Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(3):325-335

Original Article

Analysis of Copula Frailty Defective Models in Presence of Cure Fraction

Ola Abuelamayem*

Department of Statistics, Faculty of Economics and Political Science, Cairo University, Giza, Egypt.

ARTICLE INFO

Received 16.02.2023
Revised 03.03.2023
Accepted 14.04.2023
Published 15.09.2023

Key words:
Cure fraction model;
Copula;
Defective distribution;
Frailty model;
Censoring;
Maximum likelihood estimation.

ABSTRACT

Introduction: Analyzing long term survivors such as diabetic patients can't be done using the usual survival models. One approach to analyze it is using defective distribution that doesn't force a pre-assumption of cure fraction to the model. To study more than one random variable interacting together, multivariate distributions may be used. However, most of multivariate distributions have complicated forms, which make the computations difficult. Besides, it may be hard to find a multivariate distribution that fits the data properly, especially in health care field. To get over this problem, one can use copula approach. In literature, to the best of our knowledge, only one paper handled copula defective models and didn't consider the effect of covariates. In this paper, we take into consideration not only existed covariates but also unobserved ones by including frailty term.
Methods: Two new models are introduced. The first model, used Gumbel copula to take the dependence into consideration together with the observed covariates. The second one take into consideration not only the dependence but also the unobserved covariates by integrating frailty term in to the model.
Results: A diabetic retinopathy data is analyzed. The two models indicated the existence of long-term survivals through negative parameters without the need of pre-assuming the existence of it. Including frailty term to the model helped in capturing more dependence between the variables. We compared the results using goodness of fit methods, and the results suggested that the model with frailty term is the best to be used.
Conclusion: The two introduced models correctly detected the existence of cure fraction with less estimated parameters than that in mixture cure fraction models. Also, it has the advantage of not pre-assuming the existence of cure fraction to the model. comparing both models, the model with frailty term fitted the data better.

Introduction

Models in survival analysis are based on the assumption that all units in the study will face the event of interest. However, in some cases this assumption is violated. For example, in medical studies, some patients are cured and never face the recurrence of a certain disease like melanoma. In economics, unemployed person may never find a job. In finance, some banks may never face bankrupt. In demography, one may never get married and may never get

*.Corresponding Author: ola.abuelamayem@feps.edu.eg

a first child. (For more examples see, Amico and Keilegom).[1] The proportion of individuals that are susceptible to the event of interest are called cured. To analyze this type of data, different models that take into consideration the existence of cure portion are used.

One common approach is mixture cure fraction model. It was first introduced by Boag[2] to analyze breast cancer data. In this approach, the survival function is rewritten as $S(t)=p+(1-p)S^*(t)$, such that p is the cure proportion and $S^*(t)$ is the survival function of susceptible subjects. This cure model has since been applied by many using different distributions. For example, Farewell,[3] Yamaguchi,[4] Kannan et al,[5], Martinez et al,[6] Swain et al[7] and Omer et al[8] used Weibull, generalized Gamma, generalized exponential, generalized modified Weibull, generalized Gompertz and exponentiated Weibull exponential, respectively.

An alternative approach to handle cure fraction is to use defective distributions which naturally becomes a cure rate model when changing the usual domain of its parameters. Because of this change, the survival function converges to a value $p\epsilon(0,1)$ which is written as function in the estimated parameters of the distribution. Hence, it has the advantage of estimating fewer number of parameters. Also, there is no need for pre-assuming the existence of cure fraction in the model, the parameter estimates will tell whether or not there is a proportion of cured elements. One of the most commonly used defective distribution is Gompertz model, it was first introduced by Cantor and Shuster[9] to analyze survival time for leukemia patients. Gieser et al[10] extended the model to include covariates.

Other defective distributions were introduced. For example, Rocha et al[11] used the Marshall-Olkin class of distributions to generalize the defective Gompertz and defective inverse Gaussian distributions. Rocha et al[12] derived the Kumaraswamy Gompertz and Kumaraswamy inverse Gaussian distributions. Rocha et al[13] used the Marashall-Olkin family to introduce ten new defective distributions. Martinez and Achcar[14] presented the defective Dagum distribution. Hamdeni and Gasmi[15] introduced the Marshall–Olkin generalized defective Gompertz distribution.

In both cure fraction approaches; one may be interested in analyzing two lifetime variables. Assuming independence between the variables are not always realistic. For example, studying of blindness in the left and right eye, analyzing the failure time of the left and right kidney and studying the lifetime of a twine-engine plane. To account for dependency between different lifetime variables, one may use multivariate lifetime distributions. However, most of multivariate lifetime distributions have complicated forms. This, make it difficult to find a proper distribution to fit the data. Also, this complexity makes the computations difficult. To overcome this disadvantage, copula approach can be used. The copula is a function that links the multivariate distributions to their one-dimensional marginal distributions through a link function. It has the advantage of combining both ease of computation and proper fit.

A generalization for the univariate mixture cure fraction model using copula is introduced. For example, Martinez and Arachar[16] considered the Farlie Gumbel Morgenstern, Clayton and Gumbel Barnett copulas to study invasive cervical cancer data.

Archar et al[17] considered Farlie Gumbel Morgenstern and Gumbel-Barnett copulas to

analyze blindness data. Coelho-Barros et al[18] applied bivariate Weibull using two different copulas to analyze diabetic retinopathy study. Peres et al[19] applied a comparison study using fifteen different copulas and three real datasets. For defective models, recently, Peres et al[20] used Clayton copula with defective Gompertz distribution to analyze three different datasets.

Handling dependent lifetime data using copula have an implicit assumption of homogeneity between individuals. In statistical analysis, part of the heterogeneity is explainable in terms of observed covariates. However, individuals with same covariates can still have different responses, this is due to some unobserved factors like environmental or biological factors. Models which take into account the unobserved heterogeneity between individuals, are known as Copula frailty models. These models use a random variable that represents the non-observed information. See for example, Wang et al[21] and Lin et al[22].

For defective models, recently Peres et al[20] considered the case of dependent lifetime models through copula. However, they didn't take into consideration the effect of covariates. Here, we present two models, one that takes into consideration the effect of observed covariates. The second one considers the heterogeneity effect through frailty term. Both models take into consideration the association between lifetime variables through copula. Also, they detect the existence of cure fraction without pre-assuming this by using defective distributions.

Methods

We introduced two new models, one that take into consideration the effect of observed covariates. The second one considers the effect of non-observed covariates through frailty term.

We will first explain the meaning of defective distributions and copula frailty models. Then illustrate the derivation of the two introduced models.

Defective Distributions

A distribution is called defective if the survival function approaches a value $p \in (0,1)$ when we change the domain of its parameters. It has the advantage of not pre-assuming the existence of cure fraction in the model. If the estimation procedure presents a value out of the usual range of parameters, then the cure fraction exists. The proportion of the cured subjects is obtained by calculating the limit of survival function using the estimated parameters. Accordingly, defective distributions have fewer number of estimated parameters compared to that of mixture cure fraction models where p is directly estimated.

Gompertz distribution is one of the most commonly used defective model. It was introduced by Cantor and Shuster[9] to analyze medical data. It was extended by Gieser et al[10] to take into consideration the effect of covariates. Dos-Santos et al[23] estimated the parameters using Bayesian approach and compared it with maximum likelihood approach. The Gompertz distribution has a simple form with two positive parameters. For negative values of the shape parameter, the distribution becomes defective. The probability density, survival and hazard functions for the Gompertz distribution are, respectively, given by

$$f(t)=be^{\alpha t}e^{\frac{-b}{\alpha}(e^{\alpha t}-1)},\alpha>0,b>0$$

$$S(t)=e^{\frac{-b}{\alpha}(e^{\alpha t}-1)},$$

$$h(t)=be^{\alpha t}.$$

The distribution becomes defective for negative values of the parameter α. The cure proportion is calculated as follows

$$\lim_{t\to\infty}S(t)=\lim_{t\to\infty}e^{\frac{-b}{\alpha}(e^{\alpha t}-1)}=e^{\frac{b}{\alpha}}=p$$

This form of the Gompertz distribution handles only the situation when one random variable is of interest or independence between events of interest is assumed. Once dependency between two lifetime events is considered, bivariate distributions are used. Copula is one of the most widely used approaches to construct bivariate distributions taking into consideration the association between events. This will be explained in more details in the next section.

Copula frailty

Copula is a function that connects marginal distributions to define a bivariate distribution. There are two main steps to define the required bivariate distribution. First, the marginal distributions should be properly defined. Second, select a suitable copula to define the dependence structure. Nelsen[24] illustrated that different copulas with the same marginals resulted in different dependence structure.

There are different types of copulas, the most commonly used one is Archimedean copula. This is due to a number of reasons, the ease with which they can be constructed, the great variety of families of copulas and many useful properties possessed by the members of this class. (For more details see, Nelsen[24]).

By using copula, one take into consideration the dependence between different lifetime variables. Also, for a more precise analysis of the survival function, one needs to take into consideration the effect of other related factors by encountering observed covariates (x) into the model. However, we may have two individuals with the same observed covariate but with different lifetimes. This is due to the effect of some unobserved factors like environmental and biological ones. This can be taken into consideration by adding frailty term into the model which can be explained as follows

$$S(t_1,t_2|u,x_j)=C_\theta[S_{T_1}(t_1|u,x_1),S_{T_2}(t_2\,|\,u,x_2)],$$

where

$$S_{T_j}(t_j|u,x_j)=e^{-uR_0(t_j)e^{\beta_j' x_j}},\ j=1,2.$$

C_θ: Is the selected copula function,

$R_0(t)$: Is the cumulative hazard function of the selected marginal.

u: Is the frailty term.

For defective distributions, to the best of our knowledge, only peres et al[20] considered the dependency between time to events through copula. However, they didn't take into consideration neither the observed covariates nor the frailty effects. In the next section, we will generalize the model to take both effects into consideration but using different Archimedean copula.

Model

Two models are developed to generalize the existing bivariate defective model in the literature. One with only observed covariates and the other including both unobserved and observed. Gumbel copula and marginal

survival function from Gompertz distribution are considered in both derived models.

Copula model without frailty

We used in our model Gumbel copula with the following formula

$$C_\theta(s,v)=\exp\left\{-\left[\left(-\ln(s)\right)^\theta+\left(-\ln(v)\right)^\theta\right]^{\frac{1}{\theta}}\right\},$$

where
θ: dependence parameter.
Now, we rewrite our model in terms of Gumbel copula

$$S(t_1,t_2|x_1,x_2)=$$
$$\exp\left\{-\left[\left(-\ln S_{T_1}(t_1|x_1)\right)^\theta+\left(-\ln S_{T_2}(t_2|x_2)\right)^\theta\right]^{\frac{1}{\theta}}\right\},$$

where (under Gompertz distribution)

$$S_{T_j}(t_j|x_j)=e^{-R_0(t_j)e^{\beta_j x_j}},\ j=1,2$$

and

$$R_0(t)=\frac{b}{\alpha}\left[e^{\alpha t}-1\right].$$

However, Gompertz model has an identifiability problem between parameters b and e^{β}. To overcome this problem, we set b=1. (For more details, see Scudilio et al[25]).
In survival analysis, failure time for some units are not observed. This portion is referred to as censored observations. Here, we considered right random censoring. To take into consideration the censoring effect, the likelihood function for n observations has the following form

$$L=\prod_{i=1}^{n}\left[f(t_1,t_2|x_1,x_2)\right]^{\delta_{1i}\delta_{2i}}\left[\frac{-\partial S(t_1,t_1|x_1,x_2)}{\partial t_1}\right]^{\delta_{1i}(1-\delta_{2i})}$$
$$\left[\frac{-\partial S(t_1,t_1|x_1,x_2)}{\partial t_2}\right]^{\delta_{2i}(1-\delta_{1i})}$$
$$*S(t_1,t_2\mid x_1,x_2)^{(1-\delta_{1i})(1-\delta_{2i})}$$

where

$$\delta_{1i}=\begin{cases}1 if\ T_j<t_j\\ 0 otherwise\end{cases},\ j=1,2.$$

$$f(t_1,t_2|x_1,x_2)=\frac{\partial^2 S(t_1,t_2\mid x_1,x_2)}{\partial t_1\,\partial t_2}=e^{\beta_1 x_1+\beta_2 x_2+\alpha_1 t_1+\alpha_2 t_2}$$
$$\left(\alpha_1^{-1}e^{\beta_1 x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta-1}$$

$$*\left(\alpha_2^{-1}e^{\beta_2 x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta-1}e^{-k^{\frac{1}{\theta}}}\left\{(\theta-1)k^{\frac{1}{\theta}-2}+k^{\frac{2}{\theta}-2}\right\}$$
$$\frac{-\partial S(t_1,t_1|x_1,x_2)}{\partial t_1}=e^{\beta_1 x_1+\alpha_1 t_1}\left(\alpha_1^{-1}e^{\beta_1 x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta-1}e^{-k^{\frac{1}{\theta}}}k^{\frac{1}{\theta}-1}$$
$$\frac{-\partial S(t_1,t_1|x_1,x_2)}{\partial t_2}=e^{\beta_2 x_2+\alpha_2 t_2}\left(\alpha_2^{-1}e^{\beta_2 x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta-1}e^{-k^{\frac{1}{\theta}}}k^{\frac{1}{\theta}-1}$$
$$k=\left(\alpha_1^{-1}e^{\beta_1 x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta}+\left(\alpha_2^{-1}e^{\beta_2 x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta}$$

We used Kendall's tau to measure the dependency between T_1 and T_2, it has the following formula

$$\tau_\theta=\frac{\theta-1}{\theta}.$$

Copula model with frailty

To allow for heterogeneity, we encountered unobserved frailty term u into our statistical model. Here, we consider gamma distribution with the following probability density function

$$f(u)=\frac{1}{\Gamma\left(\frac{1}{\eta}\right)\eta^{\frac{1}{\eta}}}u^{\frac{1}{\eta}-1}e^{-\frac{u}{\eta}},\ \eta>0,u>0.$$

such that η represents the degree of heterogeneity.
The combination of Gumbel Copula and

gamma frailty is selected to reach a closed formula after integrating out the frailty term (see, Wang et al[21] for more details). Now, we write our model in terms of Gumbel copula after adding the frailty term

$$S(t_1,t_2|u,x_1,x_2)=$$
$$\exp\left\{-\left[\left(-\ln S_{T_1}(t_1|u,x_1)\right)^{\theta}+\left(-\ln S_{T_2}(t_2|u,x_2)\right)^{\theta}\right]^{\frac{1}{\theta}}\right\}, \quad (2)$$

where (under Gompertz distribution)

$$S_{T_j}(t_j|u,x_j)=e^{-u\,R_0(t_j)e^{\beta_j' x_j}},\ j=1,2 \quad \text{and}$$
$$R_0(t)=\frac{1}{\alpha}\left[e^{\alpha t}-1\right].$$

Model (2) is not applicable as it includes unobserved term (u). To get the joint survival function we integrate out the frailty term as follows

$$S(t_1,t_2|x_1,x_2)=$$
$$\left\{1+\eta\left[\left(\frac{1}{\alpha_1}e^{\beta_1' x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta}+\left(\frac{1}{\alpha_2}e^{\beta_2' x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta}\right]^{\frac{1}{\theta}}\right\}^{-\frac{1}{\eta}}.$$

Proof

$S(t_1,t_2|x_1,x_2)=\int_0^{\infty}S(t_1,t_2|u,x_j)f(u)du$, using change of variables technique, the result is reached.

Now, to account for censoring, we will derive the components of the likelihood function (1) as follows

$$f(t_1,t_2|x_1,x_2)=\frac{\partial^2 S(t_1,t_2\mid x_1,x_2)}{\partial t_1\,\partial t_2}=$$
$$e^{\beta_1' x_1+\beta_2' x_2+\alpha_1 t_1+\alpha_2 t_2}\left(\alpha_1^{-1}e^{\beta_1' x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta-1}$$
$$*\left(\alpha_2^{-1}e^{\beta_2' x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta-1}$$
$$*\left\{\left(1+\eta k^{\frac{1}{\theta}}\right)^{-\frac{1}{\eta}-1}(\theta-1)k^{\frac{1}{\theta}-2}+k^{2\left(\frac{1}{\theta}-1\right)}(1+\eta)\left(1+\eta k^{\frac{1}{\theta}}\right)^{-\frac{1}{\eta}-2}\right.,$$

$$\frac{-\partial S(t_1,t_1|x_1,x_2)}{\partial t_1}=e^{\beta_1' x_1+\alpha_1 t_1}\left(\alpha_1^{-1}e^{\beta_1' x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta-1}$$
$$\left(1+\eta k^{\frac{1}{\theta}}\right)^{-\frac{1}{\eta}-1}\ k^{\frac{1}{\theta}-1},$$

$$\frac{-\partial S(t_1,t_1|x_1,x_2)}{\partial t_2}=e^{\beta_2' x_2+\alpha_2 t_2}\left(\alpha_2^{-1}e^{\beta_2' x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta-1}$$
$$\left(1+\eta k^{\frac{1}{\theta}}\right)^{-\frac{1}{\eta}-1}\ k^{\frac{1}{\theta}-1},$$

$$k=\left(\alpha_1^{-1}e^{\beta_1' x_1}\left[e^{\alpha_1 t_1}-1\right]\right)^{\theta}+\left(\alpha_2^{-1}e^{\beta_2' x_2}\left[e^{\alpha_2 t_2}-1\right]\right)^{\theta}$$

We used Kendall's tau to measure the dependency between T_1 and T_2. Using the properties of Archimedean copula, it is found to be

$$\tau_{\theta,\eta}=1-\frac{2}{(\eta+2)\theta}.$$

Proof

$\tau_{\theta,\eta}=1+4\int\frac{\phi(t)}{\phi'(t)}dt$, using $\phi(t)=\left(t^{-\eta}-1\right)^{\theta}$.

matching the same result as Wang et al.[21]

To study the effect of including frailty term, we compared this model with the one using only observed covariates and the one with no covariates. We applied it into real dataset to test its applicability.

Here, we analyze a diabetic retinopathy data which was first introduced by Huster et al.[26] It consists of 197 diabetic patients. Each patient had one eye randomized to laser treatment

and the other eye received no treatment. In our analysis, T_1 corresponds to the time up to blindness for the control eye, and T_2 is the time up to blindness for treatment eye. Censoring was caused by death, dropout, or end of the study. Two covariates are included in the analysis

- x_1: age of diagnoses of diabetes.
- x_2: type of diabetes.

Results

The age at diagnosis of diabetes for patients ranges from 1 till 58 years. The mean time for control eye is 34.2 months with 43.6% censoring. While for treated eye, the mean time is 36.2 months with 35.02% censoring.

The Kaplan-Meier survival curve (Figure 1) has a long plateau at the end, indicating the possibility of cured fraction in the model. The two graphs for treated and non-treated groups time. However, the survival curve for treated eye is higher than that for non-treated one.

We used R package and nlm function to maximize the logarithm of the likelihood function (equation (1)) for both models. Model 1 (M1) corresponds to the one with only observed covariates, model 2 (M2) corresponds to the frailty one, while model 3 (M3) corresponds to the model without covariates or frailty term. Maximum likelihood estimates (ML) and 95% confidence intervals are illustrated in Table 1. It can be seen that the estimates for α_1 and α_2 are negative with upper confidence limit less than zero, accordingly all models suggest the existence of cure fraction.

Cure fractions for control eye (p_1) and treatment eye (p_2) are calculated using the following equation and presented in Table 2.

$$p_j = e^{\frac{1}{\alpha_j}}, \quad j = 1, 2.$$

It can be seen from Table 2, that models M1

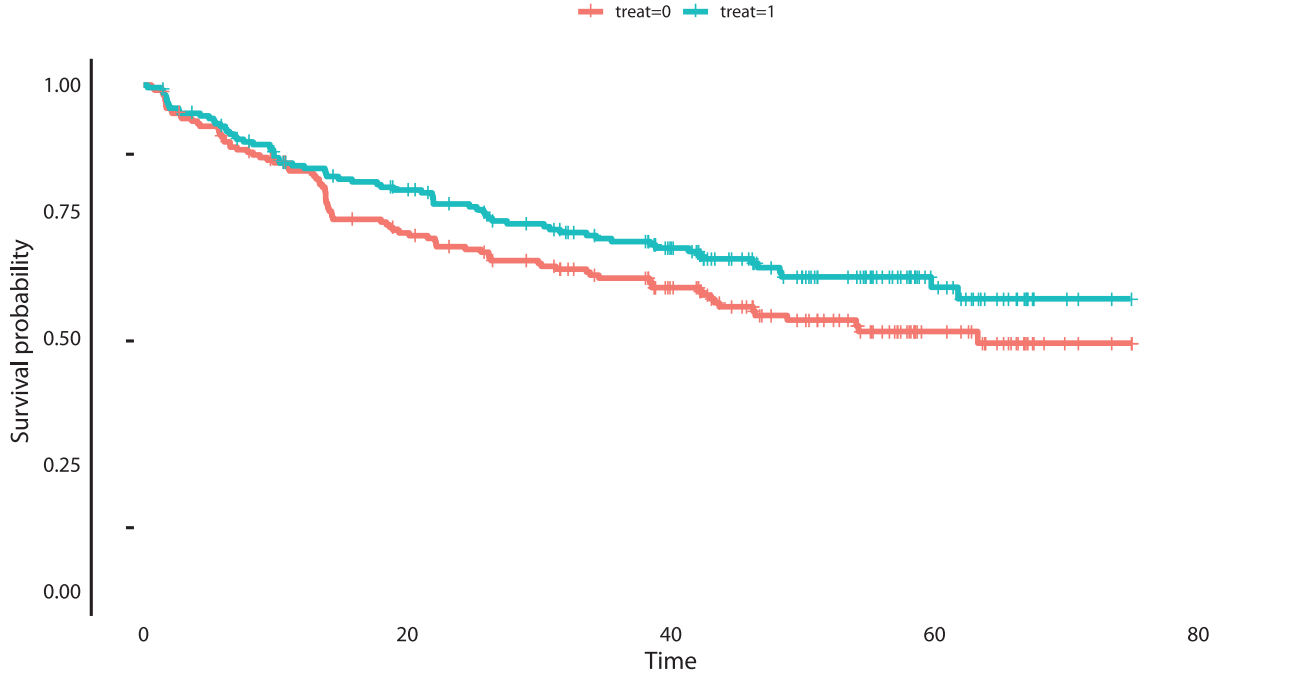

Figure 1. Kaplan-Meier survival curve for diabetic patients

around 0.5 for control eye, while model M_3 has lower estimate. All Models has higher estimate of cure fraction for treatment eye than that for control eye.

The heterogeneity variance is estimated as $\hat{\eta} = 0.58$ and the CI does not include zero which illustrates the significance of the frailty term. adding frailty term to the model resulted in capturing more association between T_1 and T_2 (higher value for Kendall's tau (Table 3)). This is logical because the data has no covariates for environmental conditions or hospitals the patients selected from.

The estimators β_{11} and β_{21} correspond to the age covariate for control eye and treatment eye, respectively. While, the estimators β_{12} and β_{22} correspond to type of diabetes for control eye and treatment eye, respectively. It can be seen that the 95% CI doesn't include zero, which illustrates that the covariates are significant in both models.

Comparing the models M_1 and M_2 using different information criteria. From Table 3, the results suggest that the model with frailty term provides a better fit (i.e. lower value for information criteria) for the data. Although Model M_3 has the lowest values, we don't recommend the use of this model. The assumption of not adding covariates is unrealistic.

Discussion

Analyzing data with long term survivors is of great importance. The usual survival data with the hidden assumption of the occurrence of the event of interest to all study units can't be used. From previous studies, some diseases like

Table 1. The maximum likelihood estimates

		α_1	α_2	θ	η	β_{11}	β_{12}	β_{21}	β_{22}
M_1	ML	-1.752	-2.078	1.170	-	0.918	0.274	0.548	0.285
	CI	(-2.26, -1.24)	(-2.71, -1.44)	(1.02,1.32)	-	(0.54,1.30)	(0.004,0.54)	(0.15,0.95)	(0.008,0.56)
M_2	ML	-1.478	-1.664	1.031	0.578	0.814	0.375	0.570	0.346
	CI	(-1.91,-1.05)	(-2.12,-1.21)	(0.87,1.19)	(0.16,0.99)	(0.34,1.29)	(0.04,0.71)	(0.09,1.05)	(0.01,0.68)
M_3	ML	-0.699	-1.489	1.139	-	-	-	-	-
	CI	(-1.34, -0.06)	(-2.24, -0.74)	(1.02, 1.26)	-	-	-	-	-

Table 2. Cure fraction estimates

		P_1	P_2
M_1	ML	0.565	0.618
M_2	ML	0.508	0.548
M_3	ML	0.239	0.511

Table 3. Information Criteria and Kendal's tau

Model	AIC	BIC	CAIC	HQIC	τ
M_1	-822.50	-796.23	-796.19	-825.18	0.145
M_2	-824.20	-797.94	-797.90	-826.89	0.248
M_3	-829.37	-819.52	-819.50	-830.37	0.122

diabetic, cancer, melanoma and Covid 19 were most likely to include long term survivors. Accordingly, accurate analysis with minimum number of parameters for this data type is very important. Our aim in this study was to accurately detect the existence of long-term survivors through the data without pre-forcing its existence to the model, and examining the effect of covariates by considering the association between events.

The study conducted by Peres etal[20] analyzed long term survivors using copula approach and defective distributions. However, they ignored the effect of covariates on survival function.

In our study, we introduced two models both taking into consideration the dependence effect using copula. One that take into consideration the effect of observed covariates. The second model analyzed the effect of unobserved covariates by adding latency term to the model. The results of the study illustrated that both models correctly detect the existence of cure fraction in the model. The one with latency term provided better fit for the data.

It is suggested for further studies to take into consideration any pre-existing information about the parameters. This can be done using Bayesian technique.

Conclusion

In this paper, we introduced two generalization for copula defective model existed in literature. The first model takes into consideration the effect of observed covariates. The second model accounts for unobserved covariates by adding frailty term. Both models detect the existence of cure fraction with less estimated parameters than that in mixture cure fraction models. Also, it has the advantage of not pre-assuming the existence of cure fraction to the model. comparing both models, the model with frailty fitted the data better.

Acknowledgments

Not applicable

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

1. Amico M, and Keilegom IV. Cure models in survival analysis . Annual Review of Statistics and Its Application. 2018; 5(1): 311-342.

2. Boag JW. Maximum likelihood estimates of patients cured by cancer therapy. Journal of the Royal Statistical Society. Series B. 1949; 11(1): 15-53.

3. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. Biometrics. 1982; 38(4): 1041-1046.

4. Yamaguchi K. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan. Journal of the American Statistical Association. 1992; 87(418): 284-292.

5. Kannan N, Kundu D, Nair P, Tripathi RC. The generalized exponential cure rate model with covariates. Journal of Applied Statistics. 2010; 37(10): 1625-1636.

6. Martinez EZ, Achcar JA, Jacome AA, Santos JS. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application on gastric cancer data. Computer Methods And Programs In Biomedicine. 2013; 112(3): 343-355.

7. Swain PK, Grover G, Goel K. Mixture and non mixture cure fraction models based on generalized Gompertz distribution under Bayesian approach. Tatra Mountains Mathematical Publication. 2016; 66(1): 121-135.

8. Omer M, Abu Bakar M, Adam M, Mustafa M. Cure models with exponentiated Weibull exponential distribution for the analysis of melanoma patients. Mathematics. 2020; 8(11): 1-15.

9. Cantor AB, Shuster JJ. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Statistics in Medicine. 1992; 11(7), 931–937.

10. Gieser PW, Chang MN, Rao PV, Shuster JJ, Pullen J. Modelling cure rates using the Gompertz model with covariate information. Statistics in Medicine. 1998; 17(8): 831–839.

11. Rocha R, Nadarajah S, Tomazella V, Louzada F. Two new defective distributions based on the Marshall–Olkin extension. Lifetime Data Analysis. 2016; 22: 216-240.

12. Rocha R, Nadarajah S, Tomazella V, Louzada F, Eudes A. New defective models based on the Kumaraswamy family of distributions with application to cancer data sets. Statistical Methods in Medical Research. (2017 a); 26(4): 1737-1755.

13. Rocha R, Nadarajah S, Tomazella V, Louzada F. A new class of defective models based on the Marshall–Olkin family of distributions for cure rate modeling. Computational Statistics and Data Analysis, (2017 b); 107, 48–63.

14. Martinez EZ, Achcar JA. A new straightforward defective distribution for survival analysis in the presence of a cure fraction. Journal of Statistical Theory and Practice. 2018; 12: 688–703.

15. Hamdeni T, Gasmi S. The Marshall–Olkin generalized defective Gompertz distribution for surviving fraction modeling. Communications in Statistics - Simulation and Computation. 2020; 51(11): 6511-6524.

16. Martinez EZ, Achcar JA. Bayesian bivariate generalized Lindley model for survival data with a cure fraction. Computer Methods and Programs in Biomedicine. 2014; 117(2), 145-157.

17. Archar JA., Martinez EZ, Tovar C, Rafael J. Bivariate lifetime modelling using copula functions in presence of mixture and non-mixture cure fraction models, censored data and covariates. Model Assisted Statistics and Applications. 2016; 11(4), 261-276.

18. Coelho-Barros EA, Archar JA., Mazucheli J. Bivariate Weibull distributions derived from copula functions in the presence of cure fraction and censored data. Journal of

Data Science. 2016; 14(2), 295-316.

19. Peres MV, Achcar JA, Martinez EZ. Bivariate lifetime models in presence of cure fraction: a comparative study with many different copula functions. Heliyon. 2020; 6(6), e03961.

20. Peres MV, Oliveira RP, Archcar JA, Martinez EZ. The bivariate defective Gompertz distribution based on Clayton copula with applications to medical data. Austrian Journal of Statistics. 2022; 51: 144-168.

21. Wang Y, Emura T, Fan T, Lo S, Wilke R. Likelihood-based inference for a frailty-copula model based on competing risks failure time data. Quality and Reliability Engineering International. 2020; 36(5): 1622-1638.

22. Lin Y, Sun L, Tseng Y, Emura T. The Pareto type I joint frailty-copula model for clustered bivariate survival data. Communication in Statistics- Simulation and Computation. 2022; DOI: 10.1080/03610918.2022.2066694.

23. Dos-Santos MR, Achcar JA, Martinez EZ. Bayesian and maximum likelihood inference for the defective Gompertz cure rate model with covariates. Cieˆncoa Natura. 2017; 39 (2): 244-258.

24. Nelsen RB. An Introduction to Copulas. 2^nd ed. New York: springer; 2006.

25. Scudilio J, Calsavara V, Rocha R, Louzada F, Tomazella V, Rodrigues A. Defective models induced by gamma frailty term for survival data with cured fraction. Journal of Applied Statistics. 2019; 46(3): 484-507.

26. Huster WJ, Brookmeyer R, Self SG. Modeling paired survival data with covariates. Biometrics. 1989; 45: 145-156.