ORIGINAL ARTICLE e-ISSN: 2349-0659 p-ISSN; 2350-0964

Flexible Parametric Survival Cure Rate Models for Pulmonary Tuberculosis Data

B. Vijai[1], P. R. Jayashree[2], C. Ponnuraja[3]

Abstract

This article mainly aims to compare Flexible parametric cure rate models using relative survival function and to predict cure fraction for tuberculosis (TB) data. In survival analysis, the Cox proportional-hazards model of time-to-event data is effective, but still there may be some benefits of using parametric models than non-parametric or semi-parametric models. Sometimes, it happens that a certain fraction of the data corresponds to subjects who are never involved in the event when assessing time-to-event data. Survival models that take this characteristic into account are typically referred to as cure rate models. Hence, in this article the parametric cure model to time-to-event (sputum conversion) on pulmonary TB data with the survival time distribution such as Weibull, Gamma, Exponential and Lognormal is developed. The objective of this article is to compare cure rate models to find the best model fitting survival time using the relative survival function and to predict cure fraction of TB data. The data were analyzed using "R-4.0.2" and STATA 15.0.0 statistical tools.

Keywords: Cure fraction, Cure rate model, Flexible parametric survival model, Relative survival function, Survival time distribution

2010 Mathematics Subject classification: 62P10, 62N02

Asian Pac. J. Health Sci., (2022); DOI: 10.21276/apjhs.2022.9.3.35

Introduction

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. Time can be days, weeks, months or years and event can be death, recurrence, and incidence relapse from remission. The term survival analysis pertains to a statistical approach to take into account the amount of time an experimental unit contributes to a study. It is the study of time between entry into observation and subsequent event.

In time-to-event research, cure models are used when not all people are expected to undergo the event of interest, or when the survival of the considered people exceeds the same level as the general population. The cure rate parametric models are a type of parametric model of survival in which a proportion of study participants or patients who will not be experiencing the defined event is presumed to occur. In a mixture cure model, these "cured" and "uncured" subjects are modelled separately with the cured individuals subject to no excess risk and the uncured individuals subject to excess risk modelled using a parametric survival distribution. A parametric survival distribution is scaled in a non-mixture model so that survival reaches the cure fraction asymptotically. Using parametric cure models, certain key points must be taken into account: the functional form of the "uncured" survival must be specified; adequate survival functions must be equipped to catch high excess hazard during the initial time of diagnosis; skewed estimates must be avoided to converge when the cure ratio is 80 percent and above. Computational difficulties need to be taken into account, especially when using the gamma distribution, make less distributional assumptions.[1]

Comprehensively applying flexible parametric survival models mainly to estimate the cure proportion and for the survival of the "not cured" or "not free from disease" or "uncured" as "not restored to health" in a population could potentially solve these needs.[2] Flexible models of the parametric approach of survival for cure rate approaches were introduced and were further protracted systematically on the way to relative survival.[3,4]

[1]Department of Statistics, DRBCCC Hindu College, Chennai, Tamil Nadu, India

[2]Department of Statistics, Presidency College, Chennai, Tamil nadu, India

[3]ICMR, Chetpet, Chennai, Tamil Nadu, India.

Corresponding Author: B. Vijai, Research Scholar, Assistant Professor, DRBCCC Hindu College, Chennai. E-mail: vijai.stats@gmail.com

How to cite this article: Vijai B, Jayashree PR, Ponnuraja C. Flexible Parametric Survival Cure Rate Models for Pulmonary Tuberculosis Data. Asian Pac. J. Health Sci., 2022;9(3):175-180.

Source of support: Nil

Conflicts of interest: None.

Received: 03/12/2021 **Revised:** 04/01/2022 **Accepted:** 12/02/2022

Materials and Methods

A total of 412 Tuberculosis (TB) patients were used which is recognized by the National Institute for TB Research-Indian Council of Medical Research, Chennai, India, to test the model and the data linked to Randomized Clinical Trial (RCT) time to data on TB sputum conversion. Accordingly, in this model, age with three categories is used for the study.

Cure rate models are survival models consisting of a cure fraction and an uncured fraction. Cure rate models to estimate the cure fraction was first developed for non-mixture cure rate model.[5] Later for mixture model were developed and it is also known as the standard cure rate model.[6] There are two types of cure rate model, mixture, and non-mixture. In the present study, the non-mixture parametric cure rate models were fitted and compared.

Relative Survival Function

The performance form of choice for researching survival of patient, especially in a population-based background is relative survival S_R (t).[7] It is defined as observed (all-cause) survival or subsistence overall survival S (t) in the patients divided by expected survival

$S^*(t)$ in the theoretical common population demographic group equivalent to patients with age, gender and possible other covariates. The relative survival model is possible to write the overall survival as

$$S(t) = S^*(t)\, S_R(t) \tag{1}$$

The chance analog of relative survival is the excess hazard rate. The diagnosis of TB among patients is associated with two elements, namely, expected hazard, $h^*(t)$ and the excess hazard $\lambda(t)$. Thus, the risk exposure or hazard $h(t)$ is given by

$$h(t) = h^*(t) + (t) \tag{2}$$

Here $S(t)$ and $h^*(t)$are believed to be recognized and are typically collected from routine data sources. One of the most widely used cure models in population-based cancer research is the mixture cure model.[8] The mixture cure model's overall survival feature incorporating relative survival can be written as

$$S(t) = S^*(t)(\pi + (1-\pi)S_u(t)) \tag{3}$$

It means that some proportion of patients, π, are treated as "cured" or completely recovered, while "uncured" or not recovered is the remainder, 1 - π. $Su(t)$ is the "uncured" survival attributes unique to the disease and is estimated by the model along with the cure proportion. There is a need to choose a parametric distribution for $Su(t)$ and a Weibull distribiution is sometimes used.[9]

Another parametric cure model used in population-based cancer research is the non-mixture cure model, which estimates the asymptotic survival function as the cure ratio. The survival function for the non-mixture model is given as

$$S(t) = S^*(t)\ \pi^{F_Z(t)} \tag{4}$$

where $F_Z(t)$ is a distribution function,a Weibull distribution can be used. The non-mixture model can be written as a mixture model as

$$\mathrm{S(t)} = S^*(t)\left(\pi + (1-\pi)\right) \tag{5}$$

This helps to measure both the cure ratio and the survival of the "uncured" Both the cure proportion and the parameters in $S_u(t)$ or $F_Z(t)$ can be allowed to differ by covariates when modelling.

Flexible Parametric Survival Model

The versatile parametric survival model is built on the log cumulative excess hazard scale.[10] Using limited cubic splines to estimate the baseline cumulative excess hazard can be found by integrating equation (2) and is represented as

$$\mathrm{H(t)} = H^*(\mathrm{t}) + \Lambda(\mathrm{t}) \tag{6}$$

where $H(t)$ is the total cumulative hazard, the cumulative hazard predicted is $H^*(t)$ and the cumulative excess hazard is $\Lambda(t)$. The reason for modelling on the log cumulative excess hazard scale instead of the log excess hazard scale is that a more stable function is the log cumulative excess hazard and its form is easier to capture. To model the accumulated excess hazard on the log scale,

$$ln(\Lambda(t)) = ln(-lnS\ (t)) = s(x\ ;\ \ ) \tag{7}$$

where $x = ln(t)$ and $s(x;\gamma_0)$ is a restricted cubic spline function, defined as

$$s(x\ ;\gamma_0) = \gamma_{00} + \gamma_{01}v_1(x) + \gamma_{02}v_2(x) + \ldots + \gamma_{0k-1}v_{k-1}(x) \tag{8}$$

where k is the number of knots and the j^{th} basis function is defined as

$$v_1(x) = x \tag{9}$$

$$v_j(\mathrm{x}) = (x-k_j)_+^3 - \lambda_j(x-k_1)_+^3 - (1-\ \lambda_j)(x-k_k)_+^3, j = 2, \ldots, k\ -1 \tag{10}$$

where k_1 and k_K the position of the first and last knot respectively and $\lambda_j = (k_K - k_j)/(k_K - k_1)$

All quadratic variables except the first one (v_1, the linear variable) are zero up to the first knot. Hence, the log cumulative excess hazard is forced to be linear before the position of the first knot. The introduction of covariates z to equation (7) can be expressed as

$$ln(\Lambda(t\ ;z)) = ln(-lnS_R(t\ ;z)) = s(x\ ;\gamma_0) + \beta^T z \tag{11}$$

This is a proportional model of excess hazards with time-dependent covariate effect model. These are extremely common and time-dependent effects typically do not need the baseline accumulated excess hazards. Fresh spline parameters are added for each time-dependent effect and different knot positions can be selected for a time-dependent effect for each new covariate z_i and thus equation (11) becomes

$$ln(\Lambda(t;z)) = s(x\ ;\gamma_0)\beta^T z + \sum_{i=1}^{D} s(x\ ;\gamma_i)z_i \tag{12}$$

where D is the number of time- dependent covariate effects and the quadratic function for the i^{th} time-dependent impact is $s(x;\square_i)$

Flexible Parametric Survival Cure Model

In the cure rate parametric model, the excess risk rate is zero when "cure" is achieved and the accumulated excess hazard will be constant after this point. The relative survival function from the versatile parametric survival model with backward splines calculated and the linear spline variable parameter restriction (γ_{01}=0) is defined as

$$S_R(t) = \exp(-\exp(\gamma_{00} + \gamma_{02}v_2(x) + \ldots + \gamma_{0k-1}v_{k-1}(x))) \tag{13}$$

which can be written as,

$$S_R(t) = \pi^{\exp(\gamma_{02}v_2(x) + \ldots + \gamma_{0k-1}v_{k-1}(x))} \tag{14}$$

where $\pi = exp(-exp(\gamma_{00}))$ In comparison to a non-mixture model, it is seen that the stable parametric cure model is a special case of a mixture cure model with

$$\pi = exp(-exp(\gamma_{00})) \tag{15}$$

And

$$F_z(t) = exp(\gamma_{02}v_2(x) + \ldots + \gamma_{0K-1}v_{K-1}(x)) \tag{16}$$

As long as the excess mortality is not negative, $F_Z(t)$ is a distribution function is rather rare distribution, but for the non-mixture model, as long as no time-dependent effects are modelled. The flexible parametric cure model can be written as a proportional model of excess hazards when covariates are introduced in the following equation,

$$S_R(t;z) = \exp(-(\exp(\gamma_{00}\beta^T z)\exp(\gamma_{02}v_2(x) + \ldots + \gamma_{0k-1}v_{k-1}(x)) + \sum_{i=1}^{D} s(x\ ;\gamma_i)z_i)) \tag{17}$$

It is seen that the perpetual parameters γ_{00} and β have been used to model the cure ratio and the time-dependent parameters are now used to model the distribution function $F_z(t)$. For each quadratic function $s(x; \gamma_i)$ the model is constructed. The restriction of a zero effect for the linear quadratic term must be incorporated. Both quadratic variables take the value 0 from the point of the last knot. It means that the constant parameter γ_{00} in equation (17) is the log cumulative excess hazard for the comparison group at and even beyond the last knot to estimate the cure ratio.

Results

In this section, the TB data is used to fit the flexible cure rate model and the results of the analysis are presented. The data collected from the National Institute for TB Research-Indian Council of Medical Research, Chennai, India, and the data linked to RCT and TB sputum conversion time to data is used to test the model. From the data, it is seen that all patients diagnosed with TB and patients who have converted from positive sputum smear to negative sputum smear called event and not experienced the event of interest or still alive and patients who have not to turn to the clinic during the treatment period were coined as censored. Among the total of 412 TB patients, 33.5% are Female, 66.5% are Male and the combined mean (SD) of different age group is 32.75 (12.09) years were found using the descriptive analysis. The patient age at diagnosis of TB ranged from 14 to 70 years with the average sputum conversion time for a female is 3.49 months and for the male is 3.29 months. The maximum follow-up period was 12 months.

In this article, the relative survival, projected relative survival and survival of the non-converted (uncured) portion of the population using flexible parametric cure models were estimated for different age groups and gender. The different components of the fitted distributions using parametric survival cure models are compared using their deviations. For different age groups and gender time-based patterns of the converted proportion, average time to conversion and S.D time to conversion were calculated for the flexible parametric cure model. The model is compared with various flexible parametric cure models and the best fit for the sputum conversion is found. The predicted relative survival using various parametric cure rate survival functions for different age groups are also studied. The percentage of sputum conversion among TB patients by sex and gender are analysed and presented in Table 1.

From Table 1 it is found that among female lower age group (≤ 30 have experienced high conversion rate than compared to elder age group and also it is reflected among male. It is also found that when the age progresses the status of sputum conversion become slower for both the genders.

Accordingly in this study, the various flexible parametric cure rate models along the covariates with Akaike information criteria (AIC) are compared and presented in Table 2. It is modelled continuously with the covariates, namely, age, sex, treatment and weight of the patients. The two main effects and interaction effects between sex and age are also studied.

Table 1: Baseline characteristics of patients by sex and age group for the time-related percentage of sputum conversion

Sex	*Age Group*	*Converted %*	*Mean_Time*	*SD_Time*
Female	≤ 30	92.0	2.65	1.693
	31–42	86.7	3.00	2.623
	≥ 43	75.0	1.33	0.516
Male	≤ 30	95.9	2.87	1.949
	31–42	93.0	2.62	1.200
	≥ 43	93.9	2.82	1.511

In Table 2, the TB data is fitted for cured patients using different life time distributions namely Weibull, Gamma, Exponential and Lognormal, by flexible parametric cure rate model. In all the models, cure fractions are estimated and compared it with their deviances. It is found that the Lognormal is the best fit as the estimated cure fraction and AIC are very smaller.

Figure 1 represents the prediction of sputum smear conversion rate using flexible parametric survival cure models, namely, Weibull, Gamma, Exponential, and Lognormal distributions.

It is found from Figure 1 that the Exponential distribution has the least conversion rate whereas the Lognormal has a higher conversion rate of sputum.

Similarly, for the sample as a whole and also for the non-converted group (uncured) group, the cure fraction and median time to conversion for different age groups concerning treatment are estimated using Flexible parametric survival distributions, namely, Weibull, Gamma, and Lognormal distributions. These results for different age group concerning with the treatment in control and trial group are tabulated in Table 3.

Table 3 gives the estimation of cure fraction and median for the non-converted group by considering different age groups with the treatment in control and trail regimen using flexible parametric survival distributions. The estimation of the cure fraction might be responsive to the option of parametric distribution. The Sputum conversion rate for the treatment period has been fitted for Weibull, Gamma and Lognormal distributions concerning with age groups. It is inferred that the Exponential distribution did not provide a good estimate of the cure fraction.

The cure fraction for Lognormal distribution is very less compared with the other two distributions. The range of cure fraction of Lognormal distribution for the age under 30 years is 0.07, for the age 31–42 is 0.14 and for the age above 43 is 0.11, respectively, for group.

The cure fraction of Gamma distribution for the age under 30 years is 0.08, for the age group 31–42 is 0.16 and for the age above 43 is 0.13. The cure fraction of Weibull distribution for the age under 30 years is 0.12, for the age 31–42 is 0.14 and for the age above 43 is 0.15. By comparing the cure fractions of all the three distributions, it is found that in younger age group Weibull distribution is the best fit of sputum conversion rate in treatment period for the non-converted group.

The graphical representation of the estimated cure fraction prediction of all the life time relative survival distributions according to different age groups for the non-converted (uncured) group is presented in Figures 2 and 3, respectively. Figure 2 is for predicting relative survival for age group ≤ 30 and Figure 3 is for the same prediction concerning age group ≥ 43. With a horizontal solid reference line, the approximate cure fraction has been established. For the entire community, the relative survival curve could be seen on the line that continually approaches a given curve at the cure fraction reference line. For the uncured, the relative survival curve is the unconverted category that shifts virtually towards zero or at the bottom of the curve.

Figure 2 shows the prediction of expected relative survival of sputum smear conversion for the sample as a whole and also for the non-converted group with age group below 30. The cure fraction is highlighted by the horizontal reference line marking the portion between the relative survival as a whole and the relative

Table 2: Comparison of flexible parametric cure models by Akaike Information Criteria (AIC) using parametric distributions with prospective analytical factors

	Weibull							
	Est.	*LB95%*	*UB95%*	*SE*	*Exp (Est)*	*LB95%*	*UB95%*	*AIC*
AGE	0.02	−0.01	0.05	0.02	1.02	0.99	1.05	1576.492
SEX_(Ref: M)	−0.48	−1.48	0.52	0.51	0.62	0.23	1.69	
Treatment_(Ref Cont)	0.77	−0.06	1.61	0.42	2.17	0.94	4.98	
Weight_Base line	−0.03	−0.10	0.05	0.04	0.97	0.90	1.05	
DST_Ref: SENS	−0.44	−1.92	1.05	0.76	0.65	0.15	2.86	
GAMMA								
AGE	0.02	−0.01	0.05	0.02	1.02	0.99	1.05	1521.28
SEX_(Ref: M)	−0.48	−1.48	0.52	0.51	0.62	0.23	1.69	
Treatment_(Ref Cont)	0.77	−0.06	1.61	0.42	2.17	0.94	4.99	
Weight_Base line	−0.03	−0.10	0.05	0.04	0.97	0.90	1.05	
DST_Ref: SENS	−0.44	−1.92	1.05	0.76	0.65	0.15	2.87	
Exponential								
AGE	0.02	−0.03	0.07	0.02	1.02	0.97	1.07	1735.408
SEX_(Ref: M)	−0.58	−1.94	0.78	0.69	0.56	0.14	2.19	
Treatment_(Ref Cont)	0.91	−0.33	2.15	0.63	2.48	0.72	8.55	
Weight_Base line	−0.06	−0.19	0.06	0.06	0.94	0.83	1.07	
DST_(Ref: SENS)	−0.78	−3.68	2.12	1.48	0.46	0.03	8.31	
lognormal								
AGE	0.02	−0.01	0.05	0.02	1.02	0.99	1.05	1487.548[#]
SEX_(Ref: M)	−0.48	−1.51	0.55	0.52	0.62	0.22	1.73	
Treatment_(Ref Cont)	0.78	−0.07	1.64	0.44	2.19	0.93	5.15	
Weight_Base line	−0.03	−0.11	0.05	0.04	0.97	0.89	1.05	
DST_Ref: SENS	−0.44	−1.98	1.09	0.78	0.64	0.14	2.98	

[#]Lowest AIC

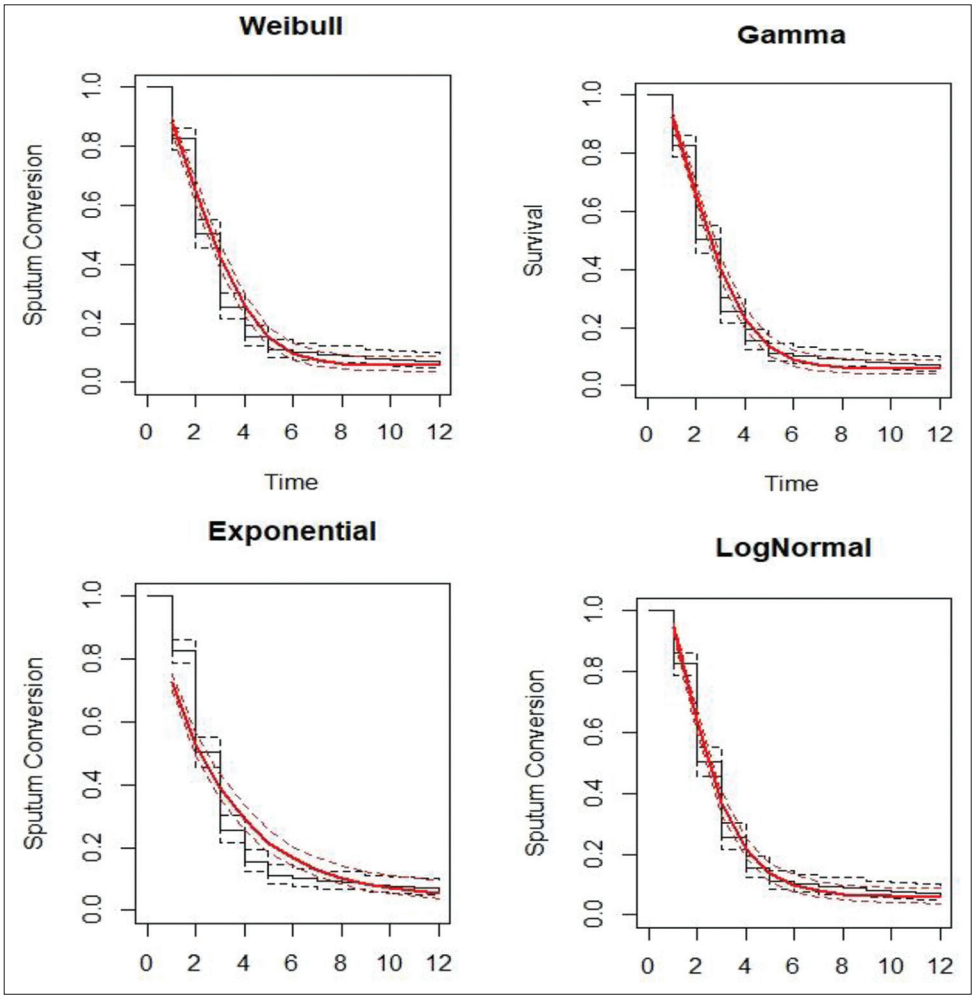

Figure 1: Prediction of sputum smear conversion rate using flexible parametric survival cure models with different distribution

Table 3: Estimation of cure fraction and median for the non-converted group for different age groups concerning treatment using flexible parametric survival distributions

Age	*Weibull*				*Gamma*				*Lognormal*			
	Treatment											
	Control_ Reg		*Trail_Reg*		*Control_Reg*		*Trail_Reg*		*Control_Reg*		*Trail_Reg*	
	Cure	*Median*	*Cure*	*Median*	*Cure*	*Median*	*Cure*	*Median*	*Cure*	*Median*	*Cure*	*Median*
≤ 30	0.12	2.46	0.09	2.38	0.08	2.35	0.08	2.35	0.07	2.32	0.07	2.32
31–42	0.14	2.51	0.10	2.43	0.16	2.54	0.10	2.41	0.14	2.55	0.08	2.38
≥ 43	0.15	2.54	0.12	2.46	0.13	2.48	0.12	2.45	0.11	2.47	0.10	2.43

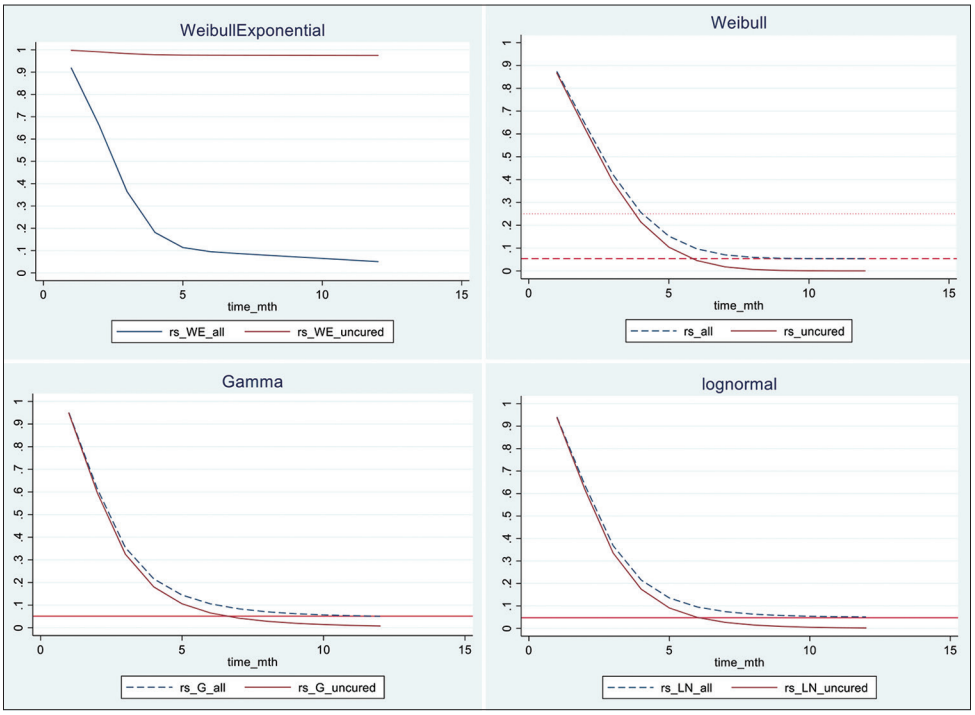

Figure 2: Predicted relative survival using various parametric survival distributions for the sample as a whole and for the non-converted (uncured) group. (Age group equal to or below 30)

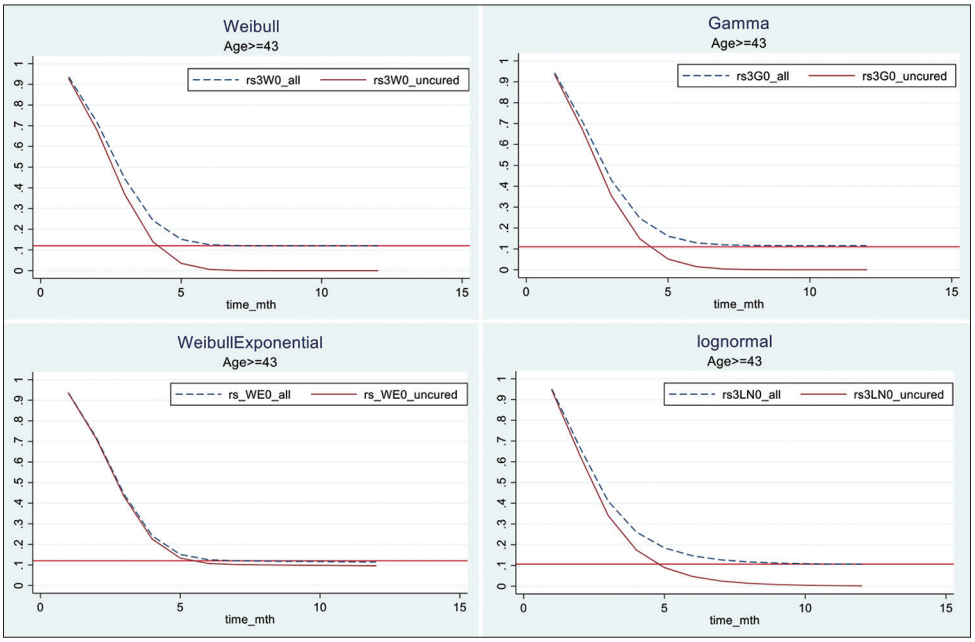

Figure 3: Predicted relative survival using various parametric survival distributions for the sample as a whole and the non-converted (uncured) group. (Age group ≥ 43)

survival of the non-converted (uncured) group as a threshold for pinpointing in all distributions. For the uncured category, the relative survival curve is virtually heading toward zero as time progresses and for some distribution at exactly zero.

Figure 3 represents the prediction of estimated relative survival of sputum smear conversion for the entire sample and also for the non-converted (uncured) population by age group more than or equal to 43. With the horizontal reference line, the cure fraction is highlighted which is slightly high as it has been anticipated that a growing pattern of cure fraction rises as per age. For the uncured category, the relative survival curve is virtually heading toward zero as time progresses for all distributions. Furthermore, the prediction of cure fraction is double the time higher than the prediction of cure fraction for the younger age group (≤ 30).

Conclusion

In this study, the relative survival, projected relative survival and survival of the non-converted (uncured) group of the population using flexible parametric cure models were investigated for different age groups and gender. The flexible parametric cure rate models along with the covariates are fitted for TB data with AIC criteria. At this juncture, data are fitted for cured patients using various distributions, namely, Weibull, Gamma, Exponential and Log Normal, respectively. It is found that Log normal is the best fit as the estimated cure fraction and AIC are very smaller when compared with other distributions.

The prediction of the relative survival cure rate fraction using various parametric cure rate survival functions for the different age groups is studied. The percentage of sputum conversion among TB patients by age-group and gender is also analysed. The average time to conversion and standard deviation time to conversion was also calculated for the flexible parametric cure model.

The various prediction of sputum smear conversion rate using flexible parametric cure models with various distributions is analyzed and represented graphically. It is observed that the predicted relative survival using various parametric survival distributions for the different age groups.

In Table 3, it is estimated that the cure fraction rate and median for the non-converted group for different age groups ≤ 30, 31-42 and ≥ 43 concerning treatment using flexible parametric survival distributions. In younger age group, the Weibull distribution has the best fit to sputum smear conversion rate in treatment period.

Overall, the flexible parametric cure model seems to provide a good fit for sputum conversion pulmonary TB data on the main factors of age and treatment over the entire follow-up period.

References

1. Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. Stat Med 2004;23:1733-47.
2. Lambert PC. Modeling of the cure fraction in survival studies. Stata J 2007;7:351-75.
3. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med 2002;21:2175-97.
4. Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. Stat Med 2007;26:5486-98.
5. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J R Stat Soc 1949;11:15-53.
6. Berkson J, Gage RP. Survival cure for cancer patients following treatment. J Am Stat Assoc 1952;47:501-15.
7. Dickman PW, Adami HO. Interpreting trends in cancer patient survival. J Intern Med 2006;260:103-17.
8. De Angelis R, Capocaccia R, Hakulinen T, Söderman B, Verdecchia A. Mixture models for cancer survival analysis. Application to population-based data with covariates. Stat Med 1999;18:441-54.
9. Lambert P, Dickman P, Weston C, Thompson J, Estimating the cure fraction in population-based cancer studies using finite mixture models. J R Statist Soc C 2010;59:35-55.
10. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. Stata J 2009;9:265-90.