Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cegh

Original Article

Risk assessment in liver transplantation patients: A shared frailty parametric approach

Dilip C. Nath^{a,*}, Atanu Bhattacharjee^b, Ramesh K. Vishwakarma^a^a Department of Statistics, Gauhati University, Guwahati 781014, India^b Division of Clinical Research and Biostatistics, Malabar Cancer Centre, Thalassery 670103, India

ARTICLE INFO

Article history:

Received 6 October 2015

Accepted 27 November 2015

Available online 23 December 2015

Keywords:

Lognormal frailty

Akaike Information Criteria

Bayesian Information Criteria

ABSTRACT

The frailty model consists of two components; the first is baseline hazard. If baseline hazard is known in advance, then, the parametric estimation approach can be used advantageously. The second component of frailty model is frailty distribution. In this paper, we compared the risk of mortality by different parametric frailty models, namely gamma, inverse Gaussian, positive stable and lognormal, with respect to Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values on liver transplantation patients. We conclude that lognormal frailty model gives slightly lower frailty variance compared to the other frailty models when Weibull distribution is considered as baseline hazard. Therapy and MELD (model for end-stage liver disease) scores may have significant role in liver transplant patients when distribution of frailty is considered as lognormal with Weibull baseline hazard for fitting the model. Therefore, there is 4.6 times more risk in getting hazard of mortality for patients receiving conventional therapy with consent for participation as compared to patients receiving new therapy with consent for participation with 95%CI [1.059, 20.199], p -value 0.042.

© 2015 INDIACLEN. Published by Elsevier, a division of Reed Elsevier India, Pvt. Ltd. All rights reserved.

1. Introduction

There are many statistical methods developed to analyze survival data. A most frequent approach to analyze survival data is by assuming that study population is homogeneous. The event times of individuals in the population, conditional on the observed covariates, are assumed to be independent. However, this cannot be assumed in all applications, because many applications encounter heterogeneous sample, i.e. individuals

with different risks and hazards. In practice, there may be an association between the time of event and few subgroups of the population since the individuals of these groups share a common trait that cannot be observed. The term frailty was introduced in order to account for unobserved heterogeneity with random effects, and association in survival models.^{1–3} This concept of frailty was later introduced to biostatistics by applying it on population mortality data. The applications of model to multivariate survival data were applied on chronic disease incidence in families.⁴

* Corresponding author. Tel.: +91 9435144485.

E-mail address: dilipc.nath@gmail.com (D.C. Nath).

<http://dx.doi.org/10.1016/j.cegh.2015.11.008>

2213-3984/© 2015 INDIACLEN. Published by Elsevier, a division of Reed Elsevier India, Pvt. Ltd. All rights reserved.

1.1. Frailty model

The frailty model is defined in terms of the conditional hazard⁵

$$h_{ij}(t_j|w_i) = h_0(t)w_i \exp(x_{ij}^T \beta) \quad (1)$$

with, $i \in I = \{1, 2, 3, \dots, C\}$ and $j \in J = \{1, 2, 3, \dots, n_{ij}\}$, where $h_0()$ is the baseline hazard function; w_i is the frailty term in the group i , x_{ij} is the vector of covariates for patient j_i in group i , β is the vector of regression coefficients, I and J are the vector of cluster and patient respectively, and C is the symbol for last cluster in the data.⁶ If the number of patients, n_{ij} , is 1 for all groups, then the univariate frailty model is obtained otherwise and the model is called the shared frailty model because all patients in the same cluster share the same frailty value w_i .^{7,8} The component of frailty model is explained in the next subsections.

1.2. Baseline hazard

Under the parametric approach, the baseline hazard is defined as a parametric function and the vector of its parameters,¹ which is estimated together with the regression coefficients (refer Table 1).

1.3. Frailty distribution

The frailty w_i is an unobservable conception of a random variable W , with probability density function $f()$ – the frailty distribution. Since w_i multiplies the hazard function, W has to be non-negative in nature. Another limitation is for identifying, similar to the zero-mean constraint of a random effect in a standard linear mixed model.^{9,10} More specifically, the mean of W is typically restricted to unity when possible (i.e., when $E(W)$ exists) in order to separate baseline hazard from the overall level of the random frailties.

Various frailty distributions have been proposed in the literature.^{9,10} Indeed, in practice, mainly the gamma distribution and the lognormal distribution are used to model the frailty term and most of the software limits the choice of frailty distribution to these cases. But the right choice of frailty distribution is essential to arrive at a good description of the dependence structure present in the data. Few results are available on comparing models with different frailty distributions.^{11,12} First we have considered Weibull distribution because this distribution is frequently used to analyze lifetime data.^{13,15} Weibull distribution is mostly used for modelling lifetimes with baseline as Weibull distribution. Hazard rate for Weibull distribution is a

monotone function,^{14,16} which increases with time to infinity when shape parameter α is greater than one and it decreases up to the value zero for $\alpha < 1$.

A near-zero failure rate implies that almost no failure will occur, which is barely feasible in the real life. The Exponential distribution can be used effectively in analysing lifetime data sets, particularly in place of gamma and Weibull.

For Exponential distribution, hazard rate increases from zero to a finite constant, when shape parameter α increases and hazard rate decreases from infinity to a finite number when α is less than one. Another distribution we have considered is Exponential power distribution. The Exponential power distribution can have increasing, decreasing or bathtub-shaped hazard rates. So, it may provide better fitting to the model. The family of the power variance function distributions was introduced as an extension of the positive stable distribution by Hougaard (1995).⁸ It contains the gamma, inverse Gaussian and positive stable distributions; they are obtained as choices for parameters at the boundary of parameter space. Hereinafter, we shall focus on the gamma, the positive stable, lognormal and the inverse Gaussian frailty distributions. In all of these, a single heterogeneity parameter (denoted either, ν or θ) indexes the degree of dependence. In the following, δ is used as a generic notation to denote either θ or ν .

1.4. Likelihood

In the parametric setting, estimation is based on the marginal likelihood, in which the frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution. Under assumptions of non-informative right-censoring and independence between the censoring time and the survival time random variables, given the covariate information, the marginal log-likelihood of the observed data $Z = \{Z_{ij}; i \in I, j \in J\}$ can be written as,¹

$$\begin{aligned} \text{lmarg}(\delta, \beta; z|\tau) = & \sum_{i=1}^C \left\{ \sum_{j=1}^{n_i} \delta_{ij} (\log(h_0(y_{ij})) + (x_{ij}^T \beta)) \right. \\ & + \log \left[(-1)^{d_i} L^{d_i} \left(\sum_{j=1}^{n_i} H_0(y_{ij}) + \exp(x_{ij}^T \beta) \right) \right] \\ & \left. - \log \left[L \left(\sum_{j=1}^{n_i} H_0(\tau_{ij}) + \exp(x_{ij}^T \beta) \right) \right] \right\}, \quad (2) \end{aligned}$$

where, $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the number of events in the i th clusters and τ is the truncation vector. The γ is used as a generic notation to denote θ . $L^{(q)}$ is the q th derivate of the Laplace transformation of the frailty distribution defined as,

$$L^{(q)} = E(\exp(-Ws)) = \int_0^\infty \exp(-ws) f(w_i) dw_i, \quad s \geq 0 \quad (3)$$

1.5. Estimation

Estimates of ϕ , β and γ are obtained by maximizing the marginal log-likelihood in Eq. (3); this can be done if one is able to compute higher order derivatives $L^{(q)}$ of the Laplace transform up to $q = \max \{d_1, d_2, \dots, d_C\}$. The explicit formulas are rather enviable.

Table 1 – Parametric for the baseline hazard.

Distribution	$h_0(t)$	$H_0(t) = \int_0^t h_0(t) dt$	Parameter space
Exponential	λ	λt	$\lambda > 0$
Weibull	$\lambda \rho t^{\rho-1}$	λt^ρ	$\lambda, \rho > 0$
This table shows the hazard and cumulative hazard functions for each of these distributions.			

2. Objective

In this paper, the aim of our study is to compare risk of mortality between control and treated group by considering other distributions that fit better than gamma and Weibull distribution on liver cirrhosis data. Therefore, we have studied the performance of gamma, inverse Gaussian and positive stable distribution as a frailty distribution and compared with the Weibull and Exponential baseline hazard distribution.^{5,17} The Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are used to compare the performance of these models on liver transplantation patients. In the next section, the description on data collection and study methodology is explained.

3. Data methodology

A total of 172 patients with cirrhosis are randomized in two different treatment groups, i.e. 86 in each group. The patients were asked to read and provide informed consent if they were interested in participating in the study. After agreeing to participate in the study, the selection tests and procedures were performed to determine eligibility of patients to participate. The study doctor reviewed the results of the screening tests to determine if a patient could participate in the study. If so, the patient was randomized to treatment or control group based on randomization procedure.

In the treated group, patients who consented for participation had received infusion treatment HFPLC as well as conventional medical treatment in the background and in control group; patients who consented for participation did not receive the infusion but only the conventional medical treatment. The duration of the study was 36 months, with 7 visits (baseline, 3 months, 6 months, 9 months, 12 months, 24 months and 36 months). The parameters under consideration were MELD (model for end-stage liver disease) scores, Therapy, Centre, Gender and Triglycerides (mg/dl). Each patient observation had been taken at seven different time points corresponding to each visit.

The data description includes the variable “Patient”, being the patient's identification code; the time from liver transplantation to death or censoring is stored in Duration variable. The status variable is 1 when patient died and 0 for censored observations. Four covariates are available: Therapy, Gender, being 1 for males and 2 for females, Centre, being 1 for T1 and 2 for T2 and MELD, being the MELD score of each patient and Triglycerides are considered into the model. The empirical data are considered from the path <http://www4.stat.ncsu.edu/~boos/var.select/pbc.html>; accessed on January 6, 2013. More details about the data can be found with Nath et al.¹⁹

3.1. Parametric frailty models

In this study, we are concentrating on parametric frailty model in which gamma, inverse Gaussian, positive stable and lognormal distribution are as frailty distribution and Weibull and Exponential baseline distributions. Therefore, we shall consider in outline few distributions that are useful in the field

of survival analysis. The distributions to be discussed here are all continuous. Throughout the literature on survival analysis, certain parametric models have been used repeatedly, such as Exponential and Weibull models. These distributions have closed form terms for survival and hazard functions. Lognormal and gamma distributions are generally less convenient computationally, but are still frequently applied.¹⁶ The pros and cons of different parametric, semi-parametric and non-parametric models and methodology for statistical inference can be found in various literatures.^{6,11,12,15} In the next subsection, we discussed some of the standard failure time models for homogeneous populations. The properties and the theoretical bases of these distributions are considered here only briefly. The distributions were studied in the simplest case of independently and identically distributed random variables. In this scenario, the random variable θ denotes the lifetime, which we are interested in making inferences about distributions described in the next subsections.

3.2. Exponential distribution

The Exponential model ($\theta \sim \text{Exp}(\lambda)$) is the simplest parametric model and assumes a constant risk over time, which reflects the property of the distribution appropriately called 'lack of memory'. It holds that

$$P(\theta > S + \delta | \theta \geq S) = P(\theta > \delta) \quad (4)$$

when, θ is interpreted as the waiting time for an event to occur relative to some initial time S , this relation implies that if T is conditioned on a failure to observe the event over some initial period of time S , the distribution of the remaining waiting time is the same as the original unconditional distribution. The parameter λ attains all positive values and distribution with $\lambda = 1$ is called the unit or standard Exponential. Therefore, the following formulae can be derived by some simple algebraic calculations:

Probability density function $f(\theta)$

$$= \lambda e^{-\lambda\theta}, \text{ Survival function } S(\theta)$$

$$= e^{-\lambda\theta}, \text{ hazard function } \lambda(\theta)$$

$$= \lambda, \text{ cumulative hazard function } H(\theta) = \lambda\theta, \text{ mean } E(\theta)$$

$$= \frac{1}{\lambda}, \text{ variance } V(\theta) = \frac{1}{\lambda^2}$$

The Exponential distribution was widely used in the early works on the reliability of electronic components and technical systems. The distribution of $c\theta$ with a positive constant c is again Exponentially distributed with parameter λ/c . The minimum of n independent Exponential random variables with parameter λ is still Exponential with parameter $n\lambda$

$$P(\min\{\theta_1, \dots, \theta_n\} \geq x) = \prod_{i=1}^n P(\theta_i \geq x) = \prod_{i=1}^n e^{-\lambda x} = e^{-n\lambda x} \quad (5)$$

The model is very sensitive to even a modest variation because it has only one adjustable parameter, the inverse of which is both mean and standard deviation. Hail et al.¹⁸ have proposed that recent works have overcome this limitation by using more flexible distributions.

3.3. Weibull distribution

The Weibull model (introduced by Waloddi Weibull in 1939) is an important generalization of the Exponential model with two positive parameters. The second parameter in the model allows great flexibility of the model and different shapes of the hazard function. The convenience of the Weibull model for empirical work stems on the one hand from this flexibility and on the other hand from the simplicity of the hazard and survival function.

Probability density function $f(x)$

$$\begin{aligned} &= \lambda \gamma \theta^{\gamma-1} e^{-\lambda \theta^\gamma}, \text{ survival function } S(\theta) \\ &= e^{-\lambda \theta^\gamma}, \text{ hazard function } \lambda(\theta) \\ &= \lambda \gamma \theta^{\gamma-1}, \text{ cumulative hazard function } \Lambda(\theta) \\ &= \lambda \theta^\gamma, \text{ mean } E(\theta) = \lambda^{-1/\gamma} \Gamma\left(1 + \frac{1}{\gamma}\right), \text{ variance } V(\theta) \\ &= \lambda^{-2/\gamma} \left(\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma\left(1 + \frac{1}{\gamma}\right)^2 \right), \end{aligned}$$

where, Γ denotes the gamma function with

$$\Gamma(k) = \int_0^\infty s^{k-1} e^{-s} ds \quad (k > 0) \quad (6)$$

We abbreviate the distribution as Weibull $(\lambda; \gamma)$. In the case of $\gamma = 1$, the Exponential distribution is obtained. The hazard function decreases monotonously from ∞ at time zero to zero at time ∞ for $\gamma < 1$, constant (Exponential distribution) for $\gamma = 1$ and it monotonously increases from zero at time zero to ∞ at time ∞ for $\gamma > 1$.

If $X \sim \text{Weibull}(\lambda; \gamma)$, then it holds that $cX \sim \text{Weibull}(\lambda c^{-\gamma}; \gamma)$, when c is a positive constant. Furthermore, the minimum of n i.i.d. variables from this distribution is Weibull $(n\lambda; \gamma)$ (minimum-stable distribution). The Weibull distribution can also be generated as the limiting distribution of the minimum of a sample from a continuous distribution with support on $[0; w)$ for some u ($0 < w < \infty$).

3.4. Gamma distribution

The gamma distribution includes the Exponential distribution as a special case. The gamma distribution is of limited use in survival analysis because the gamma models do not have closed form expressions for survival and hazard functions. Both include the incomplete gamma integral suggested¹⁸

$$I_k(\theta) = \frac{\int_0^\theta s^{k-1} e^{-s} ds}{\Gamma(k)} \quad (7)$$

Consequently, traditional maximum likelihood estimation is difficult and requires the calculation of such incomplete gamma integrals, which imposes additional numerical problems in parameter estimation. A random variable θ is gamma distributed with parameter k and λ ($\theta \sim \Gamma(k; \lambda)$), if the following holds:

Probability density function $f(x)$

$$\begin{aligned} &= \frac{1/\lambda^k \theta^{k-1} e^{-\theta/\lambda}}{\Gamma(k)} \quad k, \lambda > 0, \text{ survival function } S(x) \\ &= 1 - I_k(\lambda x), \text{ hazard function } \lambda(x) \\ &= \frac{1/\lambda^k \theta^{k-1} e^{-\theta/\lambda}}{(1 - I_k(\lambda \theta)) \Gamma(k)}, \text{ mean } E(\theta) = \frac{k}{\lambda}, \text{ variance } V(\theta) = \frac{k}{\lambda^2}, \end{aligned} \quad (8)$$

$$\text{Laplace transforms } L(s) = Ee^{-\theta s} = \left(1 + \frac{s}{\lambda}\right)^{-k} \quad (9)$$

If $k = 1$, the gamma distribution is reduced to the Exponential distribution. With integer k , it can be derived as the distribution of waiting time to the k th emission from a Poisson source with intensity parameter λ . Consequently, the sum of k independent Exponential variates with parameter λ has a gamma distribution with parameters k and λ and can be used to model lifetimes of technical systems with repeated repairing after failure.

For the gamma distribution, the Kendall's tau,⁸ which measures the association between any two event times from the same cluster in the multivariate case, can be computed as,

$$\tau = \frac{\theta}{\theta + 2} \in (0, 1) \quad (10)$$

3.5. Lognormal distribution

In the lognormal model ($\theta \sim \log N(m; s^2)$), the natural logarithm $\ln(\theta)$ of the lifetime θ is assumed to be normally distributed ($\ln(\theta) \sim N(m; s^2)$). A lognormal distribution results when the variable is the product of a large number of independent, identically distributed variables in the same way that a normal distribution results when the variable is the sum of a large number of independent, identically distributed variables. The survival and hazard functions include the incomplete normal integral

$$\varphi(x) = \int_{-\infty}^{\theta} \phi(s) ds, \quad (11)$$

where, $\phi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-(\ln(\theta)-m)^2/2s^2}$ denotes the probability density function of a standard normal distribution. Consequently, Probability density function $f(x)$

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi s \theta}} e^{-(\ln(x)-m)^2/2s^2}, \text{ Mean } E(X) \\ &= e^{m+(s^2/2)}, \text{ Variance } V(X) = e^{2m+s^2} (e^{s^2} - 1) \end{aligned}$$

The lognormal distribution may be convenient to use with non-censored data, but when this distribution is applied to censored data, the computations quickly become formidable. Unfortunately, the hazard function has a strange form; it has value zero at $\theta=0$, increases to a maximum and then decreases, approaching zero as θ heads to infinity. Because of the decreasing form of the hazard function for older ages, the distributions seem implausible as a lifetime model in most situations. Nevertheless, it makes sense if interest is focused on time periods of younger ages. Despite its unattractive features, the lognormal distribution has been widely used as failure distribution in diverse situations, such as the analysis of electrical insulation or time to occurrence of lung cancer among smokers. The lognormal distributions are in practice very close to the inverse Gaussian distributions.

3.6. Positive stable frailty

The positive stable distributions were introduced as a family with two parameters:⁸ a scale $\Delta > 0$ and the so-called index $\alpha < 1$. Imposing

$\Delta = \alpha$, the positive stable frailty distribution $PS^*(nu)$ is obtained, with $nu = 1 - \alpha$.

The associated probability density function is then

$$f(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{\Gamma(k(1-nu) + 1)}{k!} (-u^{nu-1})^k \sin((1-nu)k\pi), \quad nu \in (0, 1). \quad (12)$$

The mean and variance are both undefined. Therefore, the heterogeneity parameter nu does not correspond to the variance of the frailty term. Because of that, we intentionally call it γ instead of θ to avoid misinterpretation.

In contrast to the probability density function, the associated Laplace transform takes a very simple form,

$$L(s) = \exp(-s^{1-nu}), \quad s \geq 0 \quad (13)$$

And for $q \geq 1$,

$$L^{(q)}(s) = (-1)^q ((1-nu)s^{-nu})^q \left[\sum_{m=0}^{q-1} \omega_{q,m} s^{-m(1-nu)} \right] L(s) \quad (14)$$

where, the $\omega'_{q,m}$ are polynomials of degree m , given recursively by⁵

$$\begin{aligned} \omega_{q,0} &= 1 \\ \omega_{q,m} &= \omega_{q-1,m} + \omega_{q-1,m-1} \left\{ \frac{q-1}{1-nu} - (q-m) \right\}, \quad m = 1, 2, 3, \dots, q-2 \end{aligned} \quad (15)$$

$$\omega_{q,q-1} = (1-nu)^{1-q} \left\{ \frac{\Gamma(q-(1-nu))}{\Gamma nu} \right\} \quad (16)$$

It follows that

$$\begin{aligned} \log((-1)^q L^{(q)}(s)) &= q(\log(1-nu) - n \log(s)) \\ &+ \log \left[\sum_{m=0}^{q-1} \omega_{q,m} s^{-m(1-nu)} \right] - s^{1-nu} \end{aligned} \quad (17)$$

$$\tau = nu e(0, 1)$$

3.7. Inverse Gaussian frailty

The inverse Gaussian frailty distribution $IG^*(\theta)$ has density

$$f(u) = \frac{1}{\sqrt{2\pi\theta}} w^{-3/2} \exp\left(-\frac{(w-1)^2}{2\theta w}\right), \quad \theta > 0. \quad (18)$$

The mean and the variance are 1 and θ , respectively. For the Laplace transform, one has

$$L(s) = \exp\left(\frac{1}{\theta} (1 - \sqrt{1 + 1\theta s})\right), \quad s > 0 \quad (19)$$

And, for $q \geq 1$,

$$L^{(q)}(s) = (-1)^q (2\theta s + 1)^{-q/2} \frac{K_{q-(1/2)}\left(\sqrt{2\theta^{-1}(s + (1/2\theta))}\right)}{K_{1/2}\left(\sqrt{2\theta^{-1}(s + (1/2\theta))}\right)} L(s), \quad (20)$$

where, K is the modified Bessel function of the second kind⁸

$$K_{\sigma-1}(p) = 1/2 \int_0^{\infty} t^{\sigma-1} \exp\left\{-\frac{p}{2}\left(t + \frac{1}{t}\right)\right\} dt, \quad \sigma \in \mathbb{R}, \quad p > 0. \quad (21)$$

So that

$$K_{1/2}(p) = \sqrt{\frac{\pi}{2p}} \exp(-p), \text{ we have}$$

$$\begin{aligned} \log((-1)^q L^{(q)}(s)) &= -\frac{q}{2} \log(2\theta s + 1) \\ &+ \log(K_{\sigma-(1/2)}(z)) - \left[\frac{1}{2(\log(\pi/2z))} - z \right] \\ &+ 1/\theta (1 - \sqrt{1 + 2\theta s}) \end{aligned} \quad (22)$$

where, $z = \sqrt{2\theta^{-1}}(s + (1/2\theta))$.

With multivariate data, an inverse Gaussian distributed frailty yields a Kendall's tau given by

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2 \frac{\exp(2/\theta)}{\theta^2} \int_{2/\theta}^{\infty} \frac{\exp(-w)}{w} dw \in \left(\frac{0, 1}{2}\right). \quad (23)$$

In the next section, the criteria required to select the model is explained.

4. Model selection

Several model-selection methods have been proposed in the literature of statistics. The most commonly used methods include information and likelihood-based criteria. To compare the gamma and Weibull frailty models, we apply the information-based criteria. The most commonly used model selection criteria are the AIC and BIC. AIC is given by the expression

$$AIC = -2\log(L) + 2K \quad (24)$$

where, L is the maximum likelihood value and K is the number of parameters in the model. BIC is given by the expression

$$BIC = -2\log(L) + K \ln(N) \quad (25)$$

where, N is the total number of sample size.

5. Result and interpretation

The hazard of patient's death is modelled as a function of the patient-received Therapy, patients MELD score, Gender, Centre and Triglycerides. Clearly, each patient measurement on seven time points cannot be considered as independent. Hence, the shared frailty model was used in this situation, with clusters of size 7 corresponding to patients. We have initiated the model building by considering one covariate at a time. We fit Cox hazard model with and without the frailty. Before fitting this model, we have performed exploratory analysis for the required categorical prognostic factor with respect to the survival time (Figs. 1–4). The Kaplan-Meier estimates between treatment and control (Fig. 1) demonstrate that patients who were receiving treatment had longer survival times as compare to control group. Among male patients, the control groups worked better than cases (Fig. 2). It is important to account for heterogeneity in the Cox hazard model because each patient is considered to have different risks. We use the gamma, inverse Gaussian, positive stable and lognormal frailties as frailty distribution, and with Weibull and Exponential as baseline hazards distribution to study the effect of including frailty in our models. We model the hazard of death as a function of the patient's treatment, Gender,

Fig. 1 – Kaplan–Meier plot for Therapy.

Fig. 2 – Kaplan–Meier plot for Gender.

Centre and MELD Score and Triglycerides values. We have performed parametric proportional hazard model with therapy as the only covariate for no frailty distribution with univariate model, gamma, inverse Gaussian, positive stable and lognormal distribution with Weibull and Exponential baseline hazard distribution simultaneously. Thereafter, we included one covariate every step into the model and checked their goodness of fit and significant change in the estimates with addition of covariate. The models are depicted as follows (Tables 2 and 3):

$$\text{Model 1 : } h_{ij} = h_0(t)w_i(\beta_1 * \text{Therapy}) \quad (26)$$

$$\text{Model 2 : } h_{ij} = h_0(t)w_i(\beta_1 * \text{Therapy} + \beta_2 * \text{Gender}) \quad (27)$$

$$\text{Model 3 : } h_{ij} = h_0(t)w_i(\beta_1 * \text{Therapy} + \beta_2 * \text{Gender} + \beta_3 * \text{MELD}) \quad (28)$$

$$\text{Model 4 : } h_{ij} = h_0(t)w_i(\beta_1 * \text{Therapy} + \beta_2 * \text{Gender} + \beta_3 * \text{MELD} + \beta_4 * \text{Triglycerides}) \quad (29)$$

$$\text{Model 5 : } h_{ij} = h_0(t)w_i(\beta_1 * \text{Therapy} + \beta_2 * \text{Gender} + \beta_3 * \text{MELD} + \beta_4 * \text{Centre} + \beta_5 * \text{Triglycerides}) \quad (30)$$

Fig. 3 – Hazard plot for Therapy.

Fig. 4 – Hazard plot for Gender.

The diagnostics of model were graphically examined in Figs. 5–10. The summaries of parameter estimates for model 5 are given in Tables 4 and 5.

We performed the Wald test to test the significance of the covariate effects. The regression coefficient for therapy after adjusting for the effect of Gender, MELD and Triglycerides is not significant with p -value 0.322 with 95%CI [0.537, 1.226] and p -value 0.135 with 95%CI [0.487, 1.102] for the model without frailty distribution with Weibull and Exponential baseline hazard distribution, respectively.

For gamma frailty, the adjusted coefficient of therapy is significant p -value < 0.044 with 95%CI for HR [1.030, 10.579] for Exponential baseline hazard distribution. However, this covariate is not significant when Weibull baseline hazard distribution is considered into the model, p -value 0.079 with 95%CI [0.907, 6.057]. For inverse Gaussian frailty, the adjusted coefficient of therapy is not significant for p -value < 0.209 with 95%CI [0.464, 33.777] when Weibull baseline hazard distribution is considered, and covariate therapy is significant when Exponential baseline hazard distribution is considered into the

Table 2 – Summary of regression coefficient after considering Therapy, Gender, MELD and Triglycerides as covariates in the model.

Covariate	Baseline hazard distribution	Frailty distribution														
		No frailty distribution					Gamma frailty distribution					Inverse Gaussian frailty distribution				
		Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI for HR	p-Value
Therapy, Gender, MELD and Triglycerides	Weibull	Theta (θ)	NA	NA	NA	NA	Theta (θ)	3.906	1.343	NA	NA	Theta (θ)	19.711	42.784	NA	NA
		β_{Therapy}	−0.209	0.211	[0.537, 1.226]	0.322	β_{Therapy}	0.852	0.484	[0.907, 6.057]	0.079	β_{Therapy}	1.376	1.094	[0.464, 33.777]	0.209
		β_{Gender}	−0.005	0.185	[0.692, 1.430]	0.978	β_{Gender}	0.076	0.462	[0.436, 2.670]	0.869	β_{Gender}	−0.377	0.578	[0.221, 2.132]	0.515
		β_{MELD}	0.242	0.021	[1.221, 1.328]	<0.0001	β_{MELD}	0.114	0.041	[1.035, 1.214]	0.005	β_{MELD}	0.150	0.045	[1.064, 1.269]	0.001
		$\beta_{\text{Triglycerides}}$	0.0045	0.002	[0.997, 1.003]	0.960	$\beta_{\text{Triglycerides}}$	0.004	0.003	[0.998, 1.010]	0.215	$\beta_{\text{Triglycerides}}$	0.004	0.003	[0.998, 1.011]	0.211
		Log-likelihood: −835.912; Kendall's Tau: NA					Log-likelihood: −808.377; Kendall's Tau: 0.661					Log-likelihood: −811.698; Kendall's Tau: 0.46				
Covariate	Baseline hazard distribution	Positive stable frailty distribution					Lognormal frailty distribution									
		Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI	p-Value					
Therapy, Gender, MELD and Triglycerides	Weibull	Theta (θ)	30.3	12.4	NA	NA	Theta (θ)	5.246	2.281	NA	NA					
		β_{Therapy}	0.871	0.931	[0.386, 14.820]	0.349	β_{Therapy}	1.531	0.752	[1.059, 20.199]	0.042					
		β_{Gender}	−0.275	0.427	[0.329, 1.755]	0.520	β_{Gender}	−0.015	0.546	[0.338, 2.874]	0.978					
		β_{MELD}	0.217	0.054	[1.118, 1.380]	<0.0001	β_{MELD}	0.131	0.044	[1.047, 1.243]	0.003					
		$\beta_{\text{Triglycerides}}$	0.002	0.003	[0.995, 1.008]	0.574	$\beta_{\text{Triglycerides}}$	0.004	0.003	[0.997, 1.011]	0.315					
		Log-likelihood: −810.203; Kendall's Tau: 0.303					Log-likelihood: −804.695; Kendall's Tau: NA									
Covariate	Baseline hazard distribution	No frailty distribution					Gamma frailty distribution					Inverse Gaussian frailty distribution				
		Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI	p-Value	Parameters	Estimates	S.E	95%CI	p-Value
Therapy, Gender, MELD and Triglycerides	Exponential	Theta (θ)	NA	NA	NA	NA	Theta (θ)	5.288	1.872	NA	NA	Theta (θ)	603.307	1868.457	NA	NA
		β_{Therapy}	−0.311	0.208	[0.487, 1.102]	0.135	β_{Therapy}	1.194	0.594	[1.030, 10.579]	0.044	β_{Therapy}	2.089	0.784	[1.739, 37.546]	0.008
		β_{Gender}	0.035	0.185	[0.721, 1.488]	0.848	β_{Gender}	0.252	0.573	[0.418, 3.955]	0.660	β_{Gender}	−0.340	0.605	[0.217, 2.333]	0.575
		β_{MELD}	0.258	0.021	[1.241, 1.350]	<0.0001	β_{MELD}	0.097	0.045	[1.010, 1.203]	0.029	β_{MELD}	0.146	0.039	[1.072, 1.248]	<0.0001
		$\beta_{\text{Triglycerides}}$	−0.001	0.002	[0.996, 1.003]	0.725	$\beta_{\text{Triglycerides}}$	0.005	0.004	[0.998, 1.012]	0.154	$\beta_{\text{Triglycerides}}$	0.004	0.004	[NA, NA]	0.310
		Log-likelihood: −966.984; Kendall's Tau: NA					Log-likelihood: −819.008; Kendall's Tau: 0.726					Log-likelihood: −819.474; Kendall's Tau: 0.498				
Covariate	Baseline hazard distribution	Frailty distribution positive stable					Lognormal frailty distribution									
		Parameters	Estimates	S.E	95%CI	p-Value	Parameters	Estimates	S.E	95%CI	p-Value					
Therapy, Gender, MELD and Triglycerides	Exponential	Theta (θ)	51.0	8.00	NA	NA	Theta (θ)	NA	NA	NA	NA					
		β_{Therapy}	2.222	1.019	[1.251, 68.026]	0.029	β_{Therapy}	NA	NA	NA	NA					
		β_{Gender}	−0.330	0.619	[0.214, 2.421]	0.594	β_{Gender}	NA	NA	NA	NA					
		β_{MELD}	0.141	0.052	[1.040, 1.275]	0.007	β_{MELD}	NA	NA	NA	NA					
		$\beta_{\text{Triglycerides}}$	0.004	0.004	[NA, NA]	0.319	$\beta_{\text{Triglycerides}}$	NA	NA	NA	NA					
		Log-likelihood: −819.576; Kendall's Tau: 0.51					Log-likelihood: NA; Kendall's Tau: NA									

Table 3 – Summary of AIC and BIC values after considering Therapy, Gender, MELD and Triglycerides as covariates in the model.

Covariate	Information criteria	Baseline hazard distribution	Frailty distribution				
			No frailty distribution	Gamma frailty distribution	Inverse Gaussian frailty distribution	Positive stable frailty distribution	Lognormal frailty distribution
Therapy, Gender, MELD and Triglycerides	AIC	Weibull	1685.824	1632.753	1639.396	1636.406	1625.39
		Exponential	1725.321	1652.016	1652.948	1653.152	–
	BIC	Weibull	1718.625	1670.240	1676.883	1673.893	1662.877
		Exponential	1753.436	1684.817	1685.749	1685.953	–

Fig. 5 – Prediction of frailties for the liver transplantation data, as given by the parametric gamma frailty with Exponential baseline hazard. **Fig. 6 – Prediction of frailties for the liver transplantation data, as given by the parametric inverse Gaussian frailty with Weibull baseline hazard.**

model (p -value 0.008 with 95%CI [1.739, 37.546]). For positive stable frailty, the adjusted coefficient of therapy is not significant at p -value 0.349 with 95%CI for HR [0.386, 14.820] when Weibull baseline hazard distribution is considered. However, this covariate is significant when Exponential baseline hazard distribution is considered into the model (p -value 0.029 with 95%CI [1.251, 68.026]). For lognormal frailty, the adjusted coefficient of therapy is significant p -value 0.042 with 95%CI for HR [1.059, 20.199] when Weibull baseline hazard distribution is considered. With Weibull and Exponential baseline distribution, lognormal frailty model gives slightly lower frailty variance compared to the other frailty models. To compare all frailty models, we use the AIC and BIC methods. All five frailty models have the same number of parameters

($k = 4$) (Fig. 11). Comparing the five models, lognormal frailty model has the smallest AIC and BIC values, suggesting that it is a good fit for these covariates in data (Figs. 11–15).

The lognormal frailty model gives slightly lower frailty variance compared to the other frailty models when Weibull distribution considered baseline hazard. From this, we can conclude that Therapy and MELD scores have significant role in liver transplant patients when distribution of frailty is considered as lognormal with Weibull baseline hazard for fitting the model. Therefore, there is 4.6 times more risk getting hazard of mortality for patients receiving conventional therapy with consent for participation as compared to patients receiving new therapy with consent for participation with 95% CI [1.059, 20.199], p -value 0.042. There is 1.14 times risk of

Fig. 7 – Prediction of frailties for the liver transplantation data, as given by the parametric inverse Gaussian frailty with Exponential baseline hazard.

Fig. 9 – Prediction of frailties for the liver transplantation data, as given by the parametric positive stable frailty with Exponential baseline hazard.

Fig. 8 – Prediction of frailties for the liver transplantation data, as given by the parametric positive stable frailty with Weibull baseline hazard.

Fig. 10 – Prediction of frailties for the liver transplantation data, as given by the parametric lognormal frailty with Exponential baseline hazard.

Table 4 – Summary of regression coefficient after considering Therapy, Gender, MELD, Centre and Triglycerides as covariates in the model.

Covariate	Baseline hazard distribution	Frailty distribution														
		No frailty distribution					Gamma frailty distribution					Inverse Gaussian frailty distribution				
		Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI for HR	p-Value
Therapy, Gender, MELD, Centre and Triglycerides	Weibull	Theta (θ)	NA	NA	NA	NA	Theta (θ)	3.932	1.390	NA	NA	Theta (θ)	647.438	2023.807	NA	NA
		β_{Therapy}	−0.209	0.211	[0.537, 1.226]	0.322	β_{Therapy}	0.823	0.482	[0.885, 5.85]	0.088	β_{Therapy}	2.347	0.785	[2.243, 48.711]	0.003
		β_{Gender}	−0.005	0.185	[0.692, 1.430]	0.978	β_{Gender}	−0.297	0.613	[0.223, 2.47]	0.628	β_{Gender}	−1.616	0.871	[0.036, 1.095]	0.064
		β_{MELD}	0.242	0.021	[1.221, 1.328]	<0.0001	β_{MELD}	0.118	0.041	[1.037, 1.22]	0.004	β_{MELD}	0.143	0.038	[1.072, 1.242]	0.001
		β_{Centre}	−2.673	0.023	[0.028, 0.169]	0.457	β_{Centre}	−2.794	0.589	[0.019, 0.19]	<0.0001	β_{Centre}	−6.646	2.336	[0.000, 0.127]	0.004
		$\beta_{\text{Triglycerides}}$	0.0045	0.002	[0.997, 1.003]	0.960	$\beta_{\text{Triglycerides}}$	0.004	0.003	[0.998, 1.01]	0.196	$\beta_{\text{Triglycerides}}$	0.004	0.003	[0.996, 1.011]	0.354
		Log-likelihood: −809.263; Kendall's Tau: NA					Log-likelihood: −809.263; Kendall's Tau: 0.663					Log-likelihood: −810.193; Kendall's Tau: 0.498				
Covariate	Baseline hazard distribution	Positive stable frailty distribution					Lognormal frailty distribution									
		Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI	p-Value					
Therapy, Gender, MELD, Centre and Triglycerides	Weibull	Theta (θ)	30.3	12.4	NA	NA	Theta (θ)	5.246	2.281	NA	NA					
		β_{Therapy}	0.871	0.931	[0.386, 14.820]	0.349	β_{Therapy}	1.531	0.752	[1.059, 20.199]	0.042					
		β_{Gender}	−0.275	0.427	[0.329, 1.755]	0.520	β_{Gender}	−0.015	0.546	[0.338, 2.874]	0.978					
		β_{MELD}	0.217	0.054	[1.118, 1.38]	<0.0001	β_{MELD}	0.131	0.044	[1.047, 1.243]	0.003					
		β_{Centre}	−2.453	0.022	[0.027, 0.159]	0.457	β_{Centre}	−2.694	0.559	[0.014, 0.184]	<0.0001					
		$\beta_{\text{Triglycerides}}$	0.002	0.003	[0.995, 1.008]	0.574	$\beta_{\text{Triglycerides}}$	0.004	0.003	[0.997, 1.011]	0.315					
		Log-likelihood: −811.202; Kendall's Tau: 0.303					Log-likelihood: −805.696; Kendall's Tau: NA									
Covariate	Baseline hazard distribution	No frailty distribution					Gamma frailty distribution					Inverse Gaussian frailty distribution				
		Parameters	Estimates	S.E	95%CI for HR	p-Value	Parameters	Estimates	S.E	95%CI	p-Value	Parameters	Estimates	S.E	95%CI	p-Value
Therapy, Gender, MELD, Centre and Triglycerides	Exponential	Theta (θ)	NA	NA	NA	NA	Theta (θ)	5.484	1.995	NA	NA	Theta (θ)	603.173	1533.642	NA	NA
		β_{Therapy}	−0.311	0.208	[0.487, 1.102]	0.135	β_{Therapy}	1.175	0.603	[0.993, 10.563]	0.051	β_{Therapy}	2.340	0.786	[2.224, 48.434]	0.003
		β_{Gender}	0.035	0.185	[0.721, 1.488]	0.848	β_{Gender}	−0.079	0.785	0.198, 4.305]	0.920	β_{Gender}	−1.609	0.871	[0.036, 1.102]	0.065
		β_{MELD}	0.258	0.021	[1.241, 1.350]	<0.0001	β_{MELD}	0.099	0.046	[1.009, 1.209]	0.031	β_{MELD}	0.147	0.039	[1.074, 1.249]	<0.0001
		β_{Center}	−2.673	0.023	[0.028, 0.169]	0.457	β_{Centre}	−2.956	0.669	[0.014, 0.193]	<0.0001	β_{Centre}	−6.675	2.296	[0.000, 0.114]	0.004
		$\beta_{\text{Triglycerides}}$	−0.001	0.002	[0.996, 1.003]	0.725	$\beta_{\text{Triglycerides}}$	0.005	0.004	[0.998, 1.012]	0.139	$\beta_{\text{Triglycerides}}$	0.004	0.004	[NA, NA]	0.340
		Log-likelihood: −809.263; Kendall's Tau: NA					Log-likelihood: −820.149; Kendall's Tau: 0.733					Log-likelihood: −818.348; Kendall's Tau: 0.498				
Covariate	Baseline hazard distribution	Frailty distribution positive stable					Lognormal frailty distribution									
		Parameters	Estimates	S.E	95%CI	p-Value	Parameters	Estimates	S.E	95%CI	p-Value					
Therapy, Gender, MELD, Centre and Triglycerides	Exponential	Theta (θ)	51.0	8.00	NA	NA	Theta (θ)	NA	NA	NA	NA					
		β_{Therapy}	2.222	1.019	[1.251, 68.06]	0.029	β_{Therapy}	NA	NA	NA	NA					
		β_{Gender}	−0.330	0.619	[0.214, 2.42]	0.594	β_{Gender}	NA	NA	NA	NA					
		β_{MELD}	0.141	0.052	[1.040, 1.275]	0.007	β_{MELD}	NA	NA	NA	NA					
		β_{Centre}	−2.643	0.023	[0.025, 0.149]	0.457	β_{Centre}	−2.856	0.659	[0.013, 0.195]	<0.0001					
		$\beta_{\text{Triglycerides}}$	0.004	0.004	[NA, NA]	0.319	$\beta_{\text{Triglycerides}}$	NA	NA	NA	NA					
		Log-likelihood: −819.456; Kendall's Tau: 0.54					Log-likelihood: NA; Kendall's Tau: NA									

Table 5 – Summary of AIC and BIC values after considering Therapy, Gender, Centre, MELD and Triglycerides as covariates in the model.

Covariate	Information criteria	Baseline hazard distribution	Frailty distribution				
			No frailty distribution	Gamma frailty distribution	Inverse Gaussian frailty distribution	Positive stable frailty distribution	Lognormal frailty distribution
Therapy, Gender, Centre, MELD And Triglycerides	AIC	Weibull	1665.834	1642.653	1649.386	1626.416	1635.29
		Exponential	1735.421	1642.116	1642.958	1643.162	–
	BIC	Weibull	1728.645	1660.248	1666.893	1683.883	1652.887
		Exponential	1743.456	1686.917	1687.949	1688.853	–

Fig. 11 – Comparing the five models with respect to AIC and BIC values when covariate is considered as Therapy.

getting hazards of mortality for every one unit increase in the MELD score with 95%CI [1.047, 1.243], p -value 0.003. There is 1.4% less risk getting hazard of mortality for female patients as compared to male patients with [0.338, 2.874] p -value 0.978. The unit increase in Triglycerides has no risk on hazard of mortality with [0.997, 1.011], p -value 0.315.

6. Discussion

This paper has explored different baseline hazard and frailty distributions to identify the best suitable distributions on Liver

Cirrhosis patient's data. With Weibull and Exponential baseline distribution, lognormal frailty model gives slightly lower frailty variance compared to the other frailty models. To compare all frailty models, we use the AIC and BIC methods. All five frailty models have the same number of parameters. Comparing the five models, lognormal frailty model has the smallest AIC and BIC values, suggesting that it is a good fit for these covariates in data. The concept of frailty provides a suitable way to introduce random effects in the model to account for relationship and unobserved heterogeneity. The dependence usually arises because individuals in the same group are related to each other or because of the recurrence of

Fig. 12 – Comparing the five models with respect to AIC and BIC values when covariate is considered as Therapy and Gender.

Fig. 13 – Comparing the five models with respect to AIC and BIC values when covariate is considered as Therapy, Gender and MELD.

Fig. 14 – Comparing the five models with respect to AIC and BIC values when covariate is considered as Therapy, Gender, MELD and Triglycerides.

Fig. 15 – Comparing the five models with respect to AIC and BIC values when covariate is considered as Therapy, Gender, MELD, Centre and Triglycerides.

an event for the same individual. Multivariate frailty models have been used frequently for modelling dependence in multivariate time-to-event data. Many distributions can be chosen for the frailty, but the most common frailty distribution is the gamma distribution. The gamma distribution has been widely applied as a mixture distribution.

Conflicts of interest

The authors have none to declare.

Acknowledgment

The first author acknowledges the support of UGC-BSR for one-time grant (F-19-145/2015 (BSR)) of University Grant Commission, New Delhi.

REFERENCES

1. Munda M, Legrand C. PARFM: parametric frailty models in R. *J Stat Softw.* 2012;51(November (11)).
2. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography.* 1979;16:439–454.
3. Lancaster T. Econometric methods for the duration of unemployment. *Econometrica.* 1979;47:939–956.
4. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika.* 1978;65:141–151.
5. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag; 2003.
6. Cox DR. Regression models and life-tables. *J R Stat Soc B.* 1972;34(2):187–220.
7. Hirsch K, Wienke A. Software for semiparametric shared gamma and log-normal frailty models: an overview. *Comput Methods Progr Biomed.* 2012;107(3):582–597.
8. Hougaard P. Frailty models for survival data. *Lifetime Data Anal.* 1995;1(3):255–273.
9. Duchateau L, Janssen P. *The Frailty Model.* Springer-Verlag; 2008.
10. Cortinas Abrahantes J, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. Comparison of different estimation procedures for proportional hazards model with random effects. *Comput Stat Data Anal.* 2007;51(8):3919–3930.
11. van den Berg GJ, Drepper B. *Inference for Shared-Frailty Survival Models with Left-Truncated Data.* Working Papers 12-5, University of Mannheim, Department of Economics. 2012 <http://ideas.repec.org/p/mnh/wpaper/30729.html>.
12. Morley JE, Perry H, Miller DK. Something about frailty. *J Gerontol Ser A: Biol Sci Med Sci.* 2002;57:M698–M704;ĆMurphy SA. Asymptotic theory for the frailty model. *Ann Stat.* 1995;23:182–198.
13. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics.* 1978;34:541–554.
14. Manatunga AK, Oakes D. Parametric analysis for matched pair survival data. *Lifetime Data Anal.* 1999;5:371–387.
15. Andersen PK, Klein JP, Zhang M. Testing for centres effects in multicentre survival studies: a Monte Carlo comparison of fixed and random effects. *Stat Med.* 1999;18:1489–1500.
16. Wang ST, Klein JP, Moeschberger ML. Semi-parametric estimation of covariate effects using the positive stable frailty model. *Appl Stoch Models Data Anal.* 1995;11(2):121–133.
17. Shu Y, Klein JP. A SAS macro for the positive stable frailty model. In: *Proceedings of the Statistical Computing Section.* 1999; 47–52.
18. Habil, et al. *Frailty Models in Survival Analysis.* [Ph.D. thesis] 2007 <http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000011542>.
19. Nath DC, Bhattacharjee A, Vishwakarma RK. Cure rate modeling: an application with Bayesian approach in liver cirrhosis. *J Data Sci.* 2015;13(2015):421–442.