Journal of Biostatistics and Epidemiology

*J Biostat Epidemiol. 2022;8(3):304-313*

Original Article

# Joint Frailty Model of Recurrent and Terminal Events in the Presence of Cure Fraction using a Bayesian Approach

Zahra Arab Borzu, PhD[1*], Ahmad Reza Baghestani, PhD[2], Elaheh Talebi Ghane, PhD[3], Ali Akbar Khadem Maboudi, PhD[2], Ali Akhavan, MD[4], Anahita Saeedi, MSc[5]

[1]Department of Biostatistics, School of Health, Zahedan University of Medical Sciences, Zahedan, Iran.
[2]Department of Biostatistics, Physiotherapy Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[3]Department of Biostatistics, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran.
[4]Assistant Professor of Radiation Oncology, Isfahan University of Medical Sciences, Isfahan, Iran.
[5]Department of Biostatistics, School of Public Health & Health Sciences, University of Massachusetts, Amherst, MA, USA.

ARTICLE INFO

Received 08.05.2022
Revised 12.06.2022
Accepted 11.07.2022
Published 15.10.2022

**Key words:**
Bayesian approach;
Breast cancer;
Cure fraction;
Joint frailty model;
Recurrent event.

ABSTRACT

**Introduction:** Recurrent event data are common in many longitudinal studies. Often, a terminating event such as death can be correlated with the recurrent event process. A shared frailty model applied to account for the association between recurrent and terminal events. In some situations, a fraction of subjects experience neither recurrent events nor death; these subjects are cured.
**Methods:** In this paper, we discussed the Bayesian approach of a joint frailty model for recurrent and terminal events in the presence of cure fraction. We compared estimates of parameters in the Frequentist and Bayesian approaches via simulation studies in various sample sizes; we applied the joint frailty model in the presence of cure fraction with Frequentist and Bayesian approaches for breast cancer.
**Results:** In small sample size Bayesian approach compared to Frequentist approach had a smaller standard error and mean square error, and the coverage probabilities close to nominal level of 95%. Also, in Bayesian approach, the sampling means of the estimated standard errors were close to the empirical standard error.
**Conclusion:** The simulation results suggested that when sample size was small, the use of Bayesian joint frailty model in the presence of cure fraction led to more efficiency in parameter estimation and statistical inference.

## Introduction

In biomedical research, the event of interest may occur more than once per subject. Such events have been termed as "recurrent events". Tumor recurrences, recurrent heart attacks, and successive hospitalizations are some examples. Different methods for analyzing

*.Corresponding Author: z_arabborzoo@yahoo.om

recurrent event data were reviewed in the literature.[1-6] In many situations, a terminating event such as death can be correlated with the recurrent event process. For example, the recurrent of an event like successive hospitalizations can increase the risk of death. In this context, the common assumption of non-informative censoring of the recurrent event process by death is violated. Therefore, this dependence should be accounted in the joint modeling of recurrent events and death, and ignoring this association leads to biased estimates.[7] One common approach to solve this problem is frailty models. In frailty models, the correlation between the recurrent and terminal events is entered to the model through a latent variable. So that two event processes are independent given frailty.[8-11] In recent years, the development of new drugs and treatments has resulted in a significant number of patients to be disease free from recurrences or death after a first treatment, hence they are considered as cured. This can be achieved through Kaplan–Meier survival curve which shows a long and stable plateau with many large censored observations at the tail. Ignoring cured fraction lead to overestimation of the probability of survival because it assumes that all subjects will experience the event of interest while the follow-up period is sufficiently long.[12]

Thus the cure models are becoming increasingly important and popular for analyzing survival data in which some subjects may never experience event. More recently, several studies proposed cure rate to analysis recurrent events. Yu et al. (2008) proposed a frailty mixture cure model for hospital readmission data in which subjects with a recurrence did not belong to the cure group 13. Rondeau et al. (2013) presented a flexible model for recurrent events with cure fraction.[12] There are numerous studies that proposed joint frailty models for recurrent events and death with Frequentist and Bayesian approaches. For example, Liu et al. (2004) proposed a joint semi-parametric model for the intensity function of recurrent events and terminal event by a shared gamma frailty that allows different effects of frailty on recurrent event and death 9 . In their study, maximum likelihood estimation was carried out through a Monte Carlo EM algorithm with Metropolis-Hastings sampler in the E-step. One disadvantage of this method was the sensitivity to the starting value and thus a heavy computation which was time consuming. In order to solve these problems Sinha et al. (2008) proposed a Bayesian method based on Markov Chain Monte Carlo[14] . Paulon et al. (2018) considered a joint model for recurrent events and terminal event with nonparametric Bayesian framework.[15] Talebi et al. (2018) proposed a Bayesian joint frailty model for recurrent and terminal events. They employed the inter-recurrence dependence and the dependence between the two events.[16] Liu et al. (2016) proposed a joint frailty model for zero inflation recurrent event and death with Frequentist approach.[17] According to published manuscripts, no study considered these models with cured fraction from the Bayesian point of view. The purpose of this study on the one hand, was to estimate the effect of covariates on the times of recurrence, death and the cure model and on the other hand, to estimate the effect of small sample size in parameters estimate in both Frequentist and Bayesian approaches. This paper is organized as follows: in section 2, we recall the joint

frailty model for recurrent and death in the presence of a cure fraction. In section 3, we compare the results of the simulation study of the cure joint frailty model for Frequentist and Bayesian approaches in different sample sizes. In section 4, we focus on the application of the proposed model with Bayesian approach for a real dataset. Finally, discussion is presented in section 6.

## The model

### Notations

We defined notations and definitions that is used in the model. Let $T_{ij} = min(X_{ij}, C_i, D_i)$ be the observed follow-up time so that $X_{ij}$, $C_i$ and $D_i$ be $j$ th recurrent event time for $i$ th subject, $(i = 1, ..., N, j = 1, ..., n_i)$ right-censoring time and the death time, respectively. Similarly, the last follow-up time for subject $i$ denote by $T^*_i = min(C_i, D_i)$ which is a time of death or censoring. We considered a binary indicator for recurrent event as $\delta_{ij} = I(T_{ij} = X_{ij})$ so that if $n_i > 0$ then $\delta_{ij} = 1$ and a binary indicator for terminal event as $\Delta_i = I(T^*_i = D_i)$. Where is $I(.)$ the indicator function, and $S_{ij}$ indicates the gap time (time interval from previous to next recurrent event), so that $S_{ij} = T_{ij} - T_{i(j-1)}$ are independent.

The observation for subject $i$ is $O_i(t) = \{S_{ij}, T^*_i - \delta_{ij}, \Delta_i\}$. Based on the theory of multivariate counting processes.[9,18] and $N^{D^*}_i(t) = I(D_i \le t)$ are $N^D_i(t) = I(X_i \le t, \Delta_i = 1)$ the actual and observed death indicator by time t, respectively. Similarly, $N^{R^*}_i(t)$ and $N^R_i(t) = N^{R^*}_i(min(X_i, t))$ are the actual and observed number of recurrent events, respectively The number of recurrent events that occurs for subject $i$ in $[t, t + dt)$ is $dN^{R^*}_i(t) = N^{R^*}_i((t + dt)^-) - N^{R^*}_i(t^-)$ and $N^R_i(t) = Y_i(t)\, dN^{R^*}_j(t)$. Furthermore $Y_i(t) = 1_{(t \le T^*_i)}$ is the at risk indicator of subject $i$ at time $t$. The process history of subject $i$ up to time $t$, is represented as $H_{it} = \sigma\{Y(h), N^R_i(h), N^D_i(h), Z_i(h), \omega_i(h), 0 \le h \le t\}$. Where $Z_i(h)$ is the vector of covariates and $\omega_i(h)$ is frailty shared by recurrent event times for subject $i$. Furthermore, recurrent event processes, death and censoring times were assumed to be continuous. And the time of the occurrence of the recurrent and death was different. Occurring recurrent event at the same time with death, we considered that death has occurred first. The recurrent event and the death intensity processes at $t$ are $Y_i(t)\, h_i(t)dt = P(dN^R_i(t) = 1 \mid F_{it^-})$ and $Y_i(t)\, \lambda_i(t)dt = P(dN^D_i(t) = 1 \mid F_{it^-})$, respectively. Where $h_i(t)dt = P(dN^{R^*}_i(t) = 1 \mid Z_i(t), \omega_i, D_i \ge t)$, and $\lambda_i(t)dt = P(dN^{D^*}_i(t) = 1 \mid Zi(t), \omega_i, D_i \ge t)$.

### Estimation

To fit the proposed model, the maximum likelihood technique was used to estimate different parameters ($\beta$, $\beta^*$, $\theta$, $\alpha$) and the baseline hazard functions ($\lambda_0(t)$, $h_0(t)$) for recurrent and death times. Hence, $\Phi = (h_0(t), \lambda_0(t), \beta, \beta^*, \theta, \alpha)$ are all the parameters that should be estimated in modeling process. The marginal log-likelihood is:

$$l(\Phi) = \ln \prod_{i=1}^{N} \int_0^{\infty} L(O_i \mid \omega_i) g(\omega_i) d\omega_i$$

Where is $g(\omega_i)$ density function for shared frailty.

We modeled the event time distribution with a regression model using the Frequentist and Bayesian methods. In the Frequentist approach, the quasi-Newton procedure was

used to maximize the observed log likelihood[6] Since a simple form does not exist for the full log likelihood, we integrated out the frailties employing the Gauss–Laguerre quadrature. The using of another estimation methods for example, EM, MCEM and partial penalized likelihood (PPL) methods, lies in their implementation. So that, two methods are not readily available for joint frailty model in the presence of cure fraction in any software.

The details are provided in Appendix A. The $\hat{H}^{-1}$ was directly used as a variance estimator, where $H$ is the convergence Hessian matrix of the log likelihood. Due to positivity constraint of some parameters ($p_R$, $p_D$, $\theta > 0$), squared and exponential transformations are utilized and their standard errors are computed by the delta method.

In Bayesian approach, we assumed informative prior distributions for a set of parameters $\Phi = \{\beta, \beta^*, \alpha, \theta, p_D, p_R\}$. Normal distribution was assumed for the parameters of regression coefficients of joint model ($\beta, \beta^* \sim N(0, 1)$). Half-Cauchy priors with scale parameter 4 were adopted for the shape parameters of baseline hazard functions and variance frailty ($\theta, p_D, p_R$): half - cauchy.[4] One prior distribution for regression coefficients of cure model is Uniform. (α): Uniform(-2, 2). The resulting joint posterior for our model is given by

$$p(\beta,\beta^*,\alpha,\mathrm{p}_R,\mathrm{p}_D,\theta \mid O) \propto L(O)\times\pi(\beta)\times \pi(\beta^*)\times\pi(\alpha)\times\pi(\mathrm{p}_R)\times\pi(\mathrm{p}_D)\times\pi(\theta) \quad (7)$$

The parameter estimation is based on Laplace approximation with implementation in R software. Also, in order to have a more accurate comparison of Frequentist and Bayesian approaches, the non-informative prior distributions ($\beta, \beta^*$) ~ N(0,1000), ($\theta, p_D, p_R$) ~ gamma (0.01, 0.01) and (α): Uniform (-200, 200) were considered for parameters.

## Simulation

A simulation study was conducted to evaluate the performance of the estimators in a Bayesian framework for joint frailty model with cure fraction and to compare the estimators in Bayesian approach with Frequentist estimation in different sample sizes. We considered the right censored and used a calendar time scale for recurrent times. The effect of sample size on the performance of estimators was investigated by considering five sample sizes (20, 30, 50, 100, and 200) and generating 500 replicate datasets.

## Generating Data

For each subject $i$, we generated binary explanatory variables $Z_i(i=0.1)$, from a Bernoulli distribution with probability 0.5. The random variables was generated from gamma distribution so that $\omega_i : gamma(\frac{1}{\theta}, \frac{1}{\theta})$ with $\theta=0.5$. A fixed right-censoring time was taken as $C_i = 6 + Unif(0,6)$. We generated the gap times $S_{ij}$ from

$$\lambda_i(s_{ij} \mid \omega_i, A_i = 2) = \omega_i \lambda_0(t) \exp(\beta_1 Z_{1i} + \beta_2 Z_{2i})$$

where $\lambda_0(t) = 0.65t^{0.25}$ and death time $D_i$ generated from $h_i(t \mid \omega_i) = \omega_i h_0(t) \exp(\beta^* Z_{1i})$ where $h_0(t) = 0.4t^{0.25}$.

A death time $D_i$ was generated from the hazard function $h_i(t \mid \omega_i)$.

i) If observed time was a death time $D_i \leq C_i$ then $T_i^* = \mathrm{D}$ and $\Delta_i = 1$.
ii) If $D_i > C_i$ individual was censored then $T_i^* = C_i$ and $\Delta_i = 0$.

We applied a logistic regression for probability of cure so that:

$$p_i = \frac{1}{1+\exp(\alpha_0 + \alpha_1 Z_{i1})}$$

and set $\alpha_0 = -0.5$ and $\alpha_1 = 1$. We generated a random variable $u_i$ from uniform distribution [0,1]. If $u_i < p_i$ the individual was cured ($T^* = C_i$, $\delta_i^* = 0$). If $u_i < p_i$ Therefore, the recurrent gap times ($S_{ij}$) were formulate d from hazard function $\lambda_i(S_{ij} | \omega_i)$ and calendar times created from

$$T_{ij} = \min(C_i, D_i, \sum_{k=1}^{j} S_{ik})$$

i) If $T_{ij} < T_i^*$ then the observed time can be a recurrent event time and $\delta_{ij} = 1$.
generating continues until $T_{ij} < T_i^*$.

ii) If $T_{ij} \geq T_i^*$ individual was censored at $T_i^*$ and $\delta_{ij} = 0$.

We set $\beta_1 = 1$, $\beta_2 = -0.5$, $\beta_1^* = 0.7$, $\alpha_0 = -0.5$, $\alpha_1 = 1$.

**Results of the simulation study**

In this section, a simulation study of the Bayesian joint frailty model in the presence of cure fraction was performed with the emphasis on sample size. In doing so, we could compare the performance of estimators with Frequentist approach, which was introduced by Liu et al. (2016) with different sample sizes (20, 30, 50, 100, 200) through generating 500 replicate datasets. Of all the subjects, 50% were cured meaning that they experienced neither recurrent nor death. The average number of recurrent events was 1.44 to 1.69 with a maximum fixed of 8; the censoring rates for death had a range from 68% to 78%. The simulation results of parameters estimation are presented in Table1. It includes the estimation of parameters (Est), the mean square error (MSE), the empirical standard error (SSE), the sampling means of the estimated standard errors (SME), and coverage probability of 95% (CP) in Frequentist and Bayesian approaches. The first and second panels of Table 1 show the summarized results of Frequentist and Bayesian approaches with informative prior considering different sample sizes. In small sample size (20, 30 and 50) Bayesian approach compared to Frequentist approach had a smaller standard error and mean square error, and the coverage probabilities close to nominal level of 95%. Also, in Bayesian approach, the sampling means of the estimated standard errors were close to the empirical standard error. However, by increasing the sample size to 100 or 200, there was less bias in estimating parameters in the Frequentist approach compared to the Bayesian approach. The third panel of Table1 shows the results of Bayesian approach using non-informative prior. It is observed that when the sample size was small, using non-informative prior led to larger mean of square errors, and larger standard deviations estimates compared to the informative prior. The coverage probabilities in this case, were approximately equal to the nominal level of 95%. Furthermore, by increasing the sample size to 100 or 200, the results were close to Frequentist approach.

Table 1. Simulation Results for a generated joint frailty model with Frequentist and Bayesian approaches

| Sample size | Parameter | Frequentist | | | | | Bayesian (Informative priors) | | | | | Bayesian (Non-informative priors) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Est | SEE | $SEM^1$ | $MSE^1$ | $CP^1$ | mean | SD | $SEM^2$ | $MSE^2$ | $CP^2$ | mean | SD | $SEM^2$ | $MSE^2$ | $CP^2$ |
| N=20 | | | | | | | | | | | | | | | | |
| | $\beta_1 = 1$ | 2.60 | 5.35 | 178.33 | 30.12 | 0.91 | 0.34 | 0.48 | 0.63 | 0.66 | 0.75 | 1.17 | 1.59 | 2.70 | 2.56 | 0.95 |
| | $\beta_2 = -0.5$ | -1.28 | 5.85 | 226.70 | 34.90 | 0.94 | -0.39 | 0.48 | 0.64 | 0.24 | 0.95 | -0.83 | 1.55 | 2.23 | 2.54 | 0.94 |
| | $\beta_1^* = 0.7$ | 1.17 | 4.68 | 14.71 | 22.10 | 0.97 | 0.25 | 0.53 | 0.60 | 0.48 | 0.86 | 0.81 | 1.41 | 1.45 | 2.00 | 0.93 |
| | $P_R = 1.25$ | 1.47 | 1.10 | 0.33 | 1.25 | 0.98 | 1.34 | 0.44 | 0.43 | 0.21 | 0.94 | 1.48 | 0.67 | 0.51 | 0.50 | 0.93 |
| | $P_D = 1.25$ | 1.98 | 4.84 | 0.29 | 23.90 | 0.99 | 1.51 | 0.53 | 0.39 | 0.35 | 0.91 | 1.66 | 0.74 | 0.49 | 0.72 | 0.92 |
| | $\theta = 0.5$ | 1.60 | 10.20 | 25.03 | 1.05 | 0.99 | 0.44 | 0.59 | 0.81 | 0.35 | 0.96 | 0.54 | 0.93 | 1.41 | 0.87 | 0.93 |
| | $\alpha_0 = -0.5$ | -0.28 | 2.41 | 9.54 | 5.88 | 0.93 | -0.51 | 0.61 | 0.67 | 0.37 | 0.93 | -0.54 | 0.95 | 0.91 | 0.91 | 0.95 |
| | $\alpha_1 = -0.5$ | 1.40 | 3.35 | 17.36 | 11.30 | 0.91 | 0.89 | 0.78 | 0.95 | 0.63 | 0.95 | 1.11 | 1.37 | 1.50 | 1.90 | 0.94 |
| N=30 | | | | | | | | | | | | | | | | |
| | $\beta_1$=1 | 1.87 | 3.03 | 16.60 | 9.95 | 0.90 | 0.46 | 0.48 | 0.57 | 0.52 | 0.82 | 1.12 | 1.09 | 1.59 | 1.21 | 0.93 |
| | $\beta_2$= -0.5 | -0.64 | 3.05 | 5.01 | 9.33 | 0.94 | -0.44 | 0.49 | 0.57 | 0.24 | 0.94 | -0.72 | 1.03 | 1.24 | 1.11 | 0.95 |
| | $\beta_1^*$= 0.7 | 0.59 | 5.56 | 4.43 | 30.90 | 0.99 | 0.30 | 0.49 | 0.54 | 0.40 | 0.87 | 0.66 | 1.02 | 1.07 | 1.05 | 0.95 |
| | $P_R = 1.25$ | 1.36 | 0.51 | 0.24 | 0.27 | 0.94 | 1.28 | 0.33 | 0.34 | 0.11 | 0.94 | 1.37 | 0.42 | 0.44 | 0.19 | 0.94 |
| | $P_D = 1.25$ | 1.67 | 4.00 | 0.23 | 16.20 | 0.99 | 1.40 | 0.39 | 0.32 | 0.11 | 0.92 | 1.49 | 0.47 | 0.50 | 0.28 | 0.91 |
| | $\theta = 0.5$ | 1.12 | 8.02 | 12.87 | 64.7 | 0.99 | 0.42 | 0.50 | 0.66 | 0.26 | 0.95 | 0.53 | 0.73 | 1.42 | 0.54 | 0.95 |
| | $\alpha_0 = -0.5$ | -0.41 | 1.28 | 2.28 | 1.65 | 0.97 | -0.54 | 0.54 | 0.55 | 0.29 | 0.94 | -0.50 | 0.62 | 0.66 | 0.38 | 0.95 |
| | $\alpha_1 = 1$ | 1.22 | 1.80 | 4.42 | 3.29 | 0.95 | 0.96 | 0.69 | 0.77 | 0.48 | 0.96 | 1.03 | 0.91 | 0.91 | 0.83 | 0.96 |
| N=50 | | | | | | | | | | | | | | | | |
| | $\beta_1$=1 | 1.39 | 1.80 | 3.79 | 3.42 | 0.97 | 0.59 | 0.45 | 0.50 | 0.36 | 0.86 | 1.04 | 0.75 | 0.77 | 0.56 | 0.94 |
| | $\beta_2$= -0.5 | -0.61 | 1.25 | 2.017 | 1.53 | 0.99 | -0.48 | 0.48 | 0.48 | 0.23 | 0.95 | -0.62 | 0.75 | 0.64 | 0.58 | 0.94 |
| | $\beta_1^*$= 0.7 | 0.77 | 0.80 | 0.68 | 0.65 | 0.96 | 0.38 | 0.42 | 0.47 | 0.28 | 0.87 | 0.69 | 0.66 | 0.66 | 0.44 | 0.95 |
| | $P_R = 1.25$ | 1.29 | 0.33 | 0.18 | 0.11 | 0.94 | 1.25 | 0.26 | 0.26 | 0.07 | 0.95 | 1.29 | 0.28 | 0.28 | 0.08 | 0.95 |
| | $P_D = 1.25$ | 1.36 | 0.35 | 0.17 | 0.12 | 0.93 | 1.32 | 0.27 | 0.25 | 0.08 | 0.94 | 1.36 | 0.30 | 0.27 | 0.10 | 0.94 |
| | $\theta = 0.5$ | 0.58 | 1.04 | 4.82 | 1.09 | 0.98 | 0.43 | 0.44 | 0.53 | 0.20 | 0.97 | 0.46 | 0.50 | 0.60 | 0.26 | 0.96 |
| | $\alpha_0 = -0.5$ | -0.45 | 0.55 | 0.49 | 0.30 | 0.95 | -0.51 | 0.44 | 0.53 | 0.20 | 0.94 | -0.50 | 0.49 | 0.47 | 0.24 | 0.94 |
| | $\alpha_1 = 1$ | 1.07 | 0.88 | 1.81 | 0.79 | 0.97 | 0.95 | 0.55 | 0.61 | 0.30 | 0.97 | 0.98 | 0.65 | 0.64 | 0.42 | 0.94 |
| N=100 | | | | | | | | | | | | | | | | |
| | $\beta_1$=1 | 1.08 | 0.51 | 0.50 | 0.27 | 0.95 | 0.73 | 0.34 | 0.39 | 0.18 | 0.88 | 1.01 | 0.46 | 0.49 | 0.21 | 0.95 |
| | $\beta_2$= -0.5 | -0.52 | 0.41 | 0.40 | 0.17 | 0.94 | -0.52 | 0.35 | 0.36 | 0.13 | 0.95 | -0.59 | 0.41 | 0.40 | 0.18 | 0.93 |
| | $\beta_1^*$= 0.7 | 0.74 | 0.51 | 0.45 | 0.26 | 0.96 | 0.47 | 0.36 | 0.37 | 0.18 | 0.91 | 0.67 | 0.44 | 0.45 | 0.21 | 0.95 |
| | $P_R = 1.25$ | 1.25 | 0.19 | 0.13 | 0.03 | 0.95 | 1.22 | 0.18 | 0.18 | 0.03 | 0.95 | 1.24 | 0.19 | 0.19 | 0.03 | 0.95 |
| | $P_D = 1.25$ | 1.29 | 0.19 | 0.13 | 0.04 | 0.93 | 1.27 | 0.18 | 0.18 | 0.03 | 0.95 | 1.28 | 0.20 | 0.18 | 0.04 | 0.94 |
| | $\theta = 0.5$ | 0.49 | 0.33 | 1.56 | 0.11 | 0.94 | 0.43 | 0.30 | 0.37 | 0.09 | 0.97 | 0.44 | 0.32 | 0.40 | 0.10 | 0.96 |
| | $\alpha_0 = -0.5$ | -0.48 | 0.32 | 0.32 | 0.10 | 0.94 | -0.50 | 0.32 | 0.31 | 0.10 | 0.95 | -0.47 | 0.33 | 0.32 | 0.11 | 0.94 |
| | $\alpha_1 = 1$ | 1.01 | 0.45 | 0.44 | 0.20 | 0.95 | 0.97 | 0.42 | 0.43 | 0.18 | 0.94 | 0.94 | 0.45 | 0.44 | 0.20 | 0.93 |
| N=200 | | | | | | | | | | | | | | | | |
| | $\beta_1$=1 | 1.03 | 0.32 | 0.34 | 0.10 | 0.94 | 0.83 | 0.28 | 0.30 | 0.11 | 0.92 | 0.97 | 0.34 | 0.33 | 0.11 | 0.94 |
| | $\beta_2$= -0.5 | -0.51 | 0.29 | 0.27 | 0.08 | 0.95 | -0.53 | 0.28 | 0.26 | 0.08 | 0.94 | -0.57 | 0.32 | 0.27 | 0.11 | 0.94 |
| | $\beta_1^*$= 0.7 | 0.71 | 0.33 | 0.31 | 0.10 | 0.94 | 0.54 | 0.28 | 0.28 | 0.10 | 0.90 | 0.66 | 0.33 | 0.31 | 0.11 | 0.94 |
| | $P_R = 1.25$ | 1.25 | 0.12 | 0.09 | 0.01 | 0.96 | 1.22 | 0.12 | 0.12 | 0.01 | 0.95 | 1.23 | 0.13 | 0.13 | 0.01 | 0.95 |
| | $P_D = 1.25$ | 1.27 | 0.12 | 0.09 | 0.01 | 0.94 | 1.24 | 0.12 | 0.12 | 0.01 | 0.95 | 1.24 | 0.13 | 0.12 | 0.01 | 0.94 |
| | $\theta = 0.5$ | 0.50 | 0.23 | 0.48 | 0.05 | 0.96 | 0.43 | 0.21 | 0.25 | 0.04 | 0.95 | 0.43 | 0.22 | 0.26 | 0.05 | 0.94 |
| | $\alpha_0 = -0.5$ | -0.49 | 0.22 | 0.25 | 0.05 | 0.94 | -0.50 | 0.24 | 0.22 | 0.06 | 0.94 | -0.49 | 0.25 | 0.22 | 0.06 | 0.94 |
| | $\alpha_1 = 1$ | 1.00 | 0.29 | 0.30 | 0.08 | 0.94 | 0.96 | 0.32 | 0.30 | 0.10 | 0.94 | 0.93 | 0.33 | 0.30 | 0.11 | 0.93 |

$CP^1$, coverage probability of 95% confidence intervals; $CP^2$, coverage probability of 95% equal-tail credible intervals; Est, estimates of parameters; $MSE^1$, mean square error; $MSE^2$, Bayesian mean square error; SD, posterior standard deviation; SEE, empirical standard error; $SEM^1$, mean of standard error; $SEM^2$, sample mean of the square root of posterior variances

The simulation results suggested that when sample size was small, the use of Bayesian joint frailty model in the presence of cure fraction led to more efficiency in parameter estimation and statistical inference.

## Application

The joint frailty model in the presence of cure fraction with two approaches (Frequentist and Bayesian) was used for women with breast cancer (BC) who had a mastectomy surgery. Our real data was obtained from Ramezanzadeh Radiotherapy Center in Yazd province in Iran between the years 2004-2012 with a follow-up until 2015. This study was conducted to evaluate the risk factors associated with recurrent events and death related to BC, considering the correlation between both events. Additionally, the subjects who experienced neither relapse nor death entered the model. Since the selection of

prior for parameter in the Bayesian approach is not an easy task and needs the opinion of an expert, we considered a non-informative prior in which information of data was predominant in producing posterior distribution. The results showed that in the Bayesian approach, the estimations of parameters have less standard deviation than frequentist approach.

In this study, we considered four baseline covariates for each patient: age (>40 years old versus ≤40 years old), Lymph node status (Positive or Negative), Lymphovascular invasion (LVI: Positive or Negative) and tumor size (T3, T2 versus T1). Of 227 patients with mastectomy surgery, the number of patients with one, two and three occurrences of breast cancer (relapse) were fourteen (6.2%), four (1.8%), and one (0.4), respectively. Twenty-two (9.7%) patients died during the follow-up and one hundred ninety-seven (86.7%) patients had no recurrent events or death due to cancer. The parameters estimate of the joint frailty model in the presence of cure fraction with the Frequentist and Bayesian approaches are shown in Table 2. In a Bayesian approach, patients with LVI had a higher risk for next recurrences and death compared to those without LVI. The patients with tumor size II, were 85% less likely to have the risk of next recurrences than patients with tumor size I (II: HR=0.15, p=0.004). In contrast, the patients with tumor size III were 2.68 times more likely at risk of recurrence (III: HR=2.68, p=0.171) and 4.045 times more likely at risk of death in comparison with patients who had tumor size I (III:HR=4.045, p=0.005). The patients with Lymph Node were more likely at risk of death than patients without Lymph Node (HR=4.67, p=0.029). In the cure model, patients with tumor size II were 56.2% less likely to be cured than those with tumor size I (HR=0.438, p=0.059). The patients who had tumor size III were more likely to be cured compared to those with tumor size I (HR=1.62, p=0.343). The estimate of frailty variance in two approaches was higher than zero, which shows that there was a correlation between recurrent events and survival time.

## Discussion

In this article we proposed a Bayesian joint frailty model for recurrent and death events in the presence of cure fraction. The Frequentist approach was proposed by Liu et al.[9] In this approach the maximum likelihood estimators (MLEs) and confidence intervals are asymptotic. In addition, there are situations where the maximum likelihood estimators do not exist. On the other hand, in many medical studies, there is a small sample size in which MLEs give biased estimates for confidence intervals.[11] The Bayesian approach was presented to avoid these problems. Another advantage of the Bayesian approach is the use of previous information that is called informative prior; thus, it combines the previous information with the observed data to make posterior distribution 3. In the absence of previous knowledge, the posterior inference is based on data observations and there is a non-informative distribution.[13] Our results of simulation study suggested that when we have a small sample size, using the Bayesian cure joint frailty model with informative priors produces estimates with smaller bias and standard error estimates, and more accurate coverage probabilities compared to the Bayesian model with non-informative prior and Frequentist approaches.

Table 2. The results of Frequentist and Bayesian cure joint frailty model for breast cancer data

| | | Frequentist | | | | Bayesian with non-Informative priors | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | | Est | SE | HR | p-value | mean | SD | HR | p-value |
| Recurrent events | | | | | | | | | |
| Age | >40 | 0.30 | 0.99 | 1.36 | 0.38 | 0.12 | 0.66 | 1.31 | 0.392 |
| | <=40 | ref | - | - | - | - | - | - | - |
| LVI | positive | 1.77 | 0.82 | 5.87 | 0.039* | 2.11 | 0.70 | 8.32 | 0.004* |
| | negative | ref | - | - | - | - | - | - | - |
| Lymph Node | | 0.14 | 0.80 | 1.15 | 0.39 | 0.69 | 0.66 | 2.00 | 0.239 |
| | negative | ref | - | - | - | - | - | - | - |
| Tumor size | II: >2-5cm | -1.11 | 1.16 | 0.31 | 0.23 | -1.89 | 0.62 | 0.15 | 0.004* |
| | III: >=5cm | 0.91 | 1.81 | 2.50 | 0.35 | 0.98 | 0.76 | 2.68 | 0.171 |
| | I: <=2cm | ref | - | - | - | - | - | - | - |
| **Cancer Death** | | | | | | | | | |
| Age | <=40 | 0.54 | 0.71 | 1.72 | 0.29 | 0.76 | 0.65 | 2.14 | 0.2 |
| | >40 | ref | - | - | - | - | - | - | - |
| LVI | positive | 1.331 | 0.86 | 3.79 | 0.12 | 1.62 | 0.70 | 5.08 | 0.028* |
| | negative | ref | - | - | - | - | - | - | - |
| Lymph Node | | -0.097 | 0.83 | 0.91 | 0.39 | 1.54 | 0.67 | 4.67 | 0.029* |
| | negative | ref | - | - | - | - | - | - | - |
| Tumor size | II: >2-5cm | 0.89 | 1.35 | 2.44 | 0.32 | 0.36 | 0.44 | 1.43 | 0.286 |
| | III: >=5cm | 0.28 | 2.16 | 1.33 | 0.39 | 1.40 | 0.47 | 4.04 | 0.005* |
| | I: <=2cm | ref | - | - | - | - | - | - | - |
| $\theta$ | | 3.70 | 6.53 | - | 0.34 | 2.52 | 2.13 | - | 0.199 |
| Cure Logistic Model | | Est | SE | OR | p-value | mean | SD | OR | p-value |
| Intercept | | -0.864 | 1.63 | 0.42 | 0.347 | -1.96 | 0.37 | 0.14 | 0.001* |
| Tumor size | II: <2-5cm | -0.464 | 0.96 | 0.62 | 0.356 | -0.82 | 0.43 | 0.43 | 0.059 |
| | III: >=5cm | 0.257 | 1.37 | 1.29 | 0.392 | 0.47 | 0.86 | 1.61 | 0.343 |
| | I: <=2cm | ref | - | - | - | - | - | - | - |

Moreover, by increasing the sample size, the results of the Bayesian approach are close to Frequentist approach.

We analyzed a real data set of the patients with breast cancer, in order to compare Frequentist and non-informative Bayesian approaches. In this dataset, 86.8% of patients were cured meaning that they never experienced recurrent and death until the end of study.

Arab Borzu et al proposed a joint frailty model in the presence of cure fraction with Frequentist approach.[19]

Extensive studies were conducted in the field of joint frailty model for recurrent and terminal events.[6,7,10,14,20] Liu et al. (2016) proposed a joint frailty model for zero-inflated recurrent events in the presence of a terminal event.[9] So far, no study evaluated this model using the Bayesian approach. The strength of the present study was the presentation of the Bayesian approach of a joint frailty model in the presence of cure fraction and comparing Bayesian and Frequentist approaches using different sample sizes. There are several suggestions for further investigations. First, the other frailty distribution, such as Gaussian, can be used in the joint frailty model. Second, another approach such as EM can be used for estimation in joint frailty model with cure fraction. Third, it would be interesting to have situations where the effect of covariates on recurrent events or death is time-dependent so that the nonparametric methods like the penalized splines are applied.

## Conclusion

The current study results showed when the sample size is small, using the Bayesian joint frailty model in the presence of cure fraction has more accurate results. Furthermore, by increasing the sample size the results were close to Frequentist approach.

## Acknowledgments

The authors thank the reviewers and associate editor for their careful reading and valuable comments

## Conflicts of Interest

Authors declare that there are no conflicts of interest regarding with this article.

## References

1. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. The annals of statistics. 1982:1100-20.

2. Nielsen GG, Gill RD, Andersen PK, Sørensen TI. A counting process approach to maximum likelihood estimation in frailty models. Scandinavian journal of Statistics. 1992:25-43.

3. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. Journal of the American statistical Association. 1993;88(423):811-20.

4. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981;68(2):373-9.

5. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American statistical association. 1989;84(408):1065-73.

6. Zeng D, Lin D. Semiparametric transformation models with random effects for recurrent events. Journal of the American Statistical Association. 2007;102(477):167-80.

7. Choi S, Huang X, Ju H, Ning J. Semiparametric accelerated intensity models for correlated recurrent and terminal events. Statistica Sinica. 2017:625-43.

8. Huang CY, Wang MC. Joint modeling and estimation for recurrent event processes and failure time data. Journal of the American Statistical Association. 2004;99(468):1153-65.

9. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. Biometrics. 2004;60(3):747-56.

10. Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. Biometrics. 2007;63(1):78-87.

11. Zeng D, Lin D. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. Biometrics. 2009;65(3):746-52.

12. Rondeau V, Schaffner E, Corbiere F, Gonzalez JR, Mathoulin-Pelissier S. Cure frailty models for survival data: Application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. Statistical methods in medical research. 2013;22(3):243-60.

13. Yu B. A frailty mixture cure model with application to hospital readmission cata. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2008;50(3):386-94.

14. Sinha D, Maiti T, Ibrahim JG, Ouyang B. Current methods for recurrent events data with dependent termination: a Bayesian perspective. Journal of the American Statistical Association. 2008;103(482):866-78.

15. Paulon G, De Iorio M, Guglielmi A, Ieva F. Joint modeling of recurrent events and survival: a Bayesian non-parametric approach. Biostatistics (Oxford, England). 2018.

16. Talebi-Ghane E, Baghestani AR, Zayeri F, Rondeau V, Saeedi A. A comparison of Joint frailty model for recurrent events and death using Classical and Bayesian approaches: Application to breast cancer data. JP Journal of Biostatistics 2018;16(1):71-90.

17. Liu L, Huang X, Yaroshinsky A, Cormier JN. Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. Biometrics. 2016;72(1):204-14.

18. Andersen PK, Borgan Ø, Gill RD, Keiding N. Nonparametric estimation. Statistical models based on counting processes: Springer; 1993. p. 176-331.

19. Arab Borzu Z, Baghestani AR, Talebi Ghane E, Saeedi A, Akhavan A. Survival analysis of Iranian patients with breast cancer using joint frailty model with a cure rate. Middle East Journal of Cancer 2021; 12(2):228-34.

20. Talebi-Ghane E, Baghestani AR, Zayeri F, Rondeau V, Akhavan A. Joint frailty model for recurrent events and death in presence of cure fraction: Application to breast cancer data. Biometrical Journal 2021; 63(4):725-44.