0 Analysis and Results

The data used to validate the parametric cure model was a gastric data from 76 patients received adjuvant CRT and 125 receiving resection (surgery) alone. The Gastric cancer data were entered into excel spread sheet for cleaning and analysed with R version 3.22. R codes were used to estimate μ, σ and C of the considered distributions as it were showcase in the previous chapter.

4.1.1 Results

Table 1 is a comparison table showing the validation of considered distribution of Parametric Cure Model parameters while table 2 reveals the Comparison of Model evaluation which reveals their AIC, means, variance, expected mean time (e^{μ}), and c. Fig 1 and 2 shows the Survival Plot of lognormal Cure Model for surgery and CRT respectively. Fig 3 and 4 shows the Survival Plot of weibull Cure Model for surgery and CRT respectively. Fig 5 and 6 shows Survival Plot of log logistic Cure model for surgery and CRT respectively. Fig 7 and 8 shows Survival Plot of generalized gamma Cure model for surgery and CRT respectively. Figures 9 and 10 show Survival Plots of Examined Cure Models for surgery and CRT respectively.

Table 1: Distribution of Parametric Cure Model parameters (Surgery and CRT)

Lognormal	***1) Surgery***			**CRT**			
Parameter	Coefficients	Std-Error	P.V	Parameter	Coefficients	Std-Error	PV
C	0.3253112	0.1195319	0.006	C	0.5103684	0.0792443	0.000
Alpha	2.4705506	0.3624391	0.000	Alpha	2.8472763	0.1241601	0.000
Beta	1.2957284	0.2220184	0.000	Beta	0.4839343	0.0905987	0.000
Weibull	*2) Surgery*			**CRT**			
Parameter	Coefficients	Std-Error	PV	Parameter	Coefficients	Std-Error	PV
C	0.3710059	0.1130479	0.001	C	0.543705	0.0632408	0.000
Alpha	0.8184908	0.1347899	0.000	Alpha	2.693517	0.4223474	0.000
Beta	15.0576272	6.6330697	0.023	Beta	19.779680	1.6370140	0.000
Log-logistic	*3) Surgery*			**CRT**			
Parameter	**Coefficients**	**Std-Error**	**PV**	**Parameter**	**Coefficients**	**Std-Error**	**PV**
C	0.1219107	0.0808121	0.131	C	0.1227404	0.1196919	0.305
Alpha	1.3003154	0.1419246	0.000	Alpha	4.6355618	0.6747078	0.000
Beta	7.2741926	1.1404804	0.000	Beta	15.4666846	1.0737646	0.000
Generalized Gamma	*4) Surgery*			**CRT**			
Parameter	**Coefficients**	**Std-Error**	**PV**	**Parameter**	**Coefficients**	**Std-Error**	**PV**
C	0.4283083	0.0652524	0.000	**C**	0.6724405	0.0502801	0.000
Alpha	10.3233229	4.1420285	0.013	**Alpha**	0.5073054	0.0074725	0.000
Beta	0.6366691	0.0930672	0.000	**Beta**	140.2669372	NA	NA
Theta	0.6278258	0.1297477	0.000	**Theta**	4.0318602	0.0597890	0.000

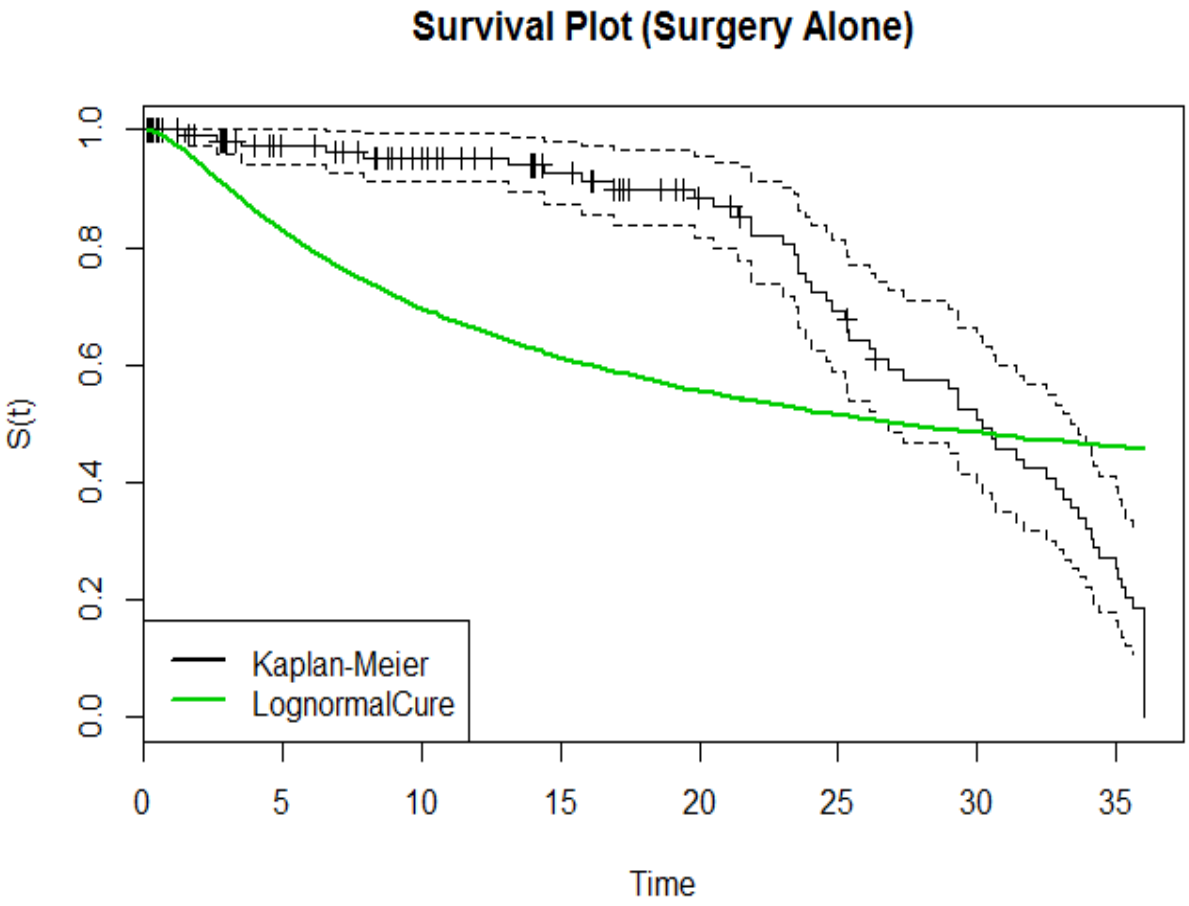


Fig 1: Survival Plot of Lognormal Cure Model (Surgery)

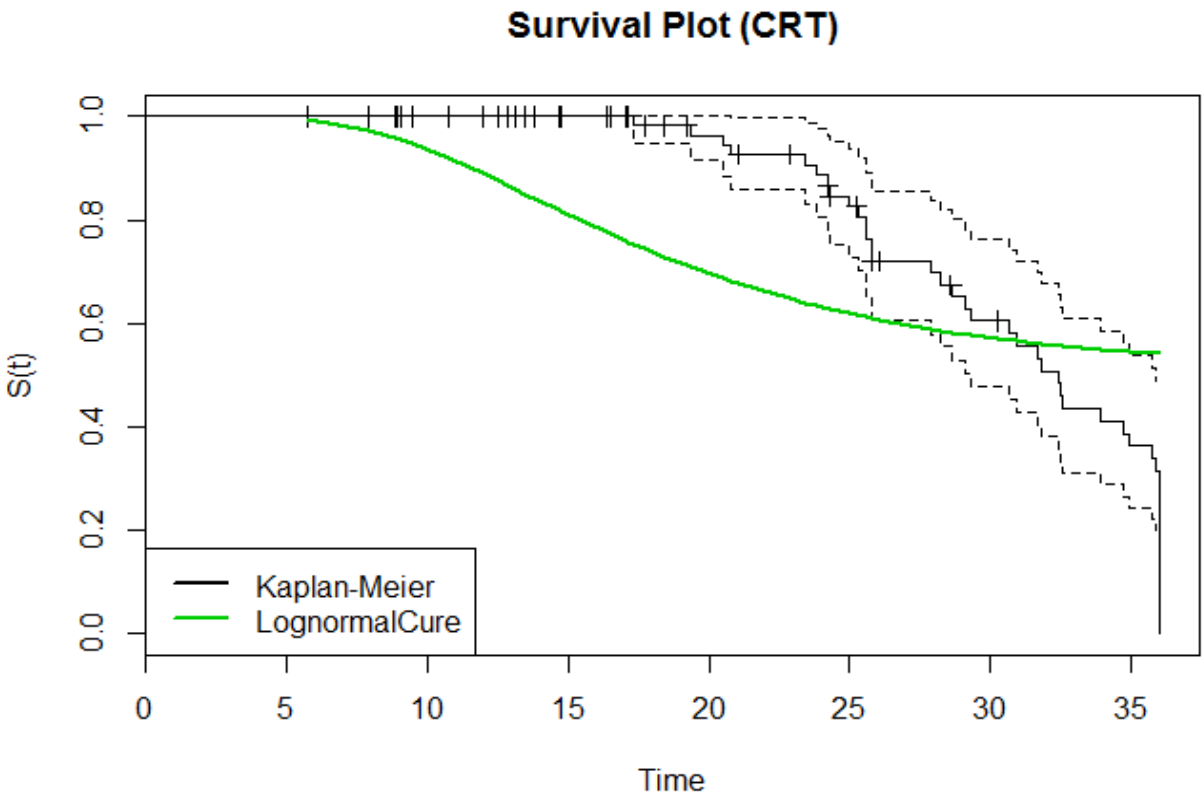


Fig 2: Survival Plot of Lognormal Cure Model (CRT)

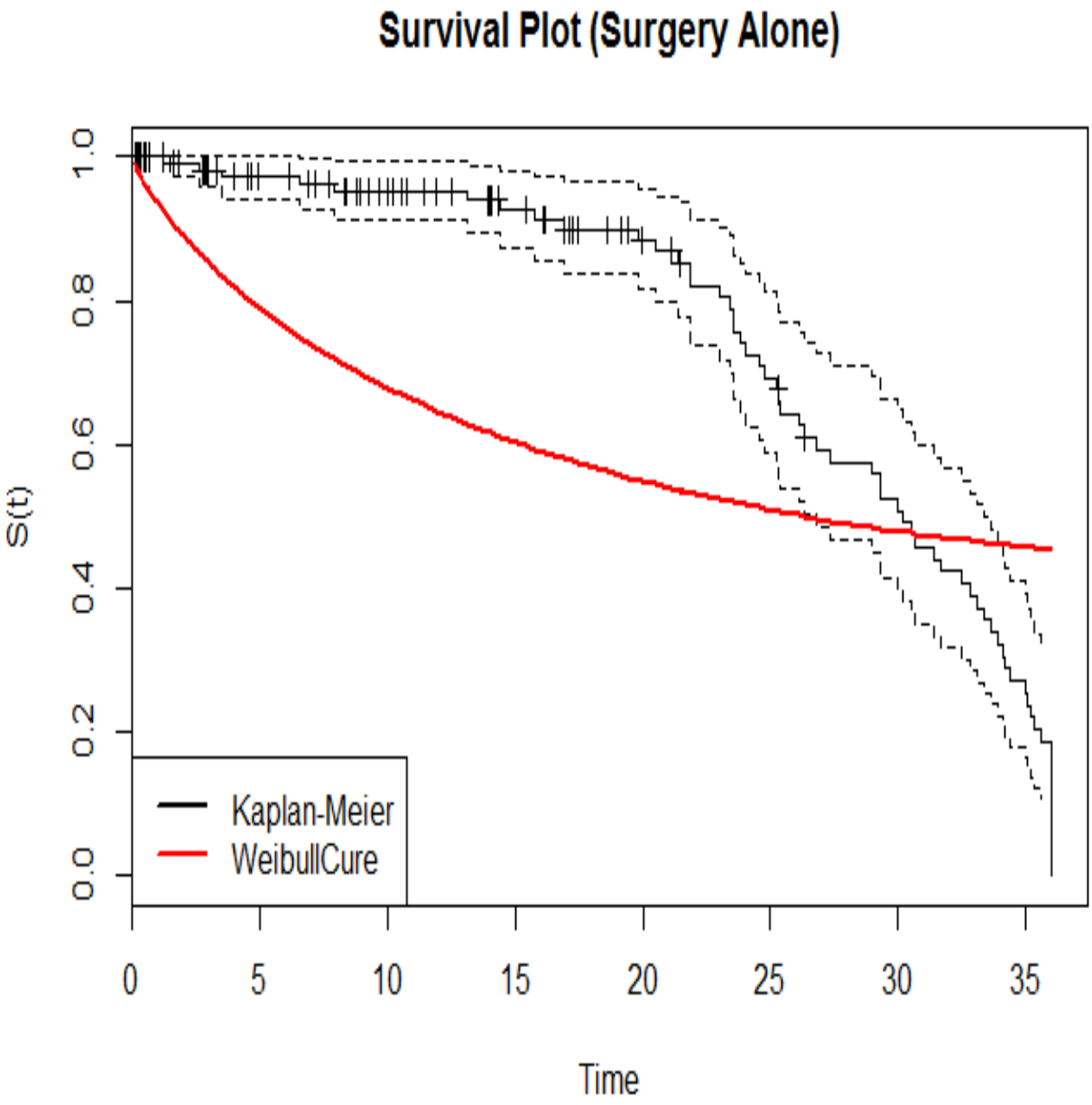


Fig 3: Survival Plot of Weibull Cure Model (Surgery)

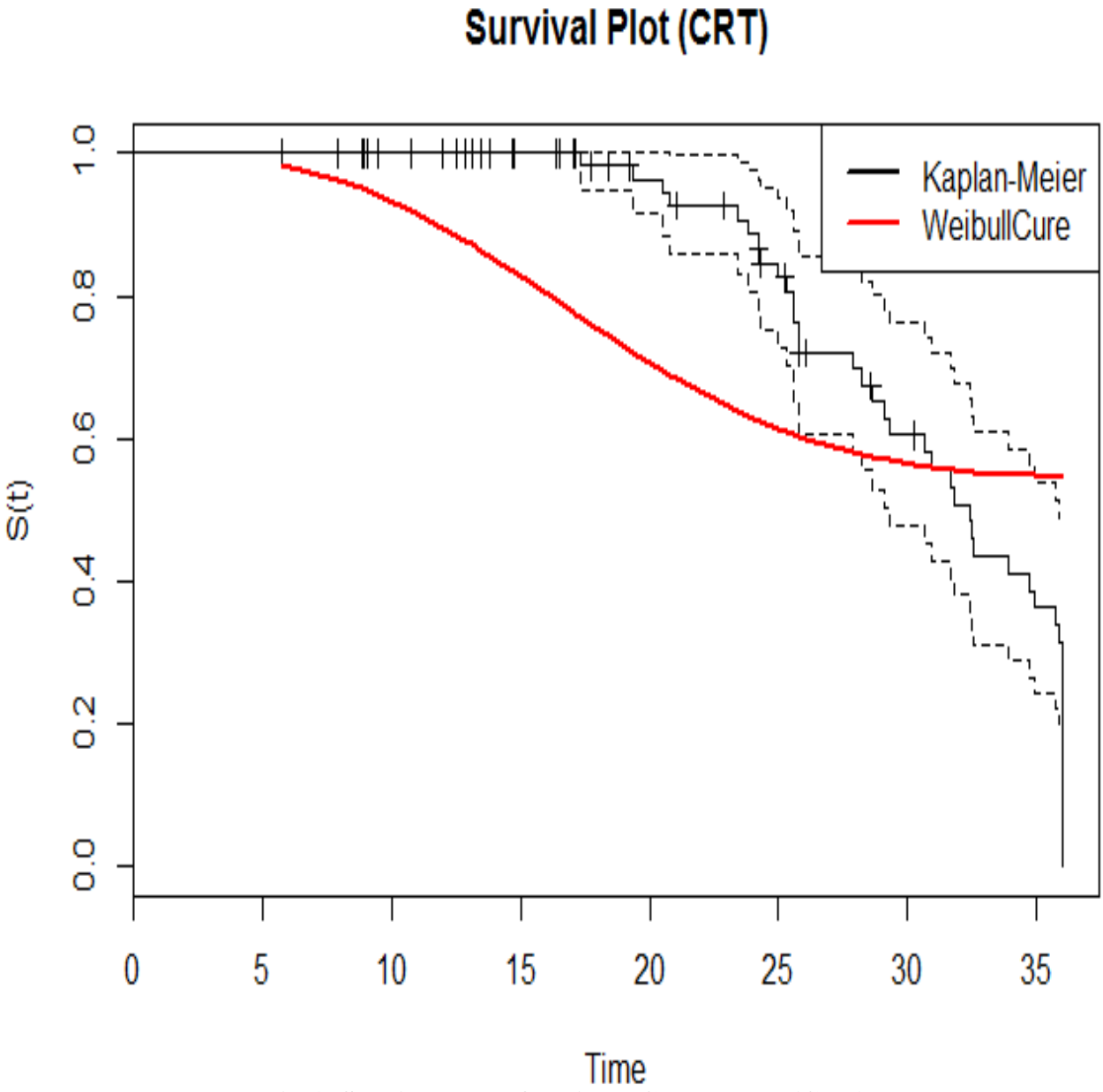


Fig 4: Survival Plot of Weibull Cure Model (CRT)

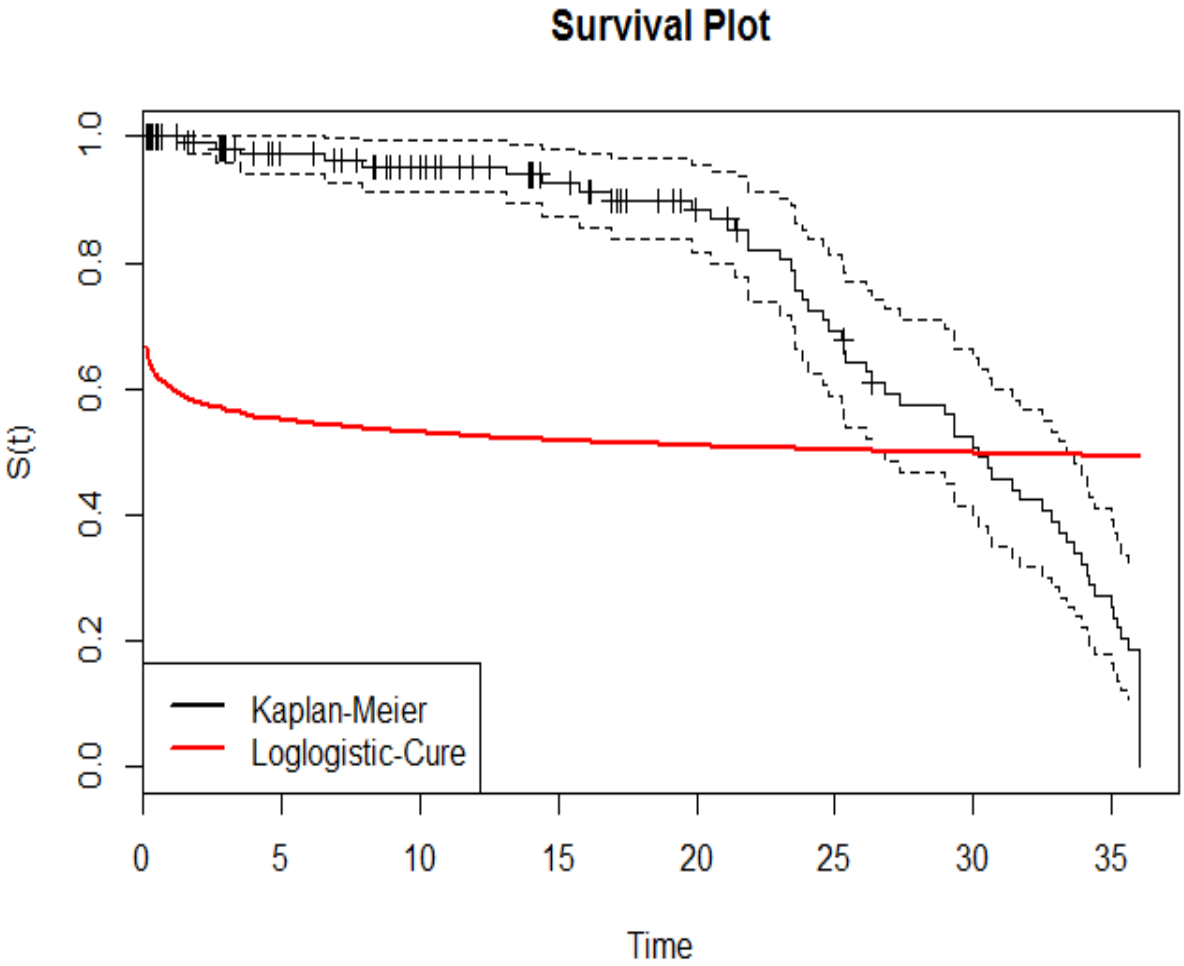


Fig 5: Survival Plot of log logistic Cure model (Surgery)

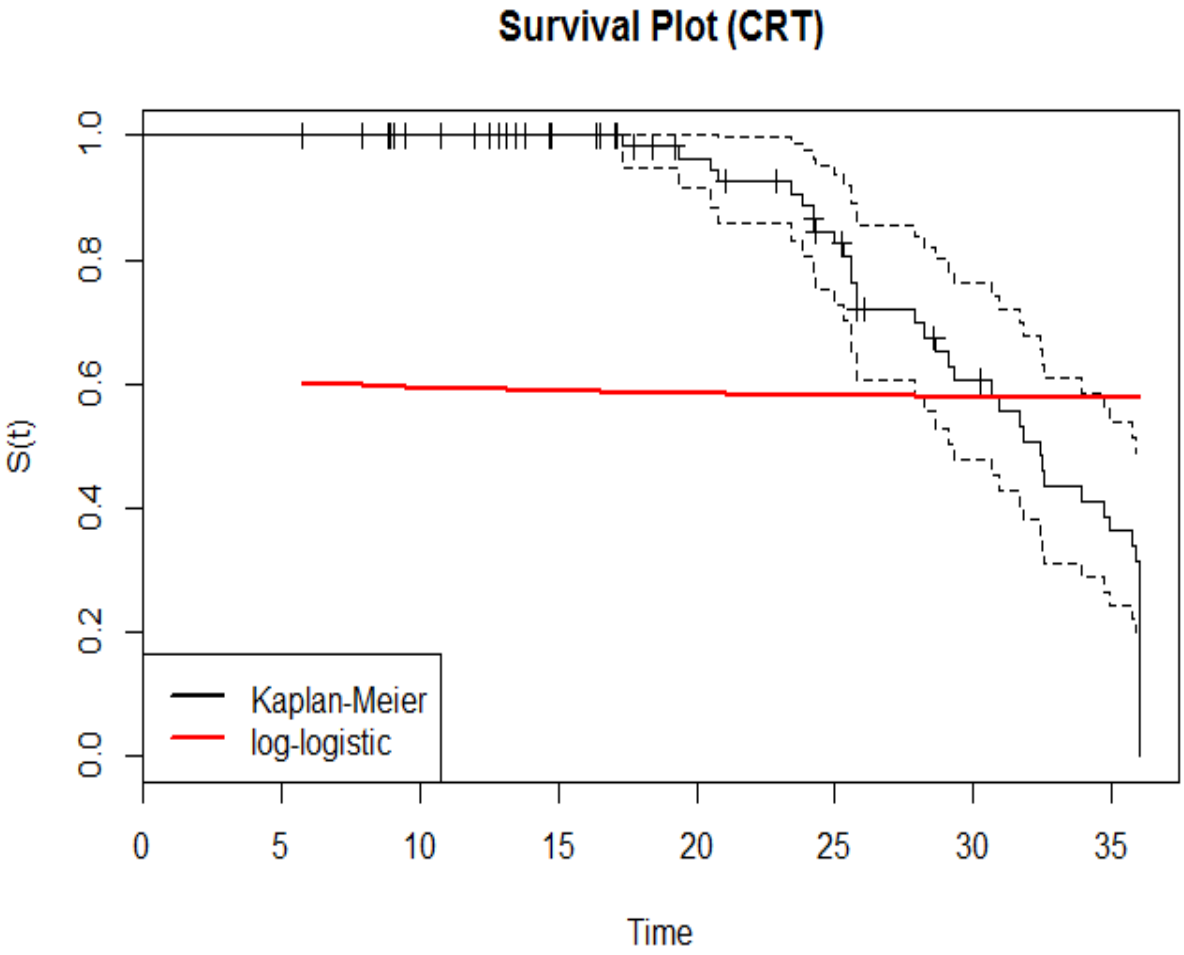


Fig 6: Survival Plot of log logistic Cure model (CRT)

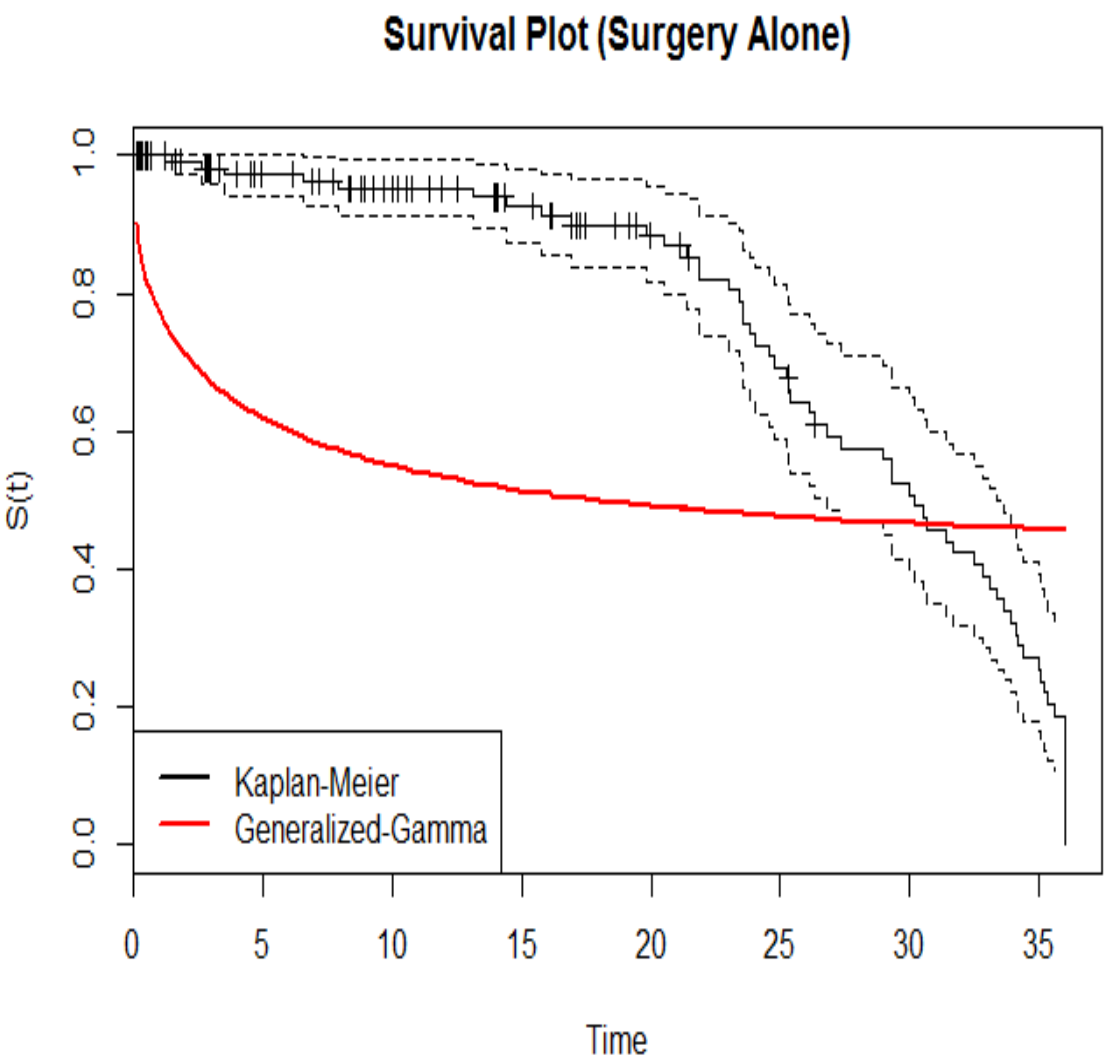


Fig 7: Survival Plot of Generalized Gamma Cure model (Surgery)

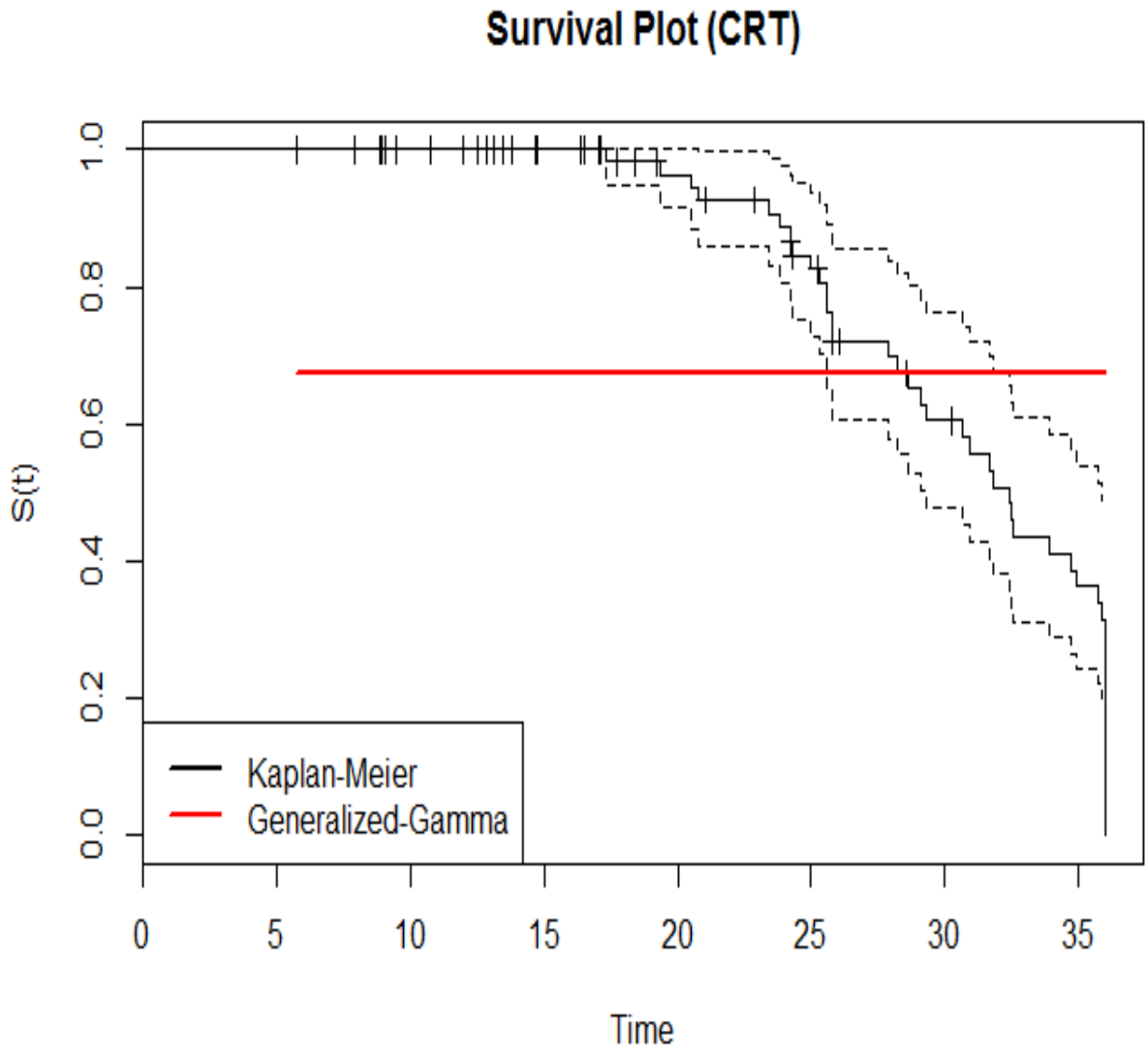


Fig 8: **Survival Plot of Generalized Gamma Cure model (CRT)**

Table 2: Comparison of Model evaluation

Model evaluation(Surgery)					
Model	AIC	Means	Variance	e^{μ}	C
Lognormal	570.7455	2.4705506	1.2957284	11.82902	0.3253112
Weibull	560.2464	0.6285831	0.0291504	1.874952	0.3710059
Log logistic	525.8647	0.5287214	0.0017197	1.696761	0.1219107
Gen-Gamma	544.3250	0.6168598	0.0197841	1.853098	0.374726
Model evaluation(CRT)					
Model	AIC	Means	Variance	e^{μ}	c
Lognormal	311.1869	2.847263	0.4839343	17.240759	0.5103684
Weibull	313.1283	0.6707620	0.0194086	1.955727	0.543705
Loglogistic	272.6705	0.5828985	0.0000403	1.791227	0.1227404
Gen-Gamma	16154.7444	0.5998050	0.0000000	1.821763	0.5998050

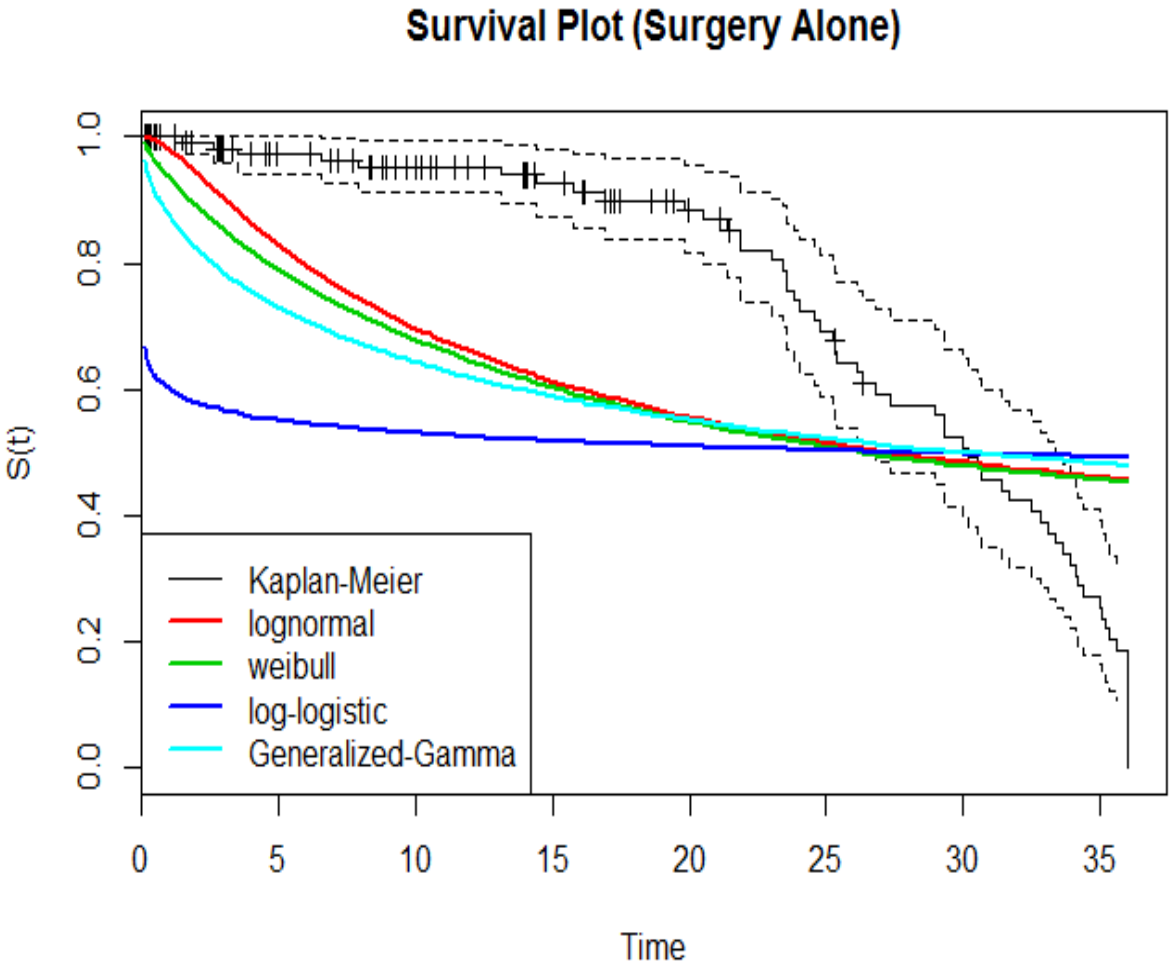
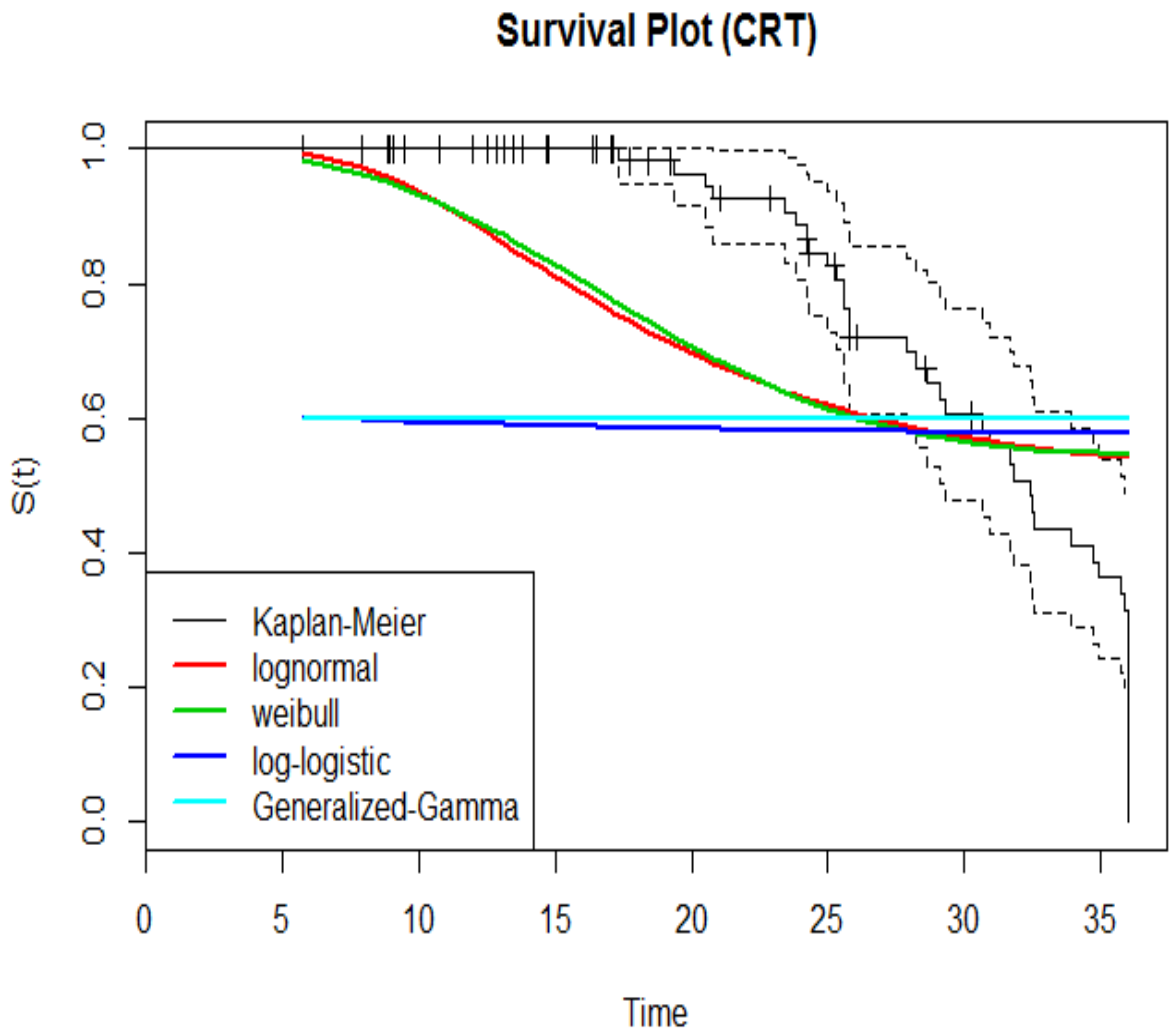

Survival Plot (Surgery Alone)
S(t)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
Time
Kaplan-Meier
lognormal
weibull
log-logistic
Generalized-Gamma

Fig 9: Survival Plots of Examined Cure Models (Surgery-Alone)



FiF

Fig. 10: Survival Plots of Examined Cure Models (CRT-Alone)

5.1 Conclusion

This study has been able to showed that generalized gamma model , gives the highest cure fraction (c), thus, it's indicated that this model has the highest proportion to cure for both surgery and chemotherapy for the data analyzed. it was discovered that from the four examined parametric cure models for both surgery and chemotherapy, log-logistic gives the minimum value for AIC , minimum means as well as cure fraction (c) which is the proportion of cure patients that we are interested in and minimum mean time to cure. The result also showed that generalized gamma model , gives the highest cure fraction (c), thus, it's indicated that this model has the highest proportion to cure for both surgery and chemotherapy for the data analyzed. From the summary of the results, we can conclude that log logistics is the flexible best model that explained the gastric cancer used for the study in terms of AIC and means and mean time to cure. But using the value of c, generalized gamma model has the highest proportion of cure which we can conclude that the model is preferred and could be the best among the considered parametric cure model.

Direction for Further Research

It was observed that the four parametric cure models were used on mixture cure model only and covariates were not included in the model.

Further research work will focus on the extensions of the flexible model which will be able to account for the non-mixture models.Female genital cancer from UCH Nigeria will be used. Thus, predictor's variable will be included in the model.

References

[1] Abu Bakar, M. R., Salah, K. A., Ibrahim, N. A., Haron, K. (2009). Bayesian Approach for Joint Longitudinal and Time-to-Event Data with Survival Fraction. Bull. Malays.Math. Sci. Soc. 32, 75-100.

[2] Berkson J. and Gage R.P. (1952). Survival curve for cancer patients following treatment.*Journal of the American Statistical Association*, 47:501–515.

[3] Betensky, R. A. and Schoenfeld, D.A. (2001). Nonparametric Estimation in a cure model with Random cure times. *Biometrics* **57**, 282-286.

[4] Boag J.W (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):15–53, 1949.

[5] Chao Cai (2013).Advanced Methodology Developments in Mixture Cure Models *University of South Carolina*

[6] Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. Journal of the American Statistical Association, 94(447), 909–919.

[7] Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. (2007). Flexible cure rate modelling under latent activation schemes. J. Amer. Statist. Assoc. 102 560–572.

[8] Fauzia Taweab, Noor Akma Ibrahim and Jayanthi Arasan (2015). A Bounded Cumulative Hazard Model with A change- Point According to a Threshold in a covariate for Right-Censored Data.

[9] Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). Bayesian survival analysis. Springer Series in Statistics. Springer-Verlag, New York.

[10] Lambert P, Dickman P, Osterlund P, et al. (2007).Temporal trends in the proportion cured for cancer of the colon and rectum: a population based study using data from the finish cancer registry. Int J Cancer;121:2052–9.

[11] Lambert P. (2007).Modeling of the cure fraction in survival studies. Stata J;7:1–25

[12] Lambert, P.C., Thompson, J.R. and Weston, C.L. (2007). Estimating and modeling the cure fraction in population based cancer survival analysis, Biostatistics, 8, 576-594

[13] Ortega Edwin M. M., Barriga Gladys D. C., Hashimoto Elizabeth M, Cancho Vicente G. and Cordeiro Gauss M (2014). A New Class of Survival Regression Models with Cure Fraction.

[14] Rodrigues, J., Cancho, V.G., de Castro, M., Louzada-Neto, F., (2009). On the unification of the long-term survival models. Statistics and Probability Letters 79,753–759.

[15] Sposto R. (2002) Cure model analysis in cancer: an application to data from the Children's Cancer Group. Stat Med;21:293–312.

[16] Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. J. Amer. Statist. Assoc., 98(464), 1063– 1078.5

[17] Wienke, A., Lichtenstein, P., and Yashin ,A.I (2003). A Bivariate frailty model with a cure fraction for modeling familial correlations in diseases .*Biometrics* 59, 1178-1183.

[18] Yakovlev, A.Y.(1994). Parametric versus nonparametric methods for estimating cure rates based on censored survival-data. Statistics in Medicine 13 (9), 983–985.

[19] Yakovlev, A.Y., Tsodikov, A.D. (1996). Stochastic Models of Tumor Latency and Their Biostatistical Applications. World Scientific, Singapore.

[20] Yakovlev, A.Y., Tsodikov, A.D., Bass, L., (1993). A stochastic-model of hormesis. Mathematical Biosciences 116 (2), 197–219.

[21] Yin, G., Ibrahim, J.G., (2005). Cure rate models: a unified approach. The Canadian Journal of Statistics 33 (4), 559–570.

[22] Zhao, G.M.A.((2008). Nonparametric and Parametric Survival Analysis of Censored Data

with Possible Violation of Method Assumptions.
[23] Yi Li and Ram C. Tiwari (2007) Mixture cure survival models with dependent censoring.