

Fitting The Hazard Ratio for Cluster Survival Data with Frailty Effect Via WinBUGS

Ayman A. Mostafa

Department of Mathematics and Statistics, Al-Imam Muhammad Ibn Saud Islamic University, P.O. BOX 90950, 11623 Riyadh (SAUDI ARABIA)

ABSTRACT

Survival analysis examines and models the times it takes for events recorded on experimental units to occur. Interest in the use of random effects (frailty) in the survival analysis setting has been increasing. Shared frailty models are of interest when one has clustered survival data and when focus is on comparing the lifetimes within clusters and further on estimating the correlation between lifetimes from the same cluster. In this paper we present an approach involving objective Bayesian reference analysis to the Frailty model with cluster survival time and sources of heterogeneity that are not captured by covariates. Early work in survival analysis ignored such random effects in survival data analysis. The objective of this study is, by means of simulation, to quantify the effect of ignoring individual heterogeneity in fitting the hazard ratio for cluster survival data. We adopt the idea behind Bayesian technique to estimate the parameters of interest using Gibbs sampling. The Gibbs sampling is proposed to simulate the Markov chain of parameters' posterior distribution dynamically, which avoids the calculation of complex integrals of the posterior using WinBUGS package. Thought the paper, the results are illustrated with survival data analyzed in the literature.

MSC 2010 Mathematics Subject Classification: 62N99, 62F15, 62M20, 65C05, 65C40, 68N15.

KEY WORDS: Survival Analysis, Regression Models, survival frailty model, Bayesian Inference, Frailty Models, BUGS.

1. INTRODUCTION

Survival data is a term used for describing data that measure the time to some event. Statistical models and methods for such data and other time-to-event data are extensively used in many fields, including the biomedical sciences, engineering, the environmental sciences, economics, actuarial sciences, management, and social sciences. Data in survival analysis are usually assumed to be univariate, with one, possibly censored, lifetime for each individual. All the standard methodology, including [27] plots and Cox analysis introduced by Cox (1972), is geared toward handling this situation. Survival data analysis, also known as failure time data, event history data, duration data or time-to-event data analysis, is one of the most significant advancements of mathematical statistics in the last few decades. Furthermore, survival (or mortality) rates are observing data measurements in time from a well-defined origin until a particular event occurs. The event is usually death (e.g., lifetime from birth to death), but it could also be a change of state (e.g., occurrence of a disease or time to failure of an electrical component). Survival analysis is an important statistical field that is required for data analysis in different disciplines. There are a large number of textbooks on the basic concepts of survival analysis, for example, (Cox and Oakes, 1984), (Andersen et al. [3]), (Collet, 2003), (Lee and Wang, 2003), (Ibrahim et al. [26]). Our objective is to model the survival time, i.e. the time to the occurrence of a given event. The event could be just about anything. Within the medical research, common examples are the time to development of a disease, response to a treatment, and of course death.

However, multivariate survival data also arise naturally in many contexts. Such data pose a problem for ordinary multivariate methods, which will have difficulty handling censored data. There are two typical ways multivariate survival data can arise. One, which may be termed the "recurrent events situation", is when several successive events of the same type are registered for each individual, for instance, repeated occurrences of ear infections. The other type of multivariate survival data which we will focus on this paper is termed the "clustered survival data situation", is when several units that may fail are collected in a cluster (see, e.g. Spiegelhalter, 2001). Described below is a biomedical study example, which involves clustered survival data situation and will be studied further in the remaining of this article.

Practical example (Diabetic Retinopathy Study; DRS)

Survival analysis has found widespread applications in medicine. It is well known that any clinical trial is an experiment carried out to gain knowledge about the relative benefits of two or more treatments. In

*Corresponding Author : Ayman A. Mostafa, Department of Mathematics and Statistics, Al-Imam Muhammad Ibn Saud Islamic University, Riyadh (SAUDI ARABIA). Email: aasaad@imamu.edu.sa

1971, National Eye Institute funded a study to test a laser treatment (laser photocoagulation) for delaying blindness in patients with DR. Photocoagulation has been used to treat DR over the last several decades (see, for example, www4.dcu.ie/cochrane/pdfs/b.dineen_poster.pdf). The study was designed as follows. Because juvenile and adult diabetes have very different courses, it is desirable to examine how the age at onset of diabetes may affect the time of blindness. One eye of each patient was randomly chosen to receive photocoagulation techniques based on two different laser types of treatment (xenon and argon), and the other eye was observed with no treatment, which is serving as untreated control. There were 1,742 patients, who participated in that study, were followed up between 1971 and 1975. For each eye, the event of interest was the time from initiation of treatment to the time when visual acuity dropped below 5/200 two visits in a row (call it ‘blindness’). One anticipates some dependence between a patient’s two eyes. Thus there is a built-in lag time of approximately 6 months (visits were every 3 months). Survival times in this dataset are therefore the actual time to blindness in months, minus the minimum possible time to event (6.5 months). Censoring was caused by death, dropout, or end of the study. The only data available are from the 197 patients defined as high-risk by DRS criteria. As mentioned above, a major goal of interest is whether the treatment by photocoagulation significantly delayed the occurrence of onset severe visual loss (blindness). One of the major goals of the study was to investigate if the time of occurrence of severe visual loss for the treated eye is longer than that for the control (untreated eye). In such example a natural clustering of study subjects such that failure times of the two eyes within the same cluster may be correlated. From this point, we assume that $T = t_{\ell m}$ denote the survival time of the m^{th} eye ($m = 1, 2$) for the ℓ^{th} patient ($\ell = 1, 2, \dots, 197$) and the main covariate is as follows:

$$Z_1 = z_{1\ell m} = \begin{cases} 1 & \text{if the } m^{\text{th}} \text{ eye of the } \ell^{\text{th}} \text{ patient was on treatment;} \\ 0 & \text{otherwise.} \end{cases} \quad \ell = 1, 2, \dots, 197; \quad m = 1, 2.$$

In survival analysis, the multiplicative hazard models (hazard ratios) provide the main principle framework for studying the association between covariates and the survival time. The hazard function, also called, the risk or intensity function, of a survival time $t_{\ell m}$ associated with a 1-vector of covariates Z_1 is defined as $h(t_{\ell m} | Z_1) = f(t_{\ell m} | Z_1) / [1 - F(t_{\ell m} | Z_1)]$, where $f(\cdot | Z_1)$ and $F(\cdot | Z_1)$ are the density function and the distribution function, respectively, of the random variable $t_{\ell m}$ conditioned on the vector of covariates Z_1 . The function $S(t_{\ell m} | Z_1) = 1 - F(t_{\ell m} | Z_1)$ is called the survival function.

The Cox Proportional Hazard model proposed by Cox (1972) gives an expression for the hazard at time $t_{\ell m}$, as the product of a baseline hazard function, $\lambda_0(t)$ (intuitively, what we have without explaining variables) and the exponential of a term linear in the predictor Z_1 . The expression is divided in two terms: one depending on time only, and the other depend only on the Z_1 . Z_1 is said to be time-independent vector of covariate (treatment indicators or prognostic factor). Here is the formula for predictor Z_1 . The β is the coefficient of the parameterized part of the model, through the flowing function.

$$\lambda(t_{\ell m} | Z_1) = \lim_{\Delta t \downarrow 0^+} \frac{1}{\Delta t} \Pr(t_{\ell m} \leq T < t_{\ell m} + \Delta t | T \geq t_{\ell m}, Z_1) = \lambda_0(t) \times \exp\{\beta z_1\} \quad (1)$$

where $\lambda_0(t)$ the unknown and unspecified nonnegative baseline hazard, β is a 1×1 vector of unknown regression parameter. Note that there is intercept β_0 in model, obviously, $\lambda(t_{\ell m} | Z_1 = 0) = \lambda_0(t_{\ell m})$, so that $\lambda_0(t_{\ell m})$ is often called the baseline hazard function. It can be interpreted as the hazard function for the population of all patients with $Z_1 = 0$. The baseline hazard function in $\lambda_0(t_{\ell m})$ in can take any shape as a function of $t_{\ell m}$. The only requirement is that $\lambda_0(t_{\ell m}) > 0$. This is the nonparametric part of the model and βz_1 is the parametric part of the model. So Cox's PHM is a semi-parametric model.

We now consider generalizations of model to allow for covariates that do not properly explain (or unobserved individuals effects). There are usually referred to as ‘frailty’ in the biomedical sciences. In many failure time studies there is a natural clustering of study subjects. The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data. In its simplest form a frailty is an unobserved proportionality factor that modifies the hazard function. Gamma frailty models are a common tool in survival analysis with random effects, dating back at least to Beard (1959), Vaupel et al. [42] and Lancaster (1979) and are considered in more detail in the

biostatistician literature (for example, Hougaard, 2000; Duchateau and Janssen, 2008) as well as in the econometric literature (for example, Lancaster, 1979; Heckman and Singer, 1984; Sinha, 1993 and van den Berg, 2001).

In some situations the extra random frailty component of proportional hazard model is required only to get a correct inference on fixed effects of covariates, whereas in other cases the distribution of the random subject effect could be one of the major interests.

In many applications, the study population cannot be assumed to be homogenous but must be considered as a heterogeneous sample, i.e., a mixture of individuals with different hazards. For example, in many applications, it is impossible to measure all relevant covariates related to the event of interest. This article focuses on the shared frailty model in the multiplicative hazard model. The shared frailty model is relevant to event times of related individuals, similar members and repeated measurements (parallel data). Individuals in a group (cluster) are assumed to share the same frailty, which is why this model is called “shared frailty model”. The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data. It is difficult for the researcher to control for all relevant patient and hospital characteristics. However, unlike the linear regression case, duration models that do not account for unobservable heterogeneity will produce biased estimates of duration dependence. A common way of approaching this problem is by letting the hazard function of a patient be modified by a proportionality factor (frailty). If we let $v_\ell > 0$ represent non-measured heterogeneity then the hazard function becomes:

$$\lambda(t_{\ell m}) = v_\ell \times \lambda_0(t_{\ell m}) \times \exp\{\beta_1 Z_1\}, t \geq 0 \quad (2)$$

Thus, for $v_\ell > 1$, patients will have a higher risk of failure while the opposite is true for values of $v_\ell < 1$.

The proportionality factor may be, for example, patient specific (unshared frailty) or hospital specific (shared frailty). Commonly, v_ℓ is assumed to be a realization of a random variable V , which has known distribution and is uncorrelated with the other covariates. Some survival studies exhibit a cluster structure. For instance in twin studies the cluster structure is induced by the twins in that it may be expected a priori that the lifetimes of the twins are correlated. It was introduced by Clayton (1978; 1991) and extensively studied in Hougaard (2000). The survival times are assumed to be conditional independent with respect to the shared (common) frailty. For recurrent events one must make assumptions about the underlying stochastic process. Two main situations are considered: first, the one where, given the frailty, the process is of renewal type. By this it means that the times between consecutive events for a particular individual are independent random variables with identical distributions. This case is very similar to the clustered data situation, since the inter-event times are identically distributed conditional on having the same frailty. Second, we consider the case where the underlying process given the frailty is Poisson. Unconditionally, this is a doubly stochastic Poisson process (or a Cox process). There are four basic approaches to analyzing multivariate survival data:

- 1) Marginal models;
- 2) Frailty models;
- 3) Random effects models for transformed times;
- 4) Dynamic models.

The marginal models are very pragmatic and do not attempt to yield any insight into the deeper structure of the processes. In marginal models the focus is on just parts of the available data, but they may still be quite useful for evaluating, say, treatment effects. On the other hand, the frailty, random effects and dynamic models are supposed to give more realistic descriptions of the full data set. The interested reader is directed to Aalen, et al. [2] for a detail discussion for the difference between the above four basic approaches.

In this paper, we consider the frailty and random effects models for Diabetic Retinopathy Study described above, using of freely available software for the analysis of complex statistical models using MCMC techniques, called WinBUGS, in the context of survival data analysis. WinBUGS (**B**ayesian inference **U**sing **G**ibbs Sampling, Spiegelhalter, Thomas, Best & Lund, 2003) is an application that can be used for the Bayesian analysis of complex models using Markov Chain Monte Carlo (MCMC) methods. WinBUGS is available free at <http://www.mrc-bsu.cam.ac.uk/bugs/> and an open-source version of the core BUGS code (OpenBUGS) is also available at <http://mathstat.helsinki.fi/openbugs/>. It turns out that OpenBUGS can become quite a powerful and flexible tool for Bayesian survival analysis, and only requires a relatively small investment on the part of the user. Once the (applied) user understands the logic of model building with WinBUGS, Bayesian analysis is conducted quite easily and many built-in features can be accessed to produce an in-depth and interactive analysis of random effect in survival data analysis. WinBUGS is a generic tool which can be used in a wide variety of situations. Thus, it cannot be expected to be quite as efficient as tailor-made MCMC implementations designed specifically for certain models. As

Griffin and Steel (2007) mention that “Although WinBUGS may not be that efficient for more complicated models, it does make Bayesian inference with stochastic frontier models easily accessible for applied researchers and its generic structure allows for a lot of flexibility in model specification”, We illustrate the use and flexibility of WinBUGS for Bayesian analysis for fitting the hazard ratio with frailty.

2. Unobserved heterogeneity model

As indicated in the introduction to this paper, from a modeling point of view we may consider the clustered survival data situation. Basically, frailty models in this context are random effects models, analogous to those well-known from linear normal model theory. However, the frailty models are better adapted to handle censored data than the normal models. The dependence may be modeled through a frailty variable, such that all survival times that are related to each other (as either multiple units in a cluster or inter-event times in a renewal process) have the same level of frailty attached to them. In eye data as exemplified above in Subsection 1.1, the dependent variable is the time to onset of blindness in patients with diabetic retinopathy. Patients were followed over several years and the pair of waiting times to blindness in the left and right eyes, respectively, was observed. In such a study one should expect some correlation between the waiting times within the patients. Here the clustering is due to the patients. The primary interest in this study was to evaluate whether laser treatment could delay onset to blindness. For this purpose one eye of each patient was randomly chosen for laser treatment while the other acted as a control. An individual level multiplicative effect $v_\ell = \exp(b_\ell)$ called the frailty is associated with the ℓ^{th} subject.

Given the unobserved v_ℓ , called the frailty, is associated with the ℓ^{th} patient, we assume the usual Cox model in (2).

2.1. Cox model with frailty through WinBUGS code program for DRS data

The distributional assumptions for the likelihood and prior elicitation are described in the following and through Display 1.

Andersen and Gill (1982) extended (2) to the counting process framework and gave elegant martingale proofs for the asymptotic properties of the associated estimators in the models try to fit models for survival data. Others that have contributed to establishing asymptotic results for the model are Tsodis (1981) and Næs (1982). We have $n = 394$ observations (eyes) under investigations (197 patients), for observation i , $i = 1, 2, \dots, n$, $I_i(t)$ is the intensity process for a counting process given covariate vector Z_1 , and $Y_i(t)$ is the at risk indicator, i.e., the set of observations still at risk at the time, t_i , of failure for observation i (i.e., unblended eye and uncensored at time point just before time t), furthermore, we observe process $N_i(t)$, to count the number of failures which occurred in the interval $[0, t]$. That process is constant and equal to zero between failures and jumps one unit at each failure time. Since the stochastic process $\{N_i(t), t \geq 0\}$ is counting process, hence, the rate of a new failure is then seen to be $I_i(t) = Y_i(t)\lambda(t | Z_1)$. The intensity $I_i(t)$ may be characterized as the probability that the event of interest occurs in the small time interval $[t, t + dt]$, given that it has not happened before. This gives approximately

$$dN(t) \approx \lambda(t | Z)dt = \lambda_0(t)\exp[\beta_1 Z_1 + b] = I(t)$$

where $dN(t)$ is the increment of $N(t)$ over the small interval $[t, t + dt]$ (i.e., number of observed failures occurring in $[t, t + dt]$). Hence, $I_i(t)$, is multiplicative intensity which can be modeled by

$$I(t) = Y(t)\lambda(t | Z) = Y(t)\lambda_0(t)\exp[\beta_1 Z_1 + b], \quad (3)$$

where the intensity process is a product of an observed process and an unobserved function. Hence the intensity process for $N_i(t)$ under model (2) is:

$$I(t)dt = Y(t)\exp[\beta_1 Z_1 + b]d\Lambda_0(t)$$

where $d\Lambda_0(t)$ represents the instantaneous probability that the subject at risk at time t has an event in the next time interval $[t, t + dt]$. As mentioned in Clayton (1994), it turned out that the martingale theory was of fundamental importance.

Hence, the natural intuitive concept of intensity (speed) process (3) is equivalent to asserting that the counting process minus the integrated intensity process is a martingale.

It is useful to consider the product-integral transformation (Gill, 2005):

$$S(t) = \prod_{(0,t]} \{1 - d\Lambda(s)\}$$

Since $d\Lambda_i(t) = \int_0^t I_i(s)ds$ is the cumulative intensity corresponding to the hazard function $I_i(t)$, we have

$$\prod_{(0,t]} \{1 - d\Lambda_i(t)\} = \exp\left(-\int_0^t I_i(s)ds\right)$$

Suppose that eyes were followed to be blinded or censored in a study. Thus we have observed data $D = \{N_i(t), Y_i(t), Z_i; i = 1, 2, \dots, n\}$ and we have unknown parameters $\beta, b, \Lambda_0(t)$. Under non-informative censoring, the likelihood of the data factorizes, with one term depending only on the censoring process and the second term:

$$L_i(D | \beta, \Lambda_i(t), b) = \exp\left(-\int_{t \geq 0} I_i(t)dt\right) \prod_{t \geq 0} [I_i(t)]^{dN_i(t)}; i = 1, 2, \dots, n. \quad (4)$$

Hence, according to the probability model under consideration, the variables D have their joint distribution given by

$$L(D | \beta, \Lambda(t), b) = \prod_{i=1}^n L_i(D | \beta, \Lambda_i(t), b) \quad (5)$$

Following the counting process notation introduced above, for subject i ($i = 1, 2, \dots, n$) represent the process counting the failures occurring up to time t , while $dN(t)$ is a small increment of $N(t)$ over the interval $[t, t+dt]$. $N(t)$ and $dN(t)$ equal 1 if the event occurs in $[0, t)$ and $[t, t+dt)$, respectively, and 0 otherwise. Under non-informative censoring, the observed counting process likelihood is proportional to (4) and is following a Poisson form. Though $dN(t)$ is at most one for all patients under investigations; the infinitesimal counting process increments, $dN(t)$, contribute to the likelihood just as with independent Poisson random variables with means $I(t)dt$ over the interval $[t, t+dt]$. Defining the model in this framework allows the intensity to be regarded as constant in that interval (Clayton, 1991). For efficient posterior computations, we implement a data augmentation approach based on the consideration that $dN_i(t)$ are independent Poisson random variables under the likelihood expression (5), i.e.

$$dN(t) \sim \text{Poisson}(I(t)dt) \quad (6)$$

After D is observed, our interest turns to the posterior distribution $P(\beta, \Lambda_0(t), b | D)$. Baye's Theorem tells us that $P(D | \beta, \Lambda_0(t), b) \propto P(\beta, \Lambda_0(t), b | D)$. Thus, whole probability model can now be expressed as the joint posterior distribution of the model parameters according to the Bayesian approach, i.e.

$$P(\beta, \Lambda_0(t), b | D) \propto L(D | \beta, \Lambda_0(t)) P(\Lambda_0(t)) P(b | \eta) P(\eta, \beta) \quad (7)$$

The focus is on inference for β . Another quantity of interest is the baseline hazard function, $\Lambda_0(t)$ which is best viewed as a process over time. As (7) has a complicated form, it is most conveniently summarized using simulation. Thus might be accomplishing using Gibbs sampler. Implementing a Gibbs sampler coded from scratch would require us to identify and then construct an effective simulation method for each of the two related full conditional posterior distributions, namely $P(\beta | D, \Lambda_0(t), b)$, $P(\Lambda_0(t) | D, \beta, b)$, $P(b | D, \Lambda_0(t), \beta)$. These are exactly the same tasks the WinBUGS software obviates, as it performs these steps internally and automatically. Various Bayesian formulations of the model (7) differ mainly in the nonparametric specification of $\Lambda_0(t)$.

A mathematically convenient choice for the distribution of the b and the one that is considered throughout this paper is the one-parameter gamma distribution is the one-parameter gamma distribution given by

$$f_b(b) = \frac{b^{\frac{1}{\theta}-1} \exp(-b/\theta)}{\theta^{-\frac{1}{\theta}} \Gamma(1/\theta)} \quad (8)$$

It is noted that from (8) $E(W) = 1$ and $\text{Var}(W) = \theta$. This gives the following interpretation: individuals in a group i with $b > 1$ ($b < 1$) are frail (strong) (higher risk, respectively lower risk). The parameter θ provides information on the variability (the heterogeneity) in the population of clusters. Bayesian inference can also be carried out using WinBUGS. A landmark in the development of Bayesian analysis is the 2009 book by Ntzoufras (simply titled ‘Bayesian Modeling Using WinBUGS’). As far as we know, it is the only book-length treatment on the topic to date. The script below implements the analysis, as we will discuss in the next Section.

Display 1 WinBUGS model specification code for the DRS Data

WinBUGS script to test the significance of frailty in Cox model in cluster data for DRS data

```

model {#Start model #1#
  #Set up data
  for( i in 1 : N ) {#Begin loop over all patients #2#
    # N = 394 is the total number of patients in the two clusters
    for( j in 1 : T ) {#Begin loop over observations within subjects #3#
      #T = 47 is the number of unique failure times for the two eyes
      #assume that obs.t[i] is the follow-up time for patient i (i = 1,2,...,394)
      #fail[N] is indicator such that failure = 1; censored = 0
      #Y[i,j]=Yi(tj) is at risk set, i.e., obs.t >= t and zero otherwise.
      #t[T+1] is the unique failure times + maximum censoring time
      Y[i , j]<- step(obs.t[i] - t[j] + eps) #4#
      # eps=0.000001;used to guard against numerical imprecision in step function
      # counting process jump = 1 to indicates event-time, obs.t in interval [t[j], t[j+1]);
      #i.e. if t[j] <= obs.t < t[j+1]
      # N[i,j]=Ni(tj)
      #fail[i] is an indicator to be failure = 1; censored = 0
      dN[i , j]<- Y[i , j] * step(t[j + 1] - obs.t[i] - eps) * fail[i] #5#
    }
  }
  for( j in 1 : T ) {#6#
    for( i in 1 : N ) {#7#
      #Idt[N,T] is the intensity process
      #beta is the regression coefficient
      # dL0[T] is increment in unknown hazard function
      dN[i , j] ~ dpois(Idt[i , j]) # Poisson likelihood #10#
      Idt[i , j]<- Y[i , j] * exp(beta * Z1[i] + b[pair[i]]) * dL0[j] # Poisson mean #11#
      # Idt[i , j]<- Y[i , j] * exp(beta * Z1[i] + epsilon[i]) * dL0[j] # if the fixed effects model is
      #used
    }
  }
  dL0[j] ~ dgamma(mu[j], c) # gamma prior distribution for Baseline Hazard Function #12#
  # c is a degree of confidence in prior guess for dL0
  #dL0.star[T] is prior guess at hazard function
  mu[j]<- dL0.star[j] * c # prior mean hazard #13#
  #mu[T] is the location parameter for Gamma (= c * dL0.star)
  S.treat[j]<- pow(exp( -sum(dL0[1:j])), exp(beta * Z1[1:197] + b[1:197])) #14#
  S.untreat[j]<- pow(exp( -sum(dL0[1:j])), exp(beta * Z1[198:394] + b[1:197])) #15#
  # S.treat[T] is survivor function for treated eye group
  # S.untreat[j] is survivor function for untreated eye group #16#
}
c ~ dgamma(0.001, 0.001) # Diffuse prior for gamma “noninformative priors” #17#
for( k in 1 : Npairs ) {# Npairs = 197 pairs of eye. #18#
  b[k] ~ dgamma(theta, theta); # Monitor this nodes #19#
  #b[k] ~ dnorm(0.0, tau.b) #Replace frailty prior with normal distribution to monitor the sensitivity for the prior #20#
}

```

```

}
theta<-1/tau.b
tau.b ~ dgamma(0.10, 0.10)
sigma <- sqrt(1 / tau.b) # Monitor this node
    #r is the prior guess at failure rate
for( j in 1 : T ){                                #21#
dL0.star[j] <- r * (t[j + 1] - t[j])          #22#
}
r ~ dgamma(0.001, 0.001) # Diffuse prior for failure rate #23#
beta ~ dnorm(0.0, 1.0E-6) # Diffuse prior for the coefficient of regression in Cox model #24#
}
#25#
#26#
#27#
#28#
#29#
#30#

```

3. An analysis of Bayesian modeling process: Diabetic retinopathy study data

In this section we illustrate the hazard ratio in Cox proportional hazard model with frailty in Diabetic Retinopathy Study (DRS) data. As mentioned in Section 1, the DRS was conducted by National Eye Institute (<http://www.nei.nih.gov>) to assess the effectiveness of laser photocoagulation (photocoagulation has been used to treat DR over the last several decades) in delaying the onset of blindness in patients with diabetic retinopathy (DR). Although this study is described in Diabetic Retinopathy Study Research Group (1976), we will give a description of this study, to reanalyze the data from Bayesian technique for fitting hazard ratio with and without frailty and compare the results. This dataset came from Therneau and Grambsch (2000; in Sect. 8.4.2), which originally came from Huster et al. [25]. We use a subset of the data set of 197 high-risk diabetes patients to assess the effectiveness of the laser treatment in delaying severe vision loss, which is provided in Professor Therneau's web page (see; <http://mayoresearch.mayo.edu/mayo/research/biostat/therneau-book.cfm>).

3.1. Outputs analysis from WinBUGS software without frailty

Here we use the model in Section 2. Running the script in Display 1 without frailty term in Cox model, i.e., eliminate (+ b[pair[i]]) in line #11# in Display 1, and eliminate its prior distribution in line #19# and line #20# and consequently eliminate lines #22#, #23# and #24#. Because this is a more complicated model with much data (see the format of the data in the Appendix), we will run two chains, starting at different points, as an aid in deciding when convergence to the posterior distribution has occurred. We do this by selecting two chains before pressing the compile button during the specification step. Also, we must give starting values for each chain, as WinBUGS cannot generate initial values of r in line #28# from the vague prior distribution we have used. Following the initial values load for the two chains, the *generate inits* button, in WinBUGS software, can be clicked to generate any other initial values required. The chains were run with a burn-in of 1,000 iterations with 20,000 retained draws and a thinning to every 10th draw (to reduce the autocorrelation of the chain). Convergence was achieved for all parameters based on the rules by Gelman and Rubin (1992). The parameter estimates are not sensitive to the choice of hyperparameters and initial values. Graphical representations of the posterior distribution can indicate problems with the performance of the MCMC algorithm. WinBUGS has a number of tools for reporting the posterior distribution. A simple summary (Table 1) can be generated and monitoring posterior mean, median and standard deviation with a 95% posterior credible interval. Parameter names are related in an obvious way to the model in Sec. 2. A fuller picture of the posterior distribution can be provided using the density option in the Sample Monitor Tool which draws a kernel density estimate of the posterior distribution of any chosen parameter, as in Figure 1. Often, the quantities of primary interest in survival data analysis is the diagnostic factor β in Cox model. The interval of beta does not contain the value zero, in Fig. 1, so the selection of that covariate is hence recommended. There are various other options for displaying the posterior distribution. For example the Compare ... menu item brings up the Comparison Tool that draws model fit for the behavior baselines described at Figure 2, with a vertex every 20 months.

Table 1. WinBUGS output for the DRS data: posterior statistics without frailty

Node	Mean	SD	MC error	2.5 %	Median	97.5 %	Start	sample
beta	-0.8034	0.1682	0.003393	-1.133	-0.8002	-0.4768	1001	20,000
c	156.1	63.5	1.089	71.72	144.0	311.0	1001	20,000
r	0.01372	0.001901	3.485E-5	0.01035	0.01359	0.01777	1001	20,000

Here the hazard ratio of the untreated eye group compared to the treated eye group is $\exp(0.8034) = 2.22$. WinBUGS automatically implements the DIC (Spiegelhalter et al. [38]) model comparison criterion. This is a portable information criterion quantity that trades off goodness-of-fit against a model complexity penalty. In hierarchical models, deciding the model complexity may be difficult and the method estimates

the “effective number of parameter”, denoted here by P_D . \bar{D} is the posterior mean of the deviance ($-2 \times \log \text{likelihood}$) and \hat{D} is a plug-in estimate of the latter based on the posterior mean of the parameters. The DIC is computed as $\text{DIC} = \bar{D} + P_D = \hat{D} + 2P_D$. Lower values of the criterion indicate better fitting models. We refer the interested reader to Berg (2004). Table 2 records the values computed, in the format given by WinBUGS. For our purposes here, we will focus only on the DIC value. The method was designed to be easy to implement using a sample from the posterior distribution and the interested reader is directed to Spiegelhalter et al. [36] for a lively discussion of its merits and its relation to the more usual Bayes factor.

Table 2 WinBUGS output for the DRS data: DIC without frailty

	\bar{D}	\hat{D}	DIC	P_D
dN	1594.0	1569.0	1620.0	25.5
total	1594.0	1569.0	1620.0	25.5

Figure 1. Posterior distribution for “beta”, “c”, and “r” without taking the frailty effect into account

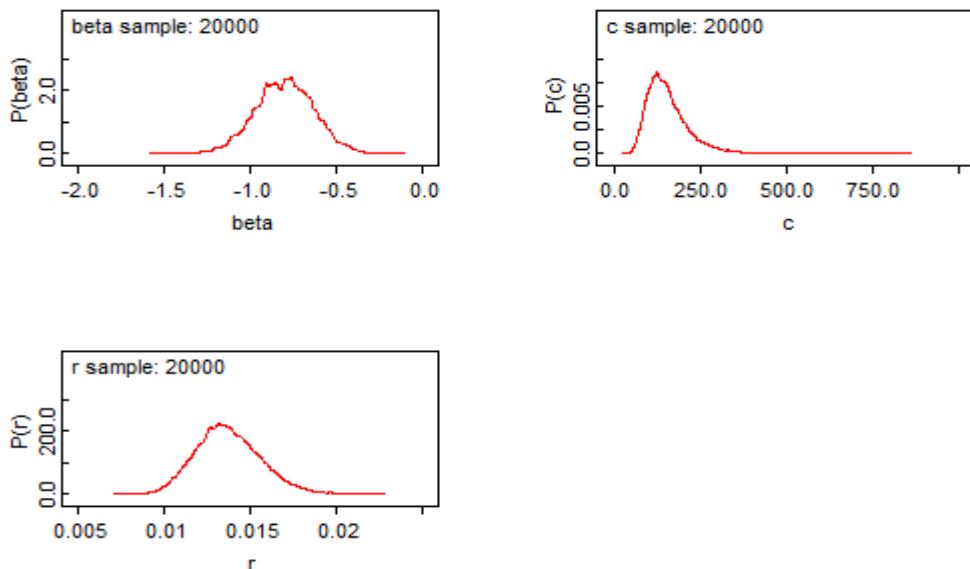
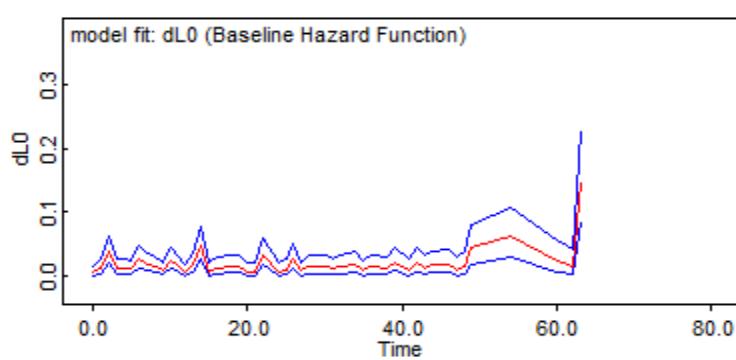


Figure 2. Monte Carlo approximations to the baseline hazard with 95% CI



3.2. Outputs analysis from WinBUGS software with gamma frailty

We are rerunning the script in Display 1 with taking frailty term in Cox model and follow the same steps has been described in the previous Subsection. We are running this script for 20,000 iterations, discarding the first 1,000 iterations to allow for convergence to the posterior distribution. The BUGS team, use the idea of parallel multiple chains to check the convergence of the Gibbs sampler, we used 2 chains, as shown in Figure 4. Table 3 shows the Monte Carlo 95% confidence intervals for the coefficient of regression in Cox model and the standard deviation for the frailty in the model (2). The DIC value is computed in case the shared gamma frailty in Table 4.

Table 3 WinBUGS output for the DRS data: posterior statistics

Node	Mean	SD	MC error	2.5 %	Median	97.5 %	Start	sample
beta	-0.9329	0.1784	0.004001	-1.295	-0.9304	-0.5874	1001	20,000
sigma	1.298	0.2665	0.01207	0.9247	1.259	1.885	1001	20,000

Table 4 WinBUGS output for the leukemia data: DIC with gamma frailty

	\bar{D}	\hat{D}	DIC	p_D
dN	1499.0	1424.0	1574.0	75.37
total	1499.0	1424.0	1574.0	75.37

Figure 3. Posterior distribution for “beta” and “sigma”, with taking the frailty effect into account

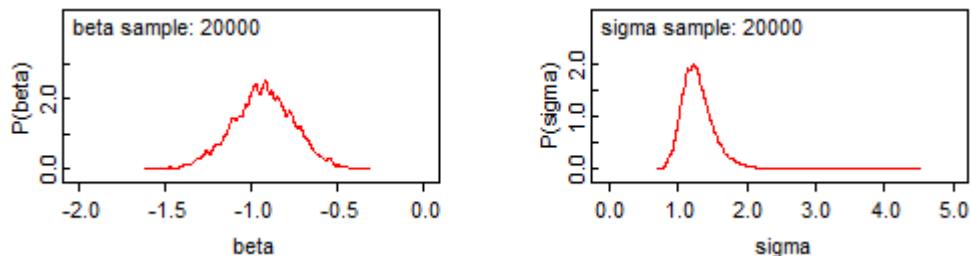


Figure 4. ACF for the iterations for each chain

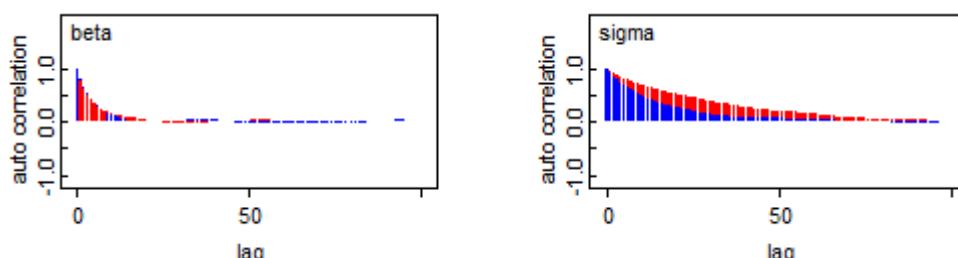


Figure 5. Monte Carlo approximations to the baseline hazard with 95% CI

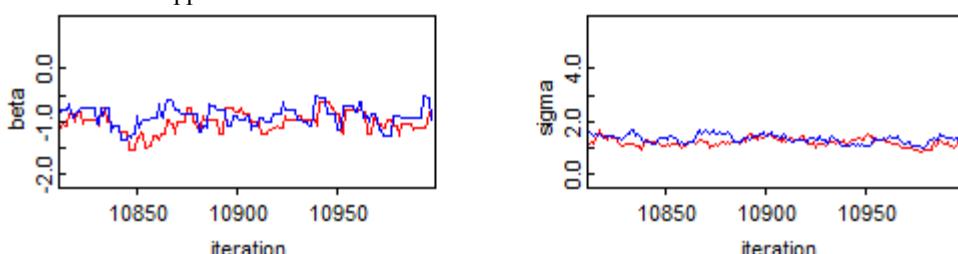
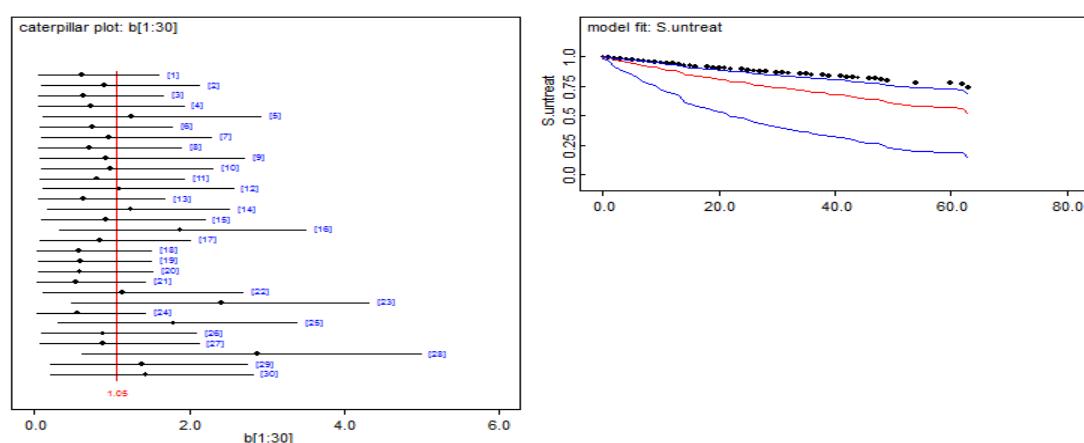


Figure 6. Plot of 95% interval frailty distribution for the first thirty pairs of eye in the sample, on the left and Monte Carlo approximations to the survival functions for the untreated eye group compared to the untreated group with 95% CI, on the write.



3.3. Outputs analysis from WinBUGS software with normal frailty

It is noted that, in Bayesian statistics, when the prior distribution is not conjugate, the posterior distribution cannot be written down in closed form. In such cases, an analyst may replace the numerically defined posterior with a distribution of a particular functional form (e.g., normal distribution for the frailty). In Script for Display 1, the line that calculates the frailty distribution (line #20#) from the gamma prior is replaced by line $b[k] \sim dnorm(0.0, tau.b)$ and repeat the analysis, we get the following tables.

Table 5. WinBUGS output for the DRS data: posterior statistics

Node	Mean	SD	MC error	2.5 %	Median	97.5 %	Start	sample
beta	-0.9549	0.1814	0.004273	-1.321	-0.9529	-0.6066	1001	20,000
sigma	1.013	0.1674	0.007753	0.6833	1.014	1.339	1001	20,000

Table 6. WinBUGS output for the DRS data: DIC with normal frailty

	\bar{D}	\hat{D}	DIC	P_D
dN	1476.0	1380.0	1573.0	96.81
total	1476.0	1380.0	1573.0	96.81

Monitoring convergence of the chains, which have been done in this article via the Brooks and Gelman (1998) convergence-diagnostic-graph. Hence, once convergence has been achieved, 10,000 observations are taken from each chain after the burn-in period to reach our goal of 20,000 observations. Inspection of the Brooks and Gelman's diagnostic graphs (Figure 8), we find the BGR (Brooks, Gelman and Rubin convergence statistic) convergent to one, this show that the convergence for the coefficient of regression β (i.e. line in the BGR plot should be close to 1.0, and the blue and green lines should be stable, as shown on the write of Fig. 8). Therefore, beyond the burn-in period, a sample of 10,000 observations from each of the two chains is drawn.

Figure 7. Estimated predictive density for the covariate of treatment effect on the left, and its convergence using quantiles button in WinBUGS software, on the right.

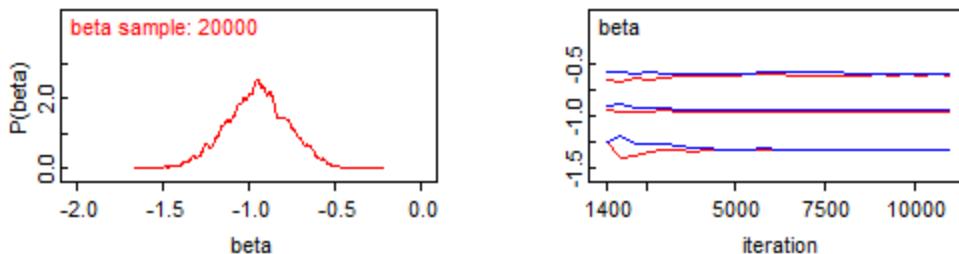


Figure 8. The trace plot and BGR Diagnostics results indicating convergence.

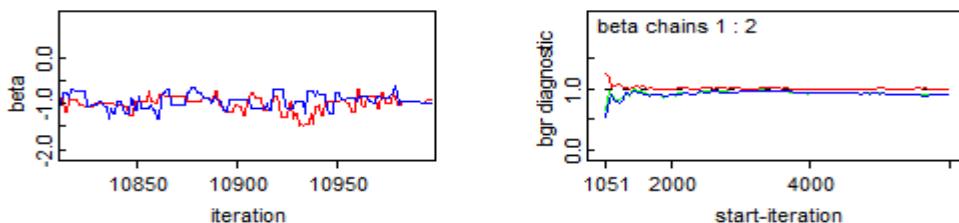
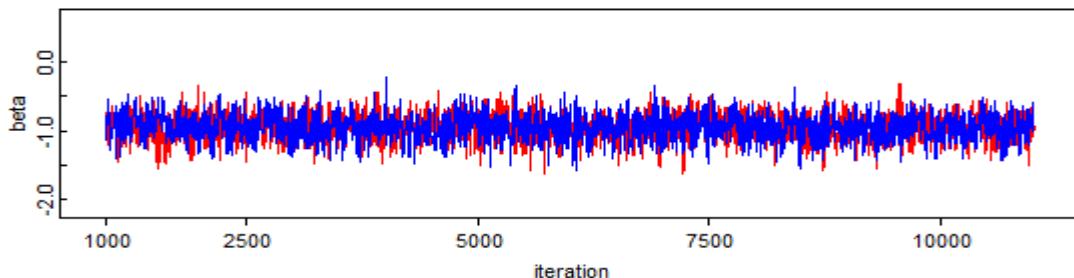


Figure 9 Evolution of the Markov Chain associated to the coefficient of the covariate of treatment effect for the DRS data



Finally; we have implemented an alternative for the frailty model (see Duchateau and Janssen, 2008; Ch. 3 Page 78), who Assume that the cluster effect is modeled by a fixed rather than by a random effect, as in the following model:

$$dN(t) \approx \lambda(t | Z)dt = \lambda_0(t) \exp[\beta_1 Z_1 + \varepsilon] = I(t)$$

where ε is the fixed effect for the i^{th} cluster. This is implemented through Display 1, by replacing #11# with the following line

```
Idt[i , j] <- Y[i , j] * exp(beta * Z1[i] + epsilon[i]) * dL0[j]
```

We rerunning the script again and we have got the DIC table as follows

Table 7 WinBUGS output for the DRS data: DIC with mixed effect frailty

	\bar{D}	\hat{D}	DIC	P_D
dN	1523.0	1436.0	1609.0	86.58
total	1523.0	1436.0	1609.0	86.58

From above, since we have run more than chains, the plots for all parameters in script Display 1 are indicative of convergence. For each node for the data set, similar set of graphs is produced to monitor convergence, independence and convergence. They are suppressed in this article for purposes of space limit. Finally; compare between Tables 2, 4, 6, and 7. For our purposes here, we will focus only the DIC value. Smaller values of the DIC suggest better models and so the addition of shared gamma frailty or shared normal frailty better than without its effect or the mixed effect frailty.

4. Conclusion and recommendation

As Griffin and Steel (2007) demonstrate that WinBUGS provides a useful framework for the Bayesian analysis of stochastic frontier models, we confirmed and also demonstrated a useful framework for the Bayesian analysis of frailty in Cox model. Although Spiegelhalter et al. 1996 explain in their user manual the usefulness of frailty using Leukemia example (see, <http://www.mrc-bsu.cam.ac.uk/bugs/documentation/exampVol1/node29.html>); in this paper, we followed his perspective and develop it using Diabetic Retinopathy Study (DRS). Leukemia example has 21 pairs of patients only, but DRS data has 197 pairs of observations. From the main assumptions of Cox proportional hazard model is that we have enough data. In this paper, we illustrate the effect of ignoring the heterogeneity through enough data (http://mayoresearch.mayo.edu/mayo/research/biostat/upload/therneau_upload/diabetes.dat). Code for converting DRS data to be suitable implemented in WinBUGS/OpenBUGS format can be using software BAUW “Bayesian Analysis Using WinBUGS”. In Web page <http://www.psychstat.org/us/sort.php/24.htm>; the authors Zhiyong Zhang and Lijuan Wang (2006) give an example on how to transform the data in a text file to WinBUGS format. We would certainly recommend the applied user of Bayesian survival analysis to experiment with WinBUGS and we hope that the BUGS team <http://www.mrc-bsu.cam.ac.uk/bugs/> replace their Leukemia example in their updated user manual with our studied example to help the new users of WinBUGS to study and develop their skills in WinBUGS through Bayesian survival data analysis.

Acknowledgements

This research was supported by (Project #321212 through Deanery of Academic Research Fund), awarded to (Faculty of Sciences) in the Population Research Unit at The University of Al-Imam Muhammad Ibn Saud Islamic by Deanery of Academic Research.

REFERENCES

- [1] [No authors listed], (1976), Diabetic Retinopathy Study Research Group. Preliminary report on effects of photocoagulation therapy. American Journal of Ophthalmology.;81(4):383–396. URL: www.ncbi.nlm.nih.gov
- [2] Aalen O. O., Borgan, Ø., Gjessing, H. K. (2008). Survival and Event History Analysis: A Process Point of View. Springer-Verlag.
- [3] Andersen, P. K. and Gill, R. D. (1982). Cox’s regression model for counting processes: A large sample study. Annals of Statistics, 10, 1100—1120.
- [4] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). Statistical models based on counting

- processes. Springer Series in Statistics. Springer-Verlag, New York.
- [5] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 10, 1100–1120.
- [6] Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N. (1993). Statistical Models Based on Counting Processes. New York: Springer-Verlag.
- [7] Beard, R.E. (1959) Note on some mathematical mortality models. In: G.E.W. Wolstenholme and M.O'Conner (eds.). *The Lifespan of Animals*. Ciba Foundation Colloquium on Ageing. Little, Brown, Boston, 302 - 311
- [8] Berg, A. Meyer, R. and Yu, J. (2004). Deviance Information Criterion for Comparing Stochastic Volatility Models. *American Statistical Association Journal of Business & Economic Statistics*, 22 (1), 107—120.
- [9] Brooks, S.P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulation. *Journal of Computational and Graphical Statistical*, 7: 434 — 455.
- [10] Brooks, S.P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulation. *Journal of Computational and Graphical Statistical*, 7: 434 — 455.
- [11] Clayton, D. (1978). A model for association in bivariate life-tables and its application in epidemiological studies of chronic disease incidence. *Biometrika*, 65: 141—151.
- [12] Clayton, D. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 47: 467–485.
- [13] Clayton, D. (1994). Bayesian analysis of frailty models. Technical Report, Medical Research Council Biostatistics Unit, Cambridge.
- [14] Clayton, D. G. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models., *Biometrics*.
- [15] Collett, D. (2003). Modelling survival data in medical research. Boca Raton: Chapman & Hall/CRC (2nd edition, texts in statistical science).
- [16] Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B*, 34, 187—220.
- [17] Cox, D. R., and Oakes, D., (1984). Analysis of survival data. London: Chapman and Hall.
- [18] Duchateau, L., Janssen, P. (2008). The Frailty Model. Springer, New York.
- [19] Gelman, A. and Rubin, D. (1992a). Inference from iterative simulation using multiple sequences. *Statistical Science* 7: 457—511.
- [20] Gelman, A. and Rubin, D. (1992b). A single from the Gibbs sampler provides a false sense of security. *Bayesian Statistics* 4, eds. J.M. Bernardo, J.O. Berger, A.P.
- [21] Gill, R. D. (2005). Product-integration. In *Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton), (6), 4246-4250. Wiley, Chichester.
- [22] Griffin, J. E, Steel, M. F. J. (2007). Bayesian stochastic frontier analysis using WinBUGS. *Journal of productivity analysis*, 27(3), 163-176.
- [23] Heckman, J.J., Singer, B. (1984) Econometric duration analysis. *Journal of Econometrics* 24, 63 – 132.
- [24] Hougaard, P. (2000). Analysis of multivariate survival data. New York: Springer- Verlag.
- [25] Huster W.J. and Brookmeyer, R. and Self. S. (1989). Modelling paired survival data with covariates, *Biometrics* 45, 145-56.
- [26] Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2005). Bayesian Survival Analysis. Springer.
- [27] Kaplan, E.L. and P. Meier, (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457– 481.
- [28] Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47, 939 – 956.
- [29] Lee, E. T. and Wang, J. W. (2003). Statistical methods for survival data Analysis. John Wiley and Sons Publication.

- [30] Næs, T. (1982). The asymptotic distribution of the estimator for the regression Parameter in Cox's regression model. *Scandinavian Journal of Statistics*, 9, 107-115.
- [31] Ntzoufras I. (2009). *Bayesian Modeling Using WinBUGS*. New York : John Wiley and Sons.
- [32] Ross, J. S., Mulvey, G. K., Stauffer, B., Patlolla, V., Bernheim, S. M., Keenan, P. S. and Krumholz, H. M. (2008). Statistical models and patient predictors of readmission for heart failure: a systematic review., *Arch Intern Med*.
- [33] Sharef, E. (2008). Nonparametric frailty models for clustered survival data. PhD thesis, Cornell, University.
- [34] Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 88: 979—983.
- [35] Spiegelhalter, D. J. (2001). Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 20 435--452.
- [36] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 64, 583–639.
- [37] Spiegelhalter, D. Thomas, A. Best, N. and Gillks, W. (1996). *BUGS 0.5: Examples Volume 1*, MRC, Biostatistics Unit, Institute of Public Health, Cambridge, UK, 1996.
- [38] Spiegelhalter, D. Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual, Version 1.4*, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK, URL: <http://www.mrc-bsu.cam.ac.uk/bugs/>.
- [39] Therneau, T. M. and Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. Springer.
- [40] Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, 9, 93–108.
- [41] van den Berg, G.J. (2001). Duration Models: specification, identification, and multiple durations. In: J.J. Heckman and E. Leamer (eds.), *Hand-book of Econometrics*, Volume V, North-Holland, Amsterdam.
- [42] Vaupel, J.W., Manton, K.G., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439 – 454.