Statistics and Applications {ISSN 2454-7395 (online)}
Volume 21, No. 2, 2023 (New Series), pp 301–318
http://www.ssca.org.in/journal

Modeling Bivariate Survival Data By Compound Frailty Distributions

David D. Hanagal[1] and Alok D. Dabade[2]
[1]*Department of Statistics, Savitribai Phule Pune University, Pune, India*
[2]*Department of Statistics, University of Mumbai, Mumbai, India*

Received: 14 March 2023; Revised: 08 April 2023; Accepted: 17 April 2023

Abstract

Share frailty models are often used to model heterogeneity in survival analysis. In these models, it is assumed that each individual from a group shares common frailty, but sometimes it may be possible that some individuals will have zero susceptibility to an event. In such cases, compound distributions are more proper to model shared frailty than usually preferred distributions, gamma, lognormal etc. In this paper we have considered compound Poisson and compound negative binomial frailty distributions with IDB as baseline distribution. Since it has increasing, decreasing, constant and bathtub shaped hazard function. MCMC approach have been used to estimate the parameters involved in the models. A real life data analysis is also considered by applying the proposed models.. . .

Key words: Bayesian model comparison; Compound negative binomial distribution; Compound Poisson distribution; IDB distribution; MCMC; Shared frailty.

AMS Subject Classifications: 62F15, 62N01, 62P10

1. Introduction

In survival data, researchers are interested to study effect of covariates on life times of individuals from a group. For example, medical practitioner in case of lung cancer patients, may be interested to study how the factors such as age, health condition of the patient and the type of tumor may affect the survival times. In experiments on the time to failure of electrical insulation, engineer is interested to find the effect of the voltage, the insulation is subject to. Also in clinical trials, the experimenter is interested to study effect of the treatment assigned to a patient on the survival time. Unfortunately, many of the times it is impossible to include all relevant covariates. May be because, we have little or no information on the individual level. For example, it is known that excretion of small amounts of albumin in the urine is a diagnostic marker for increased mortality, however we are unable to include this variable, unless we actually obtain urine and analyze samples for each individual under study. Furthermore, we may not aware the relevance of the risk factor or even that the factor we ought to include in the analysis. For example, a genetic factor as we do not know all

Corresponding Author: David D. Hanagal
Email: david.hanagal@gmail.com

possible genes having influence on survival. In other cases, it may be impossible to measure the risk factor without great financial cost or time effort. In such cases, the usual practice is to ignore such covariates. The neglect of such covariates leads to heterogeneity into the data. This heterogeneity is named as frailty by Vapuel *et al.* (1979). To address the frailty, it is necessary to include random effect term into the model. Such models are well known as frailty models.

Sometimes individuals from a group share a common frailty, for example, if we consider data on twins then for monozygotic twins, sex, any other genetically based covariates, date of birth and pre-birth environment is common. For the timings of failures of several paired human organs like kidneys, lungs, eyes, ears etc. shares common frailty because they are of same individual. In case of sequences of times of asthmatic attacks of asthma patients or in tumor diagnosis, tumor recurrence times in individual patients also has common frailty because occurrence time of an event is on same individual. In industrial applications, if we consider the breakdown times of dual generator in a power plant or failure times of two engines in a two engines airplane then common environment is shared by both the engines and generators. In such situations, shared frailty models are suggested in the literature (see Clayton (1978)).

Hanagal (2005) proposed a positive stable frailty model with bivariate exponential of Marshall-Olkin (1967) as baseline distribution. Hanagal (2006) discussed the gamma frailty regression model in the bivariate survival data and Hanagal (2007) also presented the gamma frailty regression models in the mixture distributions. Hanagal and Sharma (2013, 2015a, 2015b, 2015c) analyzed diabetic retinopathy data, acute leukaemia data and kidney infection data using shared gamma and inverse Gaussian frailty models.

In shared frailty models, it is assumed that, each individual from a group experiences an event of interest but sometimes it may be possible that some individuals are immune to a particular event *i.e.*, they are non-susceptible or they have zero susceptibility. For example, some cancer patients survive their cancer. In medicine, there are several examples of diseases primarily attacking people with particular susceptibility, for instance, a genetic kind, other people having virtually zero susceptibility of getting the disease. Another example is fertility, some couples are unable to conceive children so that the time to have first child birth for them have zero susceptibility. In case of marriages, some people never marry, some marriages are not prone to dissolve so that time to divorce for such couples have zero susceptibility. In such type of data, compound distribution having some positive mass at zero value can be a suitable choice. For example, compound Poisson distribution or compound negative binomial distribution.

Aalen (1992) considered a compound Poisson distribution as a mixture distribution in survival analysis. Also, Moger and Aalen (2005), Hanagal (2010a), Hanagal (2010b), Hanagal and Dabade (2012) and Hanagal and Kamble (2015) have considered compound Poisson frailty models. Hanagal and Dabade (2013) and Hanagal and Kamble (2016) have introduced compound negative binomial shared frailty model. Recently Hanagal (2023a, 2024a, 2024b) introduced compound Poisson frailty models based on additive hazard, correlated compound Poisson frailty models based on the hazard rate and reversed hazard rates to analyze kidney infection data and Australian twin data. Hanagal (2023b) proposed correlated compound geometric frailty models to analyze kidney infection data. More details on compound Poisson

frailty models are available in Hanagal (2011, 2019).

A random variable Z following a compound distribution is defined as,

$$Z = \begin{cases} Y_1 + Y_2 + \cdots + Y_N & ; \ N > 0 \\ 0 & ; \ N = 0. \end{cases} \tag{1}$$

where N is also random variable with some statistical distribution and $Y_1, Y_2, \cdots, Y_N$ are independent, identically gamma distributed random variables with scale parameter ν and shape parameter γ having density function,

$$f(y) = \begin{cases} \frac{\nu^\gamma}{\Gamma(\gamma)} y^{\gamma-1} e^{-\nu y} & ; \ y > 0, \nu > 0, \gamma > 0 \\ 0 & ; \ \text{otherwise.} \end{cases}$$

Here, variable Y_i represents length of i^{th} failure. If $N = 0$ frailty is not at all affecting the life times of an individual from a group and if $N > 0$ then frailty is cumulative effect of heterogeneity due to N failures.

Aalen and Tretli (1999) modelled testis cancer data using compound Poisson frailty model. A man receives damages during a critical period of their fetal development which may develop testis cancer after the hormonal process of puberty has started. The damage may be a result of the mother's exposure to environmental factors, for example an excessive estrogenic burden, and may also interact with genetic factors. Aalen and Tretli (1999) represented Y_i as size of the damage at i^{th} occasion and N be the number of damages occurred. Thus Z is now cumulative effect of damages occurred. Some other examples can be given as, in case of marriage data, Z may represents cumulative heterogeneity for not getting a perfect partner due to different unknown difficulties like, medical issues of an individual, hereditary problems etc. In case of fertility, Z may be cumulative effect due to different unknown reasons such as, effect of miss-carriages on health, male infertility, age related issues *etc.* However, Aalen and Tretli (1999) says, this point of view should not be taken too literally as a description of biological reality. The main reason for using compound frailty random variables is statistical convenience. Compound Poisson and compound negative binomial distribution both have simple and closed from expression of Laplace transform, which a requirement of any frailty model.

To complete the parametric form of the model we now make assumption on baseline distribution. Weibull distribution is one of the most widely used baseline distribution. Hazard function for Weibull distribution is a monotone function, which increases with time to infinity when shape parameter α is greater than one and it decreases up to the value zero for $\alpha < 1$. At time zero, it has a zero-failure rate implies that almost no failure will occur which is hardly feasible in real life. Also, other usually preferred baseline distributions such as, gamma, lognormal etc. has monotone hazard function. So, there is a need to have another baseline distribution which is feasible to model increasing, decreasing and bathtub shape hazard function. Hjort (1980) introduced Increasing, Decreasing, Constant and Bathtub-shaped failure rate distribution (IDB) which has all the above shapes. Also at time zero, failure rate is positive. So, we thought IDB distribution can be better than Weibull to model as baseline distribution.

For estimation of parameters of the model, we have considered MCMC technique. To check the performance of the model we have considered simulation study. Also, we have

applied the proposed models to a bivariate survival data set of McGrilchrist and Aisbett (1991) related to kidney infection and suggested the best model by using Bayesian model comparison techniques. The remainder of the paper is organized as follows, in Section 2, we provide introduction to general bivariate shared frailty model. In Section three, baseline distribution IDB is discussed. Section four and five respectively considers compound Poison and compound negative binomial shared frailty models. Section 6 is contributing to proposed models. In section 7 estimation procedure is discussed followed by simulation study and data analysis of kidney infection data in Section 8 and 9 respectively. Finally, paper concluded with Conclusion.

2. General bivariate shared frailty model

Suppose n individuals are observed for the study and let a bivariate random variable (T_{1j}, T_{2j}) be represent first and second survival time of j^{th} individual $(j = 1, 2, 3, \ldots, n)$. Also suppose that there are k observed covariates collected in a vector $\underline{X}_j = (X_{1j}, \ldots, X_{kj})$ for j^{th} individual where X_{aj} $(a = 1, 2, 3, \ldots, k)$ represent the value of a^{th} observed covariate for j^{th} individual. Here we assume that both the survival times for each individual share the same value of the covariates.

Let Z_j be represent shared frailty variable for j^{th} individual. Assuming that the frailties are acting multiplicatively on the baseline hazard function and both the survival times of individuals are conditionally independent for given frailty, the conditional hazard function and hence conditional survival function for j^{th} individual at i^{th} $(i = 1, 2)$ survival time $t_{ij} > 0$ for given frailty $Z_j = z_j$ has the form respectively,

$$h(t_{ij} \mid z_j, \underline{X}_j) = z_j h_0(t_{ij}) \eta_j \tag{2}$$

$$S(t_{ij} \mid z_j, \underline{X}_j) = e^{-z_j H_0(t_{ij}) \eta_j} \tag{3}$$

where $h_0(t_{ij})$ and $H_0(t_{ij})$ are respectively baseline hazard and cumulative baseline hazard functions at time $t_{ij} > 0$; $\eta_j = e^{\underline{X}_j \underline{\beta}}$ and $\underline{\beta}$ is a vector of order k, of regression coefficients. Under the assumption of independence, bivariate conditional survival function for given frailty $Z_j = z_j$ at time $t_{1j} > 0$ and $t_{2j} > 0$ is,

$$S(t_{1j}, t_{2j} \mid z_j, \underline{X}_j) = e^{-z_j (H_{01}(t_{1j}) + H_{02}(t_{2j})) \eta_j} \tag{4}$$

Unconditional bivariate survival function at time $t_{1j} > 0$ and $t_{2j} > 0$ is obtained by integrating over frailty variable Z_j having the probability function f(z_j), for j^{th} individual.

$$S(t_{1j}, t_{2j} \mid \underline{X}_j) = \int_{Z_j} S(t_{1j}, t_{2j} \mid z_j) f(z_j) dz_j = L_{Z_j}[(H_{01}(t_{1j}) + H_{02}(t_{2j}))\eta_j]$$

where $L_{Z_j}(.)$ is Laplace transform of frailty variable of Z_j for j^{th} individual. Thus, unconditional bivariate survival function for j^{th} individual at time $t_{1j} > 0$ and $t_{2j} > 0$ is,

$$S(t_{1j}, t_{2j} \mid \underline{X}_j) = L_{Z_j}[(H_{01}(t_{1j}) + H_{02}(t_{2j}))\eta_j] \tag{5}$$

Here onwards we represent $S(t_{1j}, t_{2j} \mid \underline{X}_j)$ as $S(t_{1j}, t_{2j})$.

Once we have unconditional survival function of bivariate random variable (T_{1j}, T_{2j}) we can obtain likelihood function and estimate the parameters of the model.

3. Baseline distribution

A continuous random variable T is said to follow three parameters Increasing, Decreasing, Constant and Bathtub-shaped (IDB) distribution if its survival function is given by,

$$S_0(t) = \begin{cases} \dfrac{e^{-\frac{\lambda t^2}{2}}}{(1+\alpha t)^{\frac{\theta}{\alpha}}} & ; \quad t>0, \alpha>0, \lambda>0, \theta>0 \\ 1 & ; \qquad \text{otherwise.} \end{cases} \tag{6}$$

Corresponding density function, hazard function and cumulative hazard function are respectively;

$$f_0(t) = \begin{cases} \dfrac{\theta+\lambda t\,(1+\alpha t)}{(1+\alpha t)^{1+\theta/\alpha}} \exp\left(-\frac{\lambda t^2}{2}\right) & ; \quad t>0, \alpha>0, \lambda>0, \theta>0 \\ 0 & ; \qquad \text{otherwise.} \end{cases} \tag{7}$$

$$h_0(t) = \begin{cases} \lambda t + \dfrac{\theta}{1+\alpha t} & ; \quad t>0, \alpha>0, \lambda>0, \theta>0 \\ 0 & ; \qquad \text{otherwise.} \end{cases} \tag{8}$$

$$H_0(t) = \begin{cases} \dfrac{\lambda t^2}{2} + \dfrac{\theta}{\alpha}\log(1+\alpha t) & ; \quad t>0, \alpha>0, \lambda>0, \theta>0 \\ 0 & ; \qquad \text{otherwise.} \end{cases} \tag{9}$$

It is easy to observe that, first term of hazard function increases and second term decreases with increase in time. So, if λ is 0 then hazard function is decreasing function and for $\theta = 0$ it is increasing in nature. From the difference between hazard function for two different time points $0 < t_1 < t_2$, $h_0(t_1) - h_0(t_2) = (t_2 - t_1)\left[\dfrac{\alpha\theta}{(1+\alpha t_1)(1+\alpha t_2)} - \lambda\right]$, we can observe that, for $\lambda \geq \alpha\theta$ hazard function is increasing function and for $0 < \lambda < \alpha\theta$ hazard function will have bathtub shape. For $\lambda = 0 = \alpha$ it has a constant hazard function.

4. Compound Poison shared frailty model

A random variable defined in (1) is said to follow compound Poisson distribution if N is Poisson distributed with mean ρ. The distribution of Z consists of two parts; a discrete part which corresponds to the probability of zero susceptibility, and a continuous part on the positive real line. The discrete part is, $P(Z=0) = e^{-\rho}$, which decreases as ρ increases and the distribution of the continuous part can be found by conditioning N and using the fact that the Y_i's are gamma distributed. It can be written as

$$f(z;\gamma,\nu,\rho) = \begin{cases} \frac{1}{z}e^{-(\rho+\nu z)} \sum\limits_{n=1}^{\infty} \dfrac{\rho^n(\nu z)^{n\gamma}}{\Gamma(n\gamma)n!} & ; \; z>0, \rho>0, \nu>0, \gamma>0 \\ 0 & ; \; \text{otherwise} \end{cases}$$

The parameter set for the compound Poisson distribution is $\rho > 0, \nu > 0, \gamma > 0$. The moments; mean, variance and Laplace transform of compound Poisson distribution are given by,

$$L_Z(s) = exp\left\{-\rho\left[1-\left(\frac{\nu}{\nu+s}\right)^{\gamma}\right]\right\} \tag{10}$$

$$E(Z) = \frac{\rho\gamma}{\nu}; \quad Var(Z) = \frac{\rho\gamma(\gamma+1)}{\nu^2}. \tag{11}$$

The shared frailty models are suffering from non-identifiability. To resolve the issue, as usual, we assume Z has expected value equal, which imposes the restriction $\nu = \rho\gamma$ on the parameters of compound Poisson distribution. Under the restriction Laplace transformation of compound Poisson distribution reduces to,

$$L_Z(s) = exp\left\{-\rho\left[1-\left(1+\frac{s}{\rho\gamma}\right)^{-\gamma}\right]\right\} \tag{12}$$

with variance $\dfrac{\gamma+1}{\rho\gamma}$. Replacing Laplace transformation in equation (5), we get the unconditional bivariate survival function for j^{th} individual at time $t_{1j} > 0$ and $t_{2j} > 0$ as,

$$S(t_{1j}, t_{2j}) = exp\left\{-\rho\left[1-\left(1+\frac{(H_{01}(t_{1j})+H_{02}(t_{2j}))\eta_j}{\rho\gamma}\right)^{-\gamma}\right]\right\} \tag{13}$$

Clayton (1978) defined a cross-ratio function given by,

$$\theta^*(t_1,t_2) = \frac{\lambda_1(t_1 \mid T_2 = t_2)}{\lambda_1(t_1 \mid T_2 > t_2)} = \frac{\lambda_2(t_2 \mid T_1 = t_1)}{\lambda_2(t_2 \mid T_1 > t_1)} = \frac{S(t_1,t_2)\dfrac{\partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2}}{\dfrac{\partial S(t_1,t_2)}{\partial t_1}\dfrac{\partial S(t_1,t_2)}{\partial t_2}}$$

where $\lambda_1(.)$ and $\lambda_2(.)$ are conditional hazard functions of T_1 and T_2. It is an association function such that,

$$\theta^*(t_1,t_2)\begin{cases} > 1 & ;\text{positive association} \\ = 1 & ;\text{no association} \\ < 1 & ;\text{negative association} \end{cases}$$

For compound Poisson shared frailty model cross-ratio function is given by,

$$\theta^*(t_1,t_2) = 1 + \sigma^2\left[1+\frac{lnS(t_1,t_2)}{\rho}\right]^{-1} \tag{14}$$

It is easy to observe that, cross ratio function is greater than one and is a function of t_1, t_2. This implies there is always positive association between the survival times t_1 and t_2. Also, it is decreasing function of $t_1 > 0$, $t_2 > 0$ and decreases from $1 + \sigma^2$ to 1.

5. Compound Negative Binomial shared frailty model

A random variable of (1) is said to follow compound negative binomial distribution if N is negative binomial variate with parameters; the number of successes, r and the probability of success, p. The probability function of N is given by,

$$P(x) = \begin{cases} \begin{pmatrix} x+r-1 \\ x \end{pmatrix} p^r q^x & ; \ x = 0, 1, \cdots; 0 < p < 1; q = 1-p, r = 1, 2, \cdots \\ 0 & ; \text{ otherwise} \end{cases}$$

Discrete part of probability function of Z is, $P(Z = 0) = p^r$ and the continuous part is given by,

$$f(z) = \begin{cases} p^r \frac{1}{z} e^{-\nu z} \sum\limits_{N=1}^{\infty} \begin{pmatrix} N+r-1 \\ N \end{pmatrix} q^N \ \frac{(\nu z)^{N\gamma}}{\Gamma(N\gamma)} & ; \ z > 0, \nu > 0, \gamma > 0, 0 < p < 1; \\ & \quad q = 1-p, r = 1, 2, \cdots \\ 0 & ; \text{ otherwise.} \end{cases}$$

The parameter set for the compound negative binomial distribution is, $r = 1, 2, \cdots$; $0 < p < 1$; $\nu > 0$ and $\gamma > 0$. The Laplace transform, mean and variance of compound negative binomial variate are respectively given by,

$$L_Z(s) = \left\{ \frac{p}{1 - q\left[1 + \frac{s}{\nu}\right]^{-\gamma}} \right\}^r \tag{15}$$

$$E(Z) = \frac{rq\gamma}{p\nu}; Var(Z) = \frac{rq\gamma(p+\gamma)}{p^2\nu^2} \tag{16}$$

Under the identifiability condition, $EZ = 1$, the restriction on parameters is $\nu = \frac{rq\gamma}{p}$. Under this restriction, Laplace transform of compound negative binomial distribution reduces to,

$$L_Z(s) = \left\{ \frac{p}{1 - q\left[1 + d\frac{ps}{rq\gamma}\right]^{-\gamma}} \right\}^r \tag{17}$$

with variance $\sigma^2 = \frac{p+\gamma}{rq\gamma}$. Replacing Laplace transform in equation (5), we get the unconditional bivariate survival function for j^{th} individual at time $t_{1j} > 0$ and $t_{2j} > 0$ as,

$$S(t_{1j}, t_{2j}) = \left\{ \frac{p}{1 - q\left[1 + \frac{p(H_{01}(t_{1j}) + H_{02}(t_{2j}))\eta}{rq\gamma}\right]^{-\gamma}} \right\}^r \tag{18}$$

For negative binomial shared frailty model cross-ratio function is given by,

$$\theta^*(t_1,t_2) = 1 - \frac{1-(\gamma+1)\left[1-pS(t_1,t_2)^{-\frac{1}{r}}\right]^{-1}}{r\gamma}$$

We can easily observe that, cross-ratio function is always positive and decreasing function of t_1, t_2. It decreases between $1-\frac{1}{r\gamma}+\frac{\gamma+1}{rq\gamma}$ to $1-\frac{1}{r\gamma}$. This implies that there is always positive association between the survival times t_1 and t_2 and it decreases as time t_1, t_2 increases.

6. Proposed models

The unconditional bivariate survival functions for compound Poisson and compound negative binomial models at time $t_{1j} > 0$ and $t_{2j} > 0$ after substituting cumulative hazard function for IDB distribution in equations (13) and (18) are,

$$S(t_{1j},t_{2j}) = exp\left\{-\rho\left[1-\left(1+\frac{\phi(t_{1j},t_{2j})\eta_j}{\rho\gamma}\right)^{-\gamma}\right]\right\} \tag{19}$$

$$S(t_{1j},t_{2j}) = p^r\left[1-q\left(1+\frac{p\phi(t_{1j},t_{2j})\eta_j}{rq\gamma}\right)^{-\gamma}\right]^{-r} \tag{20}$$

where $\phi(t_{1j},t_{2j}) = \frac{\lambda_1 t_{1j}^2}{2}+\frac{\lambda_2 t_{2j}^2}{2}+\frac{\theta_1}{\alpha_1}\log(1+\alpha_1 t_{1j})+\frac{\theta_2}{\alpha_2}\log(1+\alpha_2 t_{2j})$. Here onwards we call equation (19) and (20) as model CP and CNB respectively.

7. Likelihood specification and bayesian estimation of parameters

Suppose there are n individuals under study, whose first and second observed failure times are represented by (t_{1j},t_{2j}). Let c_{1j} and c_{2j} be the observed censoring times for j^{th} individual $(j = 1,2,3,...,n)$ for first and second recurrence times respectively. Here we assume the independence between censoring scheme and life times of individuals.

The contribution of bivariate life time random variable of j^{th} individual in likelihood function is given by,

$$L_j(t_{1j},t_{2j}) = \begin{cases} f_1(t_{1j},t_{2j}), & ;\ t_{1j} < c_{1j}, t_{2j} < c_{2j}, \\ f_2(t_{1j},c_{2j}), & ;\ t_{1j} < c_{1j}, t_{2j} > c_{2j}, \\ f_3(c_{1j},t_{2j}), & ;\ t_{1j} > c_{1j}, t_{2j} < c_{2j}, \\ f_4(c_{1j},c_{2j}), & ;\ t_{1j} > c_{1j}, t_{2j} > c_{2j}. \end{cases}$$

and likelihood function is,

$$L(\underline{\theta},\underline{\beta},\underline{\tau}) = \prod_{j=1}^{n_1} f_1(t_{1j},t_{2j})\prod_{j=1}^{n_2} f_2(t_{1j},c_{2j})\prod_{j=1}^{n_3} f_3(c_{1j},t_{2j})\prod_{j=1}^{n_4} f_4(c_{1j},c_{2j}) \tag{21}$$

where $\underline{\tau}$, $\underline{\theta} = (\alpha_1, \lambda_1, \theta_1, \alpha_2, \lambda_2, \theta_2)$ and $\underline{\beta}$ are respectively vector of frailty parameters, vector of baseline parameters and vector of regression coefficients. In compound Poisson model $\underline{\tau} = (\rho, \gamma)$ and in compound negative binomial model $\underline{\tau} = (r, p, \gamma)$. Let n_1, n_2, n_3 and n_4 be the number of pairs for which first and second failure times (t_{1j}, t_{2j}) lie in the ranges $t_{1j} < c_{1j}, t_{2j} < c_{2j}$; $t_{1j} < c_{1j}, t_{2j} > c_{2j}$; $t_{1j} > c_{1j}, t_{2j} < c_{2j}$ and $t_{1j} > c_{1j}, t_{2j} > c_{2j}$ respectively and

$$\begin{aligned} f_1(t_{1j}, t_{2j}) &= \frac{\partial^2 S(t_{1j}, t_{2j})}{\partial t_{1j} \partial t_{2j}} \quad , \quad f_2(t_{1j}, c_{2j}) = -\frac{\partial^2 S(t_{1j}, c_{2j})}{\partial t_{1j}} \\ f_3(c_{1j}, t_{2j}) &= -\frac{\partial^2 S(c_{1j}, t_{2j})}{\partial t_{2j}} \quad , \quad f_4(c_{1j}, c_{2j}) = S(c_{1j}, c_{2j}) \end{aligned}$$

These functions for CP and CNB model respectively are given by,
CP model:

$$\begin{aligned} f_1(t_{1j}, t_{2j}) &= \left[\lambda_1 t_{1j} + \frac{\theta_1}{1+\alpha_1 t_{1j}}\right]\left[\lambda_2 t_{2j} + \frac{\theta_2}{1+\alpha_2 t_{2j}}\right]\left[1 + \frac{\phi(t_{1j}, t_{2j})\eta_j}{\rho\gamma}\right]^{-(\gamma+2)} \\ &\quad \left\{\frac{\gamma+1}{\rho\gamma} + \left[1 + \frac{\phi(t_{1j}, t_{2j})\eta_j}{\rho\gamma}\right]^{-\gamma}\right\} S(t_{1j}, t_{2j})\eta_j^2 \\ f_2(t_{1j}, c_{2j}) &= \left[\lambda_1 t_{1j} + \frac{\theta_1}{1+\alpha_1 t_{1j}}\right]\left[1 + \frac{\phi(t_{1j}, c_{2j})\eta_j}{\rho\gamma}\right]^{-(\gamma+1)} S(t_{1j}, t_{2j})\eta_j \\ f_3(c_{1j}, t_{2j}) &= \left[\lambda_2 t_{2j} + \frac{\theta_2}{1+\alpha_2 t_{2j}}\right]\left[1 + \frac{\phi(c_{1j}, t_{2j})\eta_j}{\rho\gamma}\right]^{-(\gamma+1)} S(t_{1j}, t_{2j})\eta_j \\ f_4(c_{1j}, c_{2j}) &= S(t_{1j}, t_{2j}) \end{aligned}$$

CNB model:

$$\begin{aligned} f_1(t_{1j}, t_{2j}) &= \frac{p^{r+2}\eta_j^2}{rq\gamma} \frac{\left[\lambda_1 t_{1j} + \frac{\theta_1}{1+\alpha_1 t_{1j}}\right]\left[\lambda_2 t_{2j} + \frac{\theta_2}{1+\alpha_2 t_{2j}}\right]\Phi_1(t_{1j}, t_{2j})}{\left[1 + \frac{p\phi(t_{1j}, t_{2j})\eta_j}{rq\gamma}\right]^{2(\gamma+1)}\left\{1 - q\left[1 + \frac{p\phi(t_{1j}, t_{2j})\eta_j}{rq\gamma}\right]^{-\gamma}\right\}^{r+2}} \\ f_2(t_{1j}, c_{2j}) &= p^{r+1}\eta_j \frac{\lambda_1 t_{1j} + \frac{\theta_1}{1+\alpha_1 t_{1j}}}{\left[1 + \frac{p\phi(t_{1j}, c_{2j})\eta_j}{rq\gamma}\right]^{(\gamma+1)}\left\{1 - q\left[1 + \frac{p\phi(t_{1j}, c_{2j})\eta_j}{rq\gamma}\right]^{-\gamma}\right\}^{r+1}} \\ f_3(c_{1j}, t_{2j}) &= p^{r+1}\eta_j \frac{\lambda_2 t_{2j} + \frac{\theta_2}{1+\alpha_2 t_{2j}}}{\left[1 + \frac{p\phi(c_{1j}, t_{2j})\eta_j}{rq\gamma}\right]^{(\gamma+1)}\left\{1 - q\left[1 + \frac{p\phi(c_{1j}, t_{2j})\eta_j}{rq\gamma}\right]^{-\gamma}\right\}^{r+1}} \\ f_4(c_{1j}, c_{2j}) &= p^r\left[1 - q\left(1 + \frac{p\phi(c_{1j}, c_{2j})\eta_j}{rq\gamma}\right)^{-\gamma}\right]^{-r} \end{aligned}$$

where $\Phi_1(t_{1j}, t_{2j}) = q\gamma(r+1) + (\gamma+1)\left[1 + \frac{p\phi(t_{1j}, t_{2j})\eta_j}{rq\gamma}\right]^{\gamma}\left\{1 - q\left[1 + \frac{p\phi(t_{1j}, t_{2j})\eta_j}{rq\gamma}\right]\right\}$

In our study, the likelihood function (21), due to censoring, is not in a simple form and so the first order derivatives. Hence, to estimate the parameters we have to use Newton-Raphson iterative procedure, but may be due to large number of parameters MLE's are not converging. So, we moved to computational Bayesian approach which does not suffer from these difficulties.

The joint posterior density function of parameters for given failure times is given by,

$$\pi(\alpha_1, \lambda_1, \theta_1, \alpha_2, \lambda_2, \theta_2, \underline{\tau}, \underline{\beta}) \propto L(\alpha_1, \lambda_1, \theta_1, \alpha_2, \lambda_2, \theta_2, \underline{\tau}, \underline{\beta}) * g_1(\alpha_1)g_2(\lambda_1)g_3(\theta_1)g_4(\alpha_2)$$
$$g_5(\lambda_2)g_6(\theta_2)\prod_{i=1}^{f} h_i(\tau_i)\prod_{i=1}^{k} p_i(\beta_i)$$

where $g_i(.)$ $(i = 1, 2, \cdots, 6)$, $h_i(.)$ $(i = 1, 2, \cdots, f)$ and $p_i(.)$ $(i = 1, 2, \cdots, k)$ are prior density functions with known hyper parameters of corresponding arguments for baseline, frailty parameters and regression coefficients. Likelihood function $L(.)$ is given by equation (21). Here we assume that all the parameters are independently distributed.

A widely used prior for frailty parameter is the gamma distribution with mean one and large variance, $G(\phi, \phi)$, say with a small choice of ϕ and the prior for regression coefficient is the normal with mean zero and large variance say ϵ^2. Similar types of prior distributions were used in Ibrahim et al. (2001), Sahu *etal*. (1997) and Santos *et al*. (2010). So, in our study also we have used same noninformative prior for frailty parameters and regression coefficients. We have considered two different noninformative prior distributions for baseline parameters, one is $G(a_1, a_2)$ and another is $U(b_1, b_2)$. All the hyper-parameters $\phi, \epsilon^2, a_1, a_2, b_1$ and b_2 are known. Here $G(a_1, a_2)$ is gamma distribution with shape parameter a_1 and scale parameter a_2 and $U(b_1, b_2)$ represents uniform distribution over the interval b_1 to b_2. We set hyper-parameters $\phi = 0.0001, \epsilon^2 = 1000, a_1 = 1, a_2 = 0.0001, b_1 = 0$ and $b_2 = 100$.

We have fitted the Bayesian model with the above prior density functions and likelihood function (21) using Metropolis-Hastings algorithm. We have monitored convergence of Markov chain to a stationary distribution by Gelman-Rubin convergence statistic and Geweke test. Trace plots, coupling from the past plots and sample autocorrelation function plots have been used, to check the behaviour of the chain, to decide burn-in period and sample autocorrelation lag respectively.

In order to compare the proposed models, we have used Akaike Information criteria (AIC), Bayesian Information Criterion (BIC), Deviance Information Criteria (DIC) and Conditional Predictive Ordinate (CPO) plot (see Gelfand (1996)). Also, we have used the Bayes factor B_{uv} for comparison of the models M_u against M_v. To compute Bayes factor, we have considered MCMC approach given in Kass and Raftery (1995).

8. Simulation study

To evaluate the performance of the Bayesian estimation procedure we have carried out a simulation study. For the simulation purpose we have considered only one covariate X_1. It is assumed to follow normal distribution. As the Bayesian methods are time consuming, we

have generated only fifty pairs of life times. According to the assumption, for given frailty Z, life times of individuals are independent. So, the conditional survival function for an individual for given frailty $Z = z$ and a covariate X_1 at time $t > 0$ is,

$$S(t \mid z, X_1) = e^{-zH_0(t)\eta}$$

Equating $S(t \mid z, X_1)$ to a random number say R $(0 < R < 1)$ over $t > 0$ we get,

$$\psi(t) = \frac{\lambda t^2}{2} + \frac{\theta}{\alpha}\log(1 + \alpha t) + \frac{\log(R)}{z\eta} \tag{22}$$

It is not possible to express explicitly as function of t, so to generate life times we have used bisection method. Exact step-wise procedure to generate sample is:

1. Generate a random sample of size 50 from frailty distribution as shared frailty for j^{th} $(j = 1, 2, \cdots, 50)$ individual. Firstly, generate a random observation $N = n$ from Poisson distribution for CP model and from negative binomial for CNB model. If $n = 0$ then assign frailty $Z = 0$ and if $n > 0$ then generate n gamma variables X_i and assign $Z = \sum_{i=1}^{n} X_i$.

2. Generate 50 covariate values for X_1 from normal distribution and compute $\eta_j = e^{X_{1j}\beta_1}$ for j^{th} individual.

3. Generate 50 pairs of life times (t_{1j}, t_{2j}) for given frailty z_j obtained in step 1 by solving equation (22) using bisection method.

4. Generate censoring times c_{1j} and c_{2j} from exponential distribution and observe survival time for i^{th} time $t^*_{ij} = min(t_{ij}, c_{ij})$ and censoring indicator δ_{ij} for j^{th} individual ($i = 1, 2$ and $j = 1, 2, \ldots, 100$), where

$$\delta_{ij} = \begin{cases} 1, & ; \ t_{ij} \leq c_{ij} \\ 0, & ; \ t_{ij} > c_{ij} \end{cases}$$

To estimate parameters of the model using simulated data, we have generated two parallel chains for both the models using two sets of prior distributions with the different starting points using Metropolis-Hastings algorithm based on normal transition kernels. We have iterated both the chains for 10000 times. There is no effect of prior distribution on posterior summaries because estimates of parameters are nearly same and convergence rate of chains for both the prior sets is also not greatly different. Also, for both the chains the results are somewhat similar, so we present here the analysis for only one chain with $G(a_1, a_2)$ as prior for baseline parameters, for both the models.

To check the effect of sample size of chain on the posterior summary, we have generated different samples and obtained posterior summary with small, moderate and large sample sizes. We have considered sample of size 7 as small, 16 as moderate and maximum possible sample size allowed by number of iterations and autocorrelation lag as large sample size. Gelman-Rubin convergence statistic values are nearly equal to one and Geweke test values are quite small and corresponding p-values are large enough to say the chain attains stationary

Table 1: Posterior summary for simulation study of CP model

Parameter	α_1	λ_1	θ_1	α_2	λ_2	θ_2	ρ	γ	β_1	$\lvert Bias\rvert$
True values	2.2	4.5	0.5	2.2	4.5	0.5	5	0.5	0.5	-
Sample size = 7;										
Estimates	2.0370	4.3470	0.5149	2.2999	4.4598	0.3534	4.5035	0.2865	0.6043	
Standard error	0.4637	0.3185	0.2442	0.4083	0.2336	0.2108	0.2601	0.1310	0.1080	
Bias	0.1630	0.1530	0.0149	0.0999	0.0402	0.1466	0.4965	0.2135	0.1043	0.6215
Sample size = 16;										
Estimates	2.2739	4.3950	0.4292	2.2676	4.5015	0.3621	4.5087	0.3225	0.6281	
Standard error	0.4414	0.3094	0.1980	0.4130	0.2098	0.2310	0.2846	0.1591	0.1283	
Bias	0.0739	0.1050	0.0708	0.0676	0.0015	0.1379	0.4913	0.1775	0.1281	0.5783
Sample size = 85;										
Estimates	2.2268	4.4535	0.4643	2.1980	4.5278	0.4838	4.7066	0.4804	0.5494	
Standard error	0.4192	0.2878	0.2458	0.3284	0.2035	0.2277	0.2659	0.2124	0.1494	
Bias	0.0268	0.0465	0.0357	0.0020	0.0278	0.0162	0.2934	0.0196	0.0494	0.3068

Table 2: Posterior summary for simulation study of CNB model

Parameter	α_1	λ_1	θ_1	α_2	λ_2	θ_2	p	γ	β_1	$\lvert Bias\rvert$
True values	2.2	4.5	0.5	2.2	4.5	0.5	0.5	0.5	0.5	-
Sample size = 7;										
Estimates	2.0312	4.6471	0.5990	2.1994	4.4922	0.4522	0.4821	0.8097	0.4954	
Standard error	0.4799	0.1739	0.2175	0.3917	0.2917	0.1978	0.0081	0.1558	0.0569	
Bias	0.1688	0.1471	0.0990	0.0006	0.0078	0.0478	0.0179	0.3097	0.0046	0.3982
Sample size = 16;										
Estimates	2.3001	4.5286	0.5976	2.2886	4.6000	0.4931	4.4826	0.7890	0.4963	
Standard error	0.4100	0.2626	0.1993	0.4110	0.3012	0.1229	0.0112	0.1358	0.0786	
Bias	0.1001	0.0286	0.0976	0.0886	0.1000	0.0069	0.0174	0.2890	0.0037	0.3494
Sample size = 85;										
Estimates	2.2169	4.4515	0.5234	2.1933	4.5115	0.4980	0.4827	0.7582	0.4923	
Standard error	0.4866	0.2707	0.1999	0.3694	0.2571	0.2114	0.0111	0.1457	0.0797	
Bias	0.0169	0.0484	0.0234	0.0067	0.0115	0.0020	0.0173	0.2582	0.0077	0.2653

distribution. Simulated values of parameters have autocorrelation of lag k, so every k^{th} iteration is selected as a sample from posterior distribution. The posterior mean and standard error with absolute bias for different sample sizes are reported in Table 1 and Table 2 for model CP and model CNB respectively. Last column of these Tables gives norm of bias which is calculated as $\sqrt{\sum_{i=1}^{n}(\text{true parameter}_i - \text{estimated value}_i)^2}$. From these Tables, it can be observed that the estimates become closer and closer to true values as sample size increases. Also, the standard error reduces as sample size increases.

9. Analysis of kidney infection data

We fit the proposed models to kidney infection data of McGrilchrist and Aisbett (1991). The data is related to recurrence times to infection at point of insertion of the catheter for 38 kidney patients using portable dialysis equipment. For each patient, first and second recurrence times (in days) of infection from the time of insertion of the catheter until it has to be removed owing to infection is recorded. The catheter may have to be removed for reasons other than kidney infection and this regard as censoring. So, survival time for a patient given

may be first or second infection time or censoring time. After the occurrence or censoring of the first infection sufficient (ten weeks interval) time was allowed for the infection to be cured before the second time the catheter was inserted. So, the first and second recurrence times are taken to be independent apart from the common frailty component. The data consists of three risk variables age, sex and disease type GN, AN and PKD where GN, AN and PKD are short forms of Glomerulo Nephritis, Acute Nephritis and Polycystic Kidney Disease. Let T_1 and T_2 be represents first and second recurrence time to infection. Five covariates age, sex and presence or absence of disease type GN, AN and PKD are represented by X_1, X_2, X_3, X_4, and X_5. To analyze kidney infection data, success is defined as getting infection first time, so we set $r = 1$.

First, we check goodness of fit of the data for both baseline distributions and then apply the Bayesian estimation procedure. To check goodness of fit for kidney data set, we have considered Kolmogorove-Smirnov test, we have applied the test to T_1 and T_2 separately. The p-values for CP and CNB models for T_1 are $0.9996, 0.4935$ and for T_2 are $0.5111, 0.3225$ respectively.

Table 3: Posterior summary for kidney infection data set for CP model

Parameter	Estimates	S.E.	L.C.L	U.C.L
n = 250, B = 1400, k = 390				
α_1	0.721066	0.125690	0.555642	0.962101
λ_1	0.000386	0.000361	0.000068	0.001358
θ_1	0.091341	0.047561	0.026447	0.213807
α_2	0.759945	0.116958	0.570522	0.982734
λ_2	0.000329	0.000300	0.000054	0.001263
θ_2	0.050038	0.028556	0.012666	0.128006
ρ	3.455383	0.805696	2.012911	4.910539
γ	2.440900	1.195852	1.032547	5.256461
β_1	0.007370	0.116010	-0.013778	0.029776
β_2	-1.885846	0.639941	-3.153641	-0.677762
β_3	0.168584	0.547786	-0.898598	1.244601
β_4	0.786868	0.544851	-0.298998	1.820400
β_5	-0.499750	0.980033	-2.549072	1.433960

Table 4: Posterior summary for kidney infection data set for CNB model

Parameter	Estimates	S.E.	L.C.L	U.C.L
n = 242, B = 2000, k = 390				
α_1	0.748220	0.130669	0.555915	0.982400
λ_1	0.000875	0.000742	0.000134	0.002706
θ_1	0.075700	0.053959	0.018796	0.217971
α_2	0.767938	0.124498	0.562309	0.979487
λ_2	0.000658	0.000502	0.000128	0.001859
θ_2	0.041461	0.024286	0.011405	0.099163
p	0.065639	0.021011	0.041000	0.118163
γ	0.496327	0.049441	0.406921	0.591675
β_1	0.009677	0.014753	-0.016473	0.040842
β_2	-2.368412	0.662620	-3.736769	-1.133544
β_3	0.221596	0.681551	-1.067240	1.394427
β_4	0.829265	0.645573	-0.526464	1.853952
β_5	-0.426339	1.044975	-2.423081	1.464310

As in case of simulation, here also we have got same conclusion. So, we present the analysis for only one chain with $G(a_1, a_2)$ as prior for baseline parameters, for both the models. In this case we iterated chains for 99000 times. The posterior summaries for CP and CNB models are presented in Table 3 and Table 4 respectively. In these Tables, second and third column represents estimate (posterior mean) and standard error whereas last two columns represent 95% lower and upper credible limits. The notations n, B and k respectively represent sample size, burn in period and auto-correlation lag.

Table 5: AIC, BIC and DIC values for kidney infection data set

Model	WOF	CP	CNB
AIC	712.3857	711.7692	709.3664
BIC	733.6743	732.7827	730.6550
DIC	708.4433	702.7835	698.9031

Table 5 provides AIC, BIC and DIC values for three models, CP, CNB and the model with ignoring frailty, which we call as without frailty (WOF) model. AIC and BIC values for CP and WOF models are nearly same, so cannot be used for comparing models, these values for CNB model are definitely smaller amongst other models. Further, if we rank DIC values from smallest to largest then CNB model will get first rank then CP and finally WOF model. This suggest that, CP and CNB models both are better than WOF model and CNB is better than CP.

Now consider comparison criteria $D_{uv} = 2\log(B_{uv})$ for comparing u^{th} numerator model against v^{th} denominator model, where B_{uv} is Bayes factor. Negative value of D_{uv} favours denominator model. These values are provided in Table 6.

Table 6: D_{uv} values for comparing CP and CNB models

		Numerator Model	
		WOF	CP
Denominator Model	CP	-0.7609	-
	CNB	-1.9194	-2.6804

From the Table 6 we can observe that, D_{uv} values for CP against WOF and CNB against WOF models are negative indicating CP and CNB models are better than WOF model. This is also confirmed with CPO plot presented in Figures 1 and 2. Large number of positive points in plot favour CP and CNB models. This implies if we ignore frailty then we may lose more informative model.

Thus, all the comparison criteria indicate that CNB model is better than CP model. We are now in a position to say that, both the proposed models, CP and CNB are more informative than ignoring frailty and CNB model is the best model then CP for modelling frailty in kidney infection data.

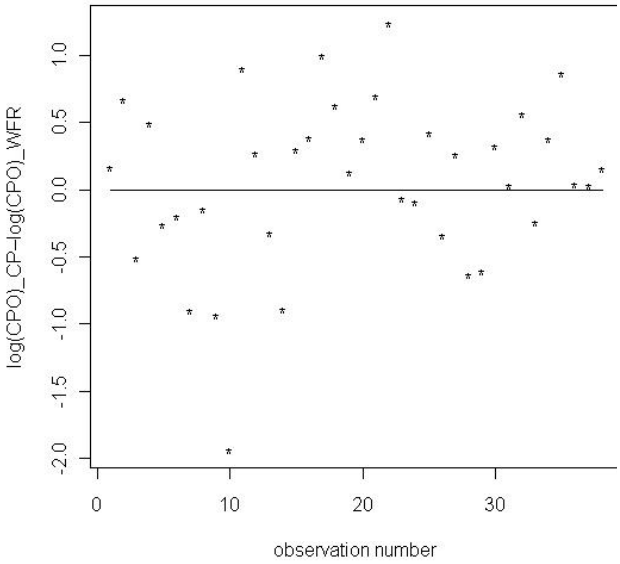

Figure 1: CPO plot for CP against WOF model

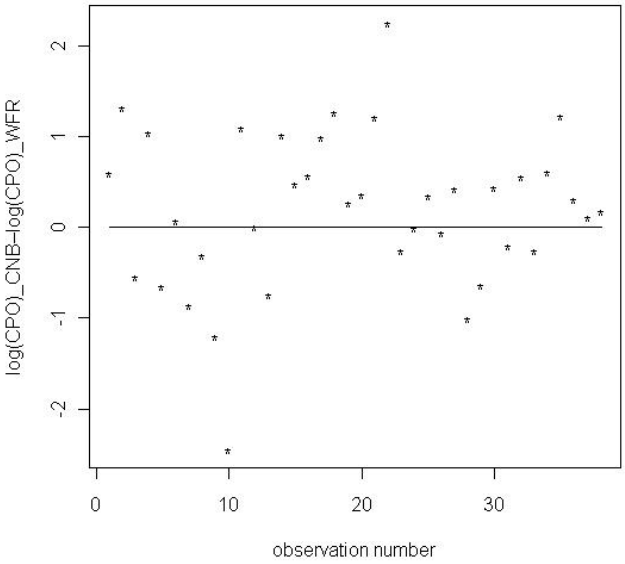

Figure 2: CPO plot for CNB against WOF model

10. Discussions

In the present paper, we have discussed compound Poisson and compound negative binomial shared frailty models. The main advantages of these models in comparison with other share frailty models is that they deal with the zero susceptibility. Further, the cross-ratio function is decreasing function of time unlike the other share frailty models, gamma and inverse Gaussian. Here we have considered IDB as baseline distribution. Even though it is an old distribution but it is more useful to model life times as it has increasing, decreasing, constant and bathtub shaped hazard function.

We have used Metropolis-Hastings algorithm to fit all the models. We analysed kidney infection data using our proposed models and the best model is suggested. We have used self-written programs in R statistical Environment to perform analysis.

The estimated frailty variances (0.4080) and (1.2118) for compound Poisson and compound negative binomial models respectively indicate that there is heterogeneity in the population of patients. Some patients are expected to be very prone to infection compared to others with the same covariate values. In continuation to this, all the model comparison criteria suggested that compound Poisson and compound negative binomial models are better than without frailty model. This indicates importance of frailty component in modelling of kidney infection data. Further comparing compound Poisson and compound negative binomial models, compound negative binomial shared frailty model is performing well for modelling of kidney infection data than compound Poisson model.

In compound negative binomial share frailty model, only one regression coefficient, β_2 is having larger ratio of its estimate to standard error and the value zero is not a credible value for the credible interval. This means, only covariate X_2 *i.e.*, Gender is significantly affecting

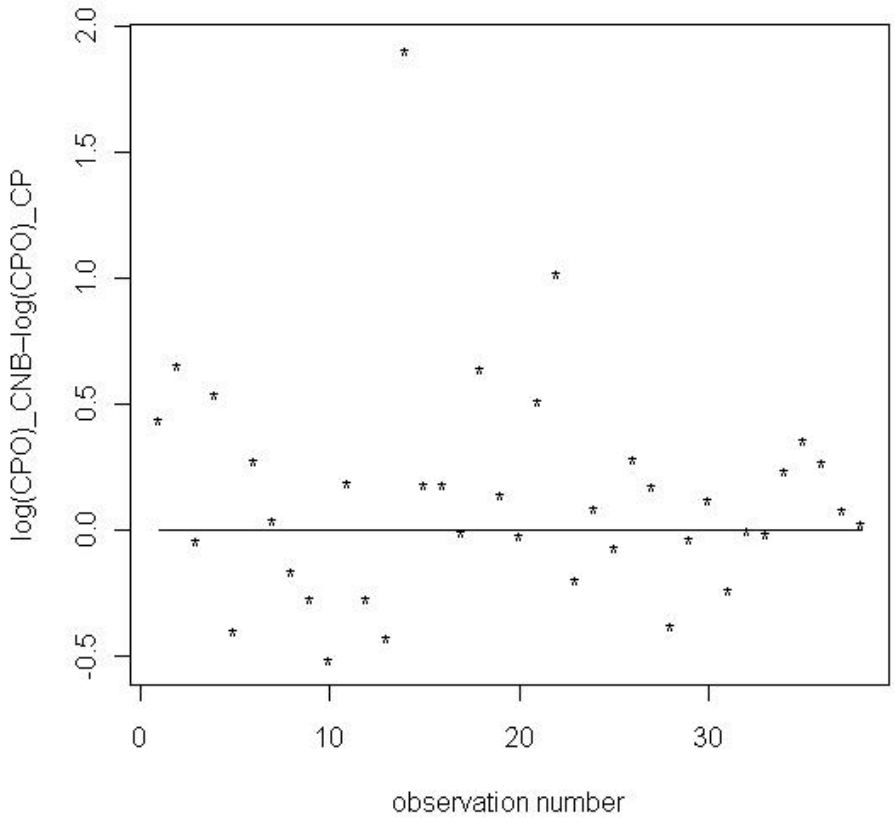

Figure 3: CPO plot for CNB against CP model

on infection rate. Negative value of β_2 indicate that the female patients have a lower risk for infection as compared to male patients. Same conclusion holds for compound Poisson share frailty models also. The estimated probability of non-susceptibility for compound negative binomial shared frailty model is 0.0656 indicating almost 6% of patients in the population are non-susceptible for kidney infection. In case of compound Poisson share frailty model, it is 3%.

In summary, this paper discussed modelling of survival times using compound frailty distributions when population consists of non-susceptible individuals.

Acknowledgements

We thank the referee for the valuable suggestions and comments which improved the earlier version the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, **2**, 951-72.

Aalen, O. O. and Tretli, S. (1999). Analyzing incidence of tests cancer by means of a frailty model. *Cancer Causes Control*, **10**, 285-92.

Clayton, D. G. (1978). A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*,**65**,141-151.

Gelfand, A. E. (1996). Model determination using sampling based methods. *Markov Chain Monte Carlo in Practice*, Eds. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., Chapman and Hall, London, 145-161.

Hanagal, D. D. (2005). A positive stable frailty regression model in bivariate survival data. *Journal of the Indian Society for Probability and Statistics*, **9**, 35-44.

Hanagal, D. D. (2006). A gamma frailty regression model in bivariate survival data. *IAPQR Transactions*, **31**, 73-83.

Hanagal, D. D. (2007). Gamma frailty regression models in mixture distributions. *Economic Quality Control*, **22**, 295-302.

Hanagal, D. D. (2010a). Modelling heterogeneity for bivariate survival data by compound Poisson distribution. *Model Assisted Statistics and Applications*, **5**, 01-09.

Hanagal, D. D. (2010b). Modeling heterogeneity for bivariate survival data by the compound Poisson distribution with random scale. *Statistics and Probability Letters*, **80**, 1781-90.

Hanagal, D. D. (2011). *Modeling Survival Data Using Frailty Models, First Edition*, CRC Press, New York.

Hanagal, D. D. and Dabade, A. D. (2012). Modeling heterogeneity in bivariate survival data by compound Poisson distribution using Bayesian approach. *International Journal of Statistics and Management systems*, **7**, 36-84.

Hanagal, D. D. and Dabade, A. D. (2013). Compound negative binomial shared frailty models for bivariate survival data. *Statistics and Probability Letters*, **83**, 2507 -2515.

Hanagal, D. D. and Kamble, A. T. (2015). Bayesian estimation in shared compound Poisson frailty models. *Journal of Reliability and Statistical Studies*, **8**, 159-180.

Hanagal, D. D. and Kamble, A. T. (2016). Bayesian Estimation in Shared Compound Negative Binomial Frailty Models. *Research & Reviews: Journal of Statistics and Mathematical Sciences*, **2**, 53-67.

Hanagal, D. D. and Sharma, R. (2013). Modeling heterogeneity for bivariate survival data by shared gamma frailty regression model. *Model Assisted Statistics and Applications*, **8**, 85-102.

Hanagal, D. D. and Sharma, R. (2015a). Bayesian inference in Marshall-Olkin bivariate exponential shared gamma frailty regression model under random censoring. *Communications in Statistics, Theory & Methods*, **44**, 24-47.

Hanagal, D. D. and Sharma, R. (2015b). Comparison of frailty models for acute leukaemia data under Gompertz baseline distribution. *Communications in Statistics, Theory & Methods*, **44**, 1338-1350.

Hanagal, D. D. and Sharma, R. (2015c). Analysis of bivariate survival data using shared inverse Gaussian frailty model. *Communications in Statistics, Theory & Methods*, **44**, 1351-1380.

Hanagal, D. D. (2019). *Modeling Survival Data Using Frailty Models, Second Edition*, Springer Nature, Singapore.

Hanagal, D. D. (2023a). Compound Poisson shared frailty models based on additive hazards. *Communications in Statistics, Theory and Methods*, **52**, 6287-6309.

Hanagal, D. D. (2023b). Correlated compound geometric frailty models based on reversed hazard rate. *Model Assisted Statistics and Applications*, **18**, 149-164.

Hanagal, D. D. (2024a). Correlated compound Poisson frailty models based on reversed hazard rate. *Communications in Statistics, Theory and Methods*, (to appear). https://doi.org/10.1080/03610926.2022.2098336.

Hanagal, D. D. (2024b). Analysis of kidney infection data using correlated compound Poisson frailty models. *Model Assisted Statistics and Applications*, **19**, (to appear).

Hjorth, U. (1980). A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates. *Technometrics*, **22**, 99-107.

Ibrahim, J. G, Ming-Hui Chen, and Sinha, D.(2001). *Bayesian Survival Analysis*. Springer Verlag.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factor.*Journal of the American Statistical Association*, **90**, 773-795.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461-466.

Moger, T. A. and Aalen, O. O. (2005). A distribution of multivariate frailty based on the compound Poisson distribution with random scale. *Lifetime Data analysis*, **11**, 41-59.

Sahu, S. K., Dey, D. K., Aslanidou, H., and Sinha, D.(1997). A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis*, **3**, 123-137.

Santos, C. A. and Achcar, J. A.(2010). A Bayesian analysis for multivariate survival data in the presence of covariates. *Journal of Statistical Theory and Applications*, **9**, 233-253.

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