REVSTAT – Statistical Journal
Volume 16, Number 2, April 2018, 231–255

SEMI-PARAMETRIC LIKELIHOOD INFERENCE FOR BIRNBAUM–SAUNDERS FRAILTY MODEL

Authors: N. Balakrishnan
– Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
bala@mcmaster.ca

Kai Liu
– Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
liuk25@math.mcmaster.ca

Received: April 2017 Revised: August 2017 Accepted: August 2017

Abstract:

- Cluster failure time data are commonly encountered in survival analysis due to different factors such as shared environmental conditions and genetic similarity. In such cases, careful attention needs to be paid to the correlation among subjects within same clusters. In this paper, we study a frailty model based on Birnbaum–Saunders frailty distribution. We approximate the intractable integrals in the likelihood function by the use of Monte Carlo simulations and then use the piecewise constant baseline hazard function within the proportional hazards model in frailty framework. Thereafter, the maximum likelihood estimates are numerically determined. A simulation study is conducted to evaluate the performance of the proposed model and the method of inference. Finally, we apply this model to a real data set to analyze the effect of sublingual nitroglycerin and oral isosorbide dinitrate on angina pectoris of coronary heart disease patients and compare our results with those based on other frailty models considered earlier in the literature.

Key-Words:

- *Birnbaum–Saunders distribution; censored data; cluster time data; frailty model; Monte Carlo simulation; piecewise constant hazards.*

AMS Subject Classification:

- 62N02.

1. INTRODUCTION

It is of natural interest in medical or epidemiological studies to examine the effects of treatments. Proportional hazards model, proposed by Cox [5], is the most popular model for the analysis of such survival data which models the hazard function as

$$h(t) = h_0(t) \exp(\boldsymbol{\beta}'\boldsymbol{x}),$$

where $t, \boldsymbol{x}$ and h_0 are the time to certain event, set of covariates and baseline hazard function, respectively. This model makes a critical assumption of independent observations from the subjects. However, correlation commonly exists in survival data due to shared environmental factors or genetic similarity. Therefore, neglecting this correlation may lead to biased results. A convenient choice for modeling these kinds of correlation in survival data is the frailty model. The terminology frailty was first introduced by Vaupel *et al.* [20], while accounting for the heterogeneity of individuals in distinct clusters. Generally speaking, the more frail an individual is, the earlier the event of interest will be. A shared frailty model introduces multiplicative random effects, which is referred to as the frailty term, in the proportional hazards model, and is defined as follows. Let $(t_{ij}, \delta_{ij}, \boldsymbol{x}_{ij}), i = 1, ..., n, j = 1, ..., m_i$, be the failure time, censoring indicator, and the covariate vector of the jth individual in the ith cluster, where δ_{ij} is 1 if t_{ij} is not censored and 0 otherwise. Let y_i be the frailty shared commonly by all the subjects in the ith cluster. Then, given y_i, t_{ij} are assumed to be independent with hazard function

$$h(t_{ij}|y_i) = y_i h_0(t_{ij}) \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij}). \tag{1.1}$$

The frailties y_i are assumed to be independent and identically distributed with a distribution, called the frailty distribution. The baseline hazard $h_0(t_{ij})$ is arbitrary. A common parametric choice of the baseline hazard is Weibull. Klein [13] proposed a non-parametric estimate of the cumulative hazard function of baseline distribution and then used a profile likelihood function.

The most prevailing choice of the frailty distribution is gamma distribution due to its mathematical simplicity and the mathematical tractability of ensuing inference [13]. It has a closed-form for the conditional likelihood function, given the observed data, so that EM algorithm can be applied effectively to obtain the maximum likelihood estimates. Another possibility is the positive stable distribution proposed by Hougaard [10]. Furthermore, Hougaard [11] derived power variance function from the positive stable distribution, which contains the preceding frailty distributions as special cases. All these distributions have simple Laplace transforms and therefore facilitates convenient computation of maximum likelihood estimates. However, there is no real biological reason for their use. Nevertheless, when the Laplace transform of the frailty distribution is unknown,

the likelihood function becomes intractable. Lognormal distribution is one such example. McGilchrist and Aisbett [16] developed a best linear unbiased prediction (BLUP) estimation method in this case of lognormal frailty model. Balakrishnan and Peng [2] proposed the generalized gamma frailty model since the generalized gamma distribution contains the gamma, Weibull, lognormal and exponential distributions all as special cases. Consequently, the generalized gamma frailty model becomes more flexible and tend to provide good fit to data as displayed by Balakrishnan and Peng [2].

The two-parameter Birnbaum–Saunders (BS) family of distributions was originally derived as a fatigue model by Birnbaum and Saunders [3] for which a more general derivation from a biological viewpoint was later provided by Desmond [7]. This distribution possesses many interesting distributional properties and shape characteristics. In the present work, we use this BS model as the frailty distribution along with a piecewise constant baseline hazard function within the proportional hazards model to come up with a flexible frailty model. The precise specification of this model is detailed in Section 2. An estimation method to obtain the maximum likelihood estimates of model parameters is presented in Section 3. A simulation study is conducted in Section 4 to assess the performance of the proposed method and then the usefulness of the proposed model and the method of inference is illustrated with a real data in Section 5. Discussions and some concluding remarks are finally made in Section 6.

2. MODEL SPECIFICATION

2.1. BS distribution as frailty distribution

The BS distribution was originally derived to model fatigue failure caused under cyclic loading [3]. The fatigue failure is due to the initiation, growth and ultimate extension of a dominant crack. It is assumed that the total crack extension Y_j due to the jth cycle, for $j = 1, ...,$ are independent and identically distributed random variables with mean μ and variance σ^2. Then, the distribution of the failure time (i.e., time for the crack to exceed a certain threshold level) is given by

$$F(t;\alpha,\beta) = \Phi\left[\frac{1}{\alpha}\left\{\left(\frac{t}{\beta}\right)^{1/2} - \left(\frac{\beta}{t}\right)^{1/2}\right\}\right], \qquad 0 < t < \infty, \quad \alpha, \beta > 0, \tag{2.1}$$

where Φ is the standard normal cumulative distribution function (CDF), and α and β are the shape and scale parameters, respectively. We now assume that the frailty random variable Y_i in (1.1) follows the BS distribution defined in (2.1).

Since $\frac{1}{\alpha}\left\{\left(\frac{T}{\beta}\right)^{1/2}-\left(\frac{\beta}{T}\right)^{1/2}\right\}$ is a standard normal random variable, the random variable T is simply given by

$$T = \beta\left\{\frac{\alpha Z}{2}+\left[\left(\frac{\alpha Z}{2}\right)^2+1\right]^{1/2}\right\}^2, \tag{2.2}$$

where $Z \sim N(0,1)$. The probability density function (PDF) of T, derived from (2.1), is given by

$$f(t;\alpha,\beta) = \frac{1}{2\sqrt{2\pi}\,\alpha\beta}\left[\left(\frac{\beta}{t}\right)^{1/2}+\left(\frac{\beta}{t}\right)^{3/2}\right]\exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)\right], \quad t>0. \tag{2.3}$$

The relation between T and Z in (2.2) enables us to obtain the mean and variance of T easily as

$$E(T) = \beta\left(1+\frac{1}{2}\,\alpha^2\right), \tag{2.4}$$

$$V(T) = (\alpha\beta)^2\left(1+\frac{5}{4}\,\alpha^2\right). \tag{2.5}$$

In the frailty model in (1.1), if the frailty term y_i is assumed to follow the BS distribution, for ensuring identifiability of model parameters, the mean of the frailty distribution needs to be set as 1. More specifically, let Y_1 be a BS random variable with shape parameter α and scale parameter β with its mean as 1. Let $Y_2 = cY_1$. Then, $E(Y_2) = cE(Y_1) = c$. Besides, we know that if $Y_1 \sim \mathrm{BS}(\alpha,\beta)$, then $cY_1 \sim \mathrm{BS}(\alpha, c\beta)$. Therefore, $Y_2 \sim \mathrm{BS}(\alpha, c\beta)$ with mean c. Then, given the frailty term y_2, the lifetime of the patients are modeled by the hazard function

$$h(t|y_2) = y_2\,h_0(t)\exp(\boldsymbol{\beta'x}) = c\,y_1\,h_0(t)\exp(\boldsymbol{\beta'x}).$$

Let us define $ch_0(t)$ to be a new baseline hazard function $h_1(t)$, which is nothing but rescaling the original baseline hazard function. Then, the model can be rewritten as

$$h(t|y_2) = y_1\,h_1(t)\exp(\boldsymbol{\beta'x}),$$

which is identical to a frailty model with frailty variable Y_1 and baseline hazard function $h_1(t) = ch_0(t)$.

Thus, the scale parameter β can be written in terms of the shape parameter α as

$$\beta = \frac{2}{2+\alpha^2}, \tag{2.6}$$

so that the variance of the frailty variable Y_i becomes

$$V(Y_i) = \frac{4\alpha^2+5\alpha^4}{\alpha^4+4\alpha^2+4}, \tag{2.7}$$

which is constrained to be in the interval $(0,5)$.

Some important discussions on inferential issues for BS distribution can be found in [1, 4, 8, 9, 15, 17, 18, 19].

2.2. Piecewise constant hazard as baseline hazard function

The baseline hazard $h_0(t)$ in (1.1) is normally assumed in the parametric setting to be that of exponential or Weibull distribution [11]. However, such a strong parametric assumption is not always desirable as the resulting inference may become non-robust. For this reason, we use a piecewise constant hazard function to approximate the baseline hazard so that it could capture inherent shape and features of the hazard function better. Let J be the number of partitions of the time interval, i.e., $0 = t^{(0)} < t^{(1)} < \cdots < t^{(J)}$, where $t^{(J)} > \max(t_{ij})$. The points $t^{(1)}, ..., t^{(J)}$ are called cut-points. The piecewise constant hazard function is then given by

$$h_0(t) = \gamma_k \qquad \text{for} \quad t^{(k-1)} \le t < t^{(k)} \quad \text{for} \quad k = 1, ..., J.$$

The corresponding cumulative hazard function is

$$H_0(t) = \sum_{q=1}^{k-1} \gamma_q \Big(t^{(q)} - t^{(q-1)} \Big) + \gamma_k \Big(t - t^{(k-1)} \Big) \qquad \text{for} \quad t^{(k-1)} \le t < t^{(k)}, \tag{2.8}$$

where γ_k is a constant hazard for interval $\big[t^{(k-1)}, t^{(k)}\big)$, $k = 1, ..., J$.

3. ESTIMATION METHOD

Let $(t_{ij}, \delta_{ij}, \boldsymbol{x}_{ij}), i = 1, ..., n, j = 1, ..., m_i$, be the failure time, censoring indicator, and the covariate vector for the jth individual in the ith cluster and y_i be the frailty term. Then, the full likelihood function of the BS frailty model is obtained from (1.1) as

$$\begin{aligned}
L &= \prod_{i=1}^{n} \int_0^\infty \Big(\prod_{j=1}^{m_i} h(t_{ij}|y_i)^{\delta_{ij}} S(t_{ij}|y_i) \Big) f(y_i)\, dy_i \\
&= \prod_{i=1}^{n} \int_0^\infty \Big[\prod_{j=1}^{m_i} \Big(y_i h_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \Big)^{\delta_{ij}} \\
&\quad \times \exp\Big(- y_i H_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \Big) \Big] f(y_i)\, dy_i \\
&= \prod_{i=1}^{n} \Big[\prod_{j=1}^{m_i} \Big(h_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \Big)^{\delta_{ij}} \\
&\quad \times \int_0^\infty y_i^{\delta_{i\cdot}} \exp\Big(- y_i \sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \Big) f(y_i)\, dy_i \Big] \\
&= \prod_{i=1}^{n} \Big[\prod_{j=1}^{m_i} \Big(h_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \Big)^{\delta_{ij}} I_i \Big],
\end{aligned} \tag{3.1}$$

where $\delta_{i\cdot} = \sum_{j=1}^{m_i} \delta_{ij}$, H_0 is the cumulative baseline hazard function with parameter $\boldsymbol{\gamma}$ as given in (2.8), f is the PDF of the BS distribution with shape parameter α and scale parameter $\beta = \frac{2}{2+\alpha^2}$ as given in (2.3), and

$$I_i = \int_0^\infty y_i^{\delta_{i\cdot}} \exp\Big(- y_i \sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big) f(y_i)\, dy_i.$$

The above expression of I_i can be rewritten as

$$I_i = \int_{-\infty}^\infty g(z_i)^{\delta_{i\cdot}} \exp\Big(- g(z_i) \sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big) f_Z(z_i)\, dz_i, \tag{3.2}$$

where f_Z is the PDF of the standard normal distribution and

$$g(z_i) = \frac{2}{2+\alpha^2}\left\{1 + \frac{\alpha^2 z_i^2}{2} + \alpha\, z_i\Big(1 + \frac{\alpha^2 z_i^2}{4}\Big)^{1/2}\right\}.$$

The maximum likelihood estimates are hard to determine due to the intractable integral in (3.2) present in the likelihood function in (3.1). A direct and convenient way is to use Monte Carlo simulation to approximate the integral in (3.2) as follows:

$$\begin{aligned} I_i &= E_Z\Big[g(Z)^{\delta_{i\cdot}} \exp\Big(- g(Z) \sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\Big] \\ &= \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}} \exp\Big(- g(z_{(k)}) \sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big), \end{aligned}$$

where $z_{(k)}, k = 1, ..., N$, are the realizations of standard normal random variable.

The log-likelihood function can then be approximated from (3.1) as

$$\begin{aligned} l = \sum_{i=1}^{n} \Bigg[& \sum_{j=1}^{m_i} \delta_{ij}\Big(\log h_0(t_{ij}) + \boldsymbol{\beta}'\boldsymbol{x}_{ij}\Big) \\ & + \log \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}} \exp\Big(- g(z_{(k)}) \sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\Bigg]. \end{aligned} \tag{3.3}$$

Once the approximate log-likelihood function is obtained as in (3.3), Fisher's score function and the Hessian matrix with respect to the parameters $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}$ can be obtained readily upon taking partial derivatives of first- and second-order, and pertinent details are presented in Appendix A. The MLEs of model parameters can then be obtained by Newton–Raphson algorithm iteratively as

$$\begin{bmatrix} \hat\alpha^{(k)} \\ \hat{\boldsymbol{\beta}}^{(k)} \\ \hat{\boldsymbol{\gamma}}^{(k)} \end{bmatrix} = \begin{bmatrix} \hat\alpha^{(k-1)} \\ \hat{\boldsymbol{\beta}}^{(k-1)} \\ \hat{\boldsymbol{\gamma}}^{(k-1)} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 l}{\partial\alpha^2} & \frac{\partial^2 l}{\partial\alpha\partial\boldsymbol{\beta}^T} & \frac{\partial^2 l}{\partial\alpha\partial\boldsymbol{\gamma}^T} \\ \frac{\partial^2 l}{\partial\alpha\partial\boldsymbol{\beta}} & \frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} & \frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^T} \\ \frac{\partial^2 l}{\partial\alpha\partial\boldsymbol{\gamma}} & \frac{\partial^2 l}{\partial\boldsymbol{\beta}^T\partial\boldsymbol{\gamma}} & \frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial l}{\partial\alpha} \\ \frac{\partial l}{\partial\boldsymbol{\beta}} \\ \frac{\partial l}{\partial\boldsymbol{\gamma}} \end{bmatrix}_{\alpha=\hat\alpha^{(k-1)}, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(k-1)}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}^{(k-1)}}.$$

The iterations need to be continued until the desired tolerance level is achieved, say, $|\hat{\boldsymbol{\theta}}_{i+1} - \hat{\boldsymbol{\theta}}_i| < 10^{-6}$. Finally, the standard errors of the estimates of $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}$ can be obtained from the inverse of the Hessian matrix evaluated at the determined MLEs.

4. SIMULATION STUDY

An extensive simulation study is carried out here to assess the performance of the proposed model and the method of estimation. We consider 4 scenarios: (1) $n = 100,\ m = 2$, (2) $n = 100,\ m = 4$, (3) $n = 100,\ m = 8$ and (4) $n = 400,\ m = 2$. Here, the clusters can be considered as hospitals and each subject as a patient in these hospitals. The patients are randomly assigned to either a treatment group or a control group with equal probability. The frailty term follows (1) the BS distribution with shape parameter $\frac{(2\sqrt{10}-2)^{1/2}}{3}$ and scale parameter $\frac{9}{8+\sqrt{10}}$, (2) gamma distribution (GA) with shape parameter 2 and scale parameter 0.5, (3) lognormal (LN) distribution with $\mu = -\frac{\log(1.5)}{2}$ and $\sigma^2 = \log(1.5)$. With these choices of parameters, the mean and variance of the frailty distribution become 1 and 0.5, respectively, for all these frailty distributions. The standard exponential distribution and the standard lognormal distribution are considered for baseline distributions. We then set $\beta = -\log(2) = -0.6931$ so that the hazard rate of patients in the treatment group is half of those in the control group. Finally, the censoring times are generated from the uniform distribution in $[0, 4.5]$.

The simulation procedure is as follows:

(**1**) Generate n frailty values from frailty distributions, i.e., $y_i, i = 1, ..., n$, and assign each subject in the same cluster with same frailty value.

(**2**) Assign each patient to treatment group or control group with probability 0.5.

(**3**) Given the frailty term, the survival function is

$$S(t_{ij}|y_i) = \exp\big(-y_i H_0(t_{ij}) \exp(\beta x_{ij})\big)$$

and the cumulative distribution function is

$$F(t_{ij}|y_i) = 1 - \exp\big(-y_i H_0(t_{ij}) \exp(\beta x_{ij})\big),$$

which follows a uniform distribution (0,1). Therefore we generate u_{ij} from Uniform(0,1) and set $F(t_{ij}|y_i) = u_{ij}$.

(**4**) Calculate the baseline cumulative hazard function, which is

$$H_0(t_{ij}) = -\frac{\log(1 - u_{ij})}{y_i \exp(\beta x_{ij})}.$$

(**5**) Solve for the lifetime according to the true baseline distribution, i.e.: for standard exponential, $t_{ij} = H_0(t_{ij})$; for standard lognormal, $t_{ij} = \exp\big(\Phi^{-1}(1 - \exp(-H_0(t_{ij})))\big)$ since $1 - \exp(-H_0(t_{ij})) = \Phi(\log(t_{ij}))$.

(**6**) Now, we generate censoring time c_{ij} from Uniform[0,4.5].

(**7**) Compare t_{ij} and c_{ij}. If $t_{ij} <= c_{ij}$, then set t_{ij} to be the observed time and the censoring indicator $\delta_{ij} = 1$. If $t_{ij} > c_{ij}$, we set c_{ij} to be our observed time and $\delta_{ij} = 0$.

We generated 1000 data sets under each setting and applied the proposed semi-parametric BS frailty model to these data sets. For comparative purposes, we fitted the simulated data sets with the parametric BS frailty model along with gamma and lognormal frailty models. Thus, we fitted 6 models for each simulated data with frailty distribution to be one of BS, gamma or lognormal, and the baseline hazard function to be either piecewise constant hazard function or Weibull hazard function. The primary parameters of interest are the treatment effect and the frailty variance, and so our attention will focus on these parameters. The estimates of the treatment effect are summarized in Figures 1 and 2, while Figures 3 and 4 demonstrate how the estimates of the frailty variance differ under different models. The horizontal black lines are the true values of the parameters of interest, while the vertical bars give 95% confidence intervals. The three numbers on the top of each plot are the rejection rate and coverage probabilities at confidence levels of 95% and 90%. The two numbers at the bottom of each plot provide bias and mean square error for the different models considered.

Figures 1 and 2 clearly show that the choice of frailty distribution has little impact on the estimate of treatment effect. When the true baseline distribution is exponential, either Weibull baseline hazard or piecewise constant hazard function will result in accurate estimation of the treatment effect. However, when the true baseline distribution is lognormal, use of piecewise constant hazard baseline distribution results in smaller bias and mean square error than when using the Weibull distribution as baseline. This reveals that misspecification of the baseline hazard function impacts the estimate of treatment effect and the semi-parametric frailty models are therefore better than the parametric frailty models based on robustness consideration.

The heterogeneity among clusters is explained by the frailty variance and so it is important to investigate the frailty variance. The estimates of frailty variance are shown in Figures 3 and 4. BS frailty model always has less mean square error than the lognormal frailty model no matter what the true frailty model is. Even though the gamma frailty model generally has smallest bias and mean square error, its coverage probabilities are quite small and considerably below the nominal level. Both parametric and semi-parametric BS frailty models have coverage probabilities close to the nominal level, and so does the lognormal frailty model. Furthermore, as the sample size gets larger, the estimates become more precise. When the sample size is small, the rejection rate is small for BS and lognormal frailty models, but they become larger when the sample size increases.

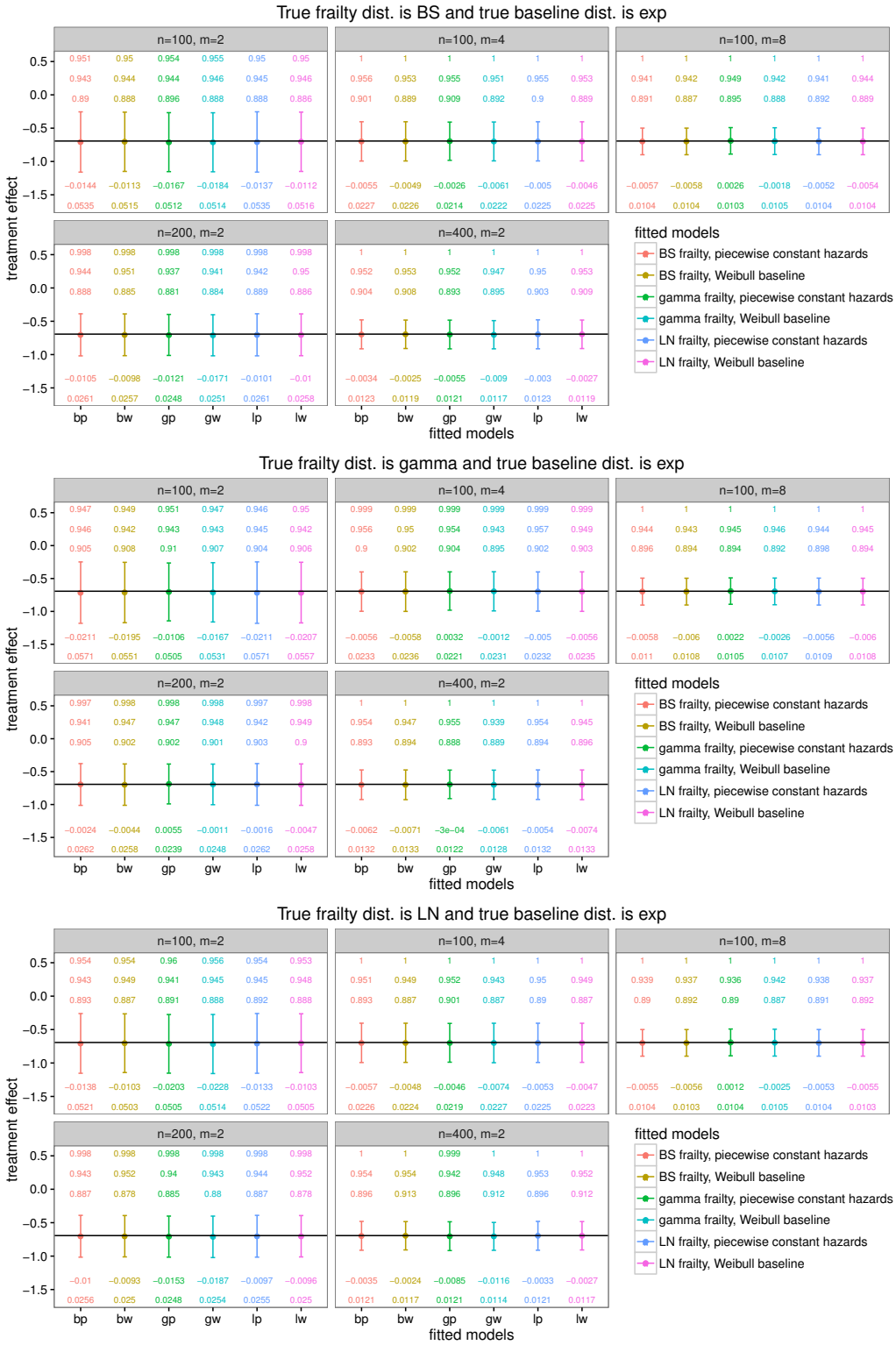

Figure 1: Estimate of treatment effect when the true baseline distribution is exponential.

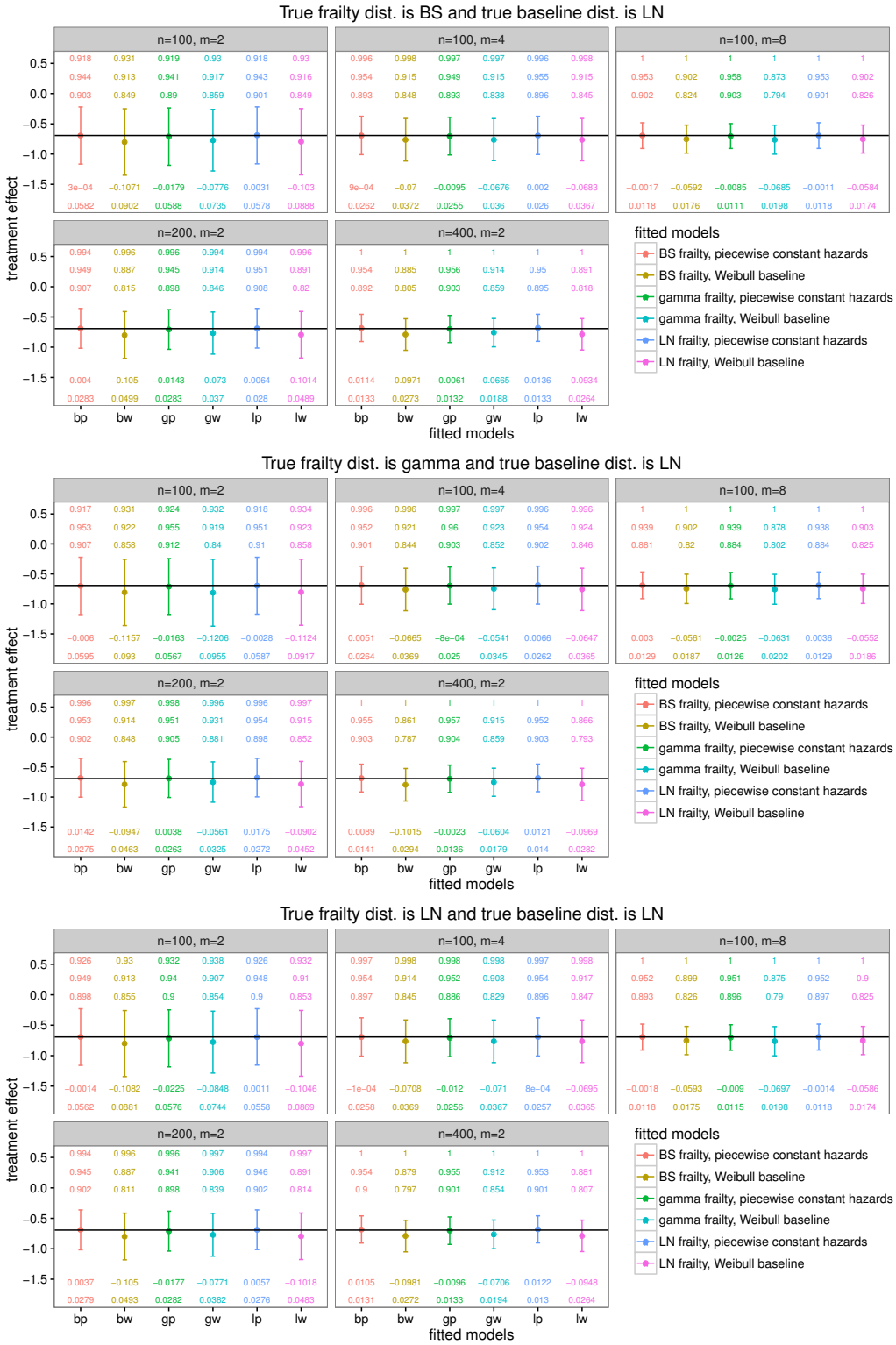

Figure 2: Estimate of treatment effect when the true baseline distribution is lognormal.

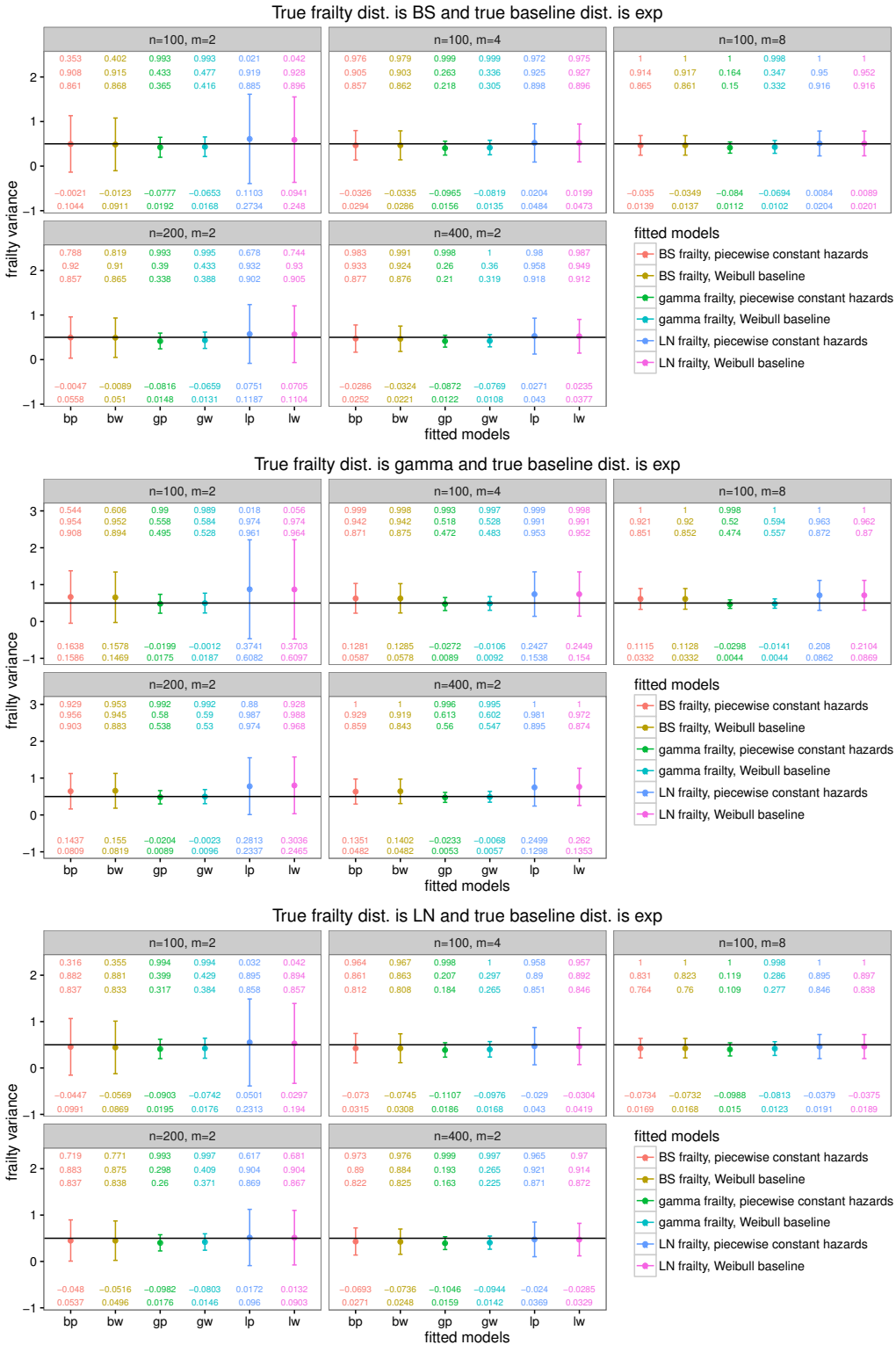

Figure 3: Estimate of frailty variance when the true baseline distribution is exponential.

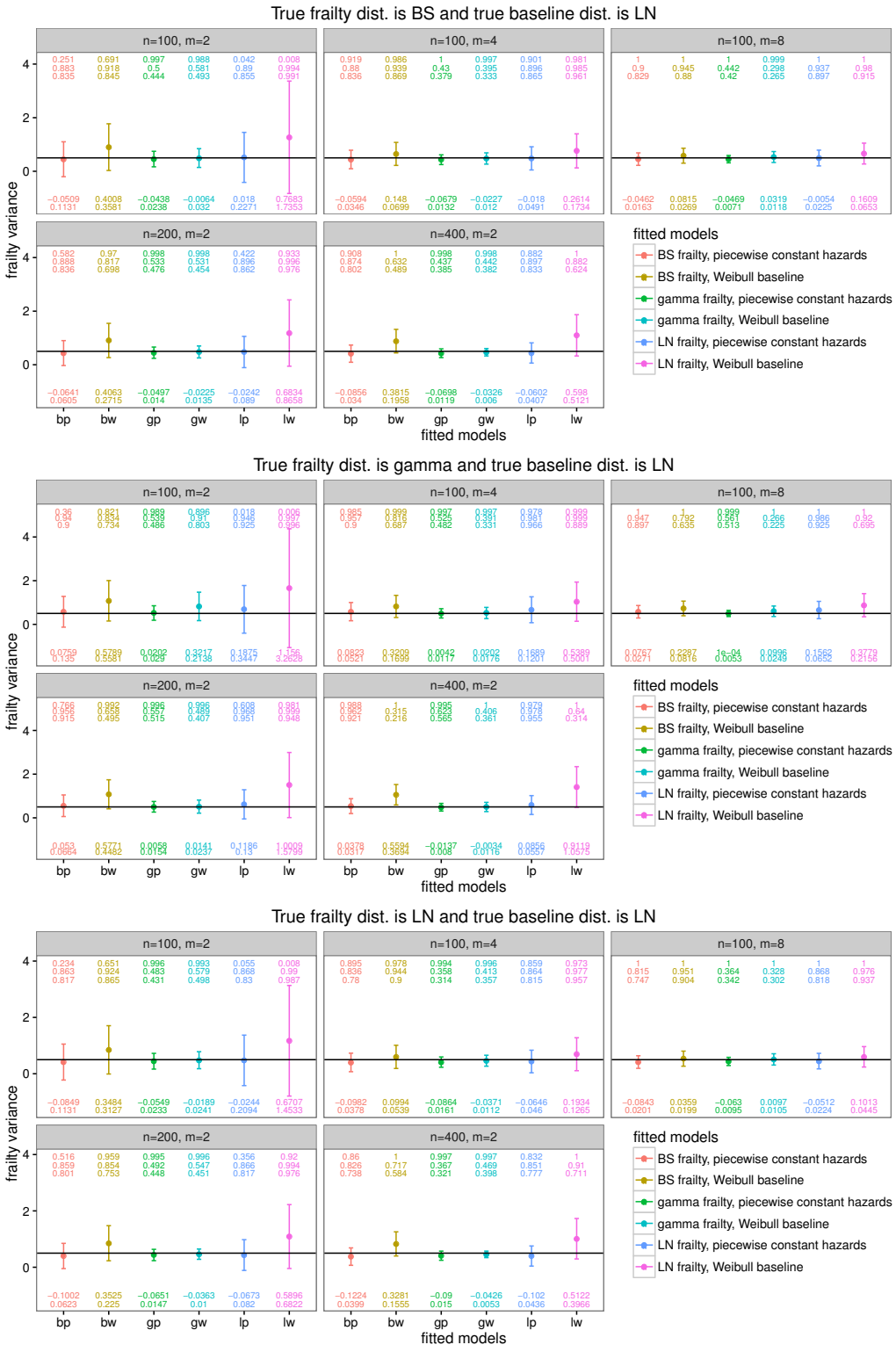

Figure 4: Estimate of frailty variance when the true baseline distribution is lognormal.

Table 1 summarizes the selection rate of the models based on the log-likelihood value. When the true baseline distribution is exponential, the models with correct frailty distributions generally have the largest selection rate except when the true frailty distribution is lognormal and the number of clusters is 100.

Table 1: Observed selection rates based on log-likelihood value.

Fitted models		True models					
baseline	frailty	BS		GA		LN	
		Exp	LN	Exp	LN	Exp	LN
$n = 100,\ m = 2$							
Weibull	BS	0.385	0.587	0.285	0.296	0.399	0.580
	GA	0.262	0.283	0.418	0.633	0.232	0.280
	LN	0.353	0.130	0.297	0.071	0.369	0.140
Piecewise	BS	0.378	0.500	0.282	0.385	0.394	0.532
	GA	0.242	0.325	0.400	0.489	0.224	0.294
	LN	0.380	0.175	0.318	0.126	0.382	0.174
$n = 100, m = 4$							
Weibull	BS	0.386	0.512	0.227	0.435	0.392	0.506
	GA	0.287	0.322	0.563	0.463	0.249	0.305
	LN	0.327	0.166	0.210	0.102	0.359	0.189
Piecewise	BS	0.385	0.449	0.224	0.289	0.381	0.455
	GA	0.282	0.338	0.571	0.590	0.244	0.286
	LN	0.333	0.213	0.205	0.121	0.375	0.259
$n = 100, m = 8$							
Weibull	BS	0.518	0.594	0.189	0.318	0.467	0.539
	GA	0.203	0.190	0.676	0.579	0.162	0.194
	LN	0.279	0.216	0.135	0.103	0.371	0.267
Piecewise	BS	0.510	0.576	0.204	0.209	0.439	0.497
	GA	0.203	0.225	0.666	0.690	0.186	0.202
	LN	0.287	0.199	0.130	0.101	0.375	0.301
$n = 200, m = 2$							
Weibull	BS	0.362	0.668	0.251	0.673	0.337	0.642
	GA	0.254	0.275	0.508	0.293	0.251	0.290
	LN	0.284	0.057	0.241	0.034	0.412	0.068
Piecewise	BS	0.339	0.518	0.268	0.356	0.354	0.548
	GA	0.284	0.376	0.501	0.597	0.250	0.334
	LN	0.377	0.106	0.231	0.047	0.396	0.118
$n = 400, m = 2$							
Weibull	BS	0.365	0.755	0.221	0.774	0.345	0.714
	GA	0.237	0.237	0.561	0.217	0.214	0.272
	LN	0.398	0.008	0.218	0.009	0.441	0.014
Piecewise	BS	0.340	0.545	0.221	0.285	0.339	0.594
	GA	0.239	0.413	0.546	0.697	0.228	0.352
	LN	0.421	0.042	0.233	0.018	0.433	0.054

Under this situation, the BS frailty models have slightly greater selection rates than the lognormal frailty model. In fact, the log-likelihood values are quite close for BS and lognormal frailty models. When the number of clusters increases, the selection rate of the lognormal frailty model increases and indeed becomes the largest in the case when the true frailty distribution is lognormal. On the other hand, when the true baseline distribution is lognormal, the parametric BS frailty model becomes more likely to be selected, especially when the number of clusters increases. Use of the piecewise constant hazard baseline function results in increasing selection probability of the true frailty distribution when the frailty distribution is gamma. However, the semi-parametric BS frailty model often has the highest selection rate when the true frailty distribution is lognormal. This suggests that the semi-parametric BS frailty model often results in MLEs with larger likelihood values than the semi-parametric lognormal frailty model and thus provide a better fit to observed data.

In summary, the choice of frailty distribution and the baseline distribution is a critical issue in frailty modeling. An inappropriate baseline distribution seems lead to larger errors in the estimation of both treatment effect and the frailty variance. However, the choice of the frailty distribution has less influence on estimating the treatment effect, but it highly impacts the estimation of frailty variance. Finally, the proposed BS frailty model provides a robust estimate of treatment effect and the frailty variance overall, and generally results in larger likelihood values among all fitted models. The R codes are available upon request from the authors.

5. ILLUSTRATION WITH A CORONARY HEART DISEASE STUDY

In this section, we fit the proposed semi-parametric BS frailty model to a real data set from Danahy *et al.* [6] concerning a study of oral administration of isosorbide dinitrate on 21 coronary heart disease patients, presented in Table 2. In the study, the patients were treated initially with sublingual nitroglycerin (SLN) and sublingual placebo (SLP) and then two tests of bike pedalling were conducted on the patients. Then, they took oral isosorbide dinitrate (OI) and oral placebo (OP) after which eight bike pedalling tests were given right after (OI0, OP0) and 1h (OI1, OP1), 3h (OI3, OP3), 5h (OI5, OP5). The times to angina pectoris were then recorded. Some of the times were censored because the patients were too exhausted (times with * are the censoring times).

Hougaard [12] studied the effects of the treatments with the proportional hazards model. In addition, several frailty models with gamma, stable and power variance function as the frailty distribution, along with non-parametric and Weibull hazard functions, were fitted to these data. The analyses carried

out demonstrated that a frailty model fitted the data better than the classical proportional hazards model, and the power variance function frailty distribution was more suitable than the gamma frailty distribution. Balakrishnan and Peng [2] analyzed the same data with a generalized gamma frailty model (GG) with both parametric and semi-parametric baseline hazard functions. These authors showed that the generalized gamma frailty model provided a better fit than the gamma, Weibull and lognormal frailty models, which are all special cases of the generalized gamma frailty model.

Table 2: Exercise times to Angina Pectoris (in seconds).

ID	SLP	SLN	OP0	OP1	OP3	OP5	OI0	OI1	OI3	OI5
1	155	431	150	172	118	143	136	445*	393*	226
2	269	259	205	287	211	207	250	306	206	224
3	408	446	221	244	147	250	215	232	258	268
4	308	349	150	290	205	210	235	248	298	207
5	135	175	87	157	135	105	129	121	110	102
6	409	523	301	357	388	388	425	580	613	514
7	455	488	342	390	441	468	441	504*	519*	484*
8	182	227	215	210	188	189	208	264	210	172
9	141	102	131	125	99	115	154	110	123	105
10	104	231	108	114	136	111	89	145	172	123
11	207	249	228	224	251	206	250	230	264	216
12	198	247	190	199	243	222	147	403	290	208
13	274	397	234	249	267	241	231	540*	370	316
14	191	251	218	194	197	223	224	432	291	212
15	156	401	199	329	197	176	152	733*	492	303
16	458	766	406	431	448	328	417	743*	566	391
17	188	199	194	168	168	159	213	250	150	180
18	258	566*	277	264	276	251	490	559*	557*	439
19	437	552	424	512	560	478	406	651	624	554
20	115	237	234	232	281	237	229	327	280	321
21	200	387	227	199	223	227	265	565*	504*	517*

We first investigate the feature of the data through the cumulative hazard plot, presented in Figure 5. The cumulative hazard after taking placebo is seen to be higher than that after taking sublingual nitroglycerin or isosorbide dinitrate. The hazard rate is increasing after taking placebo while it looks to be increasing and then decreasing after taking isosorbide dinitrate. We then fitted these data with the parametric and semi-parametric BS frailty models. The obtained results are presented in Tables 3 and 4. In addition, we fitted the parametric and semi-parametric gamma (GA), lognormal (LN) and inverse Gaussian (IG) frailty models to these data. Furthermore, for comparative purpose, we also include estimates of the generalized gamma frailty model (GG) from [2]. The minimum and maximum time to angina pectoris were 87s and 766s, respectively. Figure 6 is a histogram of observed times and it

shows that the data is sparse at the tail. So, we chose the cutpoints to be $t^{(0)} = 87, t^{(1)} = 150, t^{(2)} = 200, t^{(3)} = 250, t^{(4)} = 300, t^{(5)} = 400, t^{(6)} = 766$ to capture changes in the piecewise constant hazard baseline function.

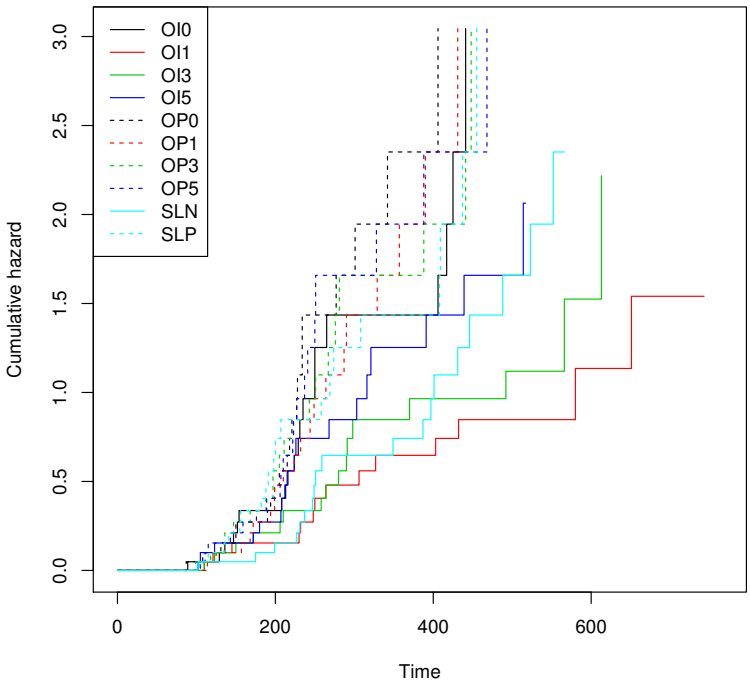

Figure 5: Cumulative hazard plot of the treatments.

Table 3: Fitted frailty models with Weibull baseline hazard function.

	BS	GA	LN	IG	GG
SLN	−1.54(0.34)	−1.51(0.34)	−1.55(0.34)	−1.43(0.34)	−1.51(0.34)
OP0	0.69(0.33)	0.69(0.33)	0.69(0.33)	0.7(0.33)	0.67(0.33)
OP1	0.12(0.32)	0.13(0.32)	0.12(0.32)	0.17(0.32)	0.11(0.33)
OP3	0.26(0.33)	0.28(0.33)	0.26(0.33)	0.24(0.33)	0.27(0.32)
OP5	0.63(0.33)	0.64(0.33)	0.63(0.33)	0.57(0.33)	0.66(0.32)
OI0	0.13(0.33)	0.15(0.33)	0.13(0.33)	0.15(0.33)	0.19(0.32)
OI1	−2.67(0.41)	−2.64(0.41)	−2.68(0.41)	−2.55(0.4)	−2.54(0.39)
OI3	−1.38(0.36)	−1.37(0.36)	−1.40(0.36)	−1.36(0.36)	−1.31(0.35)
OI5	−0.35(0.35)	−0.35(0.35)	−0.37(0.35)	−0.41(0.35)	−0.38(0.33)
$\log(p)$	1.59(0.06)	1.58(0.05)	1.59(0.06)	1.56(0.05)	1.59(0.06)
$\log(\lambda)$	−25.66(1.53)	−25.44(1.44)	−25.49(1.66)	−25.34(1.58)	−25.40(8.14)
Frailty variance	3.35(0.45)	2.51(0.07)	49.33(54.78)	10.87(1.19)	232.27(617.98)
Log−likelihood	−1121.86	−1124.86	−1122.12	−1123.00	1120.92
AIC	2267.71	2273.72	2268.23	2270.00	2267.82

Table 4: Fitted frailty models with piecewise constant baseline hazard function.

	BS	GA	LN	IG	GG
SLN	−1.38(0.34)	−1.38(0.33)	−1.38(0.34)	−1.43(0.33)	−1.37(0.34)
OP0	0.57(0.33)	0.59(0.33)	0.58(0.33)	0.50(0.32)	0.56(0.33)
OP1	−0.004(0.32)	0.01(0.32)	−0.01(0.32)	−0.07(0.31)	−0.06(0.31)
OP3	0.20(0.33)	0.22(0.32)	0.20(0.33)	0.12(0.32)	0.17(0.33)
OP5	0.50(0.32)	0.52(0.32)	0.50(0.32)	0.42(0.32)	0.48(0.32)
OI0	0.05(0.32)	0.06(0.32)	0.05(0.32)	−0.02(0.32)	0.09(0.32)
OI1	−2.18(0.38)	−2.24(0.38)	−2.17(0.39)	−2.21(0.38)	−2.16(0.38)
OI3	−1.29(0.35)	−1.32(0.35)	−1.29(0.35)	−1.34(0.35)	−1.28(0.35)
OI5	−0.39(0.34)	−0.41(0.34)	−0.40(0.34)	−0.46(0.33)	−0.41(0.41)
$\log(\gamma_1)$	−5.42(0.42)	−5.37(0.43)	−5.64(0.77)	−5.48(0.48)	−4.76(0.49)
$\log(\gamma_2)$	−3.96(0.48)	−3.95(0.46)	−4.15(0.76)	−3.98(0.44)	−3.24(0.48)
$\log(\gamma_3)$	−2.49(0.49)	−2.48(0.46)	−2.70(0.78)	−2.52(0.42)	−1.79(0.51)
$\log(\gamma_4)$	−1.83(0.52)	−1.79(0.49)	−2.05(0.80)	−1.91(0.42)	−1.14(0.57)
$\log(\gamma_5)$	−2.08(0.55)	−1.98(0.50)	−2.30(0.83)	−2.19(0.43)	−1.38(0.61)
$\log(\gamma_6)$	−0.57(0.56)	−0.43(0.50)	−0.79(0.84)	−0.71(0.41)	0.08(0.62)
Frailty variance	3.02(0.47)	2.34(0.07)	16.67(17.45)	10.52(1.17)	56.18(105.16)
Log−likelihood	−1111.88	−1116.58	−1112.16	−1111.562	−1111.39
AIC	2255.75	2265.16	2256.31	2255.12	2256.78

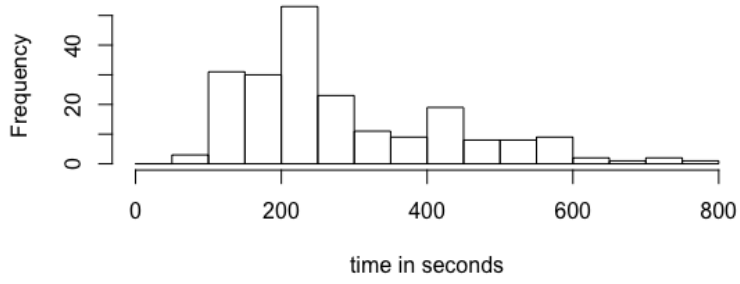

Figure 6: Histogram of observed times.

All the models result in similar estimates of the treatment effects, which are consistent with the results of Hougaard [12] and Balakrishnan and Peng [2]. Among all the frailty models fitted, the parametric BS frailty models provided the best fit since they had the smallest AIC values compared to other parametric models, even compared to the parametric generalized gamma frailty model possessing one extra shape parameter. Among the semi-parametric models, even though the inverse Gaussian model has the smallest AIC, the AIC of semi-parametric BS frailty model is quite close. Upon comparing the parametric and semi-parametric frailty models, we note that the semi-parametric frailty model has smaller AIC than its parametric counterparts. It is of interest to notice that estimates of

frailty variance are quite large for parametric lognormal and generalized gamma, and so are their standard errors. It is because we estimate the parameter of the frailty distribution (i.e., shape parameter for BS, gamma and inverse Gaussian and standard deviation of logarithm for lognormal), the estimates of frailty variance and its standard error are obtained by delta method. Small changes of estimate of parameter for lognormal distribution results in large change in estimate of its variance. The estimated CDF is presented in Figure 7. The black step curve is the non-parametric CDF obtained from the Kaplan–Meier estimates.

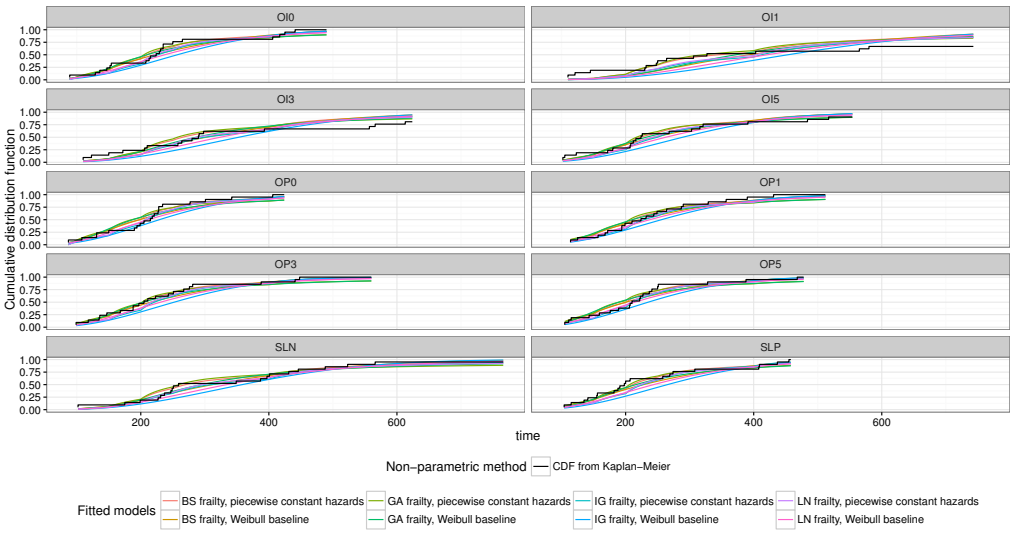

Figure 7: Fitted cumulative distribution functions.

We can see all the eight models fit the data well. To quantify the goodness-of-fit, we calculate the Kolmogorov–Smirnov distance (KSD) between the CDF of fitted models and the non-parametric CDF, presented in Table 5. It is defined as $D = \sup|\hat{F}(t) - \hat{F}_{km}(t)|$. It is seen clearly that piecewise linear baseline is better than Weibull baseline for all the models considered. This is also seen in the maximized log-likelihood and AIC values in Tables 3 and 4. Overall, the fits as measured by KSD are all quite similar with those by AIC indicating BS and IG models to be better. We also should examine the residuals to check the error. Figure 8 presents the deviance residuals, which is defined as

$$D_{r_{ij}} = \operatorname{sign}(r_{ij}) \sqrt{-2\left[r_{ij} + \delta_{ij} \log(\delta_{ij} - r_{ij})\right]},$$

where $r_{ij} = \delta_{ij} + \log(\hat{S}(t_{ij}))$. It can be seen that the deviance residuals are randomly distributed along 0. The deviance residuals should follow a standard normal distribution. For checking this, the QQ plot and envelopes of the deviance residuals are presented in Figure 9. It seems that all the models satisfy the normality assumption and the semi-parametric models are slightly better than the

parametric ones. The right tail of semi-parametric gamma frailty model deviates from the straight line more than the others. Semi-parametric BS, lognormal and inverse Gaussian frailty models are quite similar. Overall, semi-parametric BS is seen to be quite a robust model for modeling these clustered failure time data.

Table 5: KSD between estimated CDF and non-parametric CDF.

Frailty	Baseline	OI0	OI1	OI3	OI5	OP0	OP1	OP3	OP5	SLN	SLP	Overall
BS	piecewise	0.15	0.22	0.22	0.12	0.19	0.11	0.07	0.17	0.16	0.12	0.22
GA	piecewise	0.19	0.21	0.18	0.15	0.23	0.14	0.11	0.20	0.19	0.12	0.23
IG	piecewise	0.13	0.18	0.21	0.12	0.13	0.07	0.10	0.13	0.11	0.20	0.21
LN	piecewise	0.12	0.19	0.20	0.10	0.12	0.06	0.09	0.11	0.09	0.19	0.20
BS	Weibull	0.13	0.25	0.23	0.12	0.18	0.11	0.12	0.15	0.11	0.14	0.25
GA	Weibull	0.16	0.19	0.19	0.12	0.23	0.15	0.10	0.18	0.12	0.12	0.23
IG	Weibull	0.24	0.25	0.22	0.23	0.24	0.14	0.21	0.25	0.23	0.28	0.28
LN	Weibull	0.17	0.22	0.21	0.15	0.18	0.10	0.15	0.18	0.18	0.20	0.22

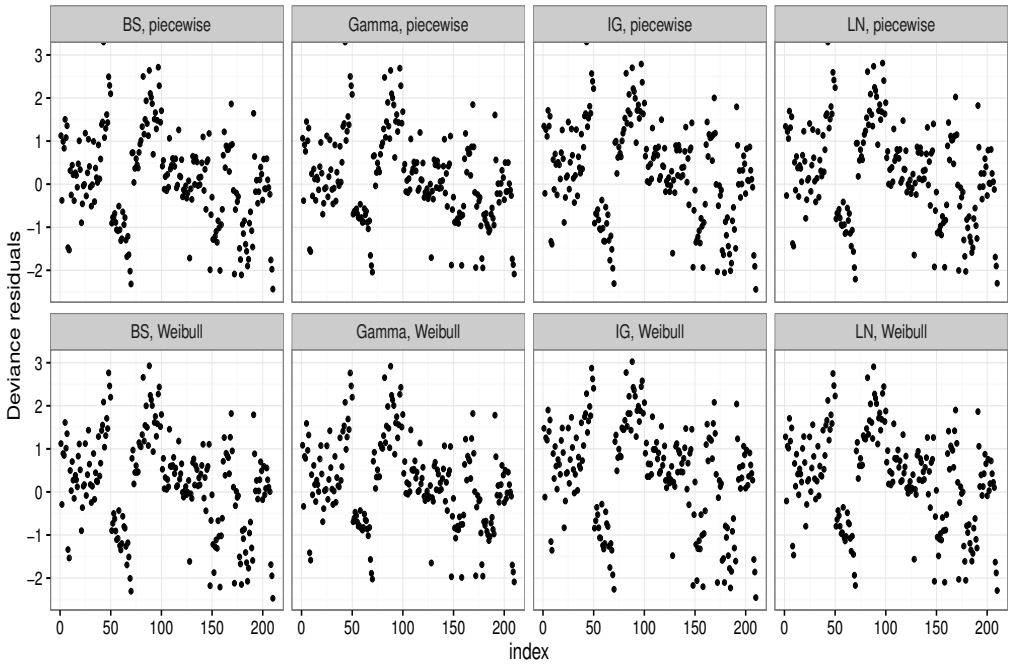

Figure 8: Deviance residuals.

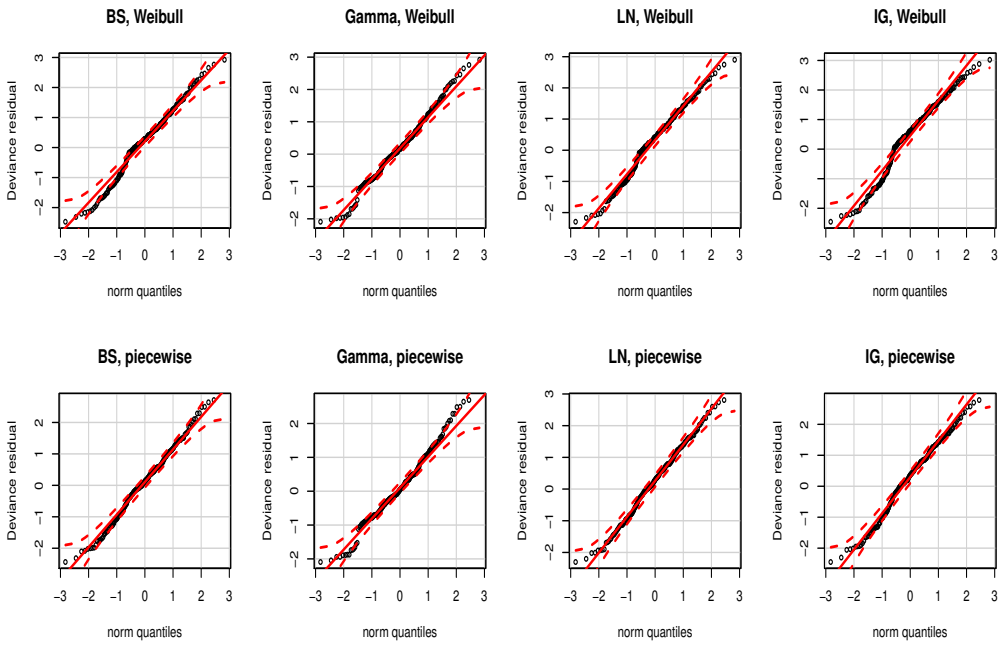

Figure 9: QQ plots for deviance residuals.

6. DISCUSSION AND CONCLUDING REMARKS

In this work, we have proposed a semi-parametric frailty model with BS frailty distribution. The non-parametric choice of baseline hazard function provides a robust and flexible way to model the data. The determination of MLEs becomes very difficult due to the intractable integrals present in the likelihood function. For this reason, Monte Carlo simulations are used to approximate the likelihood function upon exploiting the relationship between BS and standard normal distributions and then expressing those integrals as expectations of some functions of standard normal variables. From the simulation study carried out and the illustrative example analyzed, the semi-parametric BS frailty model is seen to be quite robust in estimating the covariate effects as well as the frailty variance. Interestingly, it is seen to be even better than the three-parameter generalized gamma frailty model though the latter has an extra shape parameter. It is of interest to mention that the work carried out here can be generalized in two different directions. The BS distribution can be generalized by assuming that the variable Z in (2.2) follows a standard elliptically symmetric distribution, including power exponential, Laplace, Student t and logistic distributions. Such a generalized Birnbaum–Saunders (GBS) distribution (see [14]) could be assumed for the frailty term y_i in (1.1) and then the resulting GBS frailty model could be studied in detail. Next, we could allow for the possibility of a cure of patients within the context of BS frailty model and develop the corresponding analysis. Work is currently under progress on these problems and we hope to report these findings in a future paper.

APPENDIX A — FIRST- AND SECOND-ORDER DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTION

The first- and second-order derivatives of the log-likelihood function with respect to $\alpha, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are as follows:

$$
\begin{aligned}
\frac{\partial l}{\partial \alpha} &= \sum_{i=1}^{n} \frac{1}{I_i} \frac{\partial I_i}{\partial \alpha}, \\
\frac{\partial l}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \left[\sum_{j=1}^{m_i} \delta_{ij} \boldsymbol{x_{ij}} + \frac{1}{I_i} \frac{\partial I_i}{\partial \boldsymbol{\beta}} \right], \\
\frac{\partial l}{\partial \boldsymbol{\gamma}} &= \sum_{i=1}^{n} \left[\sum_{j=1}^{m_i} \frac{\delta_{ij}}{h_0(t_{ij})} \frac{dh_0(t_{ij})}{d\boldsymbol{\gamma}} + \frac{1}{I_i} \frac{\partial I_i}{\partial \boldsymbol{\gamma}} \right]; \\
\frac{\partial^2 l}{\partial \alpha^2} &= \sum_{i=1}^{n} \left[-\frac{1}{I_i^2} \left(\frac{\partial I_i}{\partial \alpha} \right)^2 + \frac{1}{I_i} \frac{\partial^2 I_i}{\partial \alpha^2} \right], \\
\frac{\partial^2 l}{\partial \alpha \partial \boldsymbol{\beta}^T} &= \sum_{i=1}^{n} \left[-\frac{1}{I_i^2} \frac{\partial I_i}{\partial \alpha} \left(\frac{\partial I_i}{\partial \boldsymbol{\beta}} \right)^T + \frac{1}{I_i} \frac{\partial^2 I_i}{\partial \alpha \partial \boldsymbol{\beta}^T} \right], \\
\frac{\partial^2 l}{\partial \alpha \partial \boldsymbol{\gamma}^T} &= \sum_{i=1}^{n} \left[-\frac{1}{I_i^2} \frac{\partial I_i}{\partial \alpha} \left(\frac{\partial I_i}{\partial \boldsymbol{\gamma}} \right)^T + \frac{1}{I_i} \frac{\partial^2 I_i}{\partial \alpha \partial \boldsymbol{\gamma}^T} \right], \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \sum_{i=1}^{n} \left[-\frac{1}{I_i^2} \frac{\partial I_i}{\partial \boldsymbol{\beta}} \left(\frac{\partial I_i}{\partial \boldsymbol{\beta}} \right)^T + \frac{1}{I_i} \frac{\partial^2 I_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right], \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} &= \sum_{i=1}^{n} \left[-\frac{1}{I_i^2} \frac{\partial I_i}{\partial \boldsymbol{\beta}} \left(\frac{\partial I_i}{\partial \boldsymbol{\gamma}} \right)^T + \frac{1}{I_i} \frac{\partial^2 I_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \right], \\
\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} -\frac{\delta_{ij}}{h_0(t_{ij})^2} \frac{dh_0(t_{ij})}{d\boldsymbol{\gamma}} \left(\frac{dh_0(t_{ij})}{d\boldsymbol{\gamma}} \right)^T \\
&\quad + \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{\delta_{ij}}{h_0(t_{ij})} \frac{d^2 h_0(t_{ij})}{d\boldsymbol{\gamma}\, d\boldsymbol{\gamma}^T} \\
&\quad + \sum_{i=1}^{n} \left[-\frac{1}{I_i^2} \frac{\partial I_i}{\partial \boldsymbol{\gamma}} \left(\frac{\partial I_i}{\partial \boldsymbol{\gamma}} \right)^T + \frac{1}{I_i} \frac{\partial^2 I_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \right],
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial I_i}{\partial \alpha} &= \delta_{i\cdot} E_{1,i} - \left[\sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \right] E_{2,i}, \\
\frac{\partial^2 I_i}{\partial \alpha^2} &= \delta_{i\cdot}(\delta_{i\cdot} - 1) E_{3,i} - 2\delta_{i\cdot} \left[\sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \right] E_{4,i} + \delta_{i\cdot} E_{5,i} \\
&\quad + \left[\sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \right]^2 E_{6,i} - \left[\sum_{j=1}^{m_i} H_0(t_{ij}) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}) \right] E_{7,i},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 I_i}{\partial\alpha\,\partial\boldsymbol{\beta}^T} &= \left\{\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\right]E_{8,i} - (\delta_{i\cdot}+1)E_{2,i}\right\} \\
&\quad\times\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x}_{ij}\right], \\
\frac{\partial I_i}{\partial\boldsymbol{\beta}} &= -E_{9,i}\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x}_{ij}\right], \\
\frac{\partial I_i}{\partial\boldsymbol{\gamma}} &= -E_{9,i}\left[\sum_{j=1}^{m_i} \frac{dH_0(t_{ij})}{d\boldsymbol{\gamma}}\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\right], \\
\frac{\partial^2 I_i}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^T} &= E_{10,i}\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x_{ij}}\right]\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x_{ij}}\right]^T \\
&\quad - E_{9,i}\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x_{ij}}\,\boldsymbol{x_{ij}}^T\right], \\
\frac{\partial^2 I_i}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\gamma}^T} &= E_{10,i}\left[\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x_{ij}}\right]\left[\sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\frac{dH_0(t_{ij})}{d\boldsymbol{\gamma}}\right]^T \\
&\quad - E_{9,i}\left[\sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\boldsymbol{x_{ij}}\left(\frac{dH_0(t_{ij})}{d\boldsymbol{\gamma}}\right)^T\right], \\
\frac{\partial^2 I_i}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\gamma}^T} &= E_{10,i}\left[\sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\frac{dH_0(t_{ij})}{d\boldsymbol{\gamma}}\right]\left[\sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\frac{dH_0(t_{ij})}{d\boldsymbol{\gamma}}\right]^T \\
&\quad - E_{9,i}\left[\sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\,\frac{d^2H_0(t_{ij};\boldsymbol{\gamma})}{d\boldsymbol{\gamma}\,d\boldsymbol{\gamma}^T}\right];
\end{aligned}
$$

in the above expressions, the quantities $E_{l,i}$ $(l = 1, ..., 10)$ are given by

$$
\begin{aligned}
E_{1,i} &= \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}-1}\exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\frac{dg(z_{(k)})}{d\alpha}, \\
E_{2,i} &= \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}}\exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\frac{dg(z_{(k)})}{d\alpha}, \\
E_{3,i} &= \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}-2}\exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\left(\frac{dg(z_{(k)})}{d\alpha}\right)^2, \\
E_{4,i} &= \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}-1}\exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\left(\frac{dg(z_{(k)})}{d\alpha}\right)^2, \\
E_{5,i} &= \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}-1}\exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\frac{d^2g(z_{(k)})}{d\alpha^2},
\end{aligned}
$$

$$E_{6,i} = \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}} \exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\left(\frac{dg(z_{(k)})}{d\alpha}\right)^2,$$

$$E_{7,i} = \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}} \exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\frac{d^2g(z_{(k)})}{d\alpha^2},$$

$$E_{8,i} = \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}+1} \exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big)\frac{dg(z_{(k)})}{d\alpha},$$

$$E_{9,i} = \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}+1} \exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big),$$

$$E_{10,i} = \frac{1}{N}\sum_{k=1}^{N} g(z_{(k)})^{\delta_{i\cdot}+2} \exp\Big(-g(z_{(k)})\sum_{j=1}^{m_i} H_0(t_{ij})\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})\Big).$$

ACKNOWLEDGMENTS

The authors express their sincere thanks to the guest editor, Professor Sat Gupta, for extending an invitation and to the anonymous reviewers for their useful comments and suggestions on an earlier version of this manuscript which led to this improved version.

REFERENCES

[1] Balakrishnan, N.; Leiva, V. and López, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum–Saunders distribution, *Communications in Statistics–Simulation and Computation*, **36**(3), 643–656.

[2] Balakrishnan, N. and Peng, Y. (2006). Generalized gamma frailty model, *Statistics in Medicine*, **25**(16), 2797–2816.

[3] Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions, *Journal of Applied Probability*, **55**(2), 319–327.

[4] Chang, D.S. and Tang, L.C. (1993). Reliability bounds and critical time for the Birnbaum–Saunders distribution, *IEEE Transactions on Reliability*, **42**(3), 464–469.

[5] Cox, D.R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society, Series B*, **34**(2), 187–220.

[6] Danahy, D.T.; Burwell, D.T.; Aronow, W.S. and Prakash, R. (1977). Sustained hemodynamic and antianginal effect of high dose oral isosorbide dinitrate, *Circulation*, **55**(2), 381–387.

[7] Desmond, A. (1985). Stochastic models of failure in random environments, *The Canadian Journal of Statistics*, **13**(3), 171–183.

[8] Dupuis, D.J. and Mills, J.E. (1998). Robust estimation of the Birnbaum–Saunders distribution, *IEEE Transactions on Reliability*, **47**(1), 88–95.

[9] From, S.G. and Li, L. (2006). Estimation of the parameters of the Birnbaum–Saunders distribution, *Communications in Statistics — Theory and Methods*, **35**(12), 2157–2169.

[10] Hougaard, P. (1986a). A class of multivariate failure time distributions, *Biometrika*, **73**(3), 671–678.

[11] Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions, *Biometrika*, **73**(2), 387–396.

[12] Hougaard, P. (2012). *Analysis of Multivariate Survival Data*, Springer, New York.

[13] Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm, *Biometrics*, **48**(3), 795–806.

[14] Leiva, V.; Riquelme, M.; Balakrishnan, N. and Sanhueza, A. (2008). Lifetime analysis based on the generalized Birnbaum–Saunders distribution, *Computational Statistics & Data Analysis*, **52**(4), 2079–2097.

[15] Lemonte, A.J.; Cribari-Neto, F. and Vasconcellos, K.L. (2007). Improved statistical inference for the two-parameter Birnbaum–Saunders distribution, *Computational Statistics & Data Analysis*, **51**(9), 4656–4681.

[16] McGilchrist, C. and Aisbett, C. (1991). Regression with frailty in survival analysis, *Biometrics*, **47**(2), 461–466.

[17] Ng, H.K.T.; Kundu, D. and Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum–Saunders distribution, *Computational Statistics & Data Analysis*, **43**(3), 283–298.

[18] Ng, H.K.T.; Kundu, D. and Balakrishnan, N. (2006). Point and interval estimation for the two-parameter Birnbaum–Saunders distribution based on type-II censored samples, *Computational Statistics & Data Analysis*, **50**(11), 3222–3242.

[19] Rieck, J.R. (1999). A moment-generating function with application to the Birnbaum–Saunders distribution, *Communications in Statistics — Theory and Methods*, **28**(9), 2213–2222.

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