Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2020;6(1): 30-39

Original Article

Bayesian Joint Modeling of Skew-Positive Longitudinal-Survival Data Using Birnbaum-Saunders Distribution

Tohid Jafari-Koshki[1]*, Sayed Mohsen Hosseini[2], Shahram Arsang-Jang[3]

[1]Molecular Medicine Research Center, Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran.

[2]Department of Biostatistics and Epidemiology, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran.

[3]Department of Biostatistics and Epidemiology, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran.

ARTICLE INFO

Received 02.11.2019
Revised 29.12.2019
Accepted 05.02.2020
Published 10.03.2020

Key words:
Birnbaum-Saunders distribution;
Joint model;
Skew-positive

ABSTRACT

Background: There has been a great interest in joint modeling of longitudinal and survival data in recent two decades. Joint models have less restrictive assumptions in multivariate modeling and could address various research questions. This has led to their wide applications in practice. However, earlier models had normality assumption on the distribution in longitudinal part that is usually violated in real data. Hence, recent research have focused on circumventing this issue. Using various skewed distributions has been proposed and applied in the literature. Nevertheless, the flexibility of the proposed methods is limited especially when the data are skew positive.

Methods: In this paper, we introduce the use of Birnbaum-Saunders (BS) distribution in joint modeling context. BS distribution is more flexible and could cover a wide range of skew, kurtotic or bimodal data.

Results: We analyzed publicly available ddI/ddC data both with normal and BS distributions in Bayesian setting and compared their fit by Widely Applicable Information Criterion (WAIC). The joint BS model showed a better fit to the data.

Conclusion: We introduced and applied BS distribution in joint modeling of longitudinal-survival data. Using multi-parameter distributions such as BS in Bayesian setting could improve the fit of models without limitations that arise in transformation of data from original scale.

Introduction

The advent of computing facilities has made noticeable developments in statistical inference in recent decades resulting in easier multivariate modeling, less biased and more accurate assessments of phenomena. Amongst the other methods of the multivariate analysis, joint modeling has attracted great attention due to its less restrictive assumptions and various approaches of modeling, each aiming a particular research question (1, 2). The literature on joint modeling of survival and longitudinal data has been increasingly growing and many applications have been reported in cancer clinical trials and observational studies (3, 4). The interest on joint survival-longitudinal models is mainly due to limitations in modeling time-varying covariates, known as response here, and time to an event under investigation (5). Time-dependent survival models are limited in flexibility and addressing advanced questions.

Joint longitudinal- survival models date back to the late 1990s. Models proposed by Faucet et al (1996), Wulfsohn et al. (1997) and Henderson et al. (2000) were the basis for future developments (4, 6, 7). Primary models were based on the assumption of normal distribution for the error term of longitudinal response. This assumption was violated in many applications

*. Corresponding Author E-mail: tjkoshki@gmail.com

Please cite this article in press as Jafari-Koshki T, Hosseini SM, Arsang-Jang S. Bayesian Joint Modeling of Skew-Positive Longitudinal-Survival Data Using Birnbaum-Saunders Distribution. *2020; 6(1): 30-39*

and, hence, models with greater flexibility were required. Song et al. (2002) used a smooth density for random effects. Brown and Ibrahim (2003) and Rizopoulos et al. (2011) proposed the use of Dirichlet process for prior distributions in modeling multivariate longitudinal and time-to-event data (8, 9). Yu et al. (2008) used a nonlinear mixed models in prostate cancer survival data (10). A similar method was applied by Song et al. (2012) in joint models of skewed longitudinal data and cure-rate survival (11). Baghfalaki et al. (2013) proposed using normal/independent distribution for random effects (12). Tang et al (2014) applied a Bayesian semi-parametric joint model capable of modeling multivariate longitudinal responses with a greater flexibility for time (13). Using t, skew-t and scale-mixture t distributions was the principal point in some studies dealing with non-normal longitudinal outcomes in joint models (14-16). Some other distributions such as skew-normal and skew-slash are specific forms of broader family of normal/independent distributions appropriate for heavy-tailed responses (17). All these distributions lie in R. However, in many applications, they are also used for non-negative outcomes. In addition, the use of distribution with positive values such as log-normal is also problematic in data with outliers (18). All these models are specific forms of scale-mixture normal distribution (19) and could be linked to normal distribution via equation

$$Y = \mu + \sqrt{g(U)}X, \qquad [1.1]$$

where $X \sim N(0, \sigma^2)$ and U is a positive random variable, independent of X and g is a positive function. Using $g(U) = 1/U$ yields normal/independent distribution.

Birnbaum and Saunders (1969) proposed a (BS) distribution for fatigue data. The distribution is capable of modeling data with various degrees of kurtosis and positive skewness defined on positive values (20). The density function of BS distribution is given by

$$f_Y(y; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2}\left[\frac{y}{\beta} + \frac{\beta}{y} - 2\right]\right) \frac{[y+\beta]}{2\alpha\sqrt{\beta y^3}}, \quad y > 0, \alpha > 0, \beta > 0; \qquad [1.2]$$

Rieck and Nedelman (1991) used Birnbaum-Saunders (BS) distribution in regression setting by applying a log-linear transformation which results in sinh-normal (SHN) distribution (21). Suppose W follows a SHN distribution. The density function for SHN distribution with shape, location and scale parameters α, μ and σ is given by

$$f_W(w) = \frac{2}{\alpha\sigma\sqrt{2\pi}} \cosh\left(\frac{w-\mu}{\sigma}\right) exp\left\{-\frac{2}{\alpha^2} sinh^2\left(\frac{w-\mu}{\sigma}\right)\right\};\ w \in R, \alpha > 0, \mu \in R, \sigma > 0. \qquad [1.3]$$

If $Y \sim BS(\alpha, \beta)$ then $W = log(Y) \sim SHN(\alpha, log(\beta), 2)$.

The log-transformation technique is widely used in regression models for analyzing skew data arising in various fields including survival analysis, insurance data, air pollution and engineering (22-24). Many other distributions have been developed based on BS distribution. For extensions on BS distribution see (18, 22, 23, 25-27).

In this study, we propose using BS distribution in Bayesian framework for joint longitudinal-survival modeling with skew longitudinal data. We illustrate the model on a publicly available data and compare model fit to commonly used normal distribution.

In section 2, simple longitudinal submodel is explained. Joint longitudinal-survival models with normal and BS distributions in longitudinal part along with details of implementation and model selection are demonstrated in section 3-5. In sections 6, an example on real data is described and the results of joint models are compared. A discussion and possible areas of future work are included in section 7.

Longitudinal Model

Linear mixed models are among popular approaches for analysis of longitudinal data. We used following mixed model used in longitudinal part of joint models by Henderson et al (2000) where association between longitudinal and survival parts are constructed by a mean-zero bivariate normal distribution (4). Suppose y_{ij} is the longitudinal response for i-th subject measured at j-th measurement and $\boldsymbol{x}_i$ is the vector of associated covariates with coefficient vector $\boldsymbol{\beta}$. For the simplicity of notation we illustrate the model for time-invariant covariates. Then, the longitudinal process could be written as

$$y_{ij} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{b}_i + \varepsilon_{ij} = \boldsymbol{x}_i'\boldsymbol{\beta} + b_{0i} + b_{1i} \times t_{ij} + \varepsilon_{ij} \qquad [2.1]$$

where $\boldsymbol{w}_i = \begin{bmatrix} 1 & t_{ij} \end{bmatrix}'$ is the design matrix for subject *i* and $\boldsymbol{b}_i = [b_{0i} \quad b_{1i}]'$ is the associated random effects vector. Here, ε_{ij} indicates error term that is usually assumed to follow normal distribution. We will substitute this with Birnbaum-Saunders distribution in multiplicative model that is usually converted to sinh-normal (SHN) distribution in log-transformed additive model as described in the next section.

Joint Models

In the remainder of the article, we will use $\boldsymbol{x}_1$and $\boldsymbol{x}_2$ with corresponding coefficient vector $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ to represent covariates in longitudinal and survival submodels, respectively, that could have elements in common or not. Suppose a set of *n* subjects are followed over a time interval [0, T) and repeated measurements $\{y_{ij}, j = 1, \ldots, n_i\}$ were recorded, with possibly partly missing, for subject *i*, *i*=1, …, *n* at *j*-th measurement. For each subject, observed survival time, S_i, and censorship status are also recorded.

We assumed linear mixed effects model for repeatedly measured response, y_{ij}, as below

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{z}_{ij}'\boldsymbol{b}_i = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + b_{0i} + b_{1i} \times t_{ij}. \qquad [3.1]$$

Here, $\boldsymbol{x}_{1i}$ is the corresponding covariate vector for subject *i* at *j*-th measurement with coefficient vector$\boldsymbol{\beta}_1$, t_{ij} is the time for subject *i* on *j*-th measurement, $\boldsymbol{b}_i = [b_{0i} \quad b_{1i}]'$ is the associated random effects vector and ε_{ij} is the error term.

Survival time is modeled using Weibull distribution that is flexible and could cover different hazard shapes.

$$T_i \sim Weibull(\nu, \lambda_i),$$

$$h_i(t) = \nu\lambda_i t_i^{\nu-1} = \nu t_i^{\nu-1} exp(\boldsymbol{x}_{2i}'\boldsymbol{\beta}_2), \qquad [3,2]$$

T_i and $h_i(.)$ are the time to event and hazard function for the i-th subject, and $\nu > 0$ is the shape parameter. In the survival submodel, covariate vector for the i-th subject at time t, $\boldsymbol{x}_{2i}$, is introduced to the model via the subject specific rate parameter $\lambda_i = exp(\boldsymbol{x}_{2it}'\boldsymbol{\beta}_2)$ with corresponding coefficient vector $\boldsymbol{\beta}_2$. The covariates could be time invariant or time dependent. We also used Exponential distribution (i.e., $Weibull(1, \lambda_i)$) for error term to compare proposed model in simpler settings. Following Henderson et al. (2000), the association between longitudinal and survival process is introduced by using shared random effects (4). For joint model with normally distributed longitudinal process, we have

$$y_{ij} = N(\eta_{ij}, \sigma_\varepsilon^2), \quad [3.3]$$

or equivalently$\varepsilon\varepsilon$

$$y_{ij} = \eta_{ij} + \varepsilon_{ij},$$
$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2),$$

$$\eta_{ij} = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{b}_i = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + b_{0i} + b_{1i} \times t. \quad [3.4]$$

Here, $\boldsymbol{z}_{ij} = \begin{bmatrix} 1 & t_{ij} \end{bmatrix}'$ is the subject specific covariate random effects with $\boldsymbol{b_i} = [b_{0i} \quad b_{1i}]'$ as coefficients for the subject i. The two submodels are joined using shared-parameter below.

$$T_i \sim Weibull(\nu, \lambda_i),$$
$$\lambda_i(t) = exp(\boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + \boldsymbol{\gamma}_i'\boldsymbol{b}_i) = exp(\boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + \gamma_0 \times b_{0i} + \gamma_1 \times b_{1i}). \quad [3.5]$$

The vector $\boldsymbol{\gamma} = [\gamma_0 \quad \gamma_1]'$ is the joining parameter where the first element reflects the effect of the initial values of the longitudinal response on the hazard and the second, describes the effect of longitudinal trends of the response on the hazard.

Similarly, for the joint model with BS distribution, longitudinal process could be written as

$$y_{ij} \sim BS(\alpha, \delta_{ij}),$$

or

$$y_{ij} = \delta_{ij}^* \varepsilon_{ij}^*,$$

$$\varepsilon_{ij}^* \sim BS(\alpha, 1),$$

$$\delta_{ij}^* = exp(\boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{b}_i) = exp(\boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + b_{0i} + b_{1i} \times t_{ij}). \quad [3.6]$$

As for normal joint model, joining for BS model is through sharing vector $\boldsymbol{\gamma} = [\gamma_0 \quad \gamma_1]'$,

$$T_i \sim Weibull(\nu, \lambda_i),$$

$$\lambda_i = exp(\boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + \boldsymbol{\gamma}_i'\boldsymbol{b}_i) = exp(\boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + \gamma_0 \times b_{0i} + \gamma_1 \times b_{1i}). \quad [3.7]$$

To fit longitudinal BS model as in [3.4], using log transformation in [3.6] gives additive model below

$$log(y_{ij}) \sim SHN(\alpha, \delta_{ij}, 2),$$

or equivalently

$$log(y_{ij}) = \delta_{ij} + \varepsilon_{ij}^{**},$$

$$\delta_{ij} = \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{z}'_i\boldsymbol{b}_i = \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1 + b_{0i} + b_{1i} \times t_{ij}$$

$$\varepsilon_{ij}^{**} \sim SHN(\alpha, 0, 2). \qquad [3.8]$$

In both normal and BS (i.e., log-BS) joint models, random effects are modeled as

$$\boldsymbol{b}_i = [b_{0i} \quad b_{1i}]' \sim iid \; N_2(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

4. Likelihood and Priors

In this section, subscripts N and BS represent elements for Normal and BS models, respectively. Suppose D is the observations and θ represents vector of parameters. Then likelihood for normal model could be written as

$$L(\boldsymbol{\theta}_N|\boldsymbol{D}) \propto \left\{\prod_{i=1}^{n} \phi\left(y_{it}; \boldsymbol{x}'_{1i}\boldsymbol{\beta}_{1,N} + \boldsymbol{w}'_i\boldsymbol{b}_i, \sigma_\varepsilon^2\right) \times \pi\left(\boldsymbol{\beta}_{1,N}\right) \times \pi(\boldsymbol{b}_i) \times \pi(\sigma_\varepsilon^2)\right\}$$

$$\times \left\{[S_N(s_i)]^{1-\varpi_i} \times \left[h_N\left(\{s_i|\nu, \lambda_{i,N}\}\right)\right]^{\varpi_i} \times \pi\left(\boldsymbol{\beta}_{2,N}\right) \times \pi(\nu) \times \pi(\boldsymbol{\gamma}_N)\right\}, \qquad [4.1]$$

where ϖ indicates censorship; 1= censored; 0= event. The density and priors could be represented in hierarchical form as

$$y_{it}|\boldsymbol{x}'_{1i}, \boldsymbol{\beta}_{1,N}, \boldsymbol{b}_i, \sigma_\varepsilon^2 \sim N\left(\boldsymbol{x}'_{1i}\,\boldsymbol{\beta}_{1,N} + \boldsymbol{w}'_i\boldsymbol{b}_i, \sigma_\varepsilon^2\right);$$

$$\boldsymbol{b}_i|\boldsymbol{\Sigma}_0 \sim N_2(\boldsymbol{0}, \boldsymbol{\Sigma}_0),$$

$$\boldsymbol{\beta}_{1,N}|\boldsymbol{\Sigma}_{1,N} \sim N_p\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{1,N}\right),$$

$$\sigma_\varepsilon^2 \sim gamma(1,1). \qquad [4.2]$$

for normal longitudinal part that is linked to survival part with Weibull distribution as

$$h_{i,N}\left(s|\nu, \lambda_{is,N}\right) = \nu s^{\nu-1}\lambda_{i,N},$$

$$\lambda_{i,N} = exp\left(\boldsymbol{x}'_{2i}\,\boldsymbol{\beta}_{2,N} + \boldsymbol{\gamma}_N\boldsymbol{b}_i\right),$$

$$\boldsymbol{\beta}_{2,N}|\boldsymbol{\Sigma}_{2,N} \sim N_p\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{2,N}\right),$$

$$\boldsymbol{\gamma}_N|\boldsymbol{\Sigma}_{\boldsymbol{\gamma},N} \sim N_2\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma},N}\right). \qquad [4.3]$$

That contains sharing parameters vector $\boldsymbol{\gamma} = [\gamma_0 \quad \gamma_1]'$ that we assumed to follow bivariate normal distribution with diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma},N}$.

Similarly, for BS distribution (transformed to SHN distribution) we could construct the likelihood as below

$$L(\boldsymbol{\theta}_{BS}|\boldsymbol{D}) \propto \left\{\prod_{i=1}^{n} f_{SHN}\left(y_{it}; \alpha, \boldsymbol{x}'_{1i}\boldsymbol{\beta}_{1,BS} + \boldsymbol{w}'_i\boldsymbol{b}_i, 2\right) \times \pi\left(\boldsymbol{\beta}_{2,BS}\right) \times \pi(\boldsymbol{b}_i) \times \pi(\alpha)\right\}$$

$$\times \left\{[S_{BS}(s_i)]^{1-\varpi_i} \times \left[h_{BS}\left(\{s_i|\nu, \lambda_{i,BS}\}\right)\right]^{\varpi_i} \times \pi\left(\boldsymbol{\beta}_{2,BS}\right) \times \pi(\nu) \times \pi(\boldsymbol{\gamma}_{BS})\right\}, \qquad [4.4]$$

with hierarchical density and priors as

$$y_{ij}|\boldsymbol{x}'_{1it}, \boldsymbol{\beta}_{1,BS}, \boldsymbol{b}_i \sim SHN\left(\alpha, \boldsymbol{x}'_{1i}\,\boldsymbol{\beta}_{1,BS} + \boldsymbol{w}'_i\boldsymbol{b}_i, 2\right);$$

$$\boldsymbol{b}_i|\Sigma_0 \sim N_2(\boldsymbol{0}, \boldsymbol{\Sigma}_0),$$

$$\boldsymbol{\beta}_{1,BS}|\boldsymbol{\Sigma}_{1,BS} \sim N_p\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{1,BS}\right),$$

$$\alpha \sim gamma(1,1), \qquad [4.5]$$

where α > 0 is the shape parameter that was assumed to follow gamma distribution. This longitudinal part is linked to survival part in a similar way to normal model with density and priors as

$$h_{i,BS}\left(s|\nu, \lambda_{is,BS}\right) = \nu s^{\nu-1}\lambda_{i,BS},$$

$$\lambda_{is,BS} = exp\left(\boldsymbol{x}'_{2is}\,\boldsymbol{\beta}_{2,BS} + \boldsymbol{\gamma}_{BS}\boldsymbol{b}_i\right),$$

$$\boldsymbol{\beta}_{2,BS}|\boldsymbol{\Sigma}_{2,BS} \sim N_p\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{2,BS}\right),$$

$$\boldsymbol{\gamma}_{BS}|\boldsymbol{\Sigma}_{\boldsymbol{\gamma},BS} \sim N_2\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma},BS}\right). \qquad [4.6]$$

Implementation and model selection

The models were fitted by using Hamiltonian Monte Carlo (HMC) method in Stan software (mc-stan.org) that avoids issues that arise in Gibbs sampling (28). Stan is a better choice in fitting complex models with complicated posteriors that usually involve high correlation among parameters. Stan is capable of handling non-conjugate priors used to obtain closed forms of posteriors and, actually, there is no value in using such priors. Using gamma priors for the variance component, instead of commonly used inverse-gamma distribution, is a result of such flexibility. Implementation of complex models in Stan is much easier as computational issues may arise in software such as BUGS (29).

Deviance information criteria (DIC), proposed by Spiegelhalter et al. (2002), is the widely used tool for model comparison in Bayesian setting. It can be easily calculated in BUGS software and this is the main reason for its popularity. However, it is not a fully Bayesian criterion and is inefficient in many situations (30-32). For these reasons, we used two other criteria for model selection. The first one is Conditional Predictive Ordinates (CPO) originally proposed by Geisser (1980) and developed into Bayesian settings by Gelfand et al. (1992) (33, 34). The

CPO is calculated based on predictability of each observation from others as follows

$$CPO_i = f\left(y_i \middle| \boldsymbol{D}^{(-i)}\right) = \int f(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)\left(\boldsymbol{\beta} \middle| \boldsymbol{D}^{(-i)}\right) d\boldsymbol{\beta}$$

The sum of log(CPO) over all observations gives log-pseudo maximum likelihood (LPML) criterion. The larger the calculated LPML, the better the model fit is. LPML is robust against improper priors and is computationally stable (35).

The second criterion we used for model comparison is the Widely Applicable Information Criterion (WAIC) proposed by Watanabe (2010) (36). Despite DIC that uses conditioning on a single point from posterior distribution (i.e., the mean), calculations in WAIC are based on the whole samples generated from posterior distribution. WAIC has more Bayesian-theoretical background than any other criterion and is probably the best goodness-of-fit criterion in Bayesian models (30, 37). Smaller values of WAIC indicate better fit.

Analysis of ddI/ddC Data

To illustrate our model, we used ddI/ddC data from a trial that has been previously analyzed using joint models assuming normal distribution for longitudinal response (29, 38). The trial aimed to compare the efficacy and safety of two antiretroviral drugs in the treatment of patients who had failed or were intolerant of zidovudine (AZT) therapy. In brief, 467 HIV-infected patients eligible according to pre-specified criteria were randomized into two groups to receive either didanosine (ddI) or zalcitabine (ddC). The longitudinal outcome, CD4 counts, was measured at study entry the four next visits with 2 months interval. Guo and Carlin (2004) used a square root transformation on CD4 counts to apply joint model with normal error term (29). Here, we sue same transformation to provide a comparative basis. Assessing the boxplot of CD4 counts for all observations and in two groups over time (Figure 1) suggests fairly positive skewness in all visit times. As it is evident, transformation is not successful in normalizing data and using distributions capable of handling skew data could provide better fit.

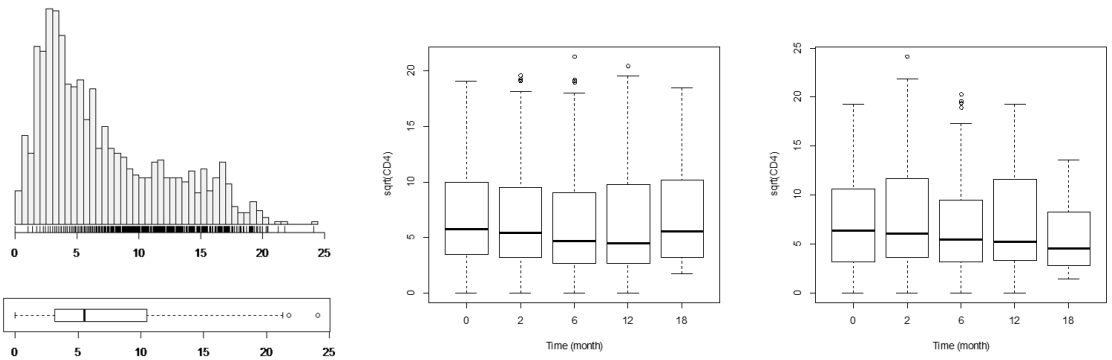

Figure 1. Square root of CD4 cell count for all aptients over all visit times (the left) and ddC group (the middle) and ddI group (the right) in each visit.

Here we consider the square root of the j-th CD4 count measurement on the i-th patient as the longitudinal outcome, $y_{ij}, j = 1, \ldots, n_i$, and i = 1, . . . , n. Main effects of four following binary explanatory variables were included in the model: Drug (1= ddI, 0= ddC), Gender (1=

male, −1= female), PrevOI (1= AIDS diagnosis at study entry, -1= no AIDS diagnosis), and Stratum (1= AZT failure, −1= AZT intolerance). The aim of study is to analyze the association between CD4 cell count and survival time with the four variables as covariates by allowing for subject-specific random effects.

Table 1. Bayesian joint analysis for ddI/ddC data.

	Exponential for survival				Weibull for survival			
	Joint normal model		Joint BS model		Joint normal model		Joint BS model	
	Estimate	95% CrI	Estimate	95% CrI	Estimate	95% CrI	Estimate	95% CrI
Parameter	Longitudinal submodel		Longitudinal submodel		Longitudinal submodel		Longitudinal submodel	
α	–		0.54	(0.51, 0.57)	–		0.54	(0.51, 0.57)
σ_{ε}^2	1.77	(1.67, 1.87)	–		1.77	(1.67, 1.87)	–	
Intercept	8.65	(8.02, 9.31)	1.87	(1.70, 2.02)	8.65	(8.03, 9.30)	1.86	(1.71, 2.01)
Time	-0.64	(-0.78, -0.49)	-0.14	(-0.19,- 0.08)	-0.64	(-0.79,- 0.49)	-0.13	(-0.19, -0.08)
Time×Drug	0.12	(-0.05, 0.32)	0.02	(-0.03,0.09)	0.12	(0.05, 0.32)	0.02	(-0.02, 0.09)
Gender	-0.09	(-0.67, 0.42)	0.03	(0.09, 0.17)	-0.10	(-0.69, 0.40)	0.04	(-0.08, 0.17)
PrevOI	-2.29	(-2.73, -1.83)	-0.35	(-0.45,- 0.24)	-2.29	(-2.75,-1.83)	-0.35	(-0.45, -0.25)
Stratum	-0.11	(-0.56, 0.27)	-0.07	(-0.17, 0.02)	-0.11	(-0.55, 0.26)	-0.07	(-0.18, 0.01)
	Survival submodel		Survival submodel		Survival submodel		Survival submodel	
ν	–		–		1.43	(1.25,1.62)	1.41	(1.25, 1.60)
Intercept	-3.75	(-4.08, 3.43)	-3.73	(-4.05,-3.42)	-4.90	(-5.52, -4.33)	-4.80	(-5.44, -4.28)
Drug	0.18	(-0.07, 0.49)	0.18	(-0.06, 0.47)	0.20	(-0.06, 0.50)	0.17	(-0.05, 0.48)
Gender	-0.15	(-0.40, 0.07)	-0.14	(-0.38, 0.07)	-0.17	(-0.42, 0.06)	-0.18	(-0.49, 0.06)
PrevOI	0.64	(0.42, 0.42)	0.62	(0.41, 0.85)	0.68	(0.46, 0.90)	0.66	(0.45, 0.87)
Stratum	0.06	(-0.07, 0.22)	0.07	(-0.06, 0.23)	0.06	(-0.08, 0.23)	0.07	(-0.05, 0.24)
γ_1	-0.02	(-0.06, 0.00)	-0.09	(-0.30, 0.11)	-0.02	(-0.06, 0.00)	-0.13	(-0.34, 0.08)
γ_2	-0.40	(-0.77, -0.05)	-0.42	(-0.99, 0.17)	-0.46	(-0.88, -0.08)	-0.50	(-1.03, 0.13)
WAIC	9327.9		5905.0		8779.7		5421.4	
LPML	-9853.8		-6432.8		-9225.4		-5814.5	

BS = Birnbaum-Saunders; WAIC = widely applicable information criterion; LPML = log-pseudo marginal likelihood; CrI = credible interval.

We used Weibull and Exponential distributions in survival part. The results of joint models using normal and Birnbaum-Saunders (SHN in transformed model) distributions are shown in Table 1. The estimated shape parameter for Weibull distribution is 1.4 in both normal and BS models. Hence the results of Weibull model is essentially same as Exponential model. The estimates in survival part are very similar as well. However, differences in estimated coefficients in longitudinal part are larger, albeit with similar significance. Using BS distribution did not alter the significance of covariates in this example. However, fit of models are different. Comparing WAIC and LPML criteria indicates better fit for Weibull distribution, as expected. Both criteria suggest better fit for joint model with BS distribution.

Discussion and future work

There have been remarkable research on joint modeling of longitudinal and time-to-event data in last two decades. Various approaches have been proposed and their strength and limitations have been discussed in the literature. All these have assumptions on the distribution and the way two processes are linked. The longitudinal part has attracted more attention in recent years and a wide range of distributions have been applied instead of traditionally used normal distribution. We proposed using Birnbaum-Saunders (BS) distribution for skew-positive data. In real dataset, BS outperformed normal distribution. The proposed model could also be modified to contain BS distribution in both submodels. There are various developments on BS distribution such as extended-BS and generalized-BS. These could have better fit in certain data. If the distribution is bimodal, using BS distribution with shape parameter of >2 would be a nice choice. If the non-normality is due to the kurtosis, BS distribution could be among the distributions with improved fit. Our aim was to introduce BS distribution in joint modeling of univariate skew data. However, it could be extended to joint modeling of multivariate and cure-rate survival data. Non-normality is common in many applications and inferences should be adjusted accordingly. BS distribution could be a preferable candidate in dealing with various violations from normality

Acknowledgment

The authors wish to thank Marjan Mansourian, PhD., for her helpful comments.

Conflict of Interest

The authors have declared no conflict of interest.

References

1.Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Longitudinal data analysis. Boca Raton, NW: CRC Press; 2008.
2.Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. Hoboken, New Jersey: John Wiley & Sons; 2012.
3.Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. Biomet. 2006;62(2):432-45.
4.Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000;1(4):465-80.
5.Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. Journal of Clinical Oncology,. 2010;28(16):2796-801.
6.Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biomet. 1997:330-9.
7.Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Statistics in Medicine,. 1996;15(15):1663-85.
8.Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat Med. 2011;30(12):1366-80.
9.Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biomet. 2003;59(2):221-8.
10.Yu M, Law NJ, Taylor JM, Sandler HM. Joint longitudinal-survival-cure models and their application to prostate cancer. Statistica Sinica. 2004;14(3):835-62.
11.Song H, Peng Y, Tu D. A new approach for joint modelling of longitudinal measurements and survival times with a cure fraction. Canadian Journal of Statistics. 2012;40(2):207-24.
12.Baghfalaki T, Ganjali M, Berridge D. Robust joint modeling of longitudinal measurements and time to event data using normal/independent distributions: A Bayesian approach. Biometrical Journal. 2013;55(6):844-65.
13.Tang N-S, Tang A-M, Pan D-D. Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computional Statisitics and Data Analysis. 2014;77:113–29.
14.Chen J, Huang Y. A Bayesian mixture of semiparametric mixed-effects joint models for skewed-longitudinal and time-to-event data. Statistics in Medicine. 2015;34:2820-43.
15.Huang Y, Yan C, Xing D, Zhang N, Chen H. Jointly Modeling Event Time and Skewed-Longitudinal Data with Missing Response and Mismeasured Covariate for AIDS Studies. Journal of Biopharmaceutical Statistics,. 2015;25:670–94.
16.Baghfalaki T, Ganjali M, Hashemi R. Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. Journal of Biopharmaceutical Statistics. 2014;24(4):834-55.
17.Baghfalaki T, Ganjali M. A Bayesian Approach for Joint Modeling of Skew-Normal Longitudinal Measurements and Time to Event Data. REVSTAT–Statistical Journal. 2015;13(2):169-91.
18.Leiva V, Barros M, Paula GA, Sanhueza A. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. Environmet. 2008;19(3):235-49.
19.Balakrishnan N, Leiva V, Sanhueza A, Labra FEV. Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. SORT-Statistics and Operations Research Transactions. 2009;33(2):171-92.
20.Birnbaum Z, Saunders SC. A new family of life distributions. Journal of Applied Probability. 1969:319-27.
21.Rieck JR, Nedelman JR. A Log-Linear Model for the Birnbaum-Saunders Distribution. Technometrics. 1991;33(1):51-60.
22.Leiva V, Vilca F, Balakrishnan N, Sanhueza A. A skewed sinh-normal distribution and its properties and application to air pollution. Communications in Statistics-Theory and Methods. 2010;39(3):426-43.
23.Vilca F, Sanhueza A, Leiva V, Christakos G. An extended Birnbaum–Saunders model and its application in the study of environmental quality in Santiago, Chile. Stochastic Environmental Research and Risk Assessment. 2010;24(5):771-82.

24.Paula GA, Leiva V, Barros M, Liu S. Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance. Applied Stochastic Models in Business and Industry. 2012;28(1):16-34.
25.Lemonte AJ. A log-Birnbaum-Saunders regression model with asymmetric errors. Journal of Statistical Computation and Simulation. 2012;82(12):1775-87.
26.Sanhueza A, Leiva V, Balakrishnan N. The generalized Birnbaum–Saunders distribution and its theory, methodology, and application. Communications in Statistics-Theory and Methods. 2008;37(5):645-70.
27.Leiva V. The Birnbaum-Saunders Distribution: Academic Press; 2015.
28.Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, Version 2.12.0. 2016.
29.Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. Americal Statistician. 2004;58(1):16-24.
30.Spiegelhalter DJ, Best NG, Carlin BP, Linde A. The deviance information criterion: 12 years on. Journal of Royal Statistical Society, Series B. 2014;76(3):485-93.
31.Plummer M. Penalized loss functions for Bayesian model comparison. Biostatistics. 2008;9:523–39.
32.van der Linde A. DIC in variable selection. Statistica Neerlandica. 2005;1:45–56.
33.Geisser S. IN DISCUSSION OF Box, George EP. Sampling and Bayes' inference in scientific modelling and robustness. Journal of Royal Statistical Society, Series A. 1980;143(4):383-430.
34.Gelfand AE, Dey DK, Chang H. Model determination using predictive distributions with implementation via sampling-based methods. DTIC Document, 1992.
35.Ibrahim JG, Chen M-H, Sinha D. Bayesian Survival Analysis: Springer-Verlag New York; 2001.
36.Watanabe S. Asymptotic equivalence of Bayes cross validation and widely application information criterion in singular learning theory. Journal of Machine Learning Research. 2010;11:3571-94.
37.Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis: Chapman & Hall/CRC Boca Raton, FL, USA; 2014.
38.Goldman A, Carlin B, Crane L, Launer C, Korvick J, Deyton L, et al. Response of CD4+ and Clinical Consequences to Treatment Using ddI or ddC in Patients with Advanced HIV Infection. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology. 1996;11:161–9.