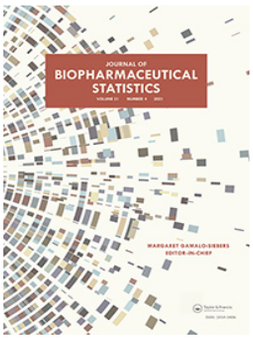

Journal of Biopharmaceutical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lbps20

A two-stage approach for joint modeling of longitudinal measurements and competing risks data

P. Mehdizadeh, Taban Baghfalaki, M. Esmailian & M. Ganjali

To cite this article: P. Mehdizadeh, Taban Baghfalaki, M. Esmailian & M. Ganjali (2021) A two-stage approach for joint modeling of longitudinal measurements and competing risks data, Journal of Biopharmaceutical Statistics, 31:4, 448-468, DOI: 10.1080/10543406.2021.1918142

To link to this article: https://doi.org/10.1080/10543406.2021.1918142

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2021, VOL. 31, NO. 4, 448–468
https://doi.org/10.1080/10543406.2021.1918142

Taylor & Francis
Taylor & Francis Group

ARTICLE

A two-stage approach for joint modeling of longitudinal measurements and competing risks data

P. Mehdizadeh[a], Taban Baghfalaki[b], M. Esmailian[a], and M. Ganjali[c]

[a]Department of Statistics and Computer Sciences, Faculty of Sciences, University of Mohaghegh Ardabili, Ardabil, Iran; [b]Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran; [c]Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

ABSTRACT

Joint modeling of longitudinal measurements and time-to-event data is used in many practical studies of medical sciences. Most of the time, particularly in clinical studies and health inquiry, there are more than one event and they compete for failing an individual. In this situation, assessing the competing risk failure time is important. In most cases, implementation of joint modeling involves complex calculations. Therefore, we propose a two-stage method for joint modeling of longitudinal measurements and competing risks (JMLC) data based on the full likelihood approach via the conditional EM (CEM) algorithm. In the first stage, a linear mixed effect model is used to estimate the parameters of the longitudinal sub-model. In the second stage, we consider a cause-specific sub-model to construct competing risks data and describe an approximation for the joint log-likelihood that uses the estimated parameters of the first stage. We express the results of a simulation study and perform this method on the "standard and new antiepileptic drugs" trial to check the effect of drug assaying on the treatment effects of lamotrigine and carbamazepine through treatment failure.

ARTICLE HISTORY
Received 3 October 2020
Accepted 28 March 2021

KEYWORDS
The conditional expected maximization algorithm; competing risks data; joint modeling; longitudinal measurements; two-stage approach

1. Introduction

In analyzing follow-up data, we are faced with a lot of issues that deal with joint modeling of longitudinal and time-to-event data. In recent years, a lot of works have been done in this regard. The standard joint model includes two sub-models; a longitudinal sub-model and a time-to-event sub-model. These two sub-models are related through an association structure that measures the relationship between the favorite outcomes in the study.

Typically, a linear-mixed effects sub-model for the longitudinal process is used for analyzing longitudinal outcome and a Cox proportional hazard sub-model Cox (1972) is considered for the time-to-event process. Wulfsohn and Tsiatis (1997), Ye et al. (2008) and Rizopoulos (2012b) considered different association structures for the longitudinal and time-to-event data. Guler et al. (2017) also described these structures in their article.

In a joint modeling framework, a great many publications utilize the full likelihood approach, based on the shared random effects models, to estimate the parameters in these models.

For example, we can refer to Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Henderson et al. (2000), Tsiatis and Davidian (2004) and Rizopoulos (2012b). Comprehensive reviews are given by Hogan and Laird (1997), Tsiatis and Davidian (2004), Diggle et al. (2008), Proust- Lima et al. (2012) and Gould et al. (2015). Sweeting and Thompson (2011) and Gould et al. (2015) conclude that the full likelihood approach is an effective procedure to study the relationship between longitudinal and time-

CONTACT Taban Baghfalaki t.baghfalaki@modares.ac.ir Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran

to-event data. Also, Brown and Ibrahim (2003), Ibrahim et al. (2004), Huang, Dagne, Wu, (2011), Baghfalaki et al. (2013), Baghfalaki, Ganjali, Berridge (2014a), Baghfalaki, Ganjali, Hashemi (2014b) and Baghfalaki and Ganjali (2015) explore joint models based on the Bayesian approach.

In order to implement the full likelihood approach, both the classical and the Bayesian paradigms are used in the literature. Although the use of the Bayesian paradigm partially simplifies computations, the existence of multi-integrals in the joint log-likelihood function and the survival function still calls for complex calculations. Therefore, a two-stage method is proposed which is performed in two stages. In the first stage, the longitudinal data are fitted and in the next stage the fitted values of the longitudinal process are used as covariates in the joint model to estimate the survival parameters. Therefore, the two-stage methods can solve the complex calculation problem in the full likelihood approaches. In a study, Wu et al. (2012) have provided a brief overview of joint models for longitudinal and survival data including the commonly used methods, the likelihood and two-stage methods. Murawska et al. (2012) proposed a two-stage joint model for nonlinear longitudinal response and a time-to-event model. Bellera et al. (2016) introduced a two-stage model in a Bayesian framework. Guler et al. (2017) investigated a two-stage model for multivariate longitudinal and survival data. Also, Huong et al. (2018) proposed a modified two-stage approach to reduce the biases in this approach. Most of the studies employing joint modeling framework focus on data with a single event time and only one failure cause. However, in medical research and some situations of interest, there are more than one possible cause of event or the censoring is informative. In such cases, the subjects of the study are at risk with more than one mutually exclusive event such as death from different causes, so that competing risks data arise spontaneously. The competing risks framework also includes settings where different possible events are not mutually exclusive but the interest lies in the first occurring event. Studies on joint modeling of longitudinal measurements and competing risks time-to-event data have grown in the past decade.

In what follows, we will refer to several papers in this regard. Huang et al. (2010) proposed a joint model for longitudinal measurements and competing risks survival data and developed a Bayesian MCMC procedure for parameter estimation and inference. Hickey et al. (2018) provided a comparison of joint models for longitudinal and competing risks data and summarized four published models. In fact they thoroughly review the literature on the implementation of joint models involving more than one event time per subject. Rajeswaran et al. (2018) proposed joint modeling of multivariate longitudinal data and competing risks, using multiphase sub-models. Baghfalaki et al. (2020) discussed a Bayesian joint modeling of ordinal longitudinal measurements and competing risks survival data for analyzing Tehran Lipid and Glucose Study.

In this paper, we propose a two-stage approach for joint modeling of longitudinal measurements and competing risks data. Because of its two-stage structure, the proposed approach facilities computational problems of shared random effects in joint models and makes it possible to use the standard packages of mixed-effects model and survival models in R software.

This paper is organized as follows: Section 2 describes the motivated data set which is analyzed in this paper. Section 3, provides a description of the sub-models and the joint modeling method and how it can be implemented. Section 4 explains how to estimate the parameters of the joint model in two stages and specifies the method. Some simulation studies are performed in Section 5. In Section 6, a real data set is analyzed by the proposed approach. The last section includes conclusions.

2. The motivating dataset

One dataset used here to demonstrate the issues in competing risks analysis is the "standard and new anti-epileptic drugs" (SANAD) study which was a non-blinded randomized-controlled trial enrolling patients with epilepsy to examine anti-epileptic drugs (AEDs). We could refer to Marson et al. (2007) in order to see the published design and analysis of this trial. Here, the withdrawal of a randomized drug is considered as the time for treatment failure that has been introduced by the International League Against Epilepsy to be one of the primary end points for the clinical trials of AEDs

(Commission on Antiepileptic Drugs, 2008). Patients may decide, due to inadequate seizure control (ISC), to switch to another AED or to begin an additional AED. Also, patients may withdraw from a treatment because of an unacceptable adverse effect (UAE). Overall analysis of treatment failure may miss differential effects of AEDs on the reasons for withdrawal, which may differ because of their relative importance for patients (Williamson et al. 2007). We use the results from a competing risks analysis of the data for pairwise lamotrigine (LTG) versus carbamazepine (CBZ).

This data set includes 605 patients whom; CBZ (n = 292) compares to LTG (n = 313). At first, Williamson et al. (2008) used a joint model framework to analyze this data set. 94 patients withdrew from the randomized drug because of UAE whereas 120 withdrew because of ISC within a maximum follow-up time of 6.6 years (median = 2.9 years). Withdrawals due to other reasons were considered as non-informative and patients were censored in these times. According to Faught (2007)'s recommendation for comparing two AEDs after regulating for titration rate, calibration is standardized at the first dose by the dose of both drugs relative to the midpoint of its maintenance dose range. The maintenance dose recommended in the SANAD trial was independently considered reasonable and the approach to calibration sensible. Therefore, these calibrated doses are taken to be the longitudinal measurements within the competing risks joint model. On average, 4.6 longitudinal measurements were recorded for patients. However, there are measurements between 1 and 15 records.

Figure 1 denotes the Kaplan-Meier estimate (with associated 95% confidence interval) of the survival function of two failure types ISC and UAE.

Figure 2 shows the longitudinal outlines for calibrated dose of drugs for some randomly selected patients over time, separated by drug type. This figure shows that LTG treatment receives higher values of calibrated dose than does CBZ treatment. Also, Figure A.1 in Supplementary Materials A is the spaghetti plots of calibrated dose values over time for all patients, separated by drug type; in fact, it

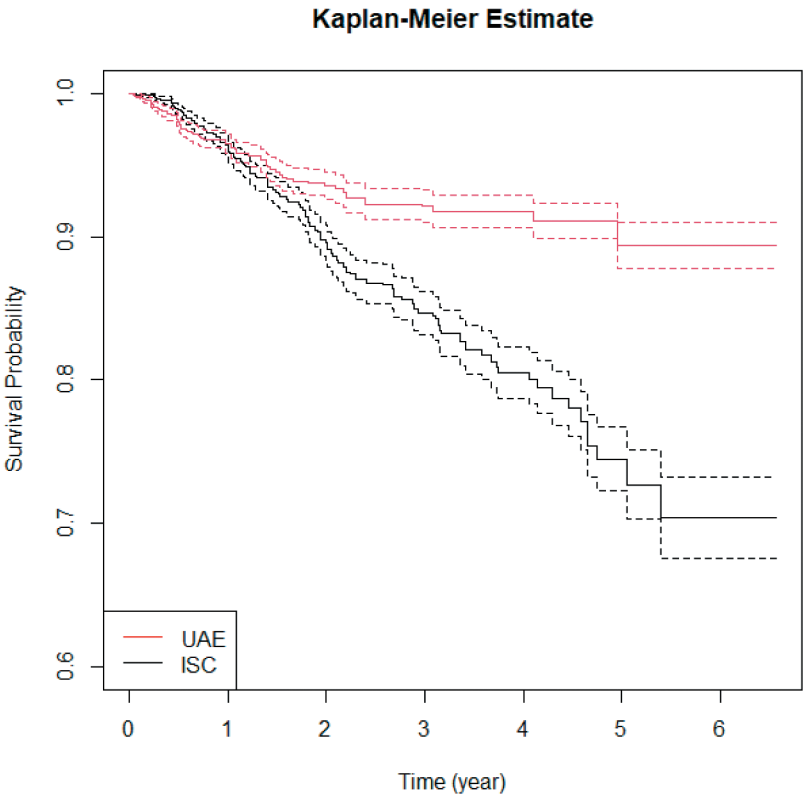

Figure 1. Kaplan-Meier estimate of the survival function for the failure types.

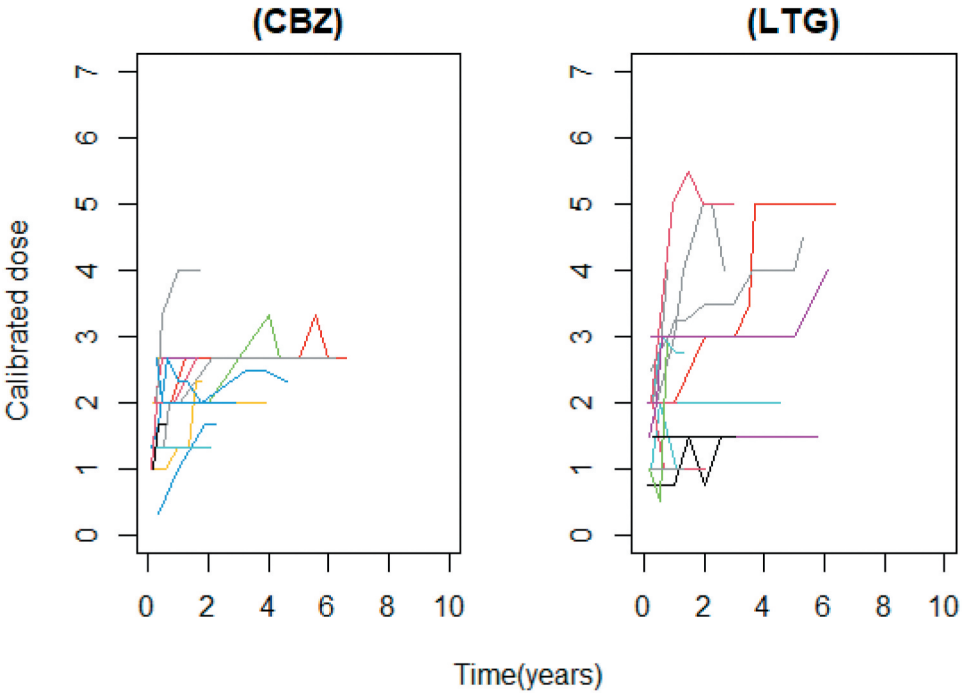

Figure 2. Spaghetti plots of calibrated dose values for 20 randomly selected patients over time separated by drug type.

is the same as Figure 2 drawn for all patients. This figure shows that the amount of the dose calibrated is close to midpoint of its maintenance dose range.

Figure A.2 in Supplementary Material A shows the spaghetti plots of calibrated dose values of drugs for all patients versus reverse time to treatment failure or censoring stratified by drug type and event time mechanism. It follows from Figure A.2 that there are a large difference in initial value of calibrated dose among patients. Furthermore, treatment failure due to ISC was observed in patients with increased calibrated dose, while treatment failure because of UAE was seen to occur in patients with decreased calibrated dose.

3. Modeling framework

Consider a longitudinal study with n individuals. These recorded longitudinal responses are measured at specific times s_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$; therefore $y_{ij} = y_i(s_{ij})$ denotes longitudinal measurement for the ith subject at time s_{ij}. Also, the observed failure time for the ith subject is the minimum of true survival time and the censoring time and it is indicated by $T_i = min(T_{i1}^*, T_{i2}^*, \ldots, T_{iK}^*, C_i)$ where, T_{ik}^* is the true survival time of subject i for each event type $k = 1, 2, \ldots, K$ and C_i is the randomly censoring time for the ith subject. Also, δ_i is defined as the event indicator, which takes value $\{0, 1, 2, \ldots, K\}$, with 0 corresponding to censoring and $1, 2, \ldots, K$ to the competing events. T_i is the censoring time $(\delta_i = 0)$ or the failure time $(\delta_i = 1, \ldots, K)$.

In order to construct a two-stage approach for joint modeling of longitudinal measurements and competing risks data, we consider two sub-models as described separately in the following. Also, association of these models is discussed.

3.1. Longitudinal sub-model

We assume a linear mixed effect sub-model to analyze longitudinal data. The longitudinal sub-model can be written as:

$$Y_i(s) = m_i(s) + \in_i(s), i = 1, 2, \ldots, n, \tag{1}$$

where, $m_i(s)$ is the mean response term for the ith subject and it is modeled as:

$$m_i(s) = X_i'(\mathbf{s})\beta + Z_i'(\mathbf{s})b_i, \tag{2}$$

and $\in_i(s) \overset{iid}{\sim} N(0, \sigma_{\in}^2)$ is the error term for the ith subject at time s. $X_i(s)$ is a $n_i \times p$ design matrix of the p observed explanatory variables and $Z_i(s)$ is a $n_i \times q$ design matrix of the q random effects for the ith subject. β is a p-dimensional vector of fixed effects and b_i is a q-dimensional vector of random effects such that $b_i \sim N_q(0, D)$, where D is a positive definite matrix.

3.2. Competing risks sub-model

We introduce a sub-model for competing risks data. The censorship function is considered to be independent of the time at which events are measured. So we assume the distribution of the competing risks failure time (T_i, δ_i) takes the form of the following cause-specific hazard frailty model:

$$h_{ik}(t|\mathcal{M}_i(t), w_i) = lim_{dt \to 0} P\{t \leq T_i < t + dt | T_i \geq t, \mathcal{M}_i(t), \delta_i = k, w_i\}/dt$$

$$= h_{0k}(t) exp\{\gamma_k' w_i + \alpha_k m_i(t)\}. \tag{3}$$

Here, $h_{0k}(t)$ is the baseline hazard function and w_i is the vector of covariates. $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ shows the history of the true unobserved longitudinal process up to time t. Also, γ_k is a p_k-dimensional vector of cause-specific regression coefficients for w_i and $\alpha = (\alpha_1, \ldots, \alpha_K)'$ is a K-dimensional vector such that α_k, $k = 1, \ldots, K$ is a coefficient for the shared time-varying covariate $m_i(t)$ that is defined in longitudinal sub-model and these two sub-models are related through it.

3.3. Joint model structure

It should be noted that, we can construct a joint model of longitudinal and time-to-event data for each event type because of the choice of cause-specified hazard for competing risks data. Therefore, to construct a joint model of longitudinal measurements and time-to-event data we need a full joint distribution of both processes. There are different factorization for this joint distribution leading into different approaches for modeling.

In the literature on joint modeling of longitudinal measurements and time-to-event data, it is common to use a shared parameter model. Usually, standard methods use a linear-mixed effects sub-model for the longitudinal process and a Cox proportional hazard sub-model (Cox 1972) for the time-to-event process.

Here, we consider a shared parameter model to join the longitudinal process to competing risks process in which the longitudinal process and the time-to-event process for the kth failure type are linked by the parameter α_k.

According to the two defined sub-models (1) and (2), the hazard rate in competing risks model depends on the true unobserved longitudinal measurements at time t. Based on the relationship between the hazard function and survival function, the vector of covariates $\mathcal{M}_i(t)$ affects both the survival and likelihood functions. Also, we suppose that the longitudinal sub-model and the survival process in the competing risks sub-model is related to the vector of random effects b_i; therefore, under these assumptions, we can write the likelihood function of the joint model.

Let $Y = (Y_1', \ldots, Y_n')'$, $Y_i = (Y_{i1}, \ldots, Y_{in_i})'$, $b = (b_1', \ldots, b_n')'$, $t = (t_1, \ldots, t_n)'$, $\delta = (\delta_1, \ldots, \delta_n)'$ and θ be the vector of all parameters in two sub-models; then given the random effects b_i, the elements of y_i and (t_i, δ_i) are independent and we have:

$$P(t_i, \delta_i, y_i|b_i; \theta) = \prod_{j=1}^{n_i} P(y_{ij}|b_i; \theta) P(t_i, \delta_i|b_i; \theta). \tag{4}$$

Therefore, the log-likelihood function of joint models based on the observed data can be written as follows:

$$l(\theta|\ y\ ,\ t\ ,\ \delta\ ) = \sum_i \log P(t_i, \delta_i, y_i; \theta)$$

$$= \sum_i \log \int_{b_i} P(t_i, \delta_i, y_i, b_i; \theta) db_i$$

$$= \sum_i \log \int_{b_i} P(t_i, \delta_i|b_i; \theta_t, \beta) P(y_i|b_i; \theta_y) P(b_i; \theta_b) db_i, \delta_i = 1, 2, \ldots, K, \tag{5}$$

where $\theta = \{\theta_y, \theta_t, \theta_b\}$ wherein $\theta_y = (\beta', \sigma^2_{\in})$ is the unknown parameters of longitudinal sub-model, $\theta_t = (\gamma_1', \ldots, \gamma_K', \alpha', \theta_{h_0}')$ is the vector of unknown parameters of competing risks sub-model and $\theta_b = vech(D)$ is the unknown parameters of the covariance matrix of random effects. Also,

$$P(y_i|\, b_i;\, \theta_y) = \prod_{j=1}^{n_i} P(y_{ij}|b_i, \theta_y),$$

and

$$P(t_i, \delta_i|b_i; \theta_t, \beta) = \prod_{k=1}^{K} [h_0(t_i) \exp\{\gamma_k' w_i + \alpha_k m_i(t_i)\}]^{I(\delta_i = k)}$$

$$\times \exp(-\sum_{k=1}^{K} \int_0^{t_i} h_0(s) \exp\{\gamma_k' w_i + \alpha_k m_i(s)\} ds).$$

To estimate the parameters of this joint model, we can write the score vector of observed data as

$$S(\theta) = \sum_i \frac{\partial}{\partial \theta'} \log \int P(t_i, \delta_i|b_i; \theta_t, \beta) P(y_i|b_i; \theta_y) P(b_i; \theta_b) db_i \tag{6}$$

$$= \sum_i \frac{\partial}{\partial \theta'} log\{P(t_i, \delta_i|b_i; \theta_t, \beta) P(y_i|b_i; \theta_y) P(b_i; \theta_b)\} P(b_i|t_i, \delta_i, y_i; \theta) db_i, \delta_i = 1, 2, \ldots, K.$$

To calculate the maximum likelihood estimation of parameters in (5), the EM algorithm can be used to obtain parameter estimation from the expected value of the complete data log-likelihood at the rth iteration of

$$Q(\theta|\theta^{(r)}) = \sum_i \int \log\{P(t_i, \delta_i, y_i, b_i; \theta)\} P(b_i|t_i, \delta_i, y_i; \theta^{(r)}) db_i \tag{7}$$

$$= \sum_i \int \{\log P(t_i, \delta_i | b_i; \theta) + \log P(y_i | b_i; \theta) + \log P(b_i; \theta)\} P(b_i | t_i, \delta_i, y_i; \theta^{(r)}) db_i.$$

Therefore, it can be seen that obtaining estimates require complex and long calculations due to multiple integrations in (7). Since multiple integrations in this equation do not have a closed form, Rizopoulos (2010) used Gaussian quadrature rules to approximate the values of integrations.

It is clear that the computations increase as the dimension of random effects and the number of quadrature points increase. Therefore, Rizopoulos (2012a) employed a pseudo-adaptive Gaussian quadrature rule to reduce the number of quadrature points and the running time. However, the algorithm is still very time-consuming when processing nonlinear longitudinal data in JM packages. Huong et al. (2018) proposed another approximation for parameter estimation and the expected value of the complete data log-likelihood function. We will use this approximation in this paper.

Let $\hat{\theta} = (\hat{\theta}_y{}', \hat{\theta}_t{}', \hat{\theta}_b{}')'$ be the estimation of the full joint model parameters in (3) and $\tilde{\theta} = (\tilde{\theta}_y{}', \tilde{\theta}_b{}')'$ be the estimator obtained from the linear mixed effect model in (1); then the expected function of the complete data log likelihood at $\hat{\theta}$ will have the form of

$$E[\log\{P(t, \delta, y, b; \hat{\theta})\}] \xrightarrow{P} \sum_i \int (\log P(t_i, \delta_i | b_i; \hat{\theta}_t, \hat{\theta}_y)$$

$$+ \log P(y_i | b_i; \hat{\theta}_b) + \log P(b_i; \hat{\theta}_b)) \times P(b_i; \tilde{b}_i, \tilde{H}_i^{-1}) db_i$$

$$\approx \sum_i \log P(t_i, \delta_i, \tilde{b}_i, \hat{\theta}_t, \tilde{\theta}_y) + \log P(y_i, \tilde{b}_i; \tilde{\theta}_y) + \log P(\tilde{b}_i; \tilde{\theta}_b), \tag{8}$$

where $\tilde{b}_i = \arg\max\{\log P(y_i, bt; \tilde{\theta}_y)\}$ and Hessian matrix, $\tilde{H}_i^{-1} = (-\partial \log P(y_i b, \tilde{\theta}_y))/(\partial b \partial b')|_{b=\tilde{b}_i}$. In the Supplementary Materials B, we express the theorem that leads to this approximation. More details and the proof of this theorem are provided by Huong et al. (2018).

4. Two-stage approach for JMLC

Based on the two sub-models defined earlier, we implement the two-stage method for joint modeling of longitudinal and competing risks data as follows:

In the first stage, we fit the linear-mixed effects model for longitudinal process as described in (1) and the coefficients of the fixed effects (β), the covariance matrix D and the predictor of the random effects b_i can be estimated. Therefore, based on the observed longitudinal measurements, we can write the fitted longitudinal sub-model of (1) as follows:

$$y_i(s_{ij}) = m_i(s_{ij}) + \in_i(s_{ij}) \tag{9}$$

$$= X_i'(s_{ij}) \overset{+}{\beta} Z_i'(s_{ij}) b_i + \in_i(s_{ij}),$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$.

In the second stage, we use the fitted values of parameters in the first step to estimate competing risks parameters, that is

$$h_{ik}(t) = h_{0k}(t) exp\{\gamma_k' w_i + \alpha_k \tilde{m}_i(t)\}, i = 1, \ldots, n, k = 1, \ldots, K, \tag{10}$$

where $\tilde{m}_i(t)$ is estimated in the first stage and is considered as a covariate in the sub-model of the competing risks data. According to the approximation (8), we obtain the estimation of parameters in the competing risks process for each failure type by maximizing the following approximation term:

$$\sum_i \log P(t_i, \delta_i, \tilde{b}_i, \hat{\theta}_t, \tilde{\theta}_y) + \log P(y_i, \tilde{b}_i; \tilde{\theta}_y) + \log P(\tilde{b}_i; \tilde{\theta}_b), \delta_i = 1, 2, \ldots K.$$

Here, the density function of event process is given by

$$P(t_i, \delta_i, \tilde{b}_i, \tilde{\theta}_y; \theta_t) = \prod_{k=1}^{K} h(t_i|\mathcal{M}_i(t_i), w_i, \tilde{\beta}; \theta_t)^{I(\delta_i=k)} S(t_i|\mathcal{M}_i(t_i), w_i, \tilde{\theta}_y; \theta_t) \tag{11}$$

$$= \prod_{K=1}^{K} [h_0(t_i) \exp\{\gamma'_k w_i + \alpha_k \tilde{m}_i(t_i)\}]^{I(\delta_i=k)}$$

$$\times \exp(-\sum_{k=1}^{K} \int_0^{t_i} h_0(s) \exp\{\gamma'_k w_i + \alpha_k \tilde{m}_i(s)\} ds).$$

Also, assuming the random effects, the density function of longitudinal data is as follows:

$$P(y_i|\tilde{b}_i; \tilde{\theta}_y) P(\tilde{b}_i; \tilde{\theta}_b) = \prod_{j=1}^{n_i} P\{y_i(s_{ij})|\tilde{b}_i; \tilde{\theta}_y\} P(\tilde{b}_i; \tilde{\theta}_b)$$

$$= \prod_{j=1}^{n_i} \frac{1}{(2\pi\tilde{\sigma}^2)^{(\frac{n_i}{2})}} \exp \frac{(-|y_i(s_{ij}) - X'_i(s_{ij})\tilde{\beta} + Z'_i(s_{ij})\tilde{b}_i||^2)}{2\tilde{\sigma}^2} \tag{12}$$

$$\times (2\pi)^{-(q_b/2)} det(\tilde{D})^{-1/2} \exp(-\tilde{b}'_i \tilde{D}^{-1} \tilde{b}_i/2).$$

4.1. Parameter estimation

Here, we present an algorithm to estimate the parameters of the joint modeling of longitudinal measurements and competing risks data based on a two-stage approach as follows:

• Stage I: By using a standard software of mixed-effects model, like the "lme" function, we estimate model (1) with all available observed longitudinal data and obtain $\tilde{\theta}_y$, $\tilde{\theta}_b$ and $\tilde{b}_i$.

• Stage II: In this stage, we use the conditional EM algorithm and estimate the parameters of the competing risks process based on the one-step Newton–Raphson method. We denote the vector of unknown parameters of the time-to-event process for each failure type as $\theta_{t_k} = (\theta_{h_{0k}}, \gamma_k, \alpha_k)$, $k = 1, \ldots, K$ and perform this stage as the following step:

– Step 1: At first, we consider initial value for the parameter of time-to-event process. We assume that there are m parameters in time-to-event vector θ_{t_k} and its first value as the vector $\theta^{(0)}_{t_k} = (\theta^{(0)}_{1k}, \ldots, \theta^{(0)}_{mk})$. Then, by using these initial values and the estimates of $\tilde{\theta}_y$, $\tilde{\theta}_b$ and $\tilde{b}_i$ in stage I, we compute the value of complete data log-likelihood as follows:

$$l(\theta^{(0)}_{t_k}) = \sum_i \log P(t_i, \delta_i, y_i, \tilde{b}_i; \theta^{(0)}_{t_k}, \tilde{\theta}_y) = \sum_i \log P(t_i, \delta_i, \tilde{b}_i; \theta^{(0)}_{t_k}, \tilde{\theta}_y) \tag{13}$$

$$+ \log P(y_i, \tilde{b}_i; \tilde{\theta}_y) + \log P(\tilde{b}_i; \tilde{\theta}_b).$$

– Step2: Now, we should update parameters. To this end, we consider the following steps:

1. Suppose that the current value of parameter vector at the rth iteration to be $\theta^{(r)}_{t_k} = (\theta^{(r)}_{1k}, \ldots, \theta^{(r)}_{mk})$ and the obtained value of the log-likelihood at this iteration to be

$$l(\theta_{t_k}^{(r)}) = \sum\nolimits_i \log p(t_i, \delta_i, y_i, \tilde{b}_i; \theta_{t_k}^{(r)}, \tilde{\theta}_y).$$

2. Evaluate a new value based on the Newton-Raphson method for the first parameter and denote it as $\theta_{1k}^{(*)}$. Then, evaluate the log-likelihood at $\theta_{t_k}^{(*)} = (\theta_{1k}^{(*)}, \theta_{2k}^{(r)}, \ldots, \theta_{mk}^{(r)})$ as $l(\theta_{t_k}^{(*)})$.
3. If $l(\theta_{t_k}^{(*)}) \geq l(\theta_{t_k}^{(r)})$, let $\theta_{t_k}^{(r)} = \theta_{t_k}^{(*)}$ else $l(\theta_{t_k}^{(r)}) = l(\theta_{t_k}^{(r)})$.
4. In the same way, based on the value of the parameter vector $\theta_{t_k}^{(r)}$, we update the second parameter and continue it to update the last parameter and obtain the $\theta_{t_k}^{(r+1)}$.

- Step 3: Repeat Step 2 until the algorithm converges numerically. The convergence criteria is considered as follows:

$$l(\theta_{t_k}^{(r+1)}) - l(\theta_{t_k}^{(r)}) \langle \in (|l(\theta_{t_k}^{(r)}| + \in)$$

where $\theta_{t_k}^{(r)}$ shows the values of parameters at the rth iteration and we let $\in = 10^{-8}$.
In order to calculate the updated values $\theta_{h_{0k}}^{(r+1)}$,$\gamma^{(r+1)}$ and $\alpha^{(r+1)}$, we have

$$S(\theta_{t_k}) = \sum_i \frac{\partial}{\partial \theta_{t_k}} \log\{p(t_i, \delta_i, \tilde{b}_i; \theta_t, \tilde{\theta}_y, \tilde{\theta}_b)\},$$

$$\hat{\theta}_{t_k}^{(r+1)} = \hat{\theta}_{t_k}^{(r)} - \left[\frac{\partial S(\hat{\theta}_{t_k}^{(r)})}{\partial \theta_{t_k}}\right]^{-1} S(\hat{\theta}_{t_k}^{(r)}). \tag{14}$$

We note that the components of the score vector corresponding to θ_{t_k} have the following forms:

$$S(\theta_{h_{0k}(t)}) = \sum_i \delta_i \frac{\partial \log h_{0k}(t_i; \theta_{h_{0k}(t)})}{\partial \theta'_{h_{0k}(t)}}$$

$$- \exp(\gamma'_k w_i) \frac{\partial}{\partial \theta_{h_{0k}(t)'}} [\int_0^{t_i} h_{0k}(s) \exp\{\alpha_k(X_i'(s)\tilde{\beta} + Z_i'(s)\tilde{b}_i)\} ds],$$

$$S(\gamma_k) = \sum_i w_i[\delta_i - \exp(\gamma'_k k w_i)] \int_0^{t_i} h_{0k}(s) \exp\{\alpha_k(X_i'(s)\tilde{\beta} + Z_i'(s)\tilde{b}_i)\} ds,$$

$$S(\alpha_k) = \sum_i \delta_i \{(X_i' i(t_i)\tilde{\beta} + Z_i'(t_i)\tilde{b}_i)\}$$

$$- \exp(\gamma' w_i) \frac{\partial}{\partial \alpha_k} [\int_0^{t_i} h_{0k}(s) \exp\{\alpha_k(X_i'(s)\tilde{\beta} + Z_i'(s)\tilde{b}_i)\} ds].$$

• Stage III: Perform the stage II for all failure types k, $k = 1, 2, \ldots, K$ and complete the competing risks process.

5. Simulation studies

5.1. Used model and data-generated model are the same

In this sub-section, we implement some simulation studies in joint modeling of longitudinal measurements and competing risks data to investigate the performance of our proposed two-stage method based on the full likelihood approach. For this purpose, we simulate $N = 200$ data from a joint model

with a linear mixed-effects model for longitudinal measurement and competing risks model with two failure type as follows:

$$Y_i(s) = m_i(s) + \in_i = \beta_0 + \beta_1 s + b_{i0} + b_{i1} s + \in_i, i = 1, 2, \ldots, n, \tag{15}$$

and

$$h_k(t_i) = h_{0k}(t_i) exp\{\gamma_k' w_i + \alpha_k m_i(t)\} = \lambda_k exp\{\gamma_{1k} w_{1i} + \alpha_k m_i(t)\}, k = 1, 2. \tag{16}$$

Here w_{1i} is the baseline covariate. $b_i = (b_{i0}, b_{i1})$ is the vector of random effects, which is assumed to have a bivariate normal distribution with mean zero and covariance matrix $\mathbf{D}$ and $h_{0k}(t_i)$ is the baseline hazard function supposed to have exponential distribution.

As we considered above, the survival times follow an exponential distribution; thus, it is required to know how we can generate survival times to simulate Cox proportional hazard (16) with time-varying covariates. The relation between the survival function $S_i(t)$, the cumulative hazard function $H_i(t)$ and the cumulative distribution $F_i(t)$ is defined as

$$S_i(t) = exp(-H_i(t)) = 1 - F_i(t). \tag{17}$$

Based on (17), we define $u = F_i(t)$ where u is a random variable with standard uniform distribution. Then, the true survival time t is obtained as a solution of

$$U = exp(-H_i(t)) = exp\left(-\int_0^t h_i(s)ds\right). \tag{18}$$

Here, we applied the Austin (2012) method and his obtained formula for generating an event time when the survival times follow an exponential distribution.

In general, suppose that $x(t)$ is a single time-varying covariate which is proportional to t: $x(t) = vt$. Also, let the other covariates, x^*, be time-invariant. Let β indicates the vector of regression coefficient associated with the vector of fixed covariate x^* and β_t be the vector of regression coefficient associated with $x(t)$. Also let the following equation be satisfied:

$$h(t|x(t)) = h_0(t) exp(\beta_t x(t) + \beta' x^*).$$

Then, the cumulative hazard function is given by: $H(t, x^*, x(t)) = \int_0^t exp(\beta_t x(u) + \beta' x^*) h_0(u) du$. Based on the exponential distribution of event time, we will have

$$H(t, x^*, x(t)) = \int_0^t exp(\beta_t x(u) + \beta' x^*)\lambda du = \lambda exp(\beta' x^*) \int_0^t exp(\beta_t x(u)) du$$

$$= \lambda exp(\beta' x^*) \int_0^t exp(\beta_t v u) du = \lambda exp(\beta' x^*) \left[\frac{1}{\beta_t v} exp(\beta_t v u)\right]_0^t$$

$$= \frac{\lambda exp(\beta' x^*)}{\beta_t v} [exp(\beta_t v t) - 1]$$

Using the equation (18), we can finally generate an event time from a survival time that follows an exponential distribution with parameter λ as follows:

$$T = \frac{1}{\beta_t v} log(1 + \frac{\beta_t v(-log(u))}{\lambda_k exp(\beta' x^*)}), \tag{19}$$

where, $u \sim U(0, 1)$.

Therefore, we can rewrite the equation (19) based on the two considered sub-models (15) and (16) as follows:

$$t_{ik} = \frac{1}{(\beta_1 + b_{i1})\alpha_k} log\left(1 + \frac{(\beta_1 + b_{i1})\,\alpha_k\,(-log(u))}{\lambda exp\,(\gamma_{1k} w_{1i} + \alpha_k(\beta_0 + b_{i0}))}\right),\ k = 1, 2. \tag{20}$$

We generate two survival times t_{i1}^* and t_{i2}^* based on (20) for $n = 200$ and 500 subjects. The observed survival time t_i is the minimum of the censoring time and the two true survival times. Here, the censoring mechanism is induced by an exponential distribution with parameter λ $(0 < \lambda < 1)$. Here, we consider the true value of parameters as: $\beta_0 = 5$, $\beta_1 = 2$, $\lambda_1 = 0.2$, $\lambda_1 = 0.4$, $\gamma_{11} = 0.5$, $\gamma_{12} = 1$, $\alpha_1 = 0.1$, $\alpha_2 = 0.08$, $D_{11} = 1$, $D_{12} = 0.5$ and $D_{22} = 1$. Also, we assume that the covariate w_{1i} has Bernoulli distribution with parameter value 0.5. Then, the observed longitudinal measurement at time point s_{ij} for the ith subject is generated by

$$y_i(s_{ij}) = m_i(s_{ij}) + \in_i(s_{ij}), j = 1, \ldots, n_i, \tag{21}$$

where, $\{0 \le s_{ij} \le t_i,\ s_{i(j+1)} = s_{ij} + 0.5\}$, t_i the observed survival time and $\in_i(s_{ij}) \sim N(0, \sigma^2)$; also, the true value of σ is 2. Moreover, for measurements taken every six months, the number of recorded longitudinal measurements for each subject, n_i, $i = 1, \ldots, n$, is determined as follows:

$$n_i = \begin{cases} 2[t_i] + 1 & \text{if } |t_i - [t_i]| \le 0.5 \\ 2[t_i] + 2 & \text{if } |t_i - [t_i]| > 0.5, \end{cases} \tag{22}$$

where the $[t_i]$ is the floor of t_i. For example, if we obtain $n_i = 6$ for a subject, s_{ij} takes the values of $0, 0.5, 1, 1.5, 2$ and 2.5.

We consider a case in which the censoring rate is about 15%. Here, based on the measurements taken every six months for both sample sizes 200 and 500, on average, two longitudinal responses are recorded for each subject and $\max n_i = 16$.

As an illustration, we also simulate a separate model and a joint model to compare with this proposed two-stage approach. We use the **lme**() function for the longitudinal measurement and the **coxph**() function in package **survival** (Therneau and Lumley (2012)) for time-to-event process to fit separate models. Also the **jointModel**() function in *JM* package (Rizopoulos (2012b)) is used to fit the joint model. We note that the joint model is fitted by providing the mixed-effects and Cox models as main argument to **jointModel**() function.

The results of the simulation studies, based on the three models for n = 200 and 500, are given in Tables 1 and table 2, respectively. These tables contain mean, standard error, bias, root-mean-square errors (RMSE) and 95% coverage probability.

Looking at the results of the simulation study, it is concluded that increasing in the sample size is an effective way of decreasing biases, RMSEs and standard deviations of parameters in the two-stage approach. This means that the proposed approach has good consistency properties. Also, the results show that the use of separate models and joint model, instead of the proposed two-stage approach, would lead to biased parameter estimates. Notably, the values of RMSEs, Biases and SEs of the time-to-event parameters and the components of covariance matrix **D** for both sample sizes show that the two-stage approach work better than the other models.

Evidently, the best inferential results are usually obtained when the model used for generating data is the same as the one used for analyzing the simulated data. Therefore, it is not surprising to see better results by using the proposed two-stage approach.

5.2. Used model and data-generated model are different

In this sub-section, we have considered some models to evaluate the robustness of the proposed two-stage approach. For this purpose, we have simulated data from some other models, such as: (1) skew-normal random effects model (SNREM), (2) multivariate t random effects model (MTREM) with normal distribution for error term, (3) skew-normal error model (SNEM), (4) skew-t error model (STEM) and (5) t error model (TEM) with normal distribution for random effects are considered to

Table 1. Results of the simulation study for proposed two-stage, separate and joint models of longitudinal measurements and competing risks data, estimate (Est.), standard error (S.E.), bias (Bias), root of mean square error (RMSE) and 95% coverage probability (CP) for N = 200 simulated data with sample size 200. Note: $D_{ij}; i; j = 1; 2$ are distinct components of the covariance matrix **D**.

Model	*parameter*	*real*	*Est.*	*Bias*	*S.E.*	*RMSE*	*CP*
Two-stage	β_0	5.00	4.970	−0.030	0.520	0.520	0.99
	β_1	2.00	1.796	0.-0.204	0.346	0.401	0.94
	σ	2.00	1.966	−0.034	0.232	0.234	0.99
	D_{11}	1.00	1.062	0.062	0.511	0.514	0.97
	D_{12}	0.50	0.278	−0.222	0.527	0.570	0.91
	D_{22}	1.00	1.254	0.254	0.946	0.977	0.92
	$\lambda 1$	0.20	0.235	0.035	0.149	0.153	0.94
	$\lambda 2$	0.40	0.412	0.012	0.209	0.208	0.97
	$\gamma 1$	1.00	0.975	−0.025	0.265	0.265	0.95
	$\gamma 2$	0.50	0.513	0.013	0.212	0.212	0.94
	$\alpha 1$	0.10	0.096	−0.004	0.089	0.089	0.95
	$\alpha 2$	0.08	0.090	0.010	0.072	0.072	0.94
Separate	β_0	5.00	5.026	0.150	0.026	0.152	0.95
	β_1	2.00	1.796	0.-0.204	0.346	0.401	0.94
	σ	2.00	1.966	−0.034	0.232	0.234	0.99
	D_{11}	1.00	1.062	0.062	0.511	0.514	0.97
	D_{12}	0.50	0.278	−0.222	0.527	0.570	0.91
	D_{22}	1.00	1.254	0.254	0.946	0.977	0.92
	$\lambda 1$	-	-	-	-	-	-
	$\lambda 2$	–	-	-	-	-	-
	$\gamma 1$	1.00	0.976	−0.024	0.272	0.272	0.94
	$\gamma 2$	0.50	0.516	0.016	0.225	0.225	0.93
	$\alpha 1$	-	–	-	-	-	-
	$\alpha 2$	–	-	-	-	-	-
Joint Model	β_0	5.00	4.973	−0.027	0.398	0.398	0.99
	β_1	2.00	1.861	−0.139	0.576	0.591	0.98
	σ	2.00	1.862	−0.138	0.257	0.291	0.94
	D_{11}	1.00	1.262	0.262	0.731	0.775	0.94
	D_{12}	0.50	0.366	−0.134	0.742	0.752	0.93
	D_{22}	1.00	1.283	0.283	1.291	1.318	0.92
	$\lambda 1$	-	-	-	-	-	-
	$\lambda 2$	–	-	-	-	-	-
	$\gamma 1$	1.00	0.997	−0.003	0.293	0.292	0.95
	$\gamma 2$	0.50	0.515	0.015	0.287	0.287	0.97
	$\alpha 1$	0.10	0.020	−0.080	0.265	0.276	0.98
	$\alpha 2$	0.08	0.045	−0.035	0.135	0.139	0.95

generate data. As we mentioned before, to analyze these data, we have considered normal distribution for the random effects and error term.

We consider a bivariate zero-mean skew-normal distribution as $SN_2(-\sqrt{2/\pi}\Delta, D, \Delta)$ for the random effects b of model (1), where $\Delta = diag(\delta_1, \delta_2)$ is a diagonal matrix of the skewness parameters (Arellano-Valle et al. 2007) and D is the covariance matrix b, where, $(\delta_1, \delta_2) = (2, 1)$, $D_{11} = 1$, $D_{12} = 0.5$ and $D_{22} = 1$. We note that the location parameter ($-\sqrt{2/\pi}\Delta$) is considered to have a zero-mean bivariate skew-normal distribution of random effects. Also, we consider a multivariate t distribution with location zero, covariance matrix D and degrees of freedom 4 for the model (2). The results of these two models are summarized in Table 3.

Let $SN(\xi, \omega, \alpha)$ be a skew-normal distribution with location parameters ξ, scale parameter ω and the shape parameter α and $ST(\xi, \omega, \alpha, \nu)$ be a skew-t distribution with location parameter ξ, scale parameter ω, shape parameter α and degrees of freedom ν. For generating data from SNEM, model (3), we have considered a zero-mean skew-normal distribution for the error as follows: $SN(-\sqrt{2/\pi}\sigma\delta, \sigma, \alpha)$ in which $\delta = \frac{\alpha}{\sqrt{(1+\alpha^2)}}$ and the values of parameters σ and α are considered to be 2 and 1, respectively.

Table 2. Results of the simulation study for proposed two-stage, separate and joint models of longitudinal measurements and competing risks data, estimate (Est.), standard error (S.E.), bias (Bias), root of mean square error (RMSE) and 95% coverage probability (CP) for N = 200 simulated data with sample size 500. Note: $D_{ij}; i; j = 1; 2$ are distinct components of the covariance matrix **D**.

Model	*parameter*	*real*	*Est.*	*Bias*	*S.E.*	*RMSE*	*CP*
Two-stage	β_0	5.00	5.013	0.013	0.098	0.098	0.95
	β_1	2.00	1.800	0.-0.200	0.192	0.277	0.81
	σ	2.00	1.995	−0.005	0.076	0.076	0.94
	D_{11}	1.00	1.022	0.022	0.325	0.325	0.95
	D_{12}	0.50	0.371	−0.129	0.341	0.364	0.94
	D_{22}	1.00	1.053	0.053	0.542	0.543	0.97
	$\lambda 1$	0.20	0.222	0.022	0.080	0.082	0.94
	$\lambda 2$	0.40	0.424	0.024	0.118	0.120	0.93
	$\gamma 1$	1.00	0.987	−0.013	0.169	0.169	0.95
	$\gamma 2$	0.50	0.500	0.000	0.143	0.143	0.96
	$\alpha 1$	0.10	0.092	−0.008	0.050	0.050	0.94
	$\alpha 2$	0.08	0.077	−0.003	0.042	0.042	0.95
Separate	β_0	5.00	5.009	0.089	0.009	0.089	0.96
	β_1	2.00	1.800	0.-0.200	0.192	0.277	0.81
	σ	2.00	1.995	−0.005	0.076	0.076	0.94
	D_{11}	1.00	1.022	0.022	0.325	0.325	0.95
	D_{12}	0.50	0.371	−0.129	0.341	0.364	0.94
	D_{22}	1.00	1.053	0.053	0.542	0.543	0.97
	$\lambda 1$	-	-	-	-	-	-
	$\lambda 2$	–	-	-	-	-	-
	$\gamma 1$	1.00	0.981	−0.019	0.230	0.229	0.99
	$\gamma 2$	0.50	0.502	0.002	0.146	0.146	0.95
	$\alpha 1$	-	–	-	-	-	-
	$\alpha 2$	–	-	-	-	-	-
Joint Model	β_0	5.00	4.999	−0.001	0.230	0.229	0.99
	β_1	2.00	1.932	−0.068	0.403	0.408	0.98
	σ	2.00	1.868	−0.132	0.317	0.343	0.99
	D_{11}	1.00	1.200	0.200	0.518	0.554	0.93
	D_{12}	0.50	0.614	0.114	0.731	0.738	0.95
	D_{22}	1.00	1.743	0.743	5.854	5.887	0.99
	$\lambda 1$	-	-	-	-	-	-
	$\lambda 2$	–	-	-	-	-	-
	$\gamma 1$	1.00	1.000	0.000	0.209	0.209	0.96
	$\gamma 2$	0.50	0.515	0.015	0.186	0.186	0.97
	$\alpha 1$	0.10	0.075	−0.125	0.122	0.125	0.97
	$\alpha 2$	0.08	0.063	−0.017	0.078	0.080	0.97

Table 3. Results of the simulation study for 200 samples of using the proposed two-stage approach with skew-normal and multivariate t random effects models (SNREM and MTREM) for generation data, 15% rate of censoring and sample size $n = 500$.

model			*SNREM*			*MTREM*	$(v = 4)$	
parameter	*Est.*	*Bias*	*S.E.*	*RMSE*	*Est.*	*Bias*	*S.E.*	*RMSE*
$\beta_0 = 5$	5.004	0.004	0.104	0.103	5.030	0.030	0.100	0.104
$\beta_1 = 2$	1.748	−0.252	0.200	0.322	1.732	−0.263	0.211	0.340
$\sigma = 2$	2.001	0.001	0.075	0.075	1.994	−0.006	0.078	0.078
$D_{11} = 1$	2.423	1.423	0.443	1.490	1.885	0.885	0.498	1.015
$D_{12} = 0.5$	0.453	−0.047	0.463	0.464	0.770	0.270	0.519	0.584
$D_{22} = 1$	1.310	0.310	0.650	0.719	1.601	0.601	0.792	0.992
$\lambda_1 = 0.2$	0.227	0.027	0.064	0.070	0.228	0.028	0.064	0.070
$\lambda_2 = 0.4$	0.429	0.029	0.103	0.107	0.448	0.048	0.105	0.115
$\gamma_1 = 1$	1.010	0.010	0.147	0.147	0.998	−0.002	0.153	0.153
$\gamma_2 = 0.5$	0.504	0.004	0.133	0.132	0.486	−0.014	0.137	0.138
$\alpha_1 = 0.1$	0.086	−0.014	0.040	0.042	0.086	−0.014	0.039	0.041
$\alpha_1 = 0.08$	0.073	−0.007	0.036	0.036	0.069	−0.011	0.034	0.036

Also, a zero-mean skew-t distribution $ST(-\sqrt{\nu/\pi}\sigma\delta\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \sigma, \alpha, \nu)$ is considered for error term to generate data from STEM, model (4), in which the values of parameters are $\sigma = 2$, $\alpha = 1$ and $\nu = 4$. We have used the *rsn*() and *rst*() functions of package *sn* (Azzalini 2020) to generate sample from skew-normal and skew-t distributions. For the TEM model, we generate data from a t-distribution with degrees of freedom 3. The results of using these three models for generation data of error models are summarized in Table 4.

The results of Tables 3 and table 4 show that normal models for the random effects and errors give good estimates and are robust. Clearly, with the exception of covariance components (D_{11}, D_{12} and D_{22}) in some cases, better robustness results than the corresponding results of the previous sub-section (Table 2) are observed. For further discussion, we have considered higher degrees of freedom for the models of MTREM, STEM and TEM to see their results. For this purpose, we have regarded $nu = 8$ and 12 for MTREM, $nu = 5$ and 8 for STEM and $nu = 5$ and 10 for TEM. The results of these models are summarized in Tables C.1, C.2 and C.3, respectively, and are given in the Supplementary Materials C. The results of Tables C.1, C.2 and C.3 show that increasing the degrees of freedom in three models MTREM, STEM and TEM leads to better results for the components of the covariance matrix.

To demonstrate more robustness of the proposed model, we consider two models where data generations are based on the two linkage terms:(1) $\alpha(\beta_0 + \beta_1 t + b_{i0})$ and (2) $\alpha(\beta_0 + \beta_1 t + b_{i1} t)$ between two sub-models of longitudinal measurements and competing risks data. These are

(1): $h_k(t_i) = \lambda_k exp\{\gamma_{1k} w_{1i} + \alpha_k(\beta_0 + \beta_1 t + b_{i0})\}$, $k = 1, 2$ and

(2): $h_k(t_i) = \lambda_k exp\{\gamma_{1k} w_{1i} + \alpha_k(\beta_0 + \beta_1 t + b_{i1} t)\}$, $k = 1, 2$.

The results of these simulation studies are given in Table 5.

It shows better robustness of model than the corresponding estimates of the previous sub-section (Table 2).

6. Application

In this section, we analyze the epileptic data using the proposed two-stage approach joint modeling of the previous sections. As defined in Section 2, calibrated dose can be considered as a response variable. We consider the following linear mixed effect model for the longitudinal data sub-model:

$$y_i(s_{ij}) = \beta_0 + \beta_1 s_{ij} + \beta_2 LTG_i + \beta_3 LTG_i s_{ij} + b_{i0} + b_{i1} s_{ij} + \in_{ij}, \tag{23}$$

where the LTG_i is a binary time-independent treatment effect that gives value 1 if patient i is randomized to LTG and zero if the patient is randomized to CBZ. $(b_{i0}, b_{i1})'$ and $\in_{ij}$ are distributed as $N_2(0, D)$ and $N(0, \sigma^2_{\in})$, respectively. Also a cause-specific hazard model for the competing risk data is defined as:

$$h_k(t_i) = \lambda_k exp\{LTG_i \gamma_k + \alpha_k m_i(t)\}, k = 1, 2, \tag{24}$$

where, the baseline hazard function is supposed to have an exponential distribution. Under this formulation, parameters γ_1 and α_1 denote the effects of treatment (LTG) and dose calibrated, respectively, on the risk of *ISCs* and γ_2 and α_2 denote the same effects for *UAEs*. We also consider a separate model and a joint model, as we did in the previous section, to fit this data set.

A summary of parameter estimates and 95% confidence intervals (CIs) for the longitudinal sub-model and the competing risks sub-model parameters are given in Table 6. In this table, AIC (Akaike information criterion), BIC (Bayesian information criterion) and computation time (using a PC with a Core(TM)i5-7200 U CPU @ 2.50 GHz processor and 8 GB of memory) are considered to be used for comparing models. Also the estimation of parameters for error term and the random effects covariance matrix are given in Table 7.

As the results of Table 6 indicate, the estimated treatment effect on calibrated dose, β_2, is non-significant for all models. The estimate of the fixed effect for time, β_1, is significant for all models. That is to say, as time passes, the average calibrated dose increases. Also the estimated effect of treatment and time interaction, β_3,

Table 4. Results of the simulation study for 200 samples of using the proposed two-stage approach with skew-normal, skew-t and t error models (SNEM, STEM and TEM) for generation data, 15% rate of censoring and sample size $n = 500$.

model			*SNEM*			*STEM*	($v = 4$)			*TEM*	($v = 3$)	
parameter	*Est.*	*Bias*	*S.E.*	*RMSE*	*Est.*	*Bias*	*S.E.*	*RMSE*	*Est.*	*Bias*	*S.E.*	*RMSE*
$\beta_0 = 5$	4.782	−0.218	0.084	0.234	5.016	0.016	0.112	0.113	5.016	0.016	0.090	0.092
$\beta_1 = 2$	1.827	−0.173	0.180	0.249	1.758	−0.242	0.245	0.343	1.806	−0.194	0.182	0.266
$\sigma = 2$	1.641	−0.359	0.062	0.364	2.404	0.404	0.187	0.445	1.661	−0.339	0.160	0.375
$D_{11} = 1$	1.027	0.027	0.254	0.255	1.081	0.081	0.667	0.670	1.267	0.267	0.985	1.018
$D_{12} = 0.5$	−0.464	−0.036	0.262	0.264	0.305	−0.195	0.693	0.718	0.250	−0.250	0.743	0.782
$D_{22} = 1$	1.018	0.018	0.427	0.426	1.496	0.496	2.649	2.688	1.309	0.309	0.912	0.961
$\lambda_1 = 0.2$	0.229	0.029	0.076	0.082	0.239	0.039	0.088	0.096	0.228	0.028	0.082	0.087
$\lambda_2 = 0.4$	0.444	0.044	0.125	0.132	0.461	0.061	0.129	0.143	0.445	0.045	0.131	0.138
$\gamma_1 = 1$	0.996	−0.004	0.169	0.168	0.967	−0.033	0.161	0.164	0.998	−0.002	0.172	0.172
$\gamma_2 = 0.5$	0.489	−0.011	0.135	0.135	0.478	−0.022	0.139	0.141	0.490	−0.010	0.141	0.141
$\alpha_1 = 0.1$	0.090	−0.010	0.049	0.050	0.083	−0.017	0.054	0.057	0.089	−0.011	0.050	0.051
$\alpha_1 = 0.08$	0.073	−0.007	0.041	0.041	0.065	−0.015	0.042	0.044	0.071	−0.009	0.043	0.044

Table 5. Results of the simulation study for 200 samples of using the proposed two-stage approach with linkage terms (1) and (2) of generation data model, 15% rate of censoring and sample size $n = 500$.

	*linkage*1				*linkage*2			
parameter	*Est.*	*Bias*	*S.E.*	*RMSE*	*Est.*	*Bias*	*S.E.*	*RMS*
$\beta_0 = 5$	5.017	0.017	0.093	0.003	5.023	0.023	0.099	0.102
$\beta_1 = 2$	1.873	−0.127	0.197	0.234	1.903	−0.097	0.192	0.215
$\sigma = 2$	1.987	−0.013	0.079	0.080	1.984	−0.016	0.075	0.076
$D_{11} = 1$	1.039	0.039	0.342	0.343	1.076	0.076	0.319	0.327
$D12 = 0.5$	0.422	−0.078	0.322	0.330	0.395	−0.105	0.349	0.364
$D22 = 1$	1.062	0.062	0.484	0.487	1.120	0.120	0.555	0.567
$\lambda_1 = 0.2$	0.248	0.048	0.078	0.092	0.254	0.054	0.092	0.107
$\lambda_2 = 0.4$	0.494	0.094	0.128	0.158	0.480	0.080	0.137	0.158
$\gamma_1 = 1$	1.002	0.002	0.152	0.152	0.992	−0.008	0.168	0.167
$\gamma_2 = 0.5$	0.481	−0.019	0.132	0.132	0.485	−0.015	0.132	0.133
$\alpha_1 = 0.1$	0.072	−0.028	0.046	0.054	0.071	−0.029	0.049	0.057
$\alpha_1 = 0.08$	0.052	−0.028	0.037	0.046	0.058	−0.022	0.040	0.046

Table 6. Parameter estimates and 95% confidence intervals (CIs) for the longitudinal sub-model and the competing risks sub-model.

	Model	Proposed	Separate	Joint model
	parameter	two-stage		
	$\beta_0(Intercept)$	1.932	1.932	1.932
Longitudinal	$(95\%\ CI)$	(1.826,2.039)	(1.826,2.039)	(1.858,2.005)
Process	$\beta_1(Time)$	0.151	0.151	0.045
	$(95\%\ CI)$	(0.085,0.217)	(0.085,0.217)	(0.013,0.077)
	$\beta_2(Treat(LTG))$	−0.087	−0.087	0.041
	$(95\%\ CI)$	(−0.236,0.061)	(−0.236,0.061)	(−0.079,0.162)
	$\beta_3(Time : Treat)$	0.214	0.214	0.404
	$(95\%\ CI)$	(0.123,0.303)	(0.123,0.303)	(0.365,0.444)
	λ_{ISC}	0.240	-	–
Event	$(95\%\ CI)$	(0.020,0.028)	–	–
Process	λ_{UAE}	0.218	-	-
	$(95\%\ CI)$	(0.174,0.262)	-	-
	γ_{ISC}	0.011	0.015	−0.134
	$(95\%\ CI)$	(−0.222,0.244)	(−0.344,0.374)	(−0.497,0.230)
	γ_{UAE}	−0.460	−0.608	−0.614
	$(95\%\ CI)$	(−0.787,-0.133)	(−1.102,-0.192)	(−1.534,0.306)
	α_{ISC}	0.546	-	0.598
	$(95\%\ CI)$	(0.487,0.605)	-	(0.443,0.751)
	α_{UAE}	−0.503	–	−0.942
	$(95\%\ CI)$	(−0.615,-0.390)	-	(−1.455,-0.428)
Comparition	*AIC*	7225.824	8260.968	7210.406
Criteria	*BIC*	7234.771	8315.179	7342.563
	Computation time	1.33 min	<1 s	4.8 min

Table 7. Parameter estimates of error term and the distinct components of the covariance matrix D (D_{11}, D_{12} and D_{22}).

parameter	Proposed	Separate	Joint model
	two-stage		
σ	0.446	0.446	0.472
D_{11}	0.711	0.711	0.748
D_{22}	0.150	0.150	0.066
D_{12}	0.062	0.062	0.157

is significant for all models. Looking at the considered hazard model (24), we conclude that the overall treatment effect on the event hazard is divided into the direct effect γ_k and the indirect effect $\alpha_k(\beta_2 + \beta_3 t_{ij})$. Hence, the direct treatment effect must be considered by adjusting for treatment-specific intercept and slope of dose titration in the hazard model. The direct treatment effect on ISC, γ_{ISC}, is non-significant for all models in Table 6. However, the direct treatment effect on UAE is significant for proposed two-stage joint

modeling approach. Hence, if LTG is tested at the same CBZ rate, the useful effect of LTG on a UAE would still be evident and the difference in seizure control between the two drugs is unclear.

The estimation of association parameter, α_{ISC} and α_{UAE} is both significant for two considered models in Table 6, which suggests that calibrated dose of drug is associated with time to treatment failure for both failure types.

Of course, the parameters have inverse signs. The clinical explanation of this fact, as Williamson et al. (2008) discussed, is that patients on higher doses are more likely to withdraw from treatment due to ISC because the reason behind their being on higher doses is that they are having continued seizures. Also, patients on higher doses are less likely to be withdrawn from treatment due to UAE since the reason that a required dose increases is that no such events have occurred. Furthermore, the estimation of baseline hazard parameters, λ_{ISC} and λ_{UAE}, is both significant in the two-stage approach. Based on the given values of AIC, BIC and computation time we can see that the proposed model performs well and it is clear that BIC of proposed model is less than those of other models. AIC of proposed model is less than that of separate model although it is a little more than the joint model. However, given that the computation time is less than the computation time of the joint model, it can be said that the proposed model has performed well compared to the other models.

6.1. Residuals analysis

In this section, we confirm the validation of the model's assumption. Residual plots are standard and useful tools for diagnosing these assumptions. For this purpose, we use subject-specific residuals to validate the continuous longitudinal data and, conditional Cox-Snell residuals (Cox and Snell (1968)) to validate competing risks data. In fact, these residuals can be used to find influential observations.

6.1.1. Residuals for the longitudinal part

Two types of residuals are usually used in the standard linear mixed effect model, the subject-specific (conditional) residuals and the marginal (population average) residuals (Nobre and Singer 2007; Verbeke and Molenberghs 2000, e.g). Here, we use the standardized subject-specific residuals for continuous longitudinal data, which is equal to the difference between the observed and fitted value divided by the square root of the residuals variance, as follows:

$$r_{ij}^{P} = \frac{Y_{ij} - \widehat{E}[Y_{ij}|b_i]}{\sqrt{\widehat{Var}[Y_{ij}|b_i]}},$$

where $\hat{E}[Y_{ij}|b_i] = x'_{ij}\hat{\beta} + z'_{ij}\hat{b}_i$ and $\widehat{Var}[Y_{ij}|b_i] = \widehat{\sigma^2_\varepsilon}$.

The standardized subject-specific residuals for the longitudinal outcomes are plotted in Figure 3. This demonstrates no pattern and confirms the goodness of the fit of the model.

6.1.2. Residuals for competing risks part

A common type of residuals for the survival models is the Cox-Snell residuals. The conditional Cox-Snell residuals $(r_{ik}^{CS}, i = 1, \ldots, n, k = 1, \ldots, K)$ which can be calculated as the estimation value of the cumulative risk function evaluated at the observed event time t_i, that is,

$$r_{ik}^{CS} = \int_0^{t_i} h_{ik}(s|\hat{\mathcal{M}}_i(s); \hat{\theta})ds \tag{25}$$

$$= \int_0^{t_i} \hat{h}_{0k}(t)exp\{\hat{\gamma}'_k x_i + \hat{\alpha}_k \hat{m}_i(s)\}ds, i = 1, \ldots, n, k = 1, \ldots, K,$$

in other words, it can be rewritten as follows:

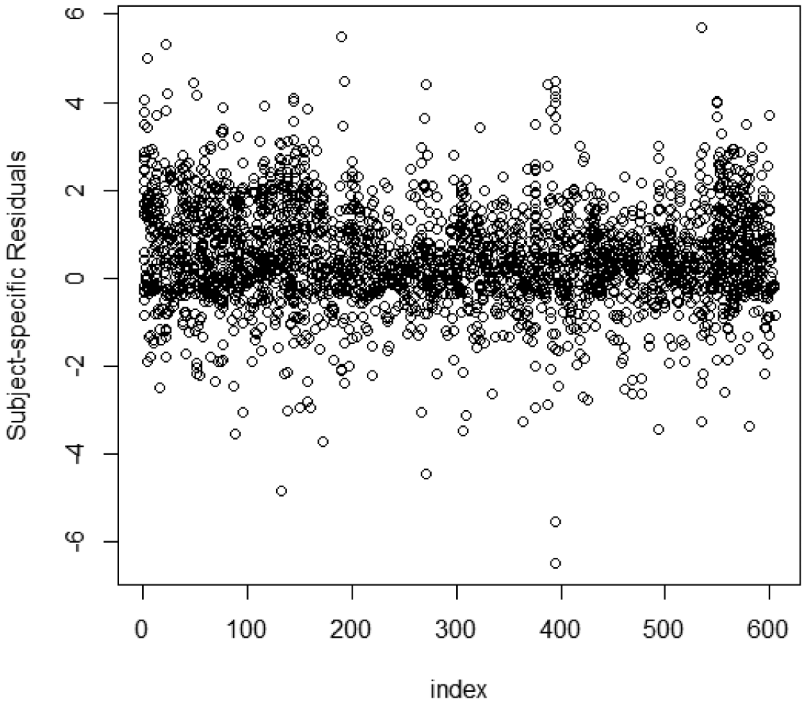

Figure 3. Subject-specific residuals for longitudinal outcomes.

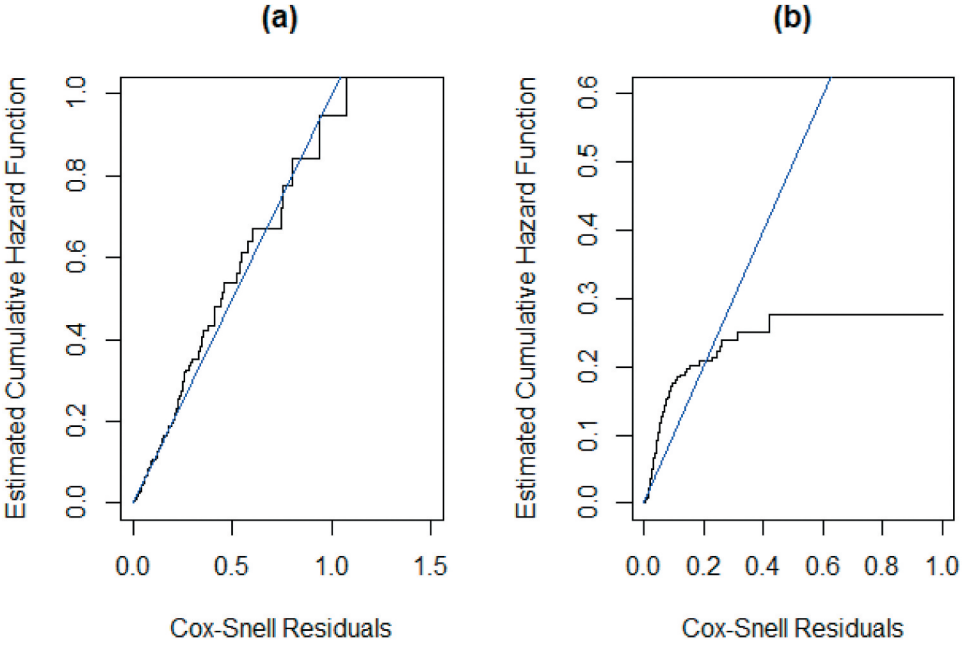

Figure 4. Cox-Snell residuals for competing risks part. (a): treatment failure due to ISC, (b): treatment failure due to UAE.

$$r_{ik}^{CS} = -log\,\hat{S}_k(t_i),\ i = 1, \ldots, n,\ k = 1, \ldots, K. \tag{26}$$

Given the probability integral transform, it is clear that the probability of failure after time t, i.e., $S(t) = Pr(T_i^* > t)$ follows a uniform distribution. Therefore, the cumulative hazard, demonstrated as $\mathcal{H}(t) = -log\,S(t)$ follows a unit exponential distribution. Therefore, $R_{ik}^{CS} \sim exp(1)$. If $S_k^{\hat{R}}(t_i)$ is the survival function of the Cox-Snell residuals R_{ik}^{CS}, then, it follows that $-logS^R(r_{ik}) = r_{ik}$. Let $S^{\hat{R}}(r_{ik})$ be the Kaplan-Meier estimate of $S^R(r_{ik})$. If the fitted survival distribution is convenient, then the plot of r_{ik}^{CS} versus $S^{\hat{R}}(r_{ik})$ should be straight line with zero intercept and unit slope. Figure 4 shows the Cox-Snell residuals for competing risks part.

As Figure 4 shows, we do not see lack of fit for state (a) and we see disparity in some places about (b) state (i.e. treatment failure due to UAE). This may be due to the fact that the number of patients who withdrew from the study because of UAE is smaller than the number of patients who withdrew due to ISC; in other words, the number of observations in the (b) state is smaller.

7. Conclusion

In this paper, a two-stage approach for joint modeling of longitudinal measurements and competing risks data was discussed. This approach overcomes the computational problems of shared random effects in joint model. The joint model was defined using the shared parameter model. Some simulation studies were conducted to illustrate the performance of the proposed approach. The results are improved by increasing the sample size. Also, the proposed approach was used to analyze the epileptic data set and the estimated treatment effects expressed a useful effect of LTG on UAEs over CBZ, but the two drugs were similar in terms of hazard for ISC. There are many tools for checking goodness of fit of the models. In this paper, we applied standardized subject-specific residuals and Cox-Snell residuals for checking goodness of fit of the longitudinal and competing risks sub-models, respectively. Here, the plots of standardized conditional residuals do not detect any lack of fit, despite the fact that the plots of Cox-Snell residuals show lack of fit for some places in the states with smaller number of observations. As a future work, Bayesian paradigm can be used to estimate the parameters. Also there is a real need to develop software that can easily use existing approaches to analyze clinical data sets that has the ability to diagnose better approaches.

ORCID

M. Esmailian http://orcid.org/0000-0001-9329-0243

References

Arellano-Valle, R. B., H. Bolfarine, and V. H. Lachos. 2007. Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics* 34 (6):663–682. doi:10.1080/02664760701236905.

Austin, P. 2012. Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in Medicine* 31 (29):3946–3958. doi:10.1002/sim.5452.

Azzalini, A. (2020). *'sn' package. The skew-Normal and related distributions such as the skew-t*. http://azzalini.stat.unipd.it/SN .

Baghfalaki, T., and M. Ganjali. 2015. A bayesian approach for joint modeling of skew-normal longitudinal measurements and time to event data. *REVSTAT Statistical Journal* 13 (2):169–191.

Baghfalaki, T., M. Ganjali, and D. Berridge. 2013. Robust joint modeling of longitudinal measurements and time to event data using normal independent distributions: A bayesian approach. *Biometrical Journal* 55 (6):844–865. doi:10.1002/bimj.201200272.

Baghfalaki, T., M. Ganjali, and D. Berridge. 2014a. Joint modeling of multivariate longitudinal mixed measurements and time to event data using a bayesian approach. *Journal of Applied Statistics* 41 (9):1934–1955. doi:10.1080/02664763.2014.898132.

Baghfalaki, T., M. Ganjali, and R. Hashemi. 2014b. Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. *Journal of Biopharmaceutical Statistics* 24 (4):834–855. doi:10.1080/10543406.2014.903657.

Baghfalaki, T., S. Kalantari, M. Ganjali, F. Hadaegh, and B. Pahlavanzadeh. 2020. Bayesian joint modeling of ordinal longitudinal measurements and competing risks survival data for analysing Tehran lipid and glucose study. *Journal of Biopharmaceutical Statistics* 30 (4):689–703.

Bellera, C., C. Proustlima, L. Joseph, P. Richaud, J. Taylor, H. Sandler, S. Mathoulin Pélissier, and S. Mathoulin-Pélissier. 2016. A two-stage model in a bayesian framework to estimate a survival endpoint in the presence of confounding by indication. *Statistical Methods in Medical Research* 27 (4):1271–1281. doi:10.1177/0962280216660127.

Brown, E., and J. Ibrahim. 2003. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Statistical Methods in Medical Research* 59 (2):221–228.

Cox, D. 1972. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34:187–220.

Cox, D., and E. Snell. 1968. A general definition of residuals. *Journal of the Royal Statistical Society, Series B 30* 248–275.

Diggle, P. J., I. Sousa, and A. G. Chetwynd. 2008. Joint modeling of repeated measurements and time?to?event outcomes: The fourth armitage lecture. *Statistics in Medicine* 27 (16):2981–2998. doi:10.1002/sim.3131.

Faucett, C., and D. C. Thomas. 1996. Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* 15 (15):1663–1685. doi:10.1002/(SICI)1097-0258-(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1.

Faught, E. 2007. Epilepsy drugs: Getting it right the first time. In *Lancet Neurol 6*, 476–478.

Gould, A. L., M. E. Boye, M. J. Crowther, J. G. Ibrahim, G. Quartey, S. Micallef, and F. Y. Bois. 2015. Joint modeling of survival and longitudinal non-survival data: Current methods and issues: Report of the dia bayesian joint modeling working group. *Statistical Medicine* 34 (14):2181–2195. doi:10.1002/sim.6141.

Guler, I., C. Faes, C. Cadarso-Suárez, L. Teixeira, A. Rodrigues, and D. Mendonca. 2017. Two-stage model for multivariate longitudinal and survival data with application to nephrology research. *Biometrical Journal* 59 (6):1204–1220. doi:10.1002/bimj.201600244.

Henderson, R., P. Diggle, and A. Dobson. 2000. Joint modeling of longitudinal measurements and event time data. *Biostatistics* 4 (4):465–480. doi:10.1093/biostatistics/1.4.465.

Hickey, G. L., P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. 2018. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181 (4):1105–1123. doi:10.1111/rssa.12348.

Hogan, J., and N. Laird. 1997. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 16 (3):239–258. doi:10.1002/(SICI)1097-0258(19970215)16:3<239::AID-SIM483>3.0.CO;2-X.

Huang, X. and R. M. Elashoff. 2010. A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics Interface* 3 (2):185–195. doi:10.4310/SII.2010.v3.n2.a6.

Huang, Y., G. Dagne, and L. Wu. 2011. Bayesian inference on joint models of hiv dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine* 30 (24):2930–2946. doi:10.1002/sim.4321

Huong, P. T. T., D. Nur, H. Pham, and A. Branford. 2018. A modified two-stage approach for joint modelling of longitudinal and time-to-event data. *Journal of Statistical Computation and Simulation* 88 (17):3379–3398. doi:10.1080/00949655.2018.1518449.

Ibrahim, J. G., M. Chen, and D. Sinha. 2004. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica* 14 (3):863–883.

Marson, A. G., A. M. Al-Kharusi, M. Alwaidh, R. Appleton, G. A. Baker, D. W. Chadwick, C. Cramp, O. C. Cockerell, P. N. Cooper, J. Doughty, et al. 2007. The sanad study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: An unblinded randomised controlled trial. *Lancet* 369 (9566):1000–1015. doi:10.1016/S0140-6736(07)60460-7.

Murawska, M., D. Rizopoulos, and E. Lesaffre. 2012. A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *Journal of Probability and Statistics* 2012:1–18. doi:10.1155/2012/194194.

Nobre, J., and J. Singer. 2007. Residuals analysis for linear mixed models. *Biometrical Journal* 6 (6):863–875. doi:10.1002/bimj.200610341.

Patsalos, P.N., Berry, D.J., Bourgeois, B.F., Cloyd, J.C., Glauser, T.A., Johannessen, S.I., Leppik, I.E., Tomson, T. and Perucca, E., 2008. Antiepileptic drugs—best practice guidelines for therapeutic drug monitoring: a position paper by the subcommission on therapeutic drug monitoring, ILAE Commission on Therapeutic Strategies. *Epilepsia*, 49(7): pp.1239–1276.

Proust-Lima, C., M. Sene, J. M. G. Taylor, and H. Jacqmin-Gadda. 2012. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods Medicine Research* 23 (1):74–90. doi:10.1177/0962280212445839.

Rajeswaran, J., E. H. Blackstone, and J. Barnard. 2018. Joint modeling of multivariate longitudinal data and competing risks using multiphase sub-models. *Statistics in Biosciences* 10 (3):651–685. doi:10.1007/s12561-018-9223-6.

Rizopoulos, D. 2010. Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* 35 (9):1–33. doi:10.18637/jss.v035.i09.
Rizopoulos, D. 2012a. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis* 56 (3):491–501. doi:10.1016/j.csda.2011.09.007.
Rizopoulos, D. 2012b. *Joint models for longitudinal and time-to-event data with applications in R*. New York: Chapman and Hall/CRC.
Sweeting, M., and S. Thompson. 2011. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal* 53 (5):750–763. doi:10.1002/bimj.201100052.
Therneau, T., and T. Lumley. 2012. Survival: Survival analysis including penalised likelihood. *R Package Version* 2:36. *http://cran.r-project.org/package=survival* .
Tsiatis, A. A., and M. Davidian. 2004. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* 14:809–834.
Verbeke, G., and G. Molenberghs (2000). Linear mixed models for longitudinal data.
Williamson, P. R., C. Tudur-Smith, J. W. Sander, and A. G. Marson. 2007. Importance of competing risks in the analysis of anti-epileptic drug failure. *Trials* 8:1–10.
Williamson, P. R., R. Kolamunnage-Dona, P. Philipson, and A. G. Marson. 2008. Joint modelling of longitudinal and competing risks data. *Statist in Medicine* 27 (30):6426–6438. doi:10.1002/sim.3451.
Wu, L., W. Liu, G. Y. Yi, and Y. Huang. 2012. Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics* 2012:1–17. doi:10.1155/2012/640153.
Wulfsohn, M. S., and A. A. Tsiatis. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics* 53 (1):330–339. doi:10.2307/2533118.
Ye, W., X. Lin, and J. Taylor. 2008. Semi parametric modeling of longitudinal measurements and time-to-event data a two stage regression calibration approach. *Biometrics* 64 (4):1238–1246. doi:10.1111/j.1541-0420.2007.00983.x.