Global Scientific JOURNALS

GSJ: Volume 10, Issue 6, June 2022, Online: ISSN 2320-9186
www.globalscientificjournal.com

Bayesian Estimation of the Colon Cancer Frailty Hazards Functions for Mixture Cure Model

Beatrice Amin Njau,[1*] Samuel Musili Mwalili,[2] George Otieno Orwa[2]

[1*] *Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI), Nairobi, Kenya*

[2] *Department of Statistics and Actuarial Sciences,JKUAT,Nairobi, Kenya.*

Abstract

In this paper, estimation of the incidence and susceptibility of colon cancer using frailty models within a Bayesian framework is presented. In contrast to previous papers, the model incorporates heterogeneity within individuals. A simulation study is carried out to evaluate the properties of the Bayesian estimates of the associated parameters in the developed model that incorporate; continuous lifetime distributions for survival data with a cured percentage, censored data, covariates, and heterogeneity within individuals. Since the previous model did not account for heterogeneity, it was unable to determine the successful recurrent or metastatic rate and survival likelihood, neither of which could be determined. The paper's findings managed to increase awareness programs for timely detection, access to care, and diagnostic accuracy. Furthermore, raising awareness of the incidence will allow researchers to investigate screening success and develop new cancer screening recommendations.

Keywords: Bayesian, frailty, heterogeneity, incidence, prevalence, screening, Survivor, susceptibility.

1 Introduction

Most developing countries are experiencing an increase in cancer burden that is still being understated (Greenlee, Murray, Bolden, & Wingo, 2000). The prevalence of colon cancer outnumbers all other cancer types, including non-melanoma skin cancer. The study's goal is to

create an underlying frailty hazard function for the mixture cure model. The statistical cure occurs when a patient population has the same mortality rate as cancer-free people.

(Vaupel, Manton, & Stallard, 1979) concentrated on life table approaches on populations whom the participants differ in their overall susceptibility to all fatalities. The methods used to investigate the impact of frailty heterogeneity on the dynamics of mortality rates(Shepard & Zeckhauser, 1984).(Shepard & Zeckhauser, 1980) Furthermore, (Vaupel et al., 1979), (Nam & Okay, 1977)findings using mortality data overestimate current life expectancy and possible gains in life expectancy from proper health and safety interventions whereas underestimating an individual's past progress in reducing mortality. In this paper, a frailty hazard mixture cure model for lifetime survival data is developed, and the parameters of the new sample are estimated using a Bayesian approach. Section 2 Review of Frailty Hazard Function of Time to Event(T) in the Mixture Cure Model, Section 3The study's Monte Carlo simulation, and Section 4 simulation of Maximum likelihood estimations. Finally, in Section 5, there is a Conclusions and future work.

2 Review of Frailty Hazard Function of Time to Event(T) in the Mixture Cure Model

The paper defines a Cancer Hazards Functions for Mixture Cure Model for simulating survival data in this section. The time until an event occurs is referred to as the survival time. For example, survival time may be defined as a patient's lifetime or the time until the recurrence of some disease in the patient. First, we will assume that we will be observing the S lifetimes of a number of patients from a potentially diversified population. As a result, we will model the population distribution as a mixture of cure model.A decomposition of the event time in the mixture modeling approach can be written as; $T = \nu T^* + (1 - \nu)\infty$. Let $T^* < \infty$ the event never occurred, indicating the presence of a cure fraction. A finite survival time, on the other

hand, corresponds to a susceptible subject who will experience the event at some point, if we indicate ν, the uncured status, such that $\nu = I\{T^* < \infty\}$ (Lu, 2008)

The proportional hazard model introduced by (Cox, 1972), is a popular semi parametric model for the hazard function as shown below:

$$\lambda(t_{ij} \mid \boldsymbol{Z_{ij}}) = \lambda(t_{ij}) \exp(\boldsymbol{\beta' Z_{ij}})$$

2.1 Frailty Hazard Model With Heterogeneity Between Individuals ($\sigma_i u_{ij}$)

The proportional hazards cure model (Cox, 1972) was proposed in this study, and the frailty hazards regression were developed for modeling susceptible and un-susceptible individuals, survival times, whereas the logistic regression models of the cure fraction specified by the two terms shown;

$$\lambda(T \mid Z_{ij}|\sigma u_{ij}) = \lim_{\mathrm{d}t_{ij}\to 0^+} P(t_{ij} \le T^* < t_{ij} + \mathrm{d}t_{ij} \mid T^* \ge t_{ij}, Z_{ij}) / \mathrm{d}t_{ij} = \lambda(t_{ij}) \exp(\beta' Z_{ij} + \sigma_i u_{ij}) \tag{1}$$

$$P(\nu_{ij} = 1 \mid X_{ij}) = \frac{\exp(\gamma' X_{ij})}{1 + \exp(\gamma' X_{ij})} \tag{2}$$

Where $\lambda(T \mid Z_{ij}|\sigma u_{ij})$ is the hazard function for a susceptible subject with p dimension covariates Z and $q-$dimension covariates X, and $\lambda(t_{ij})$ is the completely unknown underlying hazard function, and β and γ are the unknown regression parameter vectors of primary interest, u is normal distributed random effects such as N(0,1) with and σ_i the corresponding covariates . Z and X may have some common components, and that X contains 1 so that σ_i contains the intercept term. (Lu, 2008) (Lázaro, Armero, & Gómez-Rubio, 2020)

2.2 The Mixture Cure Frailty Hazard Function for Metastatic Event

This paper represent the proportion of uncured patients $P\left(X_{ij}\right) = P\left(\nu_{ij} = 1 \mid X_{ij}\right)$, which may depend on the covariate vector $X = (X_1, \ldots, X_q)$ related to the incidence by the logistic form $P\left(X_{ij}\right) = \exp\left(\gamma' X_{ij}\right) / \left(1 + \exp\left(\gamma'\right)\right)$. (Lázaro et al., 2020)

Let T be the occasion to event, defined just when $\nu = 1$ using the provisional survival function $S(t \mid \nu_{ij} = 1) = P(T > t \mid \nu_{ij} = 1)$ and Z_{ij} being the covariate vector linked with latency (Rondeau, Schaffner, Corbiere, Gonzalez, & Mathoulin-Pélissier, 2013). As a result, the comparatively small survival function seems to be:

$$S(t_{ij}) = 1 - P\left(X_{ij}\right) + P\left(X_{ij}\right) S(t_{ij} \mid \nu_{ij} = 1) \tag{3}$$

The paper, suggest using frailty mixture cure model; to model the hidden distribution. By substituting the susceptible subject's survival function by the frailty model equation yields the frailty mixture cure model shown below.

$$\begin{cases} S\left(t_{ij} \mid \sigma_i u_{ij}\right) = 1 - P\left(X_{ij}\right) + P\left(X_{ij}\right) \exp\left(-\Lambda\left(t_{ij} \mid \nu_{ij} = 1\right) \exp\left(\beta' Z_{ij} + \sigma_i u_{ij}\right)\right) \\ P\left(X_{ij}\right) = P\left(\nu_{ij} = 1 \mid X_{ij}\right) = \frac{\exp(\gamma' X)}{(1+\exp(\gamma' X_{ij}))} \\ u \sim i.i.d \mathcal{N}\left(0; \sigma^2\right) \end{cases}.$$

Where, $S\left(t_{ij} \mid \nu_{ij} = 1, \sigma_i u_{ij}\right) = \exp\left(-\Lambda\left(t_{ij} \mid \nu = 1\right) \exp\left(\beta' Z + \sigma_i u_{ij}\right)\right)$ and $\lambda\left(t_{ij} \mid \nu_{ij} = 1, \sigma_i u_{ij}\right) = \lambda_0\left(t_{ij} \mid \ \nu_{ij} = 1\right) \exp\left(\beta' Z_{ij} + \sigma_i u_{ij}\right)$.

$$\begin{cases} S\left(t_{ij} \mid \sigma_i u_{ij}\right) = 1 - P\left(X_{ij} \mid \sigma_i u_{ij}\right) + P\left(X_{ij} \mid \sigma_i u_{ij}\right) \exp\left(-\Lambda\left(t_{ij} \mid \eta_{ij} = 1\right) \exp\left(\beta' Z_{ij} + \sigma_i u_{ij}\right)\right) \\ P\left(X_{ij} \mid \sigma_i u_{ij}\right) = P\left(\nu_{ij} = 1 \mid X_{ij}\right) = \frac{\exp(\gamma' X + \psi \sigma_i u_{ij})}{(1+\exp(\gamma' X_{ij} + \psi \sigma_i u_{ij}))} \\ u \sim i.i.d \mathcal{N}\left(0; \sigma_1^2\right) \end{cases}. \tag{4}$$

from which ψ allows for the relationship of the cured percentage and the frailty-related metastatic rate. When $\psi = 0$, the cured percentage and metastatic incidents are distinct of covariates, as

are the distinct cured fractions for the same patient. The consequence of frailty on the metastases rate and the chance of designing a new event is comparable when $\psi = 1$. The metastatic rate as well as cured fraction are inextricably linked when $\psi > 0$. Significantly greater frailty raises the risk of metastatic cancer and the occurrence of a new event.Higher frailty, on the other hand, leads to greater survival between at patients but a low probability of designing a new incident when $\psi < 0$. (Rondeau et al., 2013).

$$\begin{cases} S\left(t_{ij} \mid \sigma_1 u_{1i}, \sigma_2 u_{2i}\right) = \\ 1 - P_{ij}\left(X \mid \sigma_1 u_{1i}, \sigma_2 u_{2i}\right) + P_{ij}\left(X \mid \sigma_2 u_{2i}\right) \exp\left(-\Lambda\left(t_{ij} \mid \nu_{ij} = 1\right) \exp\left(\beta' Z_{ij} + \sigma_1 u_{1i} + \sigma_2 u_{2i}\right)\right) \\ P_{ij}\left(X \mid \sigma_2 u_{2i}\right) = P\left(\nu_{ij} = 1 \mid X\right) = \frac{\exp(\gamma' X + \alpha \sigma_2 u_{2i})}{(1+\exp(\gamma' X + \alpha \sigma_2 u_{2i}))} \\ u_{1i} \sim \mathcal{N}\left(0; \sigma_1^2\right), u_{1i} \perp u_{2i} \\ u_{2i} \sim \mathcal{N}\left(0; \sigma_2^2\right) \end{cases} . \tag{5}$$

The two outcomes, u_{1i} and u_{2i}, are unrelated. The variability of the random effect u_{2i} σ_2 reflects well-known heterogeneity in cancer progression rate and cured fraction, which are correlated via a mutual random effect u_{2i} in equation 5. Furthermore, there is a random effect term u_{1i} that is independent of u_{2i} and identifies heterogeneity in event that happens times caused by random effects that are not shared with the cured percentage. (Rondeau et al., 2013). Here are some fundamental properties of S(t) and hazard fumction(λt):

- S(0) equals 1 and S(∞) equals 0.
- S(t) is a function that does not increase.
- N(t) the number of events that occurred in (0, t) $\sim Pois(\lambda t)$

2.3 Bayesian survival mixture cure frailty hazard model

For the underlying hazard function $\lambda(t)$, we will make the assumption a piecewise weibul distribution. The piecewise weibul model is useful and straightforward for modeling survival

data, and it serves as a baseline for similarities with other semi-parametric and fully parametric models. The likelihood function is constructed in the following manner. Consider J to be the countable number of time axis pieces such that $0 < r_1 < \cdots < r_J$, with $r_J > y_{il}$ for $i = 1, \ldots, n; l = 1, \ldots, L_i$. Thus, we have J intervals, $(0, r_1]\ (r_1, r_2], \ldots, (r_{J-1}, r_J]$, In which each interval contains at least one failure and a satisfactory method of assigning the data is to weigh the number of events among intervals. The piecewise weibul model is based on the assumption that, $\lambda(y) = \lambda_j$ for $y \in (r_{j-1}, r_j]\ j = 1, \ldots, J$. If the l^{th} subject in thei^{th} cluster fails or is censored in the j^{th} interval, define δ_{ilj}as 1, as either 0. When$J = 1$, that is, when there is no sections, the underlying hazard is lowered to that of a weibull distribution with $\lambda(t) \equiv \lambda_1$. By continuing to increase J, we can achieve finer sections of the time scale, permitting us to use a more adaptable structure of the underlying hazard. D also represents the demonstrated data, $\mathbf{W} = (W_1, \ldots, W_n)'$ and $\lambda = (\lambda_1, \ldots, \lambda_J)'$. The random effects $W_i (i = 1, \ldots, n)$ are generally believed to have a gamma distribution, $W_i \sim Ga(\psi, \psi)$, with mean 1 and variance ψ^{-1}. First, we suggest a promotion time cure rate frailty model with population hazard;

$$\lambda\left(t \mid \mathbf{Z}_{il}, W_i\right) = \gamma p W_i (\gamma t)^{p-1} \exp\left((-\beta' \mathbf{Z}_{il})^p + \sigma u\right) \tag{6}$$

Thus, the conditional likelihood function concerning model, $(3.1.5)$ is given by

$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid \mathbf{W}, D) = \prod_{i=1}^{n} \mathcal{L}_i\left(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid W_i, D\right)$, where

$$
\begin{aligned}
&\mathcal{L}_i\left(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{p}, \boldsymbol{\sigma}, \boldsymbol{\alpha} \mid W_i, D\right) \\
&\quad=\prod_{l=1}^{L_i}\left[\gamma p W_i(\gamma t)^{p-1} \exp\left((-\beta' \mathbf{Z}_{il})^p+\sigma u\right)\right]^{\nu_{il}} \\
&\quad\times \exp\left\{-\left[1-\exp\left\{-\Lambda\left(y_{il}\right) W_i\right\}\right] \exp\left((-\beta' \mathbf{Z}_{il})^p+\sigma u\right)\right\} \\
&=\prod_{l=1}^{L_i} \prod_{j=1}^{J}\left\{\begin{array}{l}\lambda_j^p W_i \exp\left[-\left\{\lambda_j\left(r_j-r_{j-1}\right)\right.\right. \\ \left.\left.\quad+\sum_{q=1}^{j-1} \lambda_q^p\left(r_q-r_{q-1}\right)\right\} W_i\right] \exp\left((-\beta' \mathbf{Z}_{il})^p+\sigma u\right)\end{array}\right\}^{\nu_i l \delta_{ilj}} \\
&\quad\left.+\sum_{q=1}^{j-1} \lambda_q^p\left(r_q-r_{q-1}\right)\right\} W_i\Bigg]\Bigg) \exp\left((-\beta' \mathbf{Z}_{il})^p+\sigma u\right)\Bigg\}
\end{aligned}
\tag{7}
$$

We will take non-informative priors for all the parameters where by, the likelihood functions dominate the posterior distributions. Also we will assume that β and λ are independent, and their components are independent (Yin & Ibrahim, 2005). (Don't interrogate the entire issue, we can assume that the prior distributions are independent of each other.) Specifically, we take $\beta_k \sim N\left(\mu, \sigma^2\right)$ for $k=0,1, \ldots, p$, and $\lambda_j \sim Ga(\alpha, \gamma)$ for $j=1, \ldots, J$. Furthermore, we take $W_i \sim Ga(\psi, \psi)$ and assume that $\eta \sim Ga(a, b)$, where the hyperparameters a and b are chosen to yield a large prior variance for W_i. Prior parameters β_k and λ_j according to their range, Normal and Gamma distributions are used respectively and for ψ parameters, Gamma distribution is used.(Yin & Ibrahim, 2005)

Let $[U \mid V]$ stand for the posterior distribution of U given V. For $k=0,1, \ldots, p ; j=1, \ldots, J$; and $i=1, \ldots, n$ the full conditional distributions of the parameters are given as follows:

$$\left[\beta_k \mid \boldsymbol{\beta}_{(-k)}, \boldsymbol{\lambda}, \mathbf{W}, D\right] \propto \boldsymbol{\mathcal{L}}(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid \mathbf{W}, D) \pi\left(\beta_k\right)$$

$$\left[\lambda_j \mid \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-j)}, \mathbf{W}, D\right] \propto \boldsymbol{\mathcal{L}}(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid \mathbf{W}, D) \pi\left(\lambda_j\right)$$

$$\left[W_i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W}_{(-i)}, \eta, D\right] \propto \mathcal{L}_i\left(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid W_i, D\right) W_i^{\psi-1} \exp\left(-\eta W_i\right)$$

$$[\psi \mid \mathbf{W}, D] \propto \frac{\psi^{n \psi+a-1}\left(\prod_{i=1}^n W_i\right)^{\psi-1} \exp\left\{-\psi\left(\sum_{i=1}^n W_i+b\right)\right\}}{\{\Gamma(\psi)\}^n}$$

where $\boldsymbol{\beta}_{(-k)}$ is the rest of $\boldsymbol{\beta}$ after deleting the k th component, $\boldsymbol{\lambda}_{(-j)}$ and $\mathbf{W}_{(-i)}$ are defined similarly, and $\pi\left(\beta_k\right)$ and $\pi\left(\lambda_j\right)$ are the prior densities. The complete posterior distribution W_i,

has the closed form of

$$Ga\left(\psi+\sum_{l=1}^{L_i}\nu_{il},\psi+\sum_{l=1}^{L_i}\sum_{j=1}^{J}\delta_{ilj}\left[1-\exp\left\{-\lambda_j^p\left(r_j-s_{j-1}\right)\right.\right.\right.$$
$$\left.\left.\left.-\sum_{q=1}^{j-1}\lambda_q^p\left(r_q-r_{q-1}\right)\right\}\exp\left(\left(-\beta'\mathbf{Z}_{il}\right)^p+\sigma u\right)\right)$$

3 The study's Monte Carlo simulation

In this paper the properties from (Diao & Yin, 2012) for the parameters of the mixture cure model as well as hazard frailty distributions were determined using simulation in this category. The convergence and normality of the parameters were noticed. The simulation results in the study are operate to assess the performance of proposed mixture cure model but also hazard frailty.

3.1 Trace and Density plots of the regression coefficients and distribution parameters for the PH model

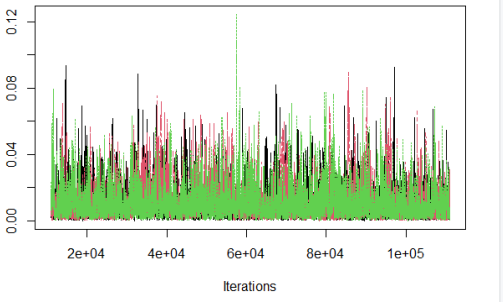
The visualization tool, such as Kernel density-based methods, suggested that when the distance between the two chains' kernel density estimation method or two parts of a single chain is close to zero, the MCMC diagnostics conclude convergence **??**. For MCMC convergence diagnosis, there are a few widely used graphical methods.

The trace plot is the most commonly used graphical convergence method of diagnosis. The trace plot is a time series graph that relates the Markov chain classifications to the iterative process numbers for each iterative process.This technique is used to show how well the Markov chain traverses the state space or how it really mixes. The trace plots show a flat bit whenever the MCMC chain has become stuck in one region of the state space, implying slow convergence. A trace plot of this category is noticed for a Metropolis-Hasting chain whenever a huge number

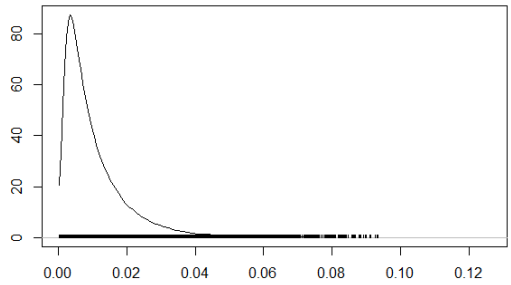
of proposals are denied repetitively.

Trace plots, on the other hand, may move very slowly if many proposals are approved repetitively in a Metropolis-Hasting chain, not going to explore the rest of the state space. Whether there are noticeable trends or modifications in the dispersion of the trace plot, normality of the data has still not been accomplished. A useful trace plot, it has been said, should represent a hairy caterpillar. If the value obtained is not in high-density region, the trace plots begin with back-to-back steps in one manner, indicating an efficient MCMC algorithm. Or else, trying to throw burn-in samples is meaningless if the trace plot shows a regular trend throughout. As shown in Figures3.1.1 below, all plots and patterns are similar.

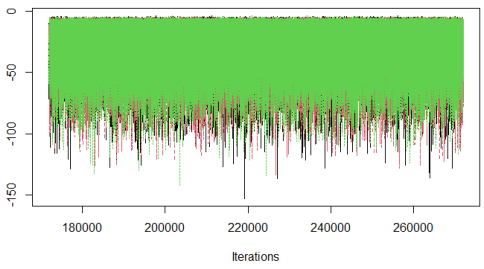
3.1.1 Trace and Density Plots



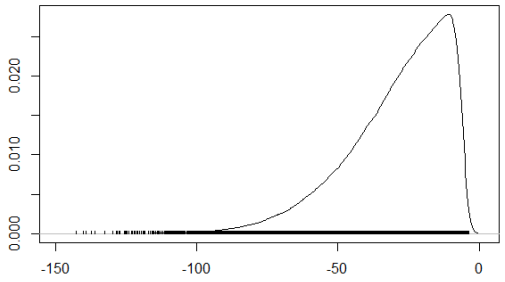
Trace of Diagnosed years



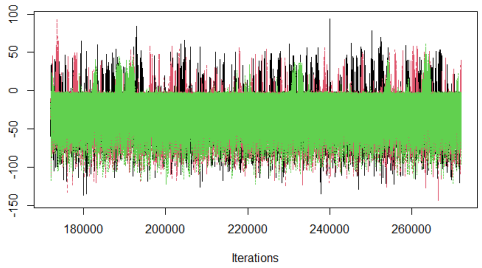
Density plots of diagnosed year



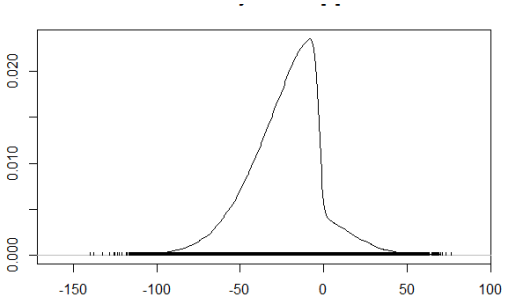
Trace of Age



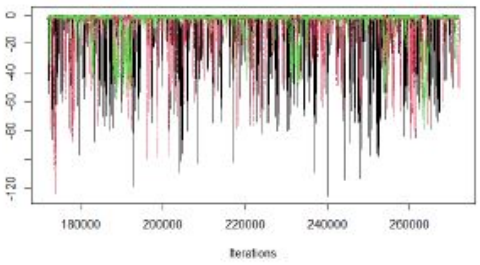
Density plots of age



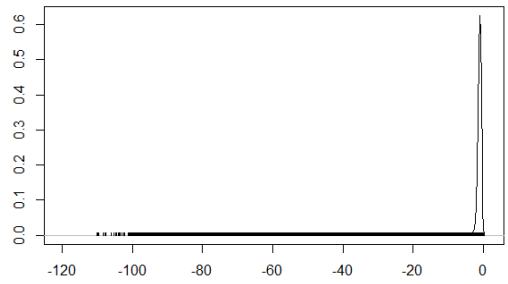
Trace of stage4



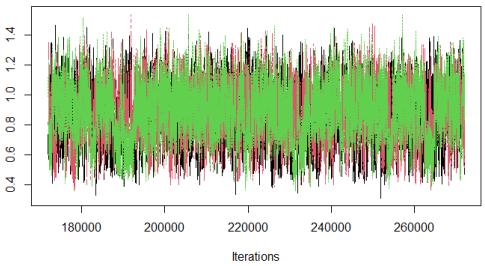
Density plots of stage4



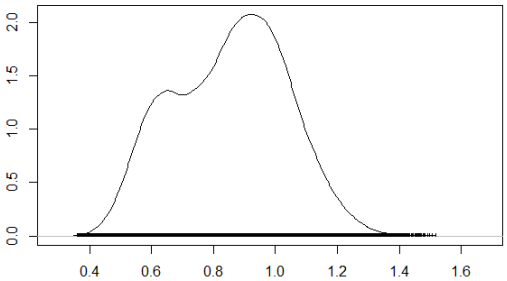
Trace of stage 3



Density plots of stage3



Trace of stage 2



Density plots of stage2

3.1.2 The Effective Sample Size.

To investigate the mixing of a Markov chain, autocorrelation and trace plots can be used. Whereby is an estimate of the difference between the sampled values' mean, which is our estimated posterior mean, and the true posterior mean. As a result, the Markov chain error can be compared to the Bayesian inference concept of a standard error. The general rule of thumb is that the Markov chain error should be less than 5% of the sample standard deviation.

Table 1: A summary of some commonly used statistical tests for convergence diagnostics is provided.

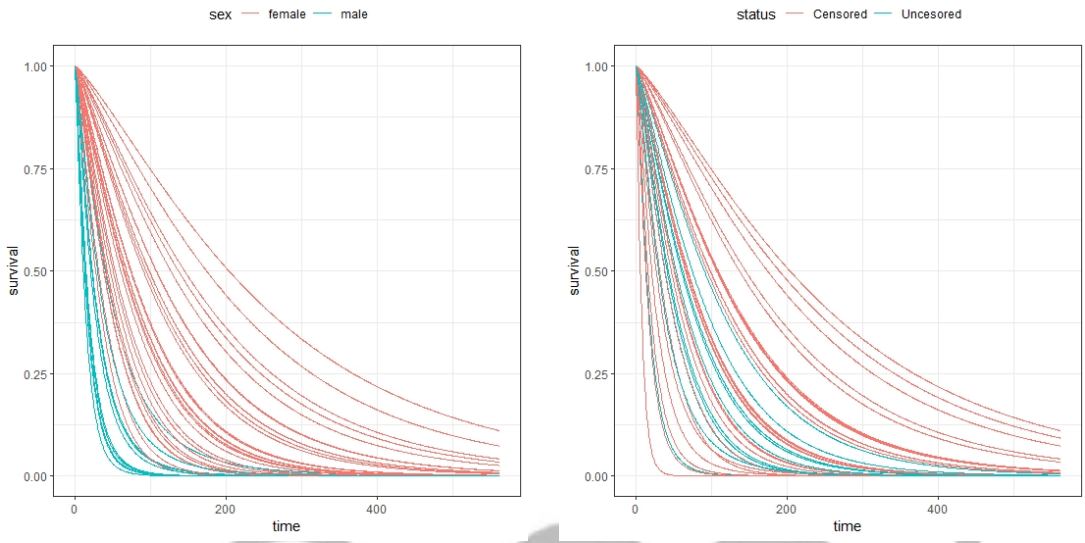
Para meter	Geweke diagnostic Pr >|z|	Raftery - Lewisl Total No.of Sample	Heidelberger - Welch P-value	Dependence Test for Stationarity	Test for Halfwidth
Stages2	-0.05275	1272706	0.0519	Passed	Passed
Stages 3	-0.74432	871200	0.6339	Passed	Failed
Stages 4	1.46866	10580	0.1000	Passed	Passed
Age	1.74905	10498	0.6587	Passed	Passed
Diagnosed year	-0.71776	61992	0.2644	Passed	Passed

3.2 Kaplan-Meier method

Kaplan-Meier is a statistical method used in simulation and time to event data analysis. The time from enrollment in a study to a specific event, such as the onset of illness, is referred to as the time to the event. The figures, 3.4, simulate survival data and fit variables such as sex, stage 2, stage 3, and stage 4. At the very top left corner, it indicates that at time "0," the beginning of our study, 100% of the patients had not experienced an event of interest.

The sex figure shows that females last longer and are more likely to survive. They are slightly higher than the male survival rate; then, in stages 2 and 3, the censored individuals outnumber the uncensored individuals; and finally, in stage 4, the uncensored individuals outnumber the censored individuals.

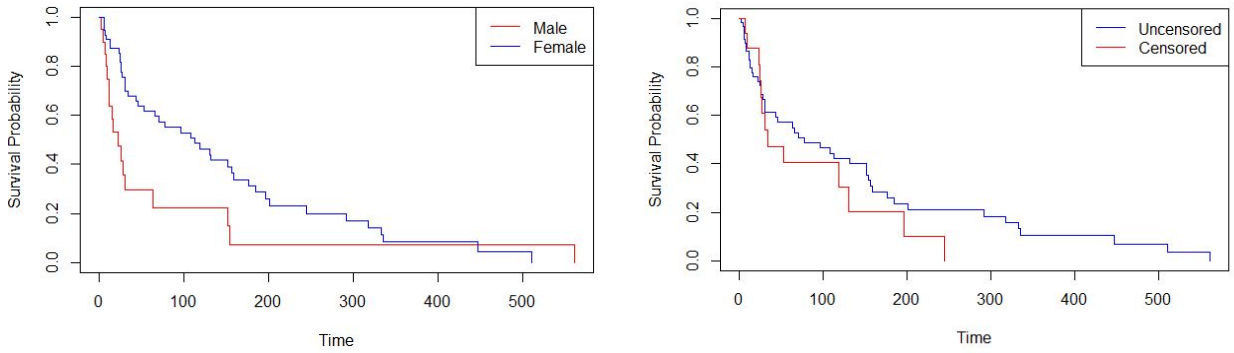
3.3 To visualize the distribution of sex and status within each individual, using ggplot2.



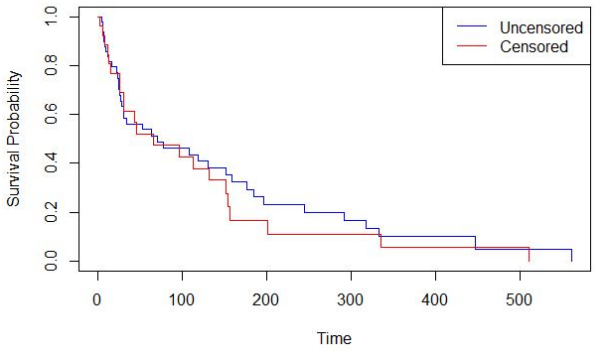
Sex Status

Figure 3.3: The posterior mean of the individual's sex and status, survival functions from the frailty mixture cure model.

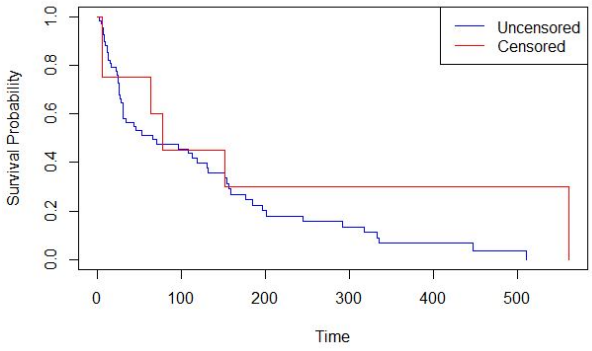
3.4 To fit the sex, stage2, stage3 and stage4 survival data using the Kaplan-Meier method



Sex Stage2



Stage3



Stage 4

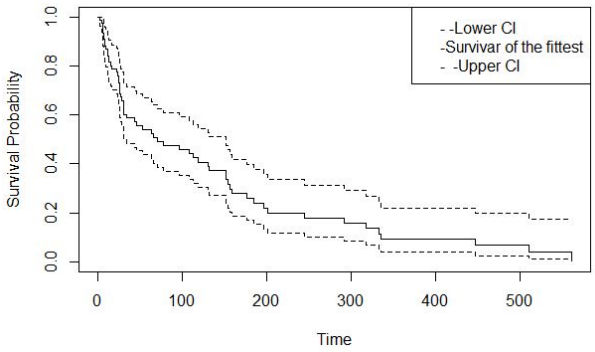


Figure 3.4: Survival of the fittest

The Kaplan-Meier survival function is a decreasing step function with a jump at each discrete event time and no censoring; the Kaplan-Meier estimator is simply the information's experimental distribution. In contrast to non-parametric estimation, independent variables can be included in parametric models.

Semiparametric models' coefficients and hazard ratios are presented in the table2 When we have a positive coefficient, this indicates that as the predictor variables rises, so the number of time to event increases. A positive coefficient indicates a shorter time or that the incident is much more likely to happen; a negative coefficient indicates a longer time and reduced hazard rates, making the experience less likely to happen.(Atuna & Adusei, 2022)

Whenever it happens to come to hazard rates, a hazard ratio significantly larger than one

Table 2: Cox proportion hazard rate Coefficients.

Covariate	Coef	exp(coef)	exp(-coef)	lower 0.95	upper 0.95	se(coef)	z	Pr(>|z|)
Intercept	0.2021	1.2240	0.8170	0.052	0.3565	0.0793	2.5495	2e-16
diagy	-1.5640	0.2092	4.7780	0.1063	0.4121	0.3457	-4.524	6.07e-06
Age	0.0013	1.0013	4.7780	0.9758	1.074	0.0136	0.096	0.924
Frail	0.0138	1.0139	0.9863	0.5978	1.7196	0.2696	0.051	0.959
Stage2	-0.2938	0.7454	1.3416	0.2547	2.1815	0.5479	-0.536	0.592
Stage3	0.0103	1.0104	0.9897	0.4142	2.4650	0.4550	0.023	0.982
Stage4	-0.3137	0.7307	1.3685	0.1683	3.1723	0.7491	-0.419	0.675
concordance	=	1	$(se = 0)$					
Likelihood	ratio	test	=351.2	on 6df,	p=<2e-16			
Wild	test	=	20.94	0n	df,	=<	0.002	
Score	(log-rank)	test	=	62.69 on	6 df,	p=< 1e-11		

clearly shows that for a one-unit increment in that specific variable, the hazard rate, which is also the probability of the event occurring, is much more likely to occur. And when the hazard rate is between0 and 1, it is assumed that there are lower hazard rates (as a result of which the event becomes less likely to happen)(Atuna & Adusei, 2022).

According to our parameter estimation, a highest value of the frailty W_i, is associated both with a relatively low cure likelihood and an increased risks of experiencing the incident.

The magnitude of coefficient points is the one to be interpreted in hazard rate coefficients, which means that the unit increase in the variable, such as stage2, is associated with only 26% increasing the hazard or 73% lower hazard rate where the event is less likely to occur.

4 Evaluating Maximum Likelihood Estimators Methods

The proposed frailty cure model's posterior distribution stability was tested using the posterior mean, standard deviation (SD), absolute bias, Naive standard error, root mean square error (RMSE), coverage probability (CP), effective number of separate simulations draws, and potential scale reduction factor (R).

Table 3: Under non-informative priors, numerical summaries of the posterior properties for the frailty hazard model, based on MCMC

Characteristicsa	Parameters					
	alpha	betac1	betac2	beta3	lambda	psi
Mean	9.968	-34.02	1.168e-01	-2.913e-02	1.701e-15	142.52
Median	9.98045	-34.04724	0.04418	-0.02279	1.6901e-15	128.0881
Mode	9.99	-34.05	0.025	-0.025	0	75
Variance	8e-04	0.01445	0.01382	0.01314	0	8172.452
Skewness	-1.95594	0.98221	-0.05962	-0.05718	2.07697	1.56377
Kurtosis	5.42496	2.21802	0.02451	0.00567	9.81136	3.42699
Minimum	9.71202	-34.41566	-0.49271	-0.55657	0	13.46436
Maximum	10	-33.12077	0.49705	0.49774	0	854.6372
SD	3.224e-02	1.304e-01	1.148e-01	2.389e-16	142.5	78.75
Naive SE	4.163e-04	1.683e-03	1.507e-03	1.482e-03	3.085e-18	1.017
Time series SE	0.001927	0.007159	0.002678	0.002739	0.000000	10.418417
2.5 percentile	9.876	-34.22	-1.898e-01	-2.598e-01	1.373e-15	36.29
Q_1	9.957	-34.10	-3.413e-02	-1.039e-01	1.545e-15	81.06
Median / Q_2	9.978	-34.03	4.408e-02	-2.736e-02	1.656e-15	125.3
Q_1	9.991	-33.95	1.211e-01	4.567e-02	1.795e-15	190.6
97 percentile	9.999	-33.68	2.708e-01	1.971e-01	2.350e-15	324.4

4.1 Evaluating the estimators' performance:

The bias of the estimators is calculated as follows:

$\text{Bias}(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta} - \theta)$

An overestimation is depicted by a positive sign, while an underestimation is demonstrated by a negative bias (Adnan & Arasan, 2018). Accuracy of the estimators: The root mean square error (RMSE) is a useful indication of overall correctness and is computed as follows:

$$\text{RMSE}(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta} - \theta)^2$$

This metric determines the accuracy of the estimates. When the RMSE is low, impact estimates are more accurate. The Naive standard error is another accuracy statistic that is calculated by dividing the posterior standard deviation by the square root of the sample size. As a result, the smaller the standard error, the larger the sample size.The Nave SE employs simulation error rather than posterior uncertainty (Naslina, Jayanthi, Syahida, & Bakri, 2020).

$$NaiveSE = \frac{PosteriorSD}{\sqrt{n}}$$

4.2 Coverage

The 95 percent coverage probability is the percentage of N simulated data sets in which the genuine estimates were used in the 95 percent confidence interval (CP). The more accurate the estimates, the closer the result is to a 95% confidence probability of coverage.The following is an explanation of CP: (Naslina et al., 2020)

$$CP = \hat{\theta} \mp 1.96 \times SE(\hat{\theta})$$

In the table 4, we demonstrated that different allocations for the number of chains affect the estimation results slightly. Then, for all chains, we must confirm that the bias for fixed initials appears to decrease as the chain number increases, while the bias for random effect estimates increases as the number of chains increases, indicating a negative association with the survival probability. Frailty heterogeneity may play a role in observed declines and reversals with age, as well as mortality differences between populations. (Miquel, Economos, Fleming, & Johnson Jr, 1980)

4.2.1 The simulation technique described below was used.

- The initial values of the model parameters are specified.
- Create a random sample of size 500 and n chains using log-likelihood equation 7
- Using the maximum likelihood method, evaluate the estimates' values.
- Determine the MSE, RMSE, BIAS, SE, and Coverage Probability for each parameter.

Table 4: Simulation Results of Maximum likelihood estimation of the Frailty cure model distribution with difference number of chains.

Parameters	Initial	Estimates	SE	RMSE	Bias	coverage%
	n=10					
α	10	9.9716	2.7905e-04	2.8392e-02	2.8346e-02	90
$\beta 1$	-34	-34.032	3.1604e-03	3.2698e-02	3.1892e-02	90
$\beta 2$	0.045	4.2511e-02	1.4758e-03	4.2814e-03	2.4888e-03	100
$\beta 3$	-0.25	-2.3234e-02	4.2777e-03	2.2678e-01	-2.2676e-01	100
λ	1.7	1.6735e-15	5.7735e-18	1.6736e-15	-1.6735e-15	90
ϕ	145	1.4593e+02	9.5776	12.5993	6.4406e-02	100
	n=20					
α	10	9.9722	2.7905e-04	2.7783e-02	2.7752e-02	90
$\beta 1$	-34	-3.4032e+01	3.1604e-03	3.3327e-02	3.2928e-02	90
$\beta 2$	0.045	4.1885e-02	1.4759e-03	3.6157e-03	3.1141e-03	95
$\beta 3$	-0.25	-2.4299e-02	4.2777e-03	2.2571e-01	-2.2570e-01	95
λ	1.7	1.6705e-15	5.7735e-18	1.6705e-15	-1.6705e-15	90
ϕ	145	146.77	9.5776	-0.77522	0.8501	95
	n=50					
α	10	9.9721	2.7905e-04	2.7938e-02	2.7891e-02	96
$\beta 1$	-34	-34.0339	3.1604e-03	3.4645e-02	3.3952e-02	94
$\beta 2$	0.045	4.3837e-02	1.4758e-03	2.8755e-03	1.1624e-03	98
$\beta 3$	-0.25	-2.3027e-02	4.2777e-03	2.2699e-01	-2.2697e-01	96
λ	1.7	1.6687e-15	5.7735e-18	1.6688e-15	-1.6688e-15	94
ϕ	145	144.6404	9.5776	13.2172	1.3596	96
	n=200					
α	10	9.9719	2.7906e-04	2.8054e-02	2.8003e-02	94
$\beta 1$	-34	-34.033e	3.1604e-03	3.3537e-02	3.2809e-02	95.5
$\beta 2$	0.045	4.3002e-02	1.4759e-03	3.6223e-03	1.9980e-03	96
$\beta 3$	-0.25	-2.4188e-02	4.2777e-03	2.2583e-01	-2.2581e-01	95
λ	1.7	1.6709e-15	5.7735e-18	1.6709e-15	-1.6709e-15	95
ϕ	145	149.71e	9.5776e	14.117e	-3.7069	94.5

5 Conclusions

For cured and uncured data, we developed an underlying frailty hazard function for the mixture cure model. This paper intends to estimate the incidence and susceptibility of colon cancer using frailty models within a Bayesian framework. In contrast to previous work, our proposed models include a proporional hazards cure model in which frailty hazards regression, random covariate, and Individual heterogeneity in mixture cure models is a broader multilevel structure with various random effect frameworks.One intriguing aspect of a models seems to be that they offer cancer patients an opportunity of someone being cured.

Random effect cure models have a privilege over minimal cure models in that they can provide understanding of the underlying similarity characteristics of the data and permit for forecasting inference. To identify patients at risk for colon cancer, a survival set of data is used. In furthermore, to predict the outcome for a particular patient. The percentage of spatial mentioning varies according to colony location. Besides that, because of the high prevalence of colon cancer in the population, these event time data are heavily censored in the early stages. As a result, it is unlikely to fail and can be regarded as cured/remitted. In modeling such grouped data sets, we make the assumption a proportion hazards model with just a surviving fraction and evaluate for dependence among colon sites utilizing spatial frailties, which also are patterned as linear combinations of positive steady random effects.

Future work

The frailty cure model can be expanded to include risk factors that compete with frailty. Two opposing events are recurring or metastatic occurrences and specific target occurrences. In colon cancer studies, the moment to repeat for colon cancer, for example, is the endpoint of interest, but uncured patients are likely to die during the follow-up period. In this case, the

moment to recurrent/metastatic exacerbation and the time to death on the same person coincide. Even if a situation of competing risks exists, it is difficult to accommodate the thing causing hazards; and their correlating log-likelihoods; in the prediction inference of such a hazard frailty cure model; because it assumes independent, and evenly restricted competing for the specific target event; as well as the presence of intersecting times to various competing different kinds.

Conflicts of interest

The author declare no possible conflicts of interest pertaining the public response, composition, or publication of the article.

Acknowledgements

I like to express my gratitude to Prof. Samuel Mwalili and Prof. George Orwa for their insightful comments and consultations on this work.This work was financially supported by Pan Africa University of Science, Technology and Innovation(PAUSTI) and data assistance from my supervisor Prof.Samwel Mwalili.

References

Adnan, W. N. A. W. M., & Arasan, J. (2018). Hypothesis testing for the parameters of log-logistic regression model with left-truncated and right-censored survival data. *Jurnal Teknologi*, *80*(3).

Atuna, L. M., & Adusei, M. (2022). Key determinants of cross-listing in sub-saharan africa. *Transnational Corporations Review*, *14*(1), 62–76.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Diao, G., & Yin, G. (2012). A general transformation class of semiparametric cure rate frailty models. *Annals of the Institute of Statistical Mathematics*, *64*(5), 959–989.

Greenlee, R. T., Murray, T., Bolden, S., & Wingo, P. A. (2000). Cancer statistics, 2000. *CA: a cancer journal for clinicians*, *50*(1), 7–33.

Lázaro, E., Armero, C., & Gómez-Rubio, V. (2020). Approximate bayesian inference for mixture cure models. *TEST*, *29*(3), 750–767.

Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics*, *60*(3), 545–574.

Miquel, J., Economos, A., Fleming, J., & Johnson Jr, J. (1980). Mitochondrial role in cell aging. *Experimental gerontology*, *15*(6), 575–591.

Nam, C., & Okay, K. (1977). Factors contributing to the mortality crossover pattern. In *Xvii general conference of the international union for the scientific study of population, mexico city.*

Naslina, A. M. N. N., Jayanthi, A., Syahida, Z. H., & Bakri, A. M. (2020). Assessing the goodness of fit of the gompertz model in the presence of right and interval censored data with covariate. *Austrian Journal of Statistics*, *49*(3), 57–71.

Rondeau, V., Schaffner, E., Corbiere, F., Gonzalez, J. R., & Mathoulin-Pélissier, S. (2013). Cure frailty models for survival data: application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. *Statistical methods in medical research*, *22*(3), 243–260.

Shepard, D. S., & Zeckhauser, R. J. (1980). Long-term effects of interventions to improve survival in mixed populations. *Journal of Chronic Diseases*, *33*(7), 413–433.

Shepard, D. S., & Zeckhauser, R. J. (1984). Survival versus consumption. *Management Science*, *30*(4), 423–439.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454.

Yin, G., & Ibrahim, J. G. (2005). A general class of bayesian survival models with zero and nonzero cure fractions. *Biometrics*, *61*(2), 403–412.