ISSN 1973-2201 STATISTICA, anno LXXXIV, n. 4, 2024
DOI: https://doi.org/10.60923/issn.1973-2201/21223

TRANSMUTED CURE MODELS - A COMPETING RISK PERSPECTIVE

C.S. Soorya
Department of Statistics, Cochin University of Science and Technology, Cochin
G. Asha [1]
Department of Statistics, Cochin University of Science and Technology, Cochin

SUMMARY

A cure rate model under a competing risk scenario where the number of competing causes follow a shifted binomial distribution with parameter p is proposed. Interestingly, the resulting distribution is the well-studied transmuted class of distribution. Few existing cure rate models are shown to be special cases of the proposed model. The identifiability issues of the model are studied in detail. Further properties of the model are investigated, and we discuss the maximum likelihood estimation of the parameter. The performance is confirmed through a simulation study using a defective Gompertz baseline and with competing causes. The Bayesian approach to the estimation of the parameter is adopted. The complexity of the likelihood function is handled through the Metropolis-Hastings algorithm. We analyse the data consisting of 8966 patients who have undergone bone marrow transplantation at the European Society for Blood and Marrow Transplantation (EBMT). The validation of the estimation algorithm is confirmed using the bootstrap technique.

Keywords: Bootstrap; Cure rate; Metropolis-Hastings algorithm; Transmuted distributions

1. INTRODUCTION

Improper or defective distributions (Rocha *et al.*, 2017) are used to model time to event data, when there is always a proportion of the individuals, commonly referred to as immunes who do not experience the event of interest. The mixture model (Berkson and Gage, 1952; Boag, 1949) and promotion time cure rate model (Yakovlev *et al.*, 1993) are the two general cure rate models that has received a lot of attention and has been worked by many authors under different perspectives. In both cases, the cure rate model has an improper distribution function. For the mixture cure rate model, the population

---

[1] Corresponding Author. E-mail: asha@cusat.ac.in

consist of two groups. The immunes and the susceptibles. If T be the failure time, then the probability of survival of the total population, $\bar{G}(t) = P[T > t]$ for $0 < p_0 < 1$ is given as

$$\bar{G}(t) = p_0 + (1 - p_0)\bar{F}_0(t), \tag{1}$$

where $\bar{F}_0(t)$ is any baseline survival function. Then, $\bar{G}(t)$ is an improper survival function and $\lim_{t\to\infty} \bar{G}(t) \to p_0$, is referred to as the cure rate.

In Cox and Oakes (1984), a competing risk perspective of the mixture cure rate model is presented. This was further studied by Tsodikov *et al.* (2003) by considering M to be an unobservable random variable denoting the number of competing causes related to the occurrence of an event of interest, with the probability mass function

$$p_m = P[M = m] \quad m = 0, 1, 2, \ldots.$$

Also, for $m > 0$, W_k, $k = 1, \ldots, m$ denote the time to event due to the k^{th} competing cause and W_k are independent and identical random variables with survival function $\bar{F}_0(t)$ that does not depend on M. The number of competing causes M and the lifetime W_k associated with a particular cause is not observable, but only the minimum lifetime among all causes is observed. It is defined as

$$T = \min(W_0, W_1, W_2, \ldots, W_m) \text{ with } P[W_0 = \infty] = 1, \tag{2}$$

which leads to a proportion p_0 (say) of the population not susceptible to the event occurrence, termed the cure fraction. Then, for a given $\bar{F}_0(t)$ the long-term survival function of the random variable T is given as

$$\bar{G}(t) = P[T > t] = \sum_{m=0}^{\infty} P[M = m]\{\bar{F}_0(t)^m\} = E\left[\bar{F}_0(t)^M\right], \tag{3}$$

where Eq. (3) is the probability generating function of M. Rodrigues *et al.* (2009) studied the model in Eq. (3) by assuming a Weibull baseline distribution for $\bar{F}_0(t)$ and M having a Conway-Maxwell (COM)- Poisson distribution. This was further studied by Balakrishnan and Pal (2013, 2015); Balakrishnan *et al.* (2024).

In this paper we propose a general class of cure rate model by extending the Rodrigues *et al.* (2009) model by letting M having a shifted Binomial distribution with probability mass function (pmf) is given as

$$p_m = P[M = m] = \binom{N-1}{m-1} p^{m-1}(1-p)^{N-m};\ m = 1, 2, \ldots, N, \tag{4}$$

where $N \geq 1$ is the total number of potential competing causes relate to the occurrence of an event of interest. It is assumed that $W_j; j = 1, 2, \ldots, N$ is the time to event due to the j^{th} cause following an improper survival function $\bar{G}_0(t)$ which implies that $\lim_{t\to\infty} \bar{G}_0(t) = c_0$, $0 < c_0 < 1$. It is interesting to see that the resulting model evolves to the general transmuted family of distributions (Rahman *et al.*, 2018).

This paper is organised as follows; in Section 2, we introduce the general class of transmuted cure rate (GTC) model, and the properties of the model are presented in Section 3. In Section 4, the identifiability issues of the proposed model are discussed. In Section 5, the proposed model is illustrated with an example, the properties and their estimation procedures of parameters are also discussed. In Section 6, the Bayesian approach to estimating parameters of GTC is explored in detail. The complexity of the likelihood function is handled through the Metropolis-Hastings algorithm. In Section 7, simulation studies are carried out to establish the effectiveness of the method of maximum likelihood and Bayesian estimation procedures. In Section 8, an application to an actual data set is detailed by using the well-analysed blood and marrow data (Fiocco *et al.*, 2005). We conclude the paper with a discussion and summary of our findings in Section 9.

2. GENERAL CLASS OF TRANSMUTED CURE RATE MODELS

Consider M to be an unobservable random variable denoting the number of competing causes related to the occurrence of an event of interest, with probability mass function (shifted Binomial).

$$P[M=m]=\binom{N-1}{m-1}p^{m-1}(1-p)^{N-m};\ m=1,2,\dots,N, \tag{5}$$

where $N \geq 1$ is the total number of potential competing causes relate to the occurrence of an event of interest. Let W_j, $j=1,2,\dots,N$ denote the time to event due to j^{th} competing cause and W_j are independent and identical random variables with a common survival function $\bar{G}_0(t)$ that is independent of M with $\bar{G}_0(\infty)=c_0$, $0 \leq c_0 < 1$. Here there is a positive probability that $W_j=\infty$, $j=1,2,\dots,N$. The number of competing causes M and the lifetime W_k associated with k^{th} cause is not observable but only the minimum lifetime among all causes is observed. The observed life time is now

$$T=\min(W_1,W_2,\dots,W_N). \tag{6}$$

Then, for a given $\bar{G}_0(t)$ the long-term survival function of the random variable T is given as

$$\begin{aligned}\bar{G}_N(t)=P[T>t]&=P[Min(W_1,W_2,\dots,W_N)>t]\\&=\sum_{m=1}^{N}[\bar{G}_0(t)]^m\binom{N-1}{m-1}p^{m-1}(1-p)^{N-m}\\&=\bar{G}_0(t)\left[1-p+p\,\bar{G}_0(t)\right]^{N-1},\end{aligned} \tag{7}$$

where $N \geq 1$. Observe that the model in Eq. (7) is an improper survival function and is formulated as a transmuted model of order N (Rahman *et al.*, 2018). Hence, we refer Eq. (7) as the general class of transmuted cure rate (GTC) model.

Let $\mathscr{G}$ be the class of baseline survival function given by

$$\mathscr{G} = \{\bar{G}_0(t,\theta_0) : \lim_{t\to\infty} \bar{G}_0(t,\theta_0) = c_0(\theta_0),\ 0 \le c_0 < 1,\ t \in R^+\}, \tag{8}$$

where $\theta_0 = (\theta_{01}, \theta_{02}, \ldots)$ is the set of parameters in the baseline distribution. From Figure 1 we see that the parameter p manages the skewness of the distribution and this motivates us to incorporate effect of covariates through the parameter p as $p(x,\beta)$, using the logistic function (Farewell, 1982) as

$$1 - p(x,\beta) = \frac{1}{1+e^{\beta x}} \text{ and } p(x,\beta) = \frac{e^{\beta x}}{1+e^{\beta x}}, \tag{9}$$

where $x = (1, x_1, \ldots, x_p)'$ be the vector of covariates and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ be the $(p+1) \times 1$ vector of regression coefficients. Define the class $\mathscr{P}$ such that

$$\mathscr{P} = \Big\{ p(x,\beta) : 0 < p(x,\beta) < 1 \Big\}. \tag{10}$$

For a given $\bar{G}_0(\cdot) \in \mathscr{G}$ and $p(\cdot) \in \mathscr{P}$, then the class of models corresponding to (7) can be written as

$$\mathscr{H} = \Big\{ \bar{G}_m(t,x) : \bar{G}_N(t,x,\theta) = \bar{G}_0(t,\theta_0)\Big[1 - p(x,\beta) + p(x,\beta)\,\bar{G}_0(t,\theta_0)\Big]^{N-1} \Big\}, \tag{11}$$

where $t \in R^+, N = 1,2,\ldots$.

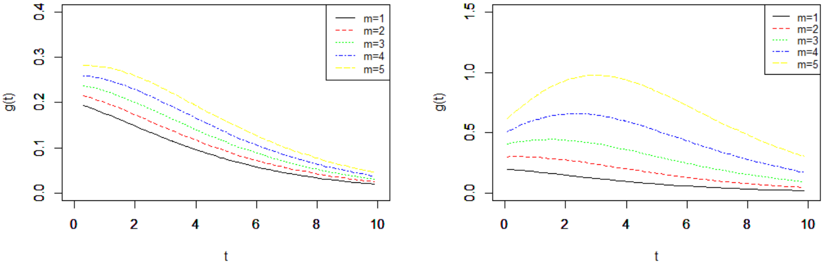
The cure fraction of GTC model is

$$c_N(x,\theta) = \lim_{t\to\infty} \bar{G}_N(t,x,\theta) = c_0(\theta_0)[1 - p(x,\beta) + p(x,\beta)\,c_0(\theta_0)]^{N-1}, \tag{12}$$

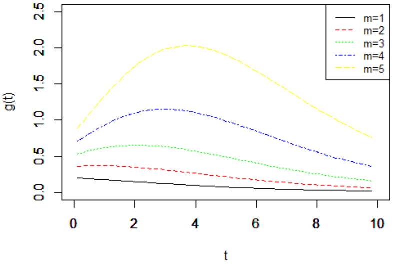
where $\lim_{t\to\infty} \bar{G}_0(t,\theta_0) = c_0(\theta_0),\ 0 \le c_0(\theta_0) < 1$. The class of GTC models reduces to some well studied distributions in literature (see Table 1).

TABLE 1
Examples of cure rate models.

N	$\bar{G}_N(t,x,\theta)$	Cure rate model
1	$\bar{G}_1(t,x,\theta) = \bar{G}_0(t,\theta_0)$	Defective cure rate model (Rocha *et al.*, 2017).
2	$\bar{G}_2(t,x,\theta) = \bar{G}_0(t,\theta_0)\big(1 - p(x,\beta) + p(x,\beta)\,\bar{G}_0(t)\big)$	Shifted Bernoulli cure rate model (Asha and Soorya, 2021).
3	$\bar{G}_3(t,x,\theta) = \bar{G}_0(t,\theta_0)\big(1 - p(x,\beta) + p(x,\beta)\,\bar{G}_0(t,\theta_0)\big)^2$	when $\bar{G}_0(\cdot)$ is proper $\bar{G}_3(\cdot)$ reduces to cubic transmuted distribution (Granzotto *et al.*, 2017; Rahman *et al.*, 2018).



(a) $p = 0.1$ (b) $p = 0.5$



(c) $p = 0.8$

Figure 1 – Plots of probability density function of improper survival function of GTC cure rate model for different values of p.

3. PROPERTIES

We proceed this Section by deriving the properties of the GTC model i Eq. (7). The density of the GTC model is obtained as

$$g_N(t,x,\theta)=\sum_{m=1}^{N} g_0(t,\theta_0)\left(m\,\bar{G}_0(t,\theta_0)^{m-1}\tbinom{N-1}{m-1}p(x,\beta)^{m-1}(1-p(x,\beta))^{N-m}\right), \quad (13)$$

where $g(t,x,\theta)$ and $g_0(t,\theta_0)$ are the probability density function of improper survival function of GTC cure rate model $\bar{G}(t,x,\theta)$ and improper baseline survival function $\bar{G}_0(t,\theta_0)$ respectively. Figure 1 shows the probability density plots of GTC model with different parameter values for p and m.

The hazard function of GTC cure rate model is given by

$$h_N(t,x,\theta)=h_0(t,\theta_0)\left[1+(N-1)\frac{p(x,\beta)\bar{G}_0(t)}{1-p(x,\beta)+p(x,\beta)\bar{G}_0(t,\theta_0)}\right], \quad (14)$$

where $h_0(t,\theta_0)$ is the hazard rate of baseline survival function $\bar{G}_0(t,\theta_0)$.

THEOREM 1. *The GTC model $\bar{G}(t,x,\theta)$, has a decreasing failure rate if the baseline distribution $\bar{G}_0(t,\theta_0)$ has a decreasing failure rate.*

PROOF. See Appendix A for the proof. □

Note that the GTC model need not exhibit the same behaviour of failure pattern as the baseline apart from the decreasing case. However, it should be mentioned that the failure rate is bounded in all cases. The following result is quite intuitive.

THEOREM 2. *The failure rate of GTC model is an increasing function of the number of competing risk N.*

PROOF. See Appendix B for the proof. □

THEOREM 3. *The cure rate of the GTC model is a decreasing function in both parameter p and N.*

PROOF. The proof is direct. □

4. IDENTIFIABILITY ISSUES OF GTC MODEL

In case of studies involving long term survivors the modelling and estimation of parameters can be challenging. Farewell (1982) discussed the issues in distinguishing censored individuals in susceptible groups from the non susceptible individual. This in turn leads to non-identification of models with a high incidence of susceptibles and long tails of the latency distribution from low incidence of susceptibles and short levels of latency distribution. Every cure rate model discussed now needs to be investigated for identifiability issues to obtain unique estimates of unknown parameters. This warrants an in-depth study of identifiability issues of the proposed GTC model. In this Section, we investigate the conditions under which GTC model is identifiable.
Note that the GTC model can be represented as

$$\bar{G_N}(t,x,\theta) = \bar{G_0}(t,\theta_0)\, U(t,x,\theta)^{N-1}, \tag{15}$$

where $U(t,x,\theta) = 1 - p(x,\beta) + p(x,\beta)\,\bar{G_0}(t,\theta_0)$. This is of the form of the standard mixture cure rate model. The identifiability of the standard mixture cure rate model has been extensively discussed by Li *et al.* (2001) and Hanin and Huang (2014).

DEFINITION 4. *The function $U(t,x,\theta)$ is identifiable within the family $\mathscr{P}$ and $\mathscr{G}$ if the equality*

$$U(t,x,\theta) = U^*(t,x,\theta^*), \tag{16}$$

where

$$U^*(t,x,\theta^*) = 1 - p^*(x,\beta^*) + p^*(x,\beta^*)\,\bar{G}_0^*(t,\theta_0^*) \tag{17}$$

for some $p(x,\beta), p^*(x,\beta^*) \in \mathscr{P}$, $\bar{G_0}(t,\theta_0), \bar{G}_0^*(t,\theta_0^*) \in \mathscr{G}$ implies that $p(x,\beta) = p^*(x,\beta^*)$ and $\bar{G_0}(t,\theta_0) = \bar{G}_0^*(t,\theta_0^*)$.

THEOREM 5. *The GTC model $\bar{G_N}(t,x,\theta)$ is identifiable if the function $U(t,x,\theta)$ is identifiable.*

PROOF. See Appendix C for the proof. □

COROLLARY 6. *The GTC model in Eq. (11) is identifiable.*

PROOF. From Theorem 1 of Li *et al.* (2001), we have that $U(t,x,\theta)$ is identifiable for $p(\cdot)$ specified as a logistic function. The result now follows from Theorem 5. □

5. ESTIMATION OF PARAMETERS

Among many estimation approaches, the maximum likelihood estimation is most popular due to the desirable large sample properties it enjoys. The likelihood approach works well for estimating the parameters in GTC model. In this Section, we discuss the likelihood and observe that it needs to be analytically solved. A simulation algorithm of the procedure is explained and illustrated through an example using the generalised Gompertz distribution as the baseline.

5.1. *The Likelihood contribution*

If Y_i and C_i denote the failure time and censoring time of the i^{th} subject, we observe $T_i = min(y_i, C_i)$ for $i = 1,2,\ldots,n$. Correspondingly, the failure-time is indicated by $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_n)$ where $\delta_i = I(T_i \geq C_i)$ where $I(.)$ is the indicator function and δ_i is equal to 1 if T_i is a time-to-event and is 0 if it is right-censored. Let $x = (1, x_1,\ldots,x_k)'$ be the vector of covariates and $\beta = (\beta_0,\beta_1,\ldots,\beta_k)$ be the $(k+1)\times 1$ vector of regression coefficient and the covariates are introduced through the parameter p. If $\theta_0 = (\theta_{01},\ldots,\theta_{0k})'$ are the parameters of the baseline distribution, then $\theta = (\theta_0,\beta)$ forms the vector of parameters for the GTC model. The observed data are of the form (t_i,δ_i,x). Suppose that the data are independently and identically distributed and come from GTC Model with density function $g_m(t_i,x;\theta)$ and survival function $\bar{G}_m(t_i,,x;\theta) \in \mathscr{H}$, the likelihood function is given by

$$L(\theta;t,x,\boldsymbol{\delta}) \propto \prod_{i=1}^{n} g_m(t_i,x;\theta)^{\delta_i}\, \bar{G}_m(t_i,x;\theta)^{(1-\delta_i)}. \tag{18}$$

For the proposed model in Eq. (7) the log-likelihood function is obtained as

$$\begin{aligned} l(\theta;t,x,\boldsymbol{\delta}) = &\sum_{i=1}^{n} \delta_i \log\left(g_0(t;\theta_0) \sum_{m=1}^{N} \left(m\, \bar{G}_0(t;\theta_0)^{m-1} \tbinom{N-1}{m-1} p^{m-1}(1-p)^{N-m} \right)\right) \\ &+ (1-\delta_i)\log\Big(\bar{G}_0(t;\theta_0)\big(1 - p(x,\beta) + p\big(x,\beta\big)\bar{G}_0(t;\theta_0)\big)^{N-1}\Big). \end{aligned} \tag{19}$$

Given the observed data $(t,\boldsymbol{\delta},x)$ we wish to find the value of θ that maximizes Eq. (19). The likelihood equations can be obtained by solving first order partial derivatives of the likelihood and equating to zero. These equations are non-linear in nature and difficult to solve. We consider the profile likelihood method. This estimation method is an alternative one when closed form of maximum likelihood estimator is not possible and iterative procedures fail to converge. The profile likelihood approach involves estimating parameters in stages. In the first stage, we fix one of the parameters in $\theta = (\theta_0,\beta)$ to estimate the remaining parameters. Then, in the second stage, the estimates obtained from the first stage are substituted back into the likelihood function, and the remaining parameters are estimated. This stepwise approach is repeated until convergence of the parameter estimates $\hat{\theta} = (\hat{\theta}_0,\hat{\beta})$, is observed. The asymptotic distribution of $\hat{\theta} = (\hat{\theta}_0,\hat{\beta})$ is multivariate normal with mean vector $\theta = (\theta_0,\beta)$ and covariance matrix $I^{-1}(\theta_0,\beta)$ which can be estimated by from the observed fisher information matrix (Patefield, 1977; Cox and Barndorff-Nielsen, 1994) evaluated at $\theta = (\hat{\theta}_0,\hat{\beta})$.

5.2. *Example - general transmuted class of Gompertz cure rate model*

It is often observed that in some clinical data on cancer survival, the hazard rate decreases exponentially with time. This motivates us to consider the Gompertz distribution as the baseline distribution in GTC model. It is specified as,

$$\bar{G}_0(t;\theta_0)=e^{\frac{b}{a}(e^{-at}-1)},\ a\neq 0,\ b>0, \tag{20}$$

where $\theta_0=(a,b)$.

If $a<0$ the survival function in Eq. (20) corresponds to the proper Gompertz distribution and if $a>0$ the distribution is an improper distribution with surviving fraction to $e^{-\frac{b}{a}}$ (Cantor, 2001).

The survival function of general transmuted class of Gompertz cure rate (GTCGC) model is

$$\bar{G}_m(t,x;\theta)=e^{\frac{b}{a}(e^{-at}-1)}\left[1-p(x,\beta)+p(x,\beta)e^{\frac{b}{a}(e^{-at}-1)}\right]^{m-1}, \tag{21}$$

where $\theta=(a,b,\beta)$, $a>0$, $b>0$, $p(x,\beta)=\frac{e^{\beta x}}{1+e^{\beta x}}$, $x=(1,x_1,\ldots,x_p)'$ and $\beta=(\beta_0,\beta_1,\ldots,\beta_p)$ are the vector of covariates and the vector of regression coefficients respectively. The cure fraction of GTCGC model is

$$c_m(\theta;x)=e^{-\frac{b}{a}}(1-p(x,\beta)+p(x,\beta)e^{-\frac{b}{a}})^{m-1}. \tag{22}$$

The probability density function and hazard functions are given by

$$\begin{aligned}g(t,x;\theta)=&b\sum_{i=1}^{m}e^{-at}\,e^{\frac{b}{a}(e^{-at}-1)}\left(i\left(e^{\frac{b}{a}(e^{-at}-1)}\right)^{i-1}\right.\\&\left.\times\binom{m-1}{i-1}p(x,\beta)^{i-1}(1-p(x,\beta))^{m-i}\right)\end{aligned} \tag{23}$$

and

$$h_m(t,x;\theta)=b\,e^{-at}\left[1+(m-1)\frac{pe^{\frac{b}{a}(e^{-at}-1)}}{1-p(x,\beta)+p(x,\beta)e^{\frac{b}{a}(e^{-at}-1)}}\right], \tag{24}$$

respectively. The likelihood function corresponding to Eq. (18) for the GTCGC in Eq. (21) model is

$$\begin{aligned}L(\theta;t,x,\delta)\propto\prod_{j=1}^{n}&\left(b\sum_{i=1}^{m}e^{-at_j}\,e^{\frac{b}{a}(e^{-at_j}-1)}\left(i\left(e^{\frac{b}{a}(e^{-at_j}-1)}\right)^{i-1}\binom{m-1}{i-1}p(x_j,\beta)^{i-1}\right.\right.\\&\left.\left.\times(1-p(x_j,\beta))^{m-i}\right)\right)^{\delta_j}\left(e^{\frac{b}{a}(e^{-at_j}-1)}\left[1-p(x_j,\beta)+p(x_j,\beta)\right.\right.\\&\left.\left.\times e^{\frac{b}{a}(e^{-at_j}-1)}\right]^{m-1}\right)^{(1-\delta_j)},\end{aligned} \tag{25}$$

where $\theta = (a, b, \beta)$ and n is the sample size. The corresponding log likelihood function is

$$
\begin{aligned}
\log L(\theta; t, x, \boldsymbol{\delta}) \propto \sum_{j=1}^{n} \delta_j \Big(b \sum_{i=1}^{m} e^{-a t_j}\, e^{\frac{b}{a}(e^{-at_j}-1)} \Big(i \Big(e^{\frac{b}{a}(e^{-at_j}-1)} \Big)^{i-1} \tbinom{m-1}{i-1} p(x_j, \beta)^{i-1} \\
\times (1 - p(x_j, \beta))^{m-i} \Big)\Big) + (1-\delta_j)\Big(e^{\frac{b}{a}(e^{-at_j}-1)} \big[1 - p(x_j, \beta) + p(x_j, \beta) \\
\times e^{\frac{b}{a}(e^{-at_j}-1)} \big]^{m-1} \Big).
\end{aligned}
\tag{26}
$$

The log likelihood can be maximized either directly or by solving the non-linear likelihood equations obtained by differentiating Eq. (26).

6. THE BAYESIAN ESTIMATION

The Bayesian approach to cure rate modelling allows one to combine prior information about the parameter with observed data and has been comprehensively presented by many authors. See Ibrahim *et al.* (2001) and Rodrigues *et al.* (2012) to mention a few.

The posterior distribution of $\theta = (a, b, \beta)$ based on the observed data is given by

$$
\pi(\hat{\theta} | \boldsymbol{t}) \propto L(\boldsymbol{t} | \theta)\, \pi(a)\, \pi(b)\, \pi(\beta), \tag{27}
$$

where $\pi(a), \pi(b)$ and $\pi(\beta)$ are the prior distribution of the parameters a, b and β, respectively. Since the Gamma family is flexible and suitable to model positive variables, we consider the prior distributions of a, b, as Gamma (m_1, n_1), Gamma (m_2, n_2). For the parameter β the prior taken as $N(\mu, \sigma)$.

The joint posterior distribution is obtained as

$$
\begin{aligned}
\log \pi(\theta; t, x, \boldsymbol{\delta}_j) \propto \sum_{j=1}^{n} \delta_i \log\Big(b \sum_{i=1}^{m} e^{-a t_j}\, e^{\frac{b}{a}(e^{-at_j}-1)} \Big(i \Big(e^{\frac{b}{a}(e^{-at_j}-1)} \Big)^{i-1} \tbinom{m-1}{i-1} \\
\times p(x, \beta)^{i-1} (1 - p(x, \beta))^{m-i} \Big)\Big] + (1-\delta_i) \log\Big(e^{\frac{b}{a}(e^{-at_j}-1)} \\
\times \Big[1 - p(x, \beta) + p(x, \beta) e^{\frac{b}{a}(e^{-at_j}-1)} \Big]^{m-1} \Big) - \log \Gamma m_1 - m_1 \log n_1 \\
+ (m_1 - 1) \log a - \frac{a}{n_1} - \log \Gamma m_2 - m_2 \log n_2 \\
+ (m_2 - 1) \log b - \frac{b}{n_2} + \log(\sigma \sqrt{2\pi}) - \frac{1}{2} \left(\frac{\beta - \mu}{\sigma} \right)^2,
\end{aligned}
\tag{28}
$$

where Γ is the gamma function, $p(x, \beta) = \frac{e^{\beta x}}{1 + e^{\beta x}}$ and $m = 1, 2, \ldots, N$.

Posterior summaries of interest will be obtained from simulated samples for the joint posterior distribution using standard MCMC procedure as the posterior is analytically intractable. Gibbs sampling algorithm is used when the complete conditional distributions have known forms. The posterior conditionals of Eq. (28) for each variable is given below

$$\begin{aligned}
\pi(a|b,p;\boldsymbol{t}) \propto & \sum_{j=1}^{n} \delta_i \log\Bigg(b \sum_{i=1}^{m} e^{-a\,t_j}\, e^{\frac{b}{a}(e^{-at_j}-1)} \Big(i \Big(e^{\frac{b}{a}(e^{-at_j}-1)} \Big)^{i-1} \tbinom{m-1}{i-1} \\
& \times (p(x,\beta)^{i-1} 1 - p(x,\beta))^{m-i} \Big) \Bigg) + (1-\delta_i)\, log \Bigg(e^{\frac{b}{a}(e^{-at_j}-1)} \\
& \times \big[1 - p(x,\beta) + p(x,\beta), e^{\frac{b}{a}(e^{-at_j}-1)} \big]^{m-1} \Bigg) - \log \Gamma m_1 - m_1 \log n_1 + \\
& + (m_1 - 1) \log a - \frac{a}{n_1},
\end{aligned}$$

$$\begin{aligned}
\pi(b|a,p_1;\boldsymbol{t}) \propto & \sum_{j=1}^{n} \delta_i \log\Bigg(b \sum_{i=1}^{m} e^{-a\,t_j}\, e^{\frac{b}{a}(e^{-at_j}-1)} \Big(i \Big(e^{\frac{b}{a}(e^{-at_j}-1)} \Big)^{i-1} \tbinom{m-1}{i-1} \\
& \times p(x,\beta)^{i-1} (1 - p(x,\beta))^{m-i} \Big) \Bigg) + (1-\delta_i) \log \Bigg(e^{\frac{b}{a}(e^{-at_j}-1)} \\
& \times \big[1 - p(x,\beta) + p(x,\beta) e^{\frac{b}{a}(e^{-at_j}-1)} \big]^{m-1} \Bigg) - \log \Gamma m_2 - m_2 \log n_2 + \\
& + (m_2 - 1) \log b - \frac{b}{n_2},
\end{aligned}$$

$$\begin{aligned}
\pi(\beta|a,b;\boldsymbol{t}) \propto & \sum_{j=1}^{n} \delta_i \log\Bigg(b \sum_{i=1}^{m} e^{-a\,t_j}\, e^{\frac{b}{a}(e^{-at_j}-1)} \Big(i \Big(e^{\frac{b}{a}(e^{-at_j}-1)} \Big)^{i-1} \tbinom{m-1}{i-1} \\
& \times (p(x,\beta)^{i-1} 1 - p(x,\beta))^{m-i} \Big) \Bigg) + (1-\delta_i) \log \Bigg(e^{\frac{b}{a}(e^{-at_j}-1)} \\
& \times \Bigg[1 - \Big(p(x,\beta) + p(x,\beta) e^{\frac{b}{a}(e^{-at_j}-1)} \Big) \Bigg]^{m-1} \Bigg) - \frac{1}{2} \Big(\frac{\beta - \mu}{\sigma} \Big)^2 \Bigg).
\end{aligned}$$

The Gibbs sampling algorithm is ruled out as the conditionals are in the above form. We resort to the Metropolis-Hastings (M-H) algorithm which enables us to sample from the posterior distribution $\pi(\theta|\boldsymbol{t})$. The algorithm is based on a candidate-generating or proposal density $k(\theta,\nu)$ such that $\int k(\theta,\nu) d\nu = 1$. The M-H algorithm is:

- Set an initial value for θ say θ_0. Set iteration number $m = 0$.
- Generate θ^* from the proposal density $k(\theta_m, \cdot)$. Draw a value u from $U(0,1)$.

- Calculate a probability of move $\alpha(\theta_m,\theta^*)$. The probability of move from θ to θ^* is the ratio
$$\alpha(\theta,\theta^*)=\min\left(\frac{\pi(\theta^*|\boldsymbol{t})k(\theta^*,\theta)}{\pi(\theta|\boldsymbol{t})k(\theta,\theta^*)},1\right).$$
Since $k(\theta,v)$ is symmetric $k(\theta^*,\theta=k(\theta,\theta^*)$. If $u\leq\alpha(\theta_m,\theta^*)$, accept the new candidate, and set $\theta_{m+1}=\theta^*$. If $u>\alpha(\theta_m,\theta^*)$, then do not accept the new candidate, and set $\theta_{m+1}=\theta_m$. Thus, at iteration $(m+1)$, θ_{m+1} either takes on the value θ^* or remains at the previous iterate value θ_m.

- Set $m=m+1$, and repeat.

- The iteration is continued untill convergence is attained.

Convergence of the samples θ_m for $m=1,2,\ldots,M$ to the invariant target posterior distribution occurs only after the Markov chain has passed the transient stage and the effect of the fixed starting value θ_0 has become negligible and can be ignored and occurs under mild regularity conditions. These regularity conditions are (a) irreducibility and (b) aperiodicity of the Markov chain. We use `R` software based on the `mhadaptive` package in the `R` software (Chivers, 2011) to develop the code and posterior summaries of interest.

For a sample size n, the `mhadaptive` takes a sample of size m from posterior distribution $\pi(\theta|t)$. This process is replicated R times. The posterior mean θ is $\hat{\theta}=\frac{1}{Rm}\sum_{i=1}^{R}\sum_{j=1}^{m}\hat{\theta}_{ij}$, where $\hat{\theta}_{ij}$ be the j^{th} sample from the i^{th} replication and R, m are the total number of replications and the iterations respectively. One can choose either posterior mean or posterior mode as the Bayesian estimate. The bias and MSE of the parameter θ is Bias$=\hat{\theta}-\theta$ and MSE$=\frac{1}{R}\sum_{i=1}^{R}\left(\frac{1}{m}\sum_{j=1}^{m}\hat{\theta ij}-\theta\right)^2$, respectively. We calculate the standard error of this estimate, we use the Bootstrap algorithm (Efron and Hastie, 2016).

7. SIMULATION STUDY

A simulation study of the performance of the maximum likelihood estimation and Bayes estimation were carried out based on simulated samples from the GTC model with three competing risks and Gompertz distribution as baseline distribution using the following simulation algorithm.

Simulation algorithm

Suppose that the time of occurrence of an event of interest has survival function $\bar{G}(t,x)$ as given in Eq. (7). The algorithm to simulate right censored failure times from $\bar{G}(t,x)$ with cure fraction $c_m(x,\theta)$ for a sample of size n is detailed below.

1. Generate n covariate values $x_1,\ldots,x_n$ from $f_X(x)$.

2. For each x_i and the values of model parameters in the baseline distribution and intensity function, compute the cure fraction $c_m(x_i;\theta)$ as in Eq. (12).

3. Generate V_i, where $V_i \sim$ Bernoulli $(1-c_m(x_i,\theta))$ for each $i = 1,2,\ldots n$.

4. For $i = 1,\cdots,n$, if $V_i = 1$ take t_i as the solution of $G(t_i,x_i) = u_i$, where u_i is generated from Uniform$(0,1-c_m(x_1,\theta))$. If $V_i = 0$, set $t_i = \infty$.

5. For each $i = 1,2,\ldots,n$, generate the censoring time $D = d_i$ from a censoring distribution say $g(d|k)$ with the associated parameter k, where k is computed to obtain the prefixed censoring rate p. To derive the value of k solve $P[D \leq T < \infty] = p$ (Wan, 2017).

6. Calculate $y_i = \min(t_i,c_i)$. If $t_i < c_i$ set the censoring indicator $\delta_i = 1$, otherwise set $\delta_i = 0$. So, clearly, those cured with $V_i = 0$ have $\delta_i = 0$.

For the simulation study the parameter values are fixed at $a = 2$, $b = 4$ and $\beta = 0.6$ with censoring proportions 20% and 40%. The simulation study was carried out 10000 times. For each simulated data set, we obtain the maximum likelihood estimates and Bayes estimates of the model parameters a, b, and β, along with the corresponding bias, mean square error and coverage probability. The mean square errors is given by

$$\text{MSE} = \frac{1}{R}\sum_{i=1}^{R}(\hat{\theta}_i-\theta)^2, \tag{29}$$

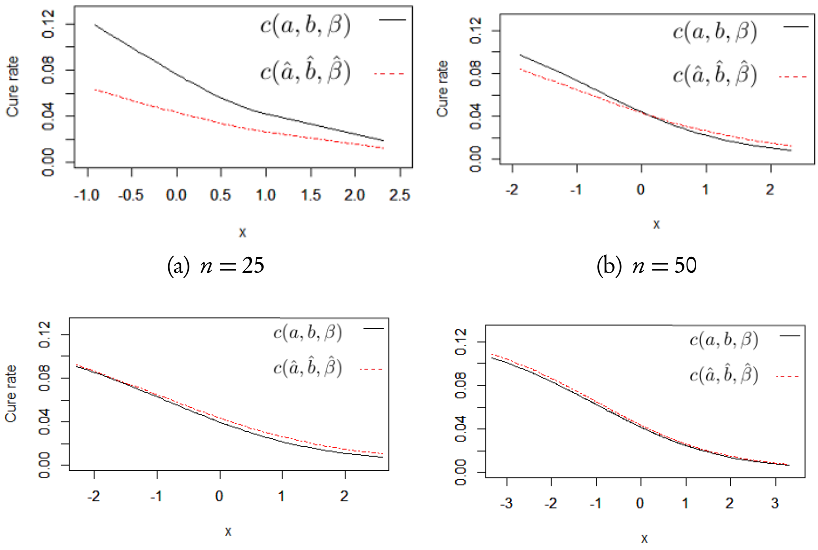
where $\hat{\theta} = \frac{1}{R}\sum_{i=1}^{R}\hat{\theta}_i$ and $\hat{\theta}_i$ be the i^{th} sample estimates of θ. The coverage probability of an estimator is the proportion of times the corresponding 95% asymptotic confidence intervals contain the parameter's actual parameter value. Four different choices of the sample size $n = 25, 50, 100$, and 500 are taken to study the behaviour of the estimates with a change in the sample size.

The results of maximum likelihood estimates and Bayes estimates with informative and reference priors are given in Table 2, Table 3, and Table 4, respectively. Now, we provide the following observations from the result presented in these Tables.

1. Table 2 provides the maximum likelihood estimator for $\theta = (a,b,\beta)$. The biases and mean square error(MSE) decreases as the sample sizes increases. We also observed that the rate of convergence improved with increasing sample size. Figure 2 and 3 show the estimated cure rate for 20% and 40% censoring respectively. From these Figures we observe that as sample size increases the empirical cure rate value $c(\hat{a},\hat{b},\hat{\beta})$ coincide with $c(a,b,\beta)$.

TABLE 2
Bias and MSE of ML estimates in GTCGC model with $(a, b, \beta) = (2, 4, 0.6)$.

Censoring Percentage		Sample Size			
		25	50	100	500
20	$\hat{a}$	2.095	1.891	2.043	1.967
	Bias	0.998	-0.108	0.043	-0.032
	MSE	1.521	0.534	0.372	0.055
	Coverage Probability	0.871	0.899	0.921	0.964
	$\hat{b}$	4.291	3.762	4.064	4.036
	Bias	0.291	-0.237	0.064	0.036
	MSE	2.286	0.620	0.295	0.056
	Coverage Probability	0.825	0.891	0.960	0.959
	$\hat{\beta}$	0.913	0.715	0.747	0.670
	Bias	0.313	0.115	0.147	0.070
	MSE	0.310	0.203	0.177	0.062
	Coverage Probability	0.828	0.842	0.903	0.956
40	$\hat{a}$	2.783	2.418	2.036	1.998
	Bias	0.783	0.418	0.036	-0.001
	MSE	1.667	1.380	0.624	0.076
	Coverage Probability	0.891	0.912	0.962	0.970
	$\hat{b}$	4.663	4.132	4.120	4.024
	Bias	0.663	0.132	0.120	0.024
	MSE	1.674	0.683	0.316	0.060
	Coverage Probability	0.888	0.901	0.971	0.981
	$\hat{\beta}$	0.874	0.672	0.653	0.606
	Bias	0.274	0.072	0.053	0.006
	MSE	0.315	0.243	0.214	0.034
	Coverage Probability	0.810	0.903	0.965	0.978



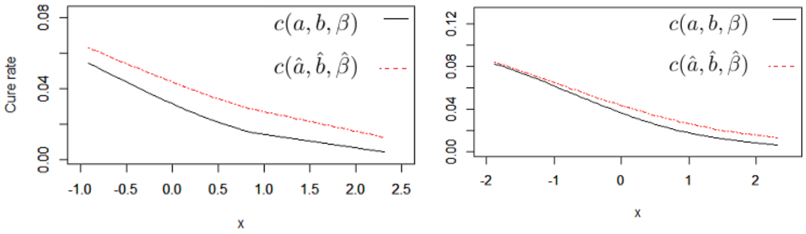
(a) $n = 25$ (b) $n = 50$

(c) $= 100$ (d) $= 500$

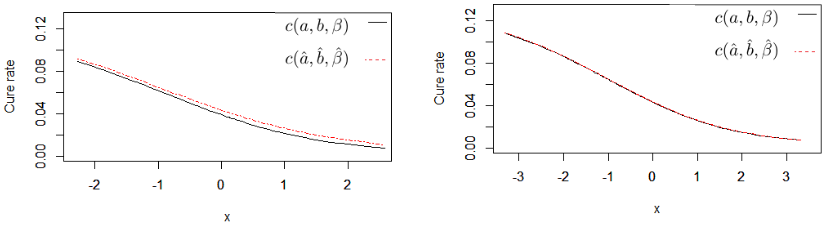
Figure 2 – Plots for cure rate for sample sizes $n = 25$, 50, 100 and 500 under 20% censoring.

2. Table 3, and Table 4 provide the Bayes estimates which obtained from informative and flat reference prior, respectively. The prior distributions were selected as Gamma$(0.1, 5)$, Gamma$(0.6, 7)$, and Normal$(0, 100)$, respectively. Randomly chosen parameter values were tested for the gamma distribution, and the results for Gamma$(0.5, 10)$, Gamma$(0.6, 7)$ are reported here. This helped rule out prior sensitivity and the Normal$(0, 100)$ was chosen for β to admit for a large variance. The simulation study shows that the biases and MSE's decreases as the sample sizes increases. Also the posterior is not prior sensitive. The prior robustness is evident from Table 3 and Table 4.

Note that the simulation study for GTCGC model with $N = 2$ is detailed in Asha and Soorya (2021).



(a) $n = 25$ (b) $n = 50$



(c) $= 100$ (d) $= 500$

Figure 3 – Plots for cure rate for sample sizes $n = 25$, 50, 100 and 500 under 40% censoring.

TABLE 3
Bias and MSE of Bayes estimates using informative prior in GTCGC model with $(a,b,\beta)=(2,4,0.6)$.

Censoring Percentage		Sample Size			
		25	50	100	500
20%	$\hat{a}$	1.700	2.860	2.195	2.121
	Bias	-0.299	0.860	0.195	0.121
	MSE	0.089	0.741	0.038	0.014
	$\hat{b}$	3.188	3.499	3.523	3.726
	Bias	-0.811	-0.500	-0.476	-0.273
	MSE	0.658	0.250	0.226	0.074
	$\hat{\beta}$	0.700	0.676	0.687	0.683
	Bias	0.100	0.076	0.087	0.083
	MSE	0.010	0.005	0.007	0.006
40%	$\hat{a}$	1.033	1.065	1.348	1.803
	Bias	-0.966	-0.934	-0.651	-0.196
	MSE	0.934	0.873	0.423	0.038
	$\hat{b}$	3.145	3.191	3.498	3.615
	Bias	-0.854	-0.808	-0.501	-0.384
	MSE	0.729	0.653	0.251	0.147
	$\hat{\beta}$	0.690	0.688	0.679	0.678
	Bias	0.090	0.088	0.079	0.078
	MSE	0.008	0.007	0.006	0.006

TABLE 4

Bias and MSE of Bayes estimates using reference prior in GTCGC model with $(a,b,\beta)=(2,4,0.6)$.

Censoring Percentage		Sample size			
		25	50	100	500
20%	$\hat{a}$	1.545	1.674	1.787	1.986
	Bias	-0.455	-0.326	-0.213	-0.014
	MSE	0.337	0.314	0.277	0.075
	$\hat{b}$	3.629	3.759	3.900	4.008
	Bias	-0.371	-0.241	-0.100	0.008
	MSE	0.417	0.262	0.158	0.064
	$\hat{\beta}$	0.505	0.514	0.529	0.586
	Bias	-0.095	-0.086	-0.071	-0.014
	MSE	0.015	0.017	0.022	0.023
40%	$\hat{a}$	1.606	1.788	1.937	2.024
	Bias	-0.394	-0.212	-0.063	0.024
	MSE	0.450	0.293	0.166	0.046
	$\hat{b}$	3.694	3.843	3.945	4.017
	Bias	-0.306	-0.157	-0.055	0.017
	MSE	0.338	0.194	0.143	0.058
	$\hat{\beta}$	0.511	0.517	0.524	0.583
	Bias	-0.089	-0.083	-0.076	-0.017
	MSE	0.014	0.018	0.022	0.024

8. ANALYSIS OF BLOOD AND MARROW TRANSPLANTATION DATA

In this Section, we analyse the data consisting of 8966 patients who have undergone bone marrow transplantation at the European Society for Blood and Marrow Transplantation (EBMT). For severe haematopoietic system problems, such as acute and chronic leukaemia, haematopoietic stem cell transplantation (HSCT) is an effective and standard treatment. Fiocco *et al.* (2005) analysed the data using the reduced rank proportional hazard model for competing risks. Data are available in the `msatate` package in `R` (Putter *et al.*, 2016). The observed time (T) is the time from transplantation to death or last follow-up ranging from 0.001 to 211.372 months. The patients still alive at the end of the study are considered as censored observations. We give a deeper analysis of the same using the GTC model.

We fit the data for the destructive negative binomial cure rate model with Weibull baseline (Rodrigues *et al.*, 2011), Generalized exponential cure rate model (Kannan *et al.*, 2010), Defective Gompertz cure rate model (Rocha *et al.*, 2017) and the proposed GTC model. After transplantation leukaemia patients are at risk of death from either of or some of or all of the listed causes which are: (i) operation, (ii) leukaemia relapse, (iii) chronic graft-versus-host disease (GvHD) (most common non-relapse complication caused by immune cells reacting against normal host tissues in the donor graft), (iv) infection, (v) haemorrhage, (vi) severe organ toxicity from high-dose chemotherapy given before the transplant. So each subject will experience either six competing causes or a combination of the causes of death. We consider defective Gompertz with $\bar{G}_0(t)$ in Eq.(20) and defective Dagum distribution (Martinez and Achcar, 2018) with $\bar{G}_0(t,\gamma,\eta,\alpha) = 1-\frac{\gamma\eta}{\eta+\gamma t^{-\alpha}}$, where $\alpha,\eta > 0$ and $0 < \gamma < 1$ as baseline distributions.

8.1. *Maximum likelihood estimation*

The maximum likelihood estimates and the corresponding standard errors (S.E), lower and upper asymptotic confidence intervals (LCL, UCL) are presented in Table 5. The LCL and UCL of the maximum likelihood estimates of the parameter θ is $\hat{\theta} - \text{S.E.}_{\hat{\theta}}\, z_{\alpha/2}$ and $\hat{\theta} + \text{S.E.}_{\hat{\theta}}\, z_{\alpha/2}$ respectively, where $\hat{\theta}$ is the maximum likelihood estimate and $\text{S.E.}_{\hat{\theta}}$ is the standard error of the parameter θ. The standard error of the parameter θ is the square root of the variance, which is obtained from the Fisher information matrix (Patefield, 1977) by plugging the estimated parameter values of θ. The model comparison by the AIC criteria evidences the GTC model (Model 4) in Table 5 for $m = 6$ and a defective Dagum baseline distribution is a good fit.

We further analysis the data using the model 4 in Table 5 in-cooperating the covariates. The covariates of the data are disease group, donor-recipient gender match, T-cell depletion and age group. It was observed that the covariates donor-recipient gender match and T-cell depletion were insignificant. An analysis was conducted using the remaining covariates namely disease sub classification with levels acute myeloid leukemia (AML),chronic myeloid leukemia (CML), acute lymphoid leukemia (ALL)

TABLE 5
Maximum likelihood estimation for different cure rate model.

Model	Parameter	Estimate	S.E	LCL	UCL	AIC
1. GTC model with defective Gompertz as baseline($m = 6$)	a	0.72	0.026	0.661	0.767	18094
	b	0.22	0.008	0.206	0.237	
	p	0.16	0.002	0.153	0.160	
2. Defective Gompertz cure rate model	a	0.74	0.012	0.735	0.784	18136
	b	0.38	0.009	0.377	0.416	
3. GTC model with defective Gompertz as baseline($m = 2$)	a	0.73	0.021	0.692	0.772	18111
	b	0.29	0.008	0.273	0.306	
	p	0.37	0.001	0.365	0.370	
4. GTC model with defective Dagum distribution as baseline($m = 6$)	γ	0.20	0.093	0.120	0.488	17423
	η	0.32	0.218	0.141	0.999	
	α	1.21	0.019	1.188	1.263	
	p	0.31	0.147	-0.147	0.431	
5. GTC model with defective Dagum distribution as baseline($m = 5$)	γ	0.17	0.053	0.168	0.379	17594
	η	0.28	0.105	0.245	0.659	
	α	1.21	0.015	1.173	1.233	
	p	0.46	0.094	0.017	0.387	
6. Destructive negative Binomial cure rate model with Weibull baseline	ϕ	4.83	1.021	2.823	6.826	17632
	ω	2.10	0.877	0.381	3.818	
	γ_1	1.09	0.540	0.036	2.149	
	γ_2	-1.21	0.491	0.326	2.150	
7. Mixture generalized Exponential cure rate model	α	0.73	0.010	0.699	0.741	18160
	λ	0.65	0.025	0.580	0.681	
	p	0.60	0.005	0.589	0.612	

and age group with levels below 20 , between 20-40 and above 40.

We adopt the usual indicator variable coding methodology and construct the dummy variables as follows:

$$\begin{aligned}
x_{1,\text{dissub:AML}} &= 1 \text{ if patient is in AML disease group,} && 0 \text{ otherwise,}\\
x_{1,\text{dissub:CML}} &= 1 \text{ if patient is in CML disease group,} && 0 \text{ otherwise,}\\
x_{2,\text{age:}\leq 20} &= 1 \text{ if patient is in below 20 age group,} && 0 \text{ otherwise,}\\
x_{2,\text{age:}20-40} &= 1 \text{ if patient is in between 20-40 age group,} && 0 \text{ otherwise.}
\end{aligned}$$

The estimates are presented in Table 6 and the cure rate in Eq. (12) of the blood and marrow transplanted subjects are obtained as

$$c_6(x,\theta) = 0.839(1 - p(x,\beta) + 0.839\, p(x,\beta))^5, \tag{30}$$

where

$$p(x,\beta) = \frac{e^{-0.247 + 0.113\, x_{1,\text{dissub:CML}} + 0.247\, x_{1,\text{dissub:AML}} - 0.799\, x_{2,\text{age:}\leq 20} - 0.106\, x_{2,\text{age:}20-40}}}{1 + e^{-0.247 + 0.113\, x_{1,\text{dissub:CML}} + 0.247\, x_{1,\text{dissub:AML}} - 0.799\, x_{2,\text{age:}\leq 20} - 0.106\, x_{2,\text{age:}20-40}}}.$$

We observe from the Table 5 that in all disease groups, the patients belonging to the age group below 20 show a higher cure rate than those who belong to the age group above 20. The patients with ALL disease and the age group below 20 (0.678) show a high cure rate, while those with AML disease and those above 40 (0.55) show a lower cure rate. Among the disease group, the AML group have a lower cure rate than the CML and ALL diseases group. The patients who belong to the age group below 20 show a higher probability of survival compared to the other age group.

TABLE 6

Maximum likelihood estimates for GTC model with defective Dagum distribution as baseline where $\beta_{1,dissub:AML}$, $\beta_{1,dissub:CML}$, $\beta_{2,age:\leq 20}$, *and* $\beta_{2,age:20-40}$ *are the regression coefficients corresponds to the covariates* $x_{1,dissub:AML}$, $x_{1,dissub:CML}$, $x_{2,age:\leq 20}$, *and* $x_{2,age:20-40}$, *respectively.*

Parameter	Estimate	S.E	$\frac{\lvert\text{Estimate}\rvert}{\text{S.E}}$	LCL	UCL
γ	0.16	0.016		0.129	0.193
η	0.26	0.029		0.203	0.316
α	1.22	0.030		1.159	1.277
β_0	-0.25	0.087	2.847	-0.417	-0.077
$\beta_{1,\text{dissub:CML}}$	0.11	0.032	3.496	0.050	0.177
$\beta_{1,\text{dissub:AML}}$	0.25	0.099	2.486	0.052	0.442
$\beta_{2,\text{age:}\leq 20}$	-0.80	0.370	2.159	-1.525	-0.074
$\beta_{2,\text{age:}20-40}$	-0.11	0.041	2.556	-0.186	-0.025

8.2. *Algorithm validation using bootstrap*

The bootstrap technique is used to check the accuracy of algorithm. Here we draw B bootstrap samples with replacement from the data and each bootstrap sample consist of 8966 observations. We obtain the the bootstrap estimate for each bootstrap sample, say $\hat{\theta}_b$, where $\hat{\theta}_b$ be the b^{th} bootstrap sample estimate of the parameter θ, $b = 1, 2, \ldots, B$ (Efron and Hastie, 2016). The estimated values are presented in Table 7 thereby confirming our claim.

TABLE 7
Bootstrap estimates for GTC cure rate models with defective Dagum baseline where $\beta_{1,dissub:AML}$, $\beta_{1,dissub:CML}$, $\beta_{2,age:\leq 20}$, *and* $\beta_{2,age:20-40}$ *are the regression coefficients corresponds to the covariates* $x_{1,dissub:AML}$, $x_{1,dissub:CML}$, $x_{2,age:\leq 20}$, *and* $x_{2,age:20-40}$, *respectively.*

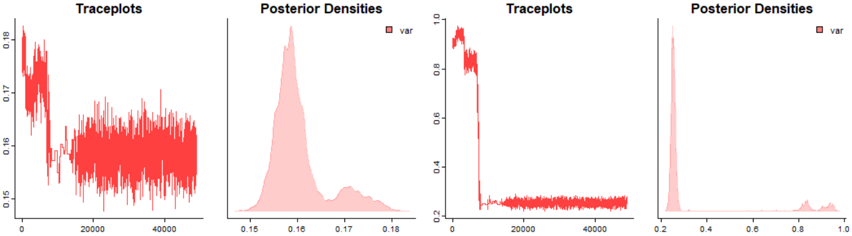
Parameter	Estimate	S.E.	$\frac{\lvert\text{Estimate}\rvert}{\text{S.E.}}$	LCL	UCL
γ	0.14	0.021		0.100	0.184
η	0.20	0.032		0.133	0.258
α	1.56	0.031		1.501	1.623
β_0	-0.21	0.089	2.302	-0.379	-0.030
$\beta_{1,\text{dissub:CML}}$	0.14	0.045	3.165	0.055	0.232
$\beta_{1,\text{dissub:AML}}$	0.27	0.097	2.756	0.077	0.457
$\beta_{2,\text{age:}\leq 20}$	-0.81	0.311	2.613	-1.423	-0.203
$\beta_{2,\text{age:}20-40}$	-0.12	0.050	2.365	-0.217	-0.020

TABLE 8
Bayesian estimates for GTC cure rate models with defective Dagum baseline.

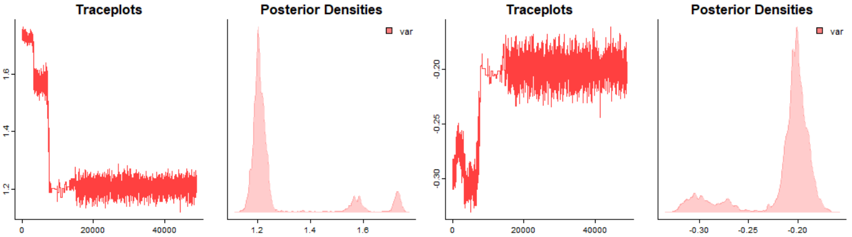
Parameter	Posterior mean(S.E)	Posterior mode(S.E)	95% Credible intervals	
			0.025	0.095
γ	0.16(0.042)	0.16(0.033)	0.153	0.176
η	0.35(0.045)	0.25(0.110)	0.235	0.938
α	1.27(0.361)	1.20(0.493)	1.168	1.732
β_0	-0.22(0.091)	-0.20(0.080)	-0.310	-0.181
$\beta_{1,\text{dissub:CML}}$	0.14(0.008)	0.14(0.009)	0.108	0.160
$\beta_{1,\text{dissub:aML}}$	0.28(0.012)	0.25(0.031)	0.245	0.353
$\beta_{4,\text{age:}\leq 20}$	-0.80(0.201)	-0.79(0.313)	-0.841	-0.773
$\beta_{4,\text{age:}20-40}$	-0.10(0.005)	-0.10(0.002)	-0.123	-0.032

8.3. *Bayesian analysis*

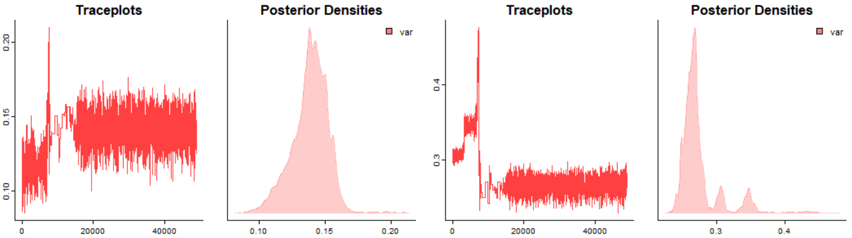
In the most of the clinical trials information regarding the experiment is available through the previous study or expert knowledge. It is only imperative that this information be



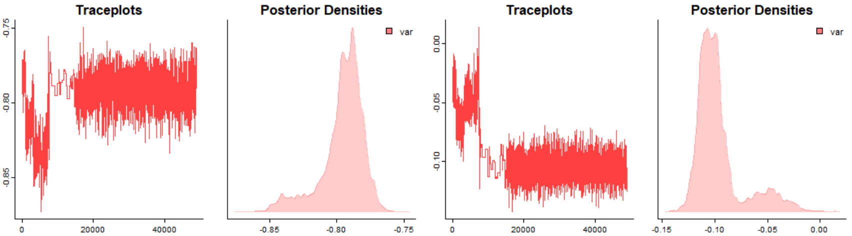
(a) γ (b) η



(c) α (d) β_0

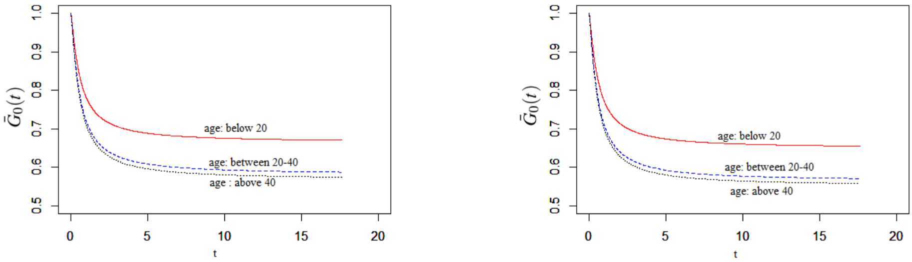


(e) $\beta_{1,\text{dissub:CML}}$ (f) $\beta_{1,\text{dissub:AML}}$



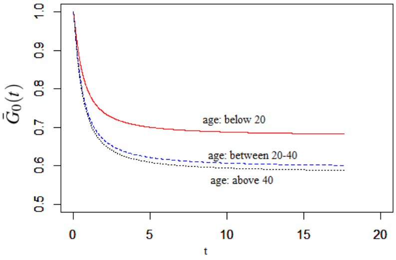
(g) $\beta_{4,\text{age:}\leq 20}$ (h) $\beta_{4,\text{age:20-40}}$

Figure 4 – Trace and density plots of parameters $\gamma, \eta, \alpha, \beta_0, \beta_{1,\text{dissub:CML}}, \beta_{1,\text{dissub:AML}}, \beta_{4,\text{age:20-40}}$ and $\beta_{4,\text{age:}\leq 20}$ in GTC cure rate model with defective Dagum distribution.



(a) Survival function of GTC model by age group (CML patients).

(b) Survival function of GTC model by age group (AML patients).



(c) Survival function of GTC model by age group (ALL patients).

Figure 5 – Plots of survival function stratified by age group for different disease types (CML, AML, ALL).

used in data analysis. This motivates us to investigate Bayesian estimation procedures. We perform Bayesian analysis on this data set. Since we do not have much prior information regarding this particular data we relay on non informative prior Uniform$(0,1)$ for the parameter γ. The flexible gamma distribution priors, Gamma$(0.1,10)$ and Gamma$(0.5,8)$, were chosen for η and α. The parameters were randomly chosen by considering naive information we obtained from the initial exploration of the data. Normal$(0,100)$ was chosen for β_0, $\beta_{1,\text{dissub:CML}}$, $\beta_{1,\text{dissub:AML}}$, $\beta_{2,\text{age:}\leq 20}$ and $\beta_{2,\text{age:}20-40}$ to admit for a large variance indicating near non information. Using Metropolis-Hastings algorithm in Section 6, we sample from the posterior distribution and the posterior summaries of interest will be obtained by using R software. The plots of marginal posterior densities are presented in Figure 4. From the plots we observe that that the chains begin to converge after approximately 20,000 iterations. To mitigate the influence of initial values and ensure that the posterior estimates are based on the stationary distribution, the first 20,000 samples were discarded as burn-in. Also we observe that the posterior densities are symmetric, so we calculate the 95% credible intervals (CI) which is also a high posterior density (HPD) interval. The standard error of the Bayesian estimates are calculated by Bootstrap algorithm in Section 6.2. The Bayesian estimates for the parameters and the corresponding S.E, CI (HPD) intervals and for the GTC model with Dagum baseline distribution is presented in Table 8.

9. DISCUSSION

We proposed a general class of transmuted cure rate models motivated by the rank transmutation of improper distribution. The identifiability issues of the model are discussed by appealing to the result in Li *et al.* (2001). The defective cure rate model and the shifted Bernoulli cure rate model can be considered special cases of the GTC model with $m = 1,2$ respectively. The cure fraction of the population modelled by the GTC model is obtained as $c_0(1-p+p\,c_0)^{m-1}$ where c_0 is the limiting value of baseline distribution. The parameter p manages the skewness of the proposed model motivates us to include covariates through p. The example of the proposed model is illustrated through the defective Gompertz distributions as a baseline distribution. The method of maximum likelihood estimation and Bayesian approaches are shown to work effectively in a synthetic data.

The application of the model is illustrated using the blood and marrow transplanted data. The survival probability of the data is modelled through some existing models and the GTC model with defective Gompertz and defective Dagum baseline distributions. The analysis results show that the GTC model with six competing risks and defective Dagum baseline distribution fits the data well. To study the effect of covariates on the long-term survival probability of blood and marrow transplanted patients, the parameter $p(\cdot)$ is linked with the covariates disease sub-classification, donor-recipient gender match, T-cell depletion and age group. The covariates gender match of donor-recipient and T-cell depletion showed no significant effect on the survival probability, while the disease sub-classification and age group are substantial.

From Figure 5 we observe that irrespective of the disease sub-group, the probability of cure of a below 20 years of age group patient has a 14% and 17% higher probability of cure than the patient in between 20-40 age group and above 40 age group, respectively. We also observe that in all age groups, the patient with AML disease shows a 4% less cure than the ALL patient, while the CML patients and AML patients did not show any significant change in their cure rate for the different age groups. This data is well studied by Fiocco *et al.* (2005) using the reduced rank proportional hazard model for competing risks. They found that a patient's age at transplantation strongly influences their death, which is consistent with our findings.

ACKNOWLEDGEMENTS

The authors disclosed the receipt of the following financial support for the research, authorship, and/or publication of this article. This work was supported by the Cochin University of Science & Technology through its USRF scheme.
We acknowledge the editor and the two reviewers for constructive comments and insightful suggestions, which have helped to improve the clarity and rigor of our manuscript.

APPENDIX

A. PROOF OF THEOREM 1

PROOF. Observe,

$$h'_N(t,x,\theta) = h'_0(t,\theta_0)\left[1+(N-1)\frac{p(x,\beta)\bar{G}_0(t\theta_0)}{1-p(x,\beta)+p(x,\beta)\bar{G}_0(t,\beta)}\right] - h_0(t,\theta_0)$$
$$(N-1)\left[\frac{g_0(t,\theta_0)p(x,\beta)(1-p(x,\beta))}{1-p(x,\beta)+p(x,\beta)\bar{G}_0(t,\theta_0)}\right] \tag{31}$$

where $h_N^{'}(t,x,\theta)$ and $h_0^{'}(t,\theta_0)$ are the first derivatives of $h_N(t,x,\theta)$ and $h_0(t,\theta_0)$ respectively with respect to t.
Let assume that $\bar{G}_0(t,\theta_0)$ has a decreasing failure rate, then from Eq. (31) it is now straightforward that $\bar{G}(t,x,\theta)$ has a decreasing failure rate. □

B. PROOF OF THEOREM 2

PROOF. Differentiate Eq. (14) with respect to N,

$$\frac{\partial h_N(t,x,\theta)}{\partial N} = h_0(t,\theta_0)\frac{p\,\bar{G}_0(t,\theta_0)}{1-p(x,\beta)+p(x,\beta)\,\bar{G}_0(t,\theta_0)}. \tag{32}$$

From Eq. (32), it is obvious that $\frac{\partial h_N(t)}{\partial N} > 0$, that is the failure rate is an increasing function of N. □

C. PROOF OF THEOREM 5

PROOF. If $U(t,x,\theta)$ is identifiable, then

$$\begin{aligned} & U(t,x,\theta) = U^*(t,x,\theta^*) \\ \Longrightarrow\ & p(x,\beta) = p(x,\beta^*) \text{ and } \bar{G_0}(t,\theta_0) = \bar{G_0^*}(t,\theta_0 *) \end{aligned}$$

which in turn implies $\bar{G_N}(t,x,\theta) = \bar{G_N^*}(t,x,\theta^*)$. Hence the proof. □

REFERENCES

G. ASHA, C. S. SOORYA (2021). *On a class of transmuted distributions to model survival data with a cure fraction*. Calcutta Statistical Association Bulletin, 73, no. 2, pp. 106–126.

N. BALAKRISHNAN, S. BARUI, F. MILIENOS (2024). *Piecewise linear approximations of baseline under proportional hazards based COM-Poisson cure models*. Communications in Statistics-Simulation and Computation, 53, no. 3, pp. 1135–1160.

N. BALAKRISHNAN, S. PAL (2013). *Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on com-Poisson family*. Computational Statistics & Data Analysis, 67, pp. 41–67.

N. BALAKRISHNAN, S. PAL (2015). *An EM algorithm for the estimation of parameters of a flexible cure rate model with generalized Gamma lifetime and model discrimination using likelihood-and information-based methods*. Computational Statistics, 30, no. 1, pp. 151–189.

J. BERKSON, R. P. GAGE (1952). *Survival curve for cancer patients following treatment*. Journal of the American Statistical Association, 47, no. 259, pp. 501–515.

J. W. BOAG (1949). *Maximum likelihood estimates of the proportion of patients cured by cancer therapy*. Journal of the Royal Statistical Society. Series B (Methodological), 11, no. 1, pp. 15–53.

A. B. CANTOR (2001). *Projecting the standard error of the Kaplan-Meier estimator*. Statistics in Medicine, 20, no. 14, pp. 2091–2097.

C. CHIVERS (2011). *MHadaptive: general Markov Chain Monte Carlo for Bayesian Inference using adaptive Metropolis-Hastings sampling*. R package version 1.1-8.

D. COX, D. OAKES (1984). *Analysis of Survival Data*. Chapman and Hall, London.

D. R. Cox, O. E. Barndorff-Nielsen (1994). *Inference and Asymptotics*, vol. 52 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton.

B. Efron, T. Hastie (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, vol. 5. Cambridge University Press, Cambridge.

V. T. Farewell (1982). *The use of mixture models for the analysis of survival data with long-term survivors*. Biometrics, 38, no. 4, pp. 1041–1046.

M. Fiocco, H. Putter, J. van Houwelingen (2005). *Reduced rank proportional hazards model for competing risks*. Biostatistics, 6, no. 3, pp. 465–478.

D. Granzotto, F. Louzada, N. Balakrishnan (2017). *Cubic rank transmuted distributions: inferential issues and applications*. Journal of statistical Computation and Simulation, 87, no. 14, pp. 2760–2778.

L. Hanin, L.-S. Huang (2014). *Identifiability of cure models revisited*. Journal of Multivariate Analysis, 130, pp. 261–274.

J. G. Ibrahim, M.-H. Chen, D. Sinha (2001). *Bayesian semiparametric models for survival data with a cure fraction*. Biometrics, 57, no. 2, pp. 383–388.

N. Kannan, D. Kundu, P. Nair, R. Tripathi (2010). *The generalized exponential cure rate model with covariates*. Journal of Applied Statistics, 37, no. 10, pp. 1625–1636.

C.-S. Li, J. M. Taylor, J. P. Sy (2001). *Identifiability of cure models*. Statistics & Probability Letters, 54, no. 4, pp. 389–395.

E. Z. Martinez, J. A. Achcar (2018). *A new straightforward defective distribution for survival analysis in the presence of a cure fraction*. Journal of Statistical Theory and Practice, 12, no. 4, pp. 688–703.

W. Patefield (1977). *On the maximized likelihood function*. Sankhyā: The Indian Journal of Statistics, Series B, 39, pp. 92–96.

H. Putter, L. de Wreede, M. Fiocco, M. Putter (2016). *mstate: An R Package for the Analysis of Competing Risks and Multi-State Models*. R package version 0.3.1. URL `https://cran.r-project.org/package=mstate`.

M. M. Rahman, B. Al-Zahrani, M. Q. Shahbaz (2018). *A general transmuted family of distributions*. Pakistan Journal of Statistics and Operation Research, 14, no. 2, pp. 451–469.

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

J. RODRIGUES, V. G. CANCHO, M. DE CASTRO, N. BALAKRISHNAN (2012). *A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data*. Statistical Methods in Medical Research, 21, no. 6, pp. 585–597.

J. RODRIGUES, M. DE CASTRO, N. BALAKRISHNAN, V. G. CANCHO (2011). *Destructive weighted Poisson cure rate models*. Lifetime Data Analysis, 17, no. 3, pp. 333–346.

J. RODRIGUES, M. DE CASTRO, V. G. CANCHO, N. BALAKRISHNAN (2009). *COM–Poisson cure rate survival models and an application to a cutaneous melanoma data*. Journal of Statistical Planning and Inference, 139, no. 10, pp. 3605–3611.

A. TSODIKOV, J. IBRAHIM, A. YAKOVLEV (2003). *Estimating cure rates from survival data: an alternative to two-component mixture models*. Journal of the American Statistical Association, 98, no. 464, pp. 1063–1078.

F. WAN (2017). *Simulating survival data with predefined censoring rates for proportional hazards models*. Statistics in medicine, 36, no. 5, pp. 838–854.

A. Y. YAKOVLEV, B. ASSELAIN, V. BARDOU, A. FOURQUET, T. HOANG, A. ROCHEFEDIERE, A. TSODIKOV (1993). *A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer*. Biometrie et Analyse De Donnees Spatio-Temporelles, 12, pp. 66–82.