Journal of Statistical Software

July 2016, Volume 71, Issue 3. doi: 10.18637/jss.v071.i03

JMFit: A SAS Macro for Joint Models of Longitudinal and Survival Data

Danjie Zhang
Gilead Sciences, Inc.

Ming-Hui Chen
University of Connecticut

Joseph G. Ibrahim
University of North Carolina

Mark E. Boye
Eli Lilly and Company

Wei Shen
Eli Lilly and Company

Abstract

Joint models for longitudinal and survival data now have a long history of being used in clinical trials or other studies in which the goal is to assess a treatment effect while accounting for a longitudinal biomarker such as patient-reported outcomes or immune responses. Although software has been developed for fitting the joint model, no software packages are currently available for simultaneously fitting the joint model and assessing the fit of the longitudinal component and the survival component of the model separately as well as the contribution of the longitudinal data to the fit of the survival model. To fulfill this need, we develop a SAS macro, called JMFit. JMFit implements a variety of popular joint models and provides several model assessment measures including the decomposition of AIC and BIC as well as ΔAIC and ΔBIC recently developed in Zhang, Chen, Ibrahim, Boye, Wang, and Shen (2014). Examples with real and simulated data are provided to illustrate the use of JMFit.

Keywords: AIC, BIC, patient-reported outcome (PRO), shared parameter model, time-varying covariates.

1. Introduction

The joint analysis of longitudinal and time-to-event outcomes has been widely published in statistical journals. One popular approach in joint modeling of longitudinal and survival data is based on shared random effects, where the longitudinal model and survival model share common random effects and these random effects then induce correlation between the longitudinal and survival components of the model. This family of joint models is also called

the "shared parameter models" (SPMs). There are two basic formulations of SPMs. The first is the "time trajectory model", denoted by SPM1, where one essentially substitutes the polynomial time trajectory function from the longitudinal model into the hazard function of the survival model, and in this case, the trajectory function acts like a time-varying covariate in the survival model. The second formulation, denoted by SPM2, is to directly include the random effects as covariates in the survival model. There are several R packages available in fitting joint models based on shared random effects, including **JM** (Rizopoulos 2012), **JMbayes** (Rizopoulos 2016), and **joineR** (Philipson, Sousa, Diggle, Williamson, Kolamunnage-Dona, and Henderson 2012). There is also a Stata module `stjm` (Crowther 2012; Crowther, Abrams, and Lambert 2013), which estimates shared random effects models. In addition, another R package, **lcmm** (Proust-Lima, Philipps, and Liquet 2016), estimates joint models based on shared latent classes.

One important issue in the joint modeling of longitudinal and survival data concerns the separate contribution of the model components to the overall goodness-of-fit of the joint model. Recently, Zhang *et al.* (2014) derived a novel decomposition of the AIC and BIC criteria into additive components that will allow us to assess the goodness of fit for each component of the joint model. Such a decomposition leads to the development of ΔAIC and ΔBIC, which quantify the change of AIC and BIC in fitting the survival data with and without using the longitudinal data. Thus, ΔAIC and ΔBIC can be used to determine the importance of the longitudinal data relative to the model fit of the survival data. In addition, ΔAIC and ΔBIC are also very useful in assessing whether a linear trajectory or quadratic trajectory is more suitable and also facilitating a direct comparison between SPM1s and SPM2s. These measures will help the data analyst in not only assessing each component of the joint model but also in determining the contribution of the longitudinal measures to the fit of the survival data. These newly developed model assessment criteria are not available in any of these packages or module mentioned before. We mention here that the methodology for ΔAIC and ΔBIC was fully developed in Zhang *et al.* (2014), but our goal here is the novel implementation of this methodology into user-friendly software along with a class of joint models for jointly analyzing longitudinal and time-to-event data.

This paper introduces `JMFit`, a SAS macro, that will allow us to fit the SPM1, SPM2, time-varying covariates, and two-stage models as well as to assess the goodness-of-fit of each of the longitudinal and survival components in the joint model. A detailed analysis of the longitudinal and survival data from a cancer clinical trial as well as an analysis of the simulated data are carried out to illustrate the functionality of `JMFit`. A detailed description of `JMFit` is given in Appendix A.

2. The models and model assessment

2.1. The joint models

Suppose that there are n subjects. For the ith subject, let $y_i(t)$ denote the longitudinal measure, which is observed at time $t \in \{a_{i1}, a_{i2}, \ldots, a_{im_i}\}$, where $0 \leq a_{i1} < a_{i2} < \cdots < a_{im_i}$ and $m_i \geq 1$. Here, $y_i(0)$ denotes the baseline value of the longitudinal measure. Let t_i and δ_i be the failure time and the censoring indicator such that $\delta_i = 1$ if t_i is a failure time and 0 if t_i is right-censored for the ith subject. We further let $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i$ denote a

p_L-dimensional vector of time-dependent covariates and a p_S-dimensional vector of baseline covariates, respectively. The joint model for (y_i, t_i) consists of the longitudinal component and the survival component.

For the longitudinal component, a mixed effects regression model is assumed for $y_i(t)$, which takes the form:

$$y_i(a_{ij}) = \boldsymbol{\theta}_i^\top \boldsymbol{g}(a_{ij}) + \boldsymbol{\gamma}^\top \boldsymbol{x}_i(t) + \epsilon_i(a_{ij}), \tag{1}$$

where $\boldsymbol{g}(a_{ij}) = (1, a_{ij}, a_{ij}^2, \ldots, a_{ij}^q)'$ is a polynomial vector of order q for $j = 1, \ldots, m_i$, $\boldsymbol{\theta}_i$ is a $(q+1)$-dimensional vector of random effects, and $\boldsymbol{\gamma}$ is a p-dimensional vector of regression coefficients. In (1), we further assume $\boldsymbol{\theta}_i \sim N(\boldsymbol{\theta}, \Omega)$, where $\boldsymbol{\theta}$ is the $(q+1)$-dimensional vector of the overall effects, Ω is a $(q+1) \times (q+1)$ positive definite covariance matrix with the lower triangle consisting of $\{\Omega_{00}, \Omega_{10}, \Omega_{11}, \ldots, \Omega_{qq}\}$, $\epsilon_i(a_{ij}) \sim N(0, \sigma^2)$, and $\boldsymbol{\theta}_i$ and $\epsilon_i(a_{ij})$ are independent. We note that in (1), if $q = 1$, $\boldsymbol{g}(a_{ij}) = (1, a_{ij})^\top$ and $\boldsymbol{\theta}_i^\top \boldsymbol{g}(a_{ij})$ represents a linear trajectory, and if $q = 2$, $\boldsymbol{g}(a_{ij}) = (1, a_{ij}, a_{ij}^2)^\top$ and $\boldsymbol{\theta}_i^\top \boldsymbol{g}(a_{ij})$ leads to a quadratic trajectory.

For the failure time t_i, we assume that the hazard function takes the form

$$\lambda(t|\lambda_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(t), \boldsymbol{z}_i) = \lambda_0(t) \exp\{\beta \boldsymbol{\theta}_i^\top \boldsymbol{g}(t) + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\} \tag{2}$$

or

$$\lambda(t|\lambda_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(t), \boldsymbol{z}_i) = \lambda_0(t) \exp\{\boldsymbol{\beta}^\top \boldsymbol{\theta}_i + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\} \tag{3}$$

where $\lambda_0(t)$ is the baseline hazard function, β is a one-dimensional regression coefficient in (2) and $\boldsymbol{\beta}$ is a $(q+1)$-dimensional vector of the regression coefficients in (3). Note that in (2) or (3), $\boldsymbol{\theta}_i$ and $\boldsymbol{g}(t)$ are the parameters and the functions from the longitudinal component of the joint model in (1) while λ_0, β (or $\boldsymbol{\beta}$), and $\boldsymbol{\alpha}$ are the only parameters pertaining to the survival component. As shown in Section 3.1, β (or $\boldsymbol{\beta}$) controls the association between the longitudinal marker and the time-to-event. A value of $\beta = 0$ (or $\boldsymbol{\beta} = \mathbf{0}$) implies no association between the longitudinal marker and the time-to-event. The joint model with hazard function specified in (2) is the trajectory model, denoted by SPM1, while the one with hazard function given by (3) is denoted by SPM2. Under SPM1, a positive value of β implies that a larger current value of the longitudinal marker is associated with a larger instantaneous hazard, whereas a negative value of β implies that a larger current value of the longitudinal marker is associated with a smaller instantaneous hazard.

2.2. The construction of the piecewise constant baseline hazard function

Assuming $\lambda_0(t)$ to be a piecewise constant baseline hazard function, we partition the time axis into J intervals with $0 = s_0^J < s_1^J < s_2^J < \ldots < s_{J-1}^J < s_J^J = \infty$. Then we assign a constant baseline hazard to each of the J intervals, that is,

$$\lambda_0(t) = \lambda_j, \quad t \in (s_{j-1}^J, s_j^J] \text{ for } j = 1, 2, \ldots, J. \tag{4}$$

Let $t_1^* \leq t_2^* \leq \cdots \leq t_{n^*}^*$ be the n^* event times of the t_i's, where $n^* = \sum_{i=1}^n \delta_i$. We consider four algorithms to construct the s_j^J's.

Algorithm 1: Equally-spaced quantile partition (ESQP)

Step 1: Compute $p_j = j/J$ for $j = 1, \ldots, J-1$.

Step 2: Let $n_j = [p_j n^*]$, which is the integer part of $p_j n^*$.

Step 3: Set

$$s_j^J = \begin{cases} (t^*_{n_j} + t^*_{n_j+1})/2 & \text{if } n_j = p_j n^*, \\ t^*_{n_j+1} & \text{if } n_j < p_j n^*, \end{cases} \tag{5}$$

for $j = 1, \ldots, J-1$.

The ESQP is a popular approach to construct the piecewise constant hazard function, which is discussed in Ibrahim, Chen, and Sinha (2001, Chapter 5) and also implemented in the R package **JM** developed by Rizopoulos (2010). We note that s_j^J is the p_jth quantile of the t_i^*'s and (5) is implemented in the SAS `UNIVARIATE` procedure as the default option for computing quantiles. We also note that the ESQP algorithm does not yield nested partitions in the sense that $\{s_j^{J_1},\ j = 1, \ldots, J_1\}$ is not necessarily a subset of $\{s_j^{J_2},\ j = 1, \ldots, J_2\}$ when $J_1 < J_2$. In order to construct nested partitions, we propose the following three bi-sectional quantile partition algorithms.

Algorithm 2: Left bi-sectional quantile partition (LBSQP)

Step 1: Decompose J into two parts:

$$J = 2^K + M, \tag{6}$$

where K and M are integers, and $M < 2^K$. In (6), $K = [\log J / \log 2]$, and then $M = J - 2^K$.

Step 2: Compute

(i) $a_k = k/2^K$, for $k = 1, \ldots, 2^K - 1$; and
(ii) $b_m = (2m-1)/2^{K+1}$, for $m = 1, \ldots, M(\geq 1)$.

Step 3: Sort $\{a_1, \ldots, a_{2^K-1}, b_1, \ldots, b_M\}$ in ascending order and the resulting ordered $J-1$ values are denoted by $p_1 \leq p_2 \leq \cdots \leq p_{J-1}$.

Step 4: Use Steps 2 and 3 of Algorithm 1 to compute $\{s_j^J,\ j = 1, \ldots, J-1\}$.

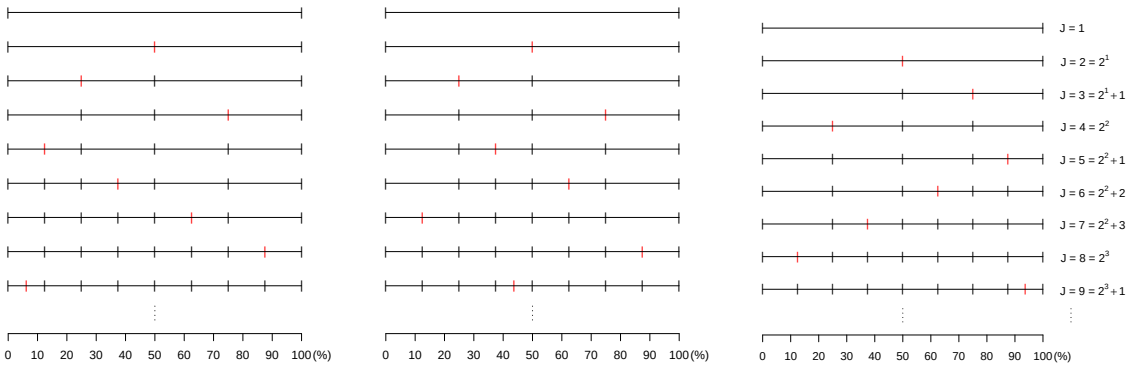
Algorithm 3: Middle bi-sectional quantile partition (MBSQP)

Step 1: The same as Algorithm 2.

Step 2: Compute

(i) $a_k = k/2^K$, for $k = 1, \ldots, 2^K - 1$; and
(ii) for $m = 1, \ldots, M(\geq 1)$,
  (a) $b_m = (2^K - m)/2^{K+1}$, for $m = 1, 3, 5, \ldots$; and
  (b) $b_m = (2^K + m - 1)/2^{K+1}$, for $m = 2, 4, 6, \ldots$.

Steps 3 and 4: The same as Algorithm 2.



(a) LBSQP (b) MBSQP (c) RBSQP

Figure 1: An illustration of three bi-sectional partition methods.

Algorithm 4: Right bi-sectional quantile partition (RBSQP)

Step 1: The same as Algorithm 2.

Step 2: Compute

(i) $a_k = k/2^K$, for $k = 1, \ldots, 2^K - 1$; and
(ii) $b_m = (2^{K+1} - (2m-1))/2^{K+1}$, for $m = 1, \ldots, M(\geq 1)$.

Steps 3 and 4: The same as Algorithm 2.

If there are ties in $\{s_j^J,\ j = 1, \ldots, J\}$, say $s_\ell^J = s_{\ell+1}^J$, the interval $(s_\ell^J, s_{\ell+1}^J]$ is undefined. Thus we only use distinct values of the s_j^J. Let J_t denote the number of ties in $\{s_j^J,\ j = 1, \ldots, J\}$. Then the number of distinct intervals reduces to $J - J_t$.

Figure 1 shows how the partition intervals are constructed based on LBSQP, MBSQP, and RBSQP.

Notice that when $J = 2^K$, $K = 1, 2, \ldots$, ESQP, LBSQP, MBSQP, and RBSQP yield the same partition. LBSQP, MBSQP, and RBSQP are desirable when there are more events at the beginning, in the middle, and at the end of the follow-up period, respectively. Another advantage of LBSQP, MBSQP, and RBSQP is that the resulting partitions are nested and, hence, the log-likelihood of the joint model increases in J when the longitudinal component remains fixed.

2.3. The joint likelihood

We rewrite (1) as follows:

$$\boldsymbol{y}_i = W_i(\boldsymbol{\theta}_i^\top, \boldsymbol{\gamma}^\top)^\top + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{y}_i = (y_i(a_{i1}), \ldots, y_i(a_{im_i}))^\top$, $W_i = ((g(a_{ij})^\top, \boldsymbol{x}_i(a_{ij})^\top)^\top, j = 1, \ldots, m_i)^\top$, and $\boldsymbol{\epsilon}_i = (\epsilon_i(a_{i1}), \ldots, \epsilon_i(a_{im_i}))^\top \sim N(\boldsymbol{0}, \sigma^2 I_{m_i})$. The complete-data likelihood function of the longitudinal measures for the ith subject is given by

$$L(\boldsymbol{\gamma}, \sigma^2 | \boldsymbol{y}_i, W_i, \boldsymbol{\theta}_i) = \frac{1}{(2\pi\sigma^2)^{\frac{m_i}{2}}} \exp\Big\{ -\frac{1}{2\sigma^2}(\boldsymbol{y}_i - W_i(\boldsymbol{\theta}_i^\top, \boldsymbol{\gamma}^\top)^\top)^\top (\boldsymbol{y}_i - W_i(\boldsymbol{\theta}_i^\top, \boldsymbol{\gamma}^\top)^\top)\Big\}, \quad (7)$$

for $i = 1, \ldots, n$. Note that the density of $\boldsymbol{\theta}_i$ is given by

$$f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) = \frac{|\Omega|^{-\frac{1}{2}}}{(2\pi)^{\frac{q+1}{2}}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta})^\top \Omega^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta})\Big\}. \tag{8}$$

Let $\boldsymbol{\varphi} = (\boldsymbol{\lambda}, \beta, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{\theta}, \Omega)$. Using (2) (or (3)), (7), and (8), the observed-data likelihood function for $(\boldsymbol{y}_i, t_i, \delta_i)$ for the ith subject is given by

$$L(\boldsymbol{\varphi}|\boldsymbol{y}_i, t_i, \delta_i, \boldsymbol{z}_i, W_i) = \int L(\boldsymbol{\lambda}, \beta, \boldsymbol{\alpha}|t_i, \delta_i, \boldsymbol{z}_i, \boldsymbol{\theta}_i, \boldsymbol{g}) L(\boldsymbol{\gamma}, \sigma^2|\boldsymbol{y}_i, W_i, \boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) d\boldsymbol{\theta}_i, \tag{9}$$

where the complete-data likelihood function for the survival component is written as

$$\begin{aligned} L(\boldsymbol{\lambda}, \beta, \boldsymbol{\alpha}|t_i, \delta_i, \boldsymbol{z}_i, \boldsymbol{\theta}_i, \boldsymbol{g}) = {} & \big[\lambda(t_i|\lambda_0, \beta, \boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(t_i), \boldsymbol{z}_i)\big]^{\delta_i} \\ & \times \exp\Big\{ -\int_0^{t_i} \lambda(u|\lambda_0, \beta, \boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(u), \boldsymbol{z}_i) du\Big\}, \end{aligned} \tag{10}$$

for $i = 1, \ldots, n$. In (2) (or (3)), when $\beta = 0$ (or $\boldsymbol{\beta} = \mathbf{0}$), the hazard function reduces to $\lambda(t|\lambda_0, \boldsymbol{\alpha}, \boldsymbol{z}_i) = \lambda_0(t) \exp(\boldsymbol{\alpha}^\top \boldsymbol{z}_i)$. In this case, we fit the survival data alone and the likelihood function in (10) for the ith subject reduces to

$$L_0(\boldsymbol{\lambda}, \boldsymbol{\alpha}|t_i, \delta_i, \boldsymbol{z}_i) = \{\lambda_0(t_i) \exp(\boldsymbol{\alpha}^\top \boldsymbol{z}_i)\}^{\delta_i} \exp\Big[-\exp(\boldsymbol{\alpha}^\top \boldsymbol{z}_i)\Big\{ \int_0^{t_i} \lambda_0(u) du\Big\}\Big]. \tag{11}$$

Letting $D_{\text{obs}} = \{(\boldsymbol{y}_i, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i),\ i = 1, \ldots, n\}$ denote the observed data, the joint likelihood for all subjects is given by

$$L(\boldsymbol{\varphi}|\boldsymbol{g}, D_{\text{obs}}) = \prod_{i=1}^{n} L(\boldsymbol{\varphi}|\boldsymbol{y}_i, t_i, \delta_i, \boldsymbol{z}_i, W_i). \tag{12}$$

2.4. AIC (BIC) decomposition and ΔAIC (ΔBIC)

Write $\boldsymbol{\varphi}_1 = (\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\theta}, \Omega)$ and $\boldsymbol{\varphi}_2 = (\boldsymbol{\lambda}, \beta, \boldsymbol{\alpha})$. Let $f(\boldsymbol{\theta}_i|\boldsymbol{y}_i, W_i, \boldsymbol{\varphi}_1)$ be the conditional density of the random effects $\boldsymbol{\theta}_i$ given $\boldsymbol{y}_i$, and also let $L(\boldsymbol{\varphi}_1|\boldsymbol{y}_i, W_i) = \int L(\boldsymbol{\gamma}, \sigma^2|\boldsymbol{y}_i, W_i, \boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) d\boldsymbol{\theta}_i$, which is the likelihood function corresponding to the marginal distribution of $\boldsymbol{y}_i$. Following Zhang *et al.* (2014), the joint likelihood given in (12) can be decomposed as

$$L(\boldsymbol{\varphi}|\boldsymbol{g}, D_{\text{obs}}) = L_{\text{Long}}(\boldsymbol{\varphi}_1|\boldsymbol{g}, D_{\text{obs}}) L_{\text{Surv|Long}}(\boldsymbol{\varphi}_2|\boldsymbol{g}, \boldsymbol{\varphi}_1, D_{\text{obs}}), \tag{13}$$

where $L_{\text{Long}}(\boldsymbol{\varphi}_1|\boldsymbol{g}, D_{\text{obs}}) = \prod_{i=1}^n L(\boldsymbol{\varphi}_1|\boldsymbol{y}_i, W_i)$ and $L_{\text{Surv|Long}}(\boldsymbol{\varphi}_2|\boldsymbol{g}, \boldsymbol{\varphi}_1, D_{\text{obs}}) = \prod_{i=1}^n \int L(\boldsymbol{\varphi}_2|t_i, \delta_i, \boldsymbol{z}_i, \boldsymbol{\theta}_i, \boldsymbol{g}) f(\boldsymbol{\theta}_i|\boldsymbol{y}_i, W_i, \boldsymbol{\varphi}_1) d\boldsymbol{\theta}_i$. Using (13), the decomposition of the total Akaike Information Criterion (AIC) (Akaike 1973) developed in Zhang *et al.* (2014) is given as

$$\text{AIC} = \text{AIC}_{\text{Long}} + \text{AIC}_{\text{Surv|Long}},$$

where $\text{AIC} = -2 \log L(\hat{\boldsymbol{\varphi}}|\boldsymbol{g}, D_{\text{obs}}) + 2 \dim(\boldsymbol{\varphi})$, $\text{AIC}_{\text{Long}} = -2 \log L_{\text{Long}}(\hat{\boldsymbol{\varphi}}_1|\boldsymbol{g}, D_{\text{obs}}) + 2 \dim(\boldsymbol{\varphi}_1)$, $\text{AIC}_{\text{Surv|Long}} = -2 \log L_{\text{Surv|Long}}(\hat{\boldsymbol{\varphi}}_2|\boldsymbol{g}, \hat{\boldsymbol{\varphi}}_1, D_{\text{obs}}) + 2 \dim(\boldsymbol{\varphi}_2)$, and $\hat{\boldsymbol{\varphi}}$, $\hat{\boldsymbol{\varphi}}_1$, and $\hat{\boldsymbol{\varphi}}_2$ are the maximum likelihood estimates (MLEs) of $\boldsymbol{\varphi}$, $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$. Similarly, the total Bayesian Information Criterion (BIC) (Schwarz 1978) for the joint model can be decomposed into

$$\text{BIC} = \text{BIC}_{\text{Long}} + \text{BIC}_{\text{Surv|Long}},$$

where BIC = AIC+dim($\boldsymbol{\varphi}$)(log n−2), $\text{BIC}_{\text{Long}} = \text{AIC}_{\text{Long}}$+dim($\boldsymbol{\varphi}_1$) (log n−2), $\text{BIC}_{\text{Surv|Long}} = \text{AIC}_{\text{Surv|Long}} + \dim(\boldsymbol{\varphi}_2)(\log n - 2)$. Using the decompositions of AIC and BIC, Zhang *et al.* (2014) proposed two new model assessment criteria given by

$$\Delta\text{AIC} = \text{AIC}_{\text{Surv},0} - \text{AIC}_{\text{Surv|Long}},$$
$$\Delta\text{BIC} = \text{BIC}_{\text{Surv},0} - \text{BIC}_{\text{Surv|Long}},$$

where

$$\text{AIC}_{\text{Surv},0} = -2\sum_{i=1}^{n} \log L_0(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}|t_i, \delta_i, \boldsymbol{z}_i) + 2\dim(\boldsymbol{\lambda}, \boldsymbol{\alpha}),$$
$$\text{BIC}_{\text{Surv},0} = -2\sum_{i=1}^{n} \log L_0(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}|t_i, \delta_i, \boldsymbol{z}_i) + \dim(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \log n,$$

and $L_0(\boldsymbol{\lambda}, \boldsymbol{\alpha}|t_i, \delta_i, \boldsymbol{z}_i)$ is defined by (11). The ΔAIC or ΔBIC measure the gain of the fit in the survival component due to the longitudinal data with a penalty for the additional parameters in the survival component of the joint model. The model with a large value of ΔAIC (ΔBIC) is more preferred.

3. The SAS macro JMFit

3.1. Design

The SAS macro `JMFit` has been developed to assess model fit in joint models of longitudinal and survival data. In fact, it can fit five models, including the two types of joint models with linear and quadratic trajectories, as well as the time-varying covariates model. The macro `JMFit` consists of five submacros, `SPM1L`, `SPM1Q`, `SPM2L`, `SPM2Q`, and `TVC`, corresponding to the five models, respectively. The `MODEL` argument of `JMFit` specifies one of the following five models to be fitted:

`SPM1L`: SPM1 with Linear trajectory. The hazard function has the form

$$\lambda(t|\lambda_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}_i, \boldsymbol{g}(t), \boldsymbol{z}_i) = \lambda_0(t) \exp\{\beta(\theta_{0i} + \theta_{1i}t) + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\}.$$

`SPM1Q`: SPM1 with Quadratic trajectory. The hazard function has the form

$$\lambda(t|\lambda_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}_i, \boldsymbol{g}(t), \boldsymbol{z}_i) = \lambda_0(t) \exp\{\beta(\theta_{0i} + \theta_{1i}t + \theta_{2i}t^2) + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\}.$$

`SPM2L`: SPM2 with Linear trajectory. The hazard function has the form

$$\lambda(t|\lambda_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}_i, \boldsymbol{g}(t), \boldsymbol{z}_i) = \lambda_0(t) \exp\{\beta_1\theta_{0i} + \beta_2\theta_{1i} + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\},$$

where θ_{0i} and θ_{1i} are subject-level random intercept and random slope.

`SPM2Q`: SPM2 with Quadratic trajectory. The hazard function has the form

$$\lambda(t|\lambda_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}_i, \boldsymbol{g}(t), \boldsymbol{z}_i) = \lambda_0(t) \exp\{\beta_1\theta_{0i} + \beta_2\theta_{1i} + \beta_3\theta_{2i} + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\},$$

where θ_{0i}, θ_{1i}, and θ_{2i} are random effects (i.e., random intercept, random slope, and random quadratic coefficient).

`TVC`: Time-Varying Covariates model. The hazard function has the form

$$\lambda(t|\lambda_0, \beta, \boldsymbol{\alpha}, \boldsymbol{z}_i, y_i(t)) = \lambda_0(t) \exp\{\beta y_i(t) + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\},$$

where $y_i(t) = y_i(a_{ij})$ for $a_{ij} \le t < a_{i,j+1}$ for $j = 1, \ldots, m_i$, where $a_{i,m_i+1} = \infty$. Note that the TVC model is a "non-joint" model, and the use of this model has great potential for bias (Fisher and Lin 1999).

We provide the two versions of SPM1L, SPM1Q, SPM2L, and SPM2Q with the `TS` argument. If `TS` is missing or equal to `0`, the joint model will be fit; while `TS = 1` yields the corresponding two-stage model. Similar to the method in Tsiatis, DeGruttola, and Wulfsohn (1995), (i) we first fit the linear mixed model specified in (1) to the longitudinal data alone and then obtain the estimates of $\boldsymbol{\theta}_i$, denoted by $\hat{\boldsymbol{\theta}}_i$; and (ii) we replace $\boldsymbol{\theta}_i$ in (2) or (3) with the estimate $\hat{\boldsymbol{\theta}}_i$ at the second stage. The only difference between the joint model and the corresponding two-stage model is that $\boldsymbol{\theta}_i$ in (2) or (3) is replaced with the estimate $\hat{\boldsymbol{\theta}}_i$. This two-stage approach may potentially lead to biased and inefficient estimates (Ibrahim, Chu, and Chen 2010).

The number of intervals J (≥ 1) for the piecewise constant baseline hazard function needs to be specified in the `NPIECES` argument. For the `PARTITION` argument, `1` represents ESQP, `2` represents LBSQP, `3` represents MBSQP, and `4` represents RBSQP.

`JMFit` automatically produces a rich text file (RTF) including five tables: (i) Number of Subjects; (ii) Fit Statistics; (iii) Survival Parameter Estimates (Survival Alone); (iv) Parameter Estimates; and (v) Hazard Ratios & λ Estimates.

3.2. Implementation details

If the observed longitudinal measures are sparse, the full trajectories of longitudinal measures might not be well estimated. For example, in the case of fitting a quadratic trajectory, the sign of the estimated second-order coefficient could be incorrect if the longitudinal measures were observed only within the first half of the follow-up period, leading to incorrect extrapolation when the observed progression time was far beyond the time of the last observed longitudinal measure.

Let $t_{\max,i} = \max_{1\le j\le m_i}\{a_{ij}\}$. When $t > t_{\max,i}$, $y_i(t)$ is never observed and no longitudinal data are available to estimate the trajectory $\boldsymbol{\theta}_i^\top \boldsymbol{g}(t)$ for $t > t_{\max,i}$. Under SPM1, the extrapolation of the trajectory $\boldsymbol{\theta}_i^\top \boldsymbol{g}(t)$ beyond $t_{\max,i}$ may lead to a survival component of the joint model that fits the survival data poorly. In addition, such an extrapolation also causes a severe convergence problem in the SAS `NLMIXED` (SAS Institute Inc. 2011b) procedure especially when $t_{\max,i} \ll t_i$ for many subjects. To circumvent these issues, for SPM1, we modify the hazard function in (2) as

$$\lambda(t \mid \lambda_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}_i, \boldsymbol{g}, \boldsymbol{z}_i) = \lambda_0(t) \exp\{\beta \boldsymbol{\theta}_i^\top \boldsymbol{g}(t - [t - t^*_{\max,i}]_+) + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\}, \tag{14}$$

or

$$\begin{aligned} &\lambda(t \mid \lambda_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}_i, \boldsymbol{g}, \boldsymbol{z}_i) \\ &= \lambda_0(t) \exp\{\beta \boldsymbol{\theta}_i^\top \boldsymbol{g}(t - [t - t^*_{\max,i}]_+) \times \frac{\tau - (t^*_{\max,i} + [t - t^*_{\max,i}]_+)}{\tau - t^*_{\max,i}} + \boldsymbol{\alpha}^\top \boldsymbol{z}_i\}, \end{aligned} \tag{15}$$

where $[t - t^*_{\max,i}]_+ = \max(t - t^*_{\max,i}, 0)$, $t^*_{\max,i} = t_{\max,i} + w \times \max(t_i - t_{\max,i}, 0)$, w ($\in [0, 1]$) is the proportion of $\max(t_i - t_{\max,i}, 0)$, and $\tau = \max_{1\le i\le n}\{t_i\}$, which is the last follow-up survival

time. If $w = 0$ (default), $t_{\max,i}$ will be the starting point of the modified extrapolation of the trajectory; while $w = 1$ implies that the trajectory extends to t_i with no $t_{\max,i}$ adjustment. This modification can also be applied to the TS model corresponding to SPM1L or SPM1Q. There are two arguments called `TMAXI` and `WEIGHT`. If `TMAXI` is missing or equal to 0, no $t_{\max,i}$ adjustment will be applied. If `TMAXI = 1`, the $t_{\max,i}$ adjustment based on hazard function given in (14) will be applied with weight given in the `WEIGHT` argument, that is, the trajectory will become flat after $t^*_{\max,i}$. If `TMAXI = 2`, the $t_{\max,i}$ adjustment based on hazard function given in (15) will be applied with weight given in the `WEIGHT` argument, that is, starting at $t^*_{\max,i}$, the trajectory will linearly go down to 0 at the last follow-up survival time. An "optimal" choice of weight w in conjunction with `TMAXI` option may be determined by either $\text{AIC}_{\text{Surv|Long}}$ or $\text{BIC}_{\text{Surv|Long}}$. The purpose of defining $t_{\max,i}$ is that we create an extrapolation of the longitudinal measures so that the trajectory function can be well estimated. This extrapolation is needed when there are few longitudinal measures at later points. We note here that the $t_{\max,i}$ method corresponds to a "prediction carried forward" approach, which may potentially induce bias in the estimation. We also note that this is not an issue for the SPM2, in which the hazard function is independent of $\boldsymbol{g}(t)$.

The `OPTIONS` argument allows users to specify options (e.g., integration method, optimization technique, and convergence criteria) that are available in the `PROC NLMIXED` statement. If `OPTIONS` is missing, `JMFit` will use adaptive Gaussian quadrature to approximate the integral of the likelihood over the random effects, perform a quasi-Newton optimization, and apply a relative gradient convergence criterion of 10^{-8}. All the methods mentioned above are default methods in `PROC NLMIXED`. The Riemann integral is used to compute the cumulative hazard function for the trajectory models. Each time interval is divided into 200 subintervals.

A big challenge in fitting joint models using the SAS `NLMIXED` procedure is convergence. Poor initial values may lead to the failure of convergence in `NLMIXED`. To address this issue, we first fit the longitudinal data alone using the SAS `MIXED` procedure to obtain the estimates, $\hat{\boldsymbol{\gamma}}$, $\hat{\sigma}^2$, $\hat{\boldsymbol{\theta}}$, $\hat{\Omega}$, $\hat{\boldsymbol{\theta}}_i$, for the parameters in the longitudinal component of the joint model. Using $(\hat{\boldsymbol{\theta}}_i,\ i = 1, \ldots, n)$ to replace $(\boldsymbol{\theta}_i,\ i = 1, \ldots, n)$ in (2) (or (3)), we fit the survival data alone to obtain the estimates, $\hat{\beta}$ (or $\hat{\boldsymbol{\beta}}$), $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\lambda}}$, for the parameters in the survival component of the joint model. Finally, these estimates are used as the initial values for the joint model.

`JMFit` does not exclude any longitudinal measures for the joint models. If one wishes to exclude the longitudinal measures observed after the survival time t_i, those longitudinal measures should be pre-excluded in the input longitudinal data for `JMFit`. "CAUTION: Longitudinal measures are observed after the survival time." will be given at the end of the output file if there are any longitudinal measures observed after the survival time.

4. Examples

4.1. The IBCSG data

To illustrate how `JMFit` works, we use the data from a clinical trial in premenopausal women with node-positive breast cancer conducted by the IBCSG (International Breast Cancer Study Group 1996). Each participant was randomly assigned in a 2×2 factorial design to receive: (A) cyclophosphamide, methotrexate, and fluorouracil for 6 consecutive courses on months 1 to 6 (CMF6); (B) CMF6 plus three single courses of reintroduction CMF given on months 9,

Number of Subjects	
Subjects_in_coping	1015
Subjects_in_os	1015
Subjects_Used	1015

Figure 2: The number of subjects for coping.

12, and 15; (C) CMF for three consecutive courses on months 1 to 3 (CMF3); or (D) CMF3 plus three single courses of reintroduction CMF given on months 6, 9, and 12. Four indicators of patients' quality of life (QOL), appetite, perceived coping, mood, and physical well-being were collected in this trial. We consider a subset of the IBCSG data, which consists of 1015 patients with at least three values of each longitudinal measure and six binary covariates, including treatment (A: `1` = `icyc`, `0` = `reint`, and `0` = `intera`; B: `1` = `icyc`, `1` = `reint`, and `1` = `intera`; C: `0` = `icyc`, `0` = `reint`, and `0` = `intera`; and D: `0` = `icyc`, `1` = `reint`, and `0` = `intera`), age (`1` = '$>$ 40' and `0` = '$\leq$ 40'), the number of positive nodes of the tumor (`1` = '$>$ 4' and `0` = '$\leq$ 4'), and estrogen receptor (ER) status (`1` = positive and `0` = negative). To satisfy the normality assumption for each of the four QOL indicators, following Chi and Ibrahim (2006), the corresponding observed value of each QOL was transformed to $\sqrt{100 - \text{QOL}}$ so that smaller value reflects better QOL. For the subset of 1015 patients, after a median follow-up of 9.68 years (interquartile range, 8.10–11.04 years), 296 patients died. We note that QOL indicators were collected only up to 1.65 years (about 18 months). By applying joint models in this study, we compare the longitudinal QOL indicators in terms of their contributions to the fit of overall survival (OS) via ΔAIC and ΔBIC, which are computed using `JMFit`.

Suppose we fit the trajectory model with a linear trajectory as considered by Chi and Ibrahim (2006) and Zhu, Ibrahim, Chi, and Tang (2012) and $J = 9$ for coping. Then we create two data sets named as "`coping`" and "`os`" for the macro `JMFit`'s `LONG` and `SURV` options, respectively. The `MODEL` option is set to "`SPM1L`" and `TS` is set to 0. We assign 2 to `TMAXI` for $t_{\max,i}$ adjustment with `WEIGHT`=0.5. The `NPIECES` option is set to 9 and the `PARTITION` option is set to 2 for the LBSQP algorithm.

`JMFit` is called:

```
%JMFit(LONG = coping, SURV = os, MODEL = SPM1L, TS = 0, TMAXI = 2,
  WEIGHT = 0.5, NPIECES = 9, PARTITION = 2);
```

The RTF file '`Output for coping under SPM1L with J=9 (Partition=2).rtf`' is generated by `JMFit`, which lists five tables. Table "Number of Subjects" given in Figure 2 shows that there are 1015 patients in data sets `coping` and `os`, respectively, and 1015 subjects are used, implying that the IDs of the subjects in these two data sets match.

Figure 3 shows the "Fit Statistics" table, which consists of the log likelihood, AIC_{Long} (BIC_{Long}), AIC_{Surv} (BIC_{Surv}), and ΔAIC (ΔBIC). Inside the `JMFit` macro, `PROC NLMIXED` provides the log likelihood, AIC and BIC, and `PROC IML` (SAS Institute Inc. 2011a) is used to compute AIC_{Long} and BIC_{Long}.

Next, the table, titled "Survival Parameter Estimates (Survival Alone)" shown in Figure 4, is obtained by fitting the survival data alone. For this table, the estimate, the standard error

Fit Statistics			
Log Likelihood	-8069.53		
AIC	16195.07	BIC	16332.90
AIC_{Long}	13830.46	BIC_{Long}	13889.54
$AIC_{Surv\|Long}$	2364.60	$BIC_{Surv\|Long}$	2443.36
$AIC_{Surv,0}$	2400.95	$BIC_{Surv,0}$	2474.79
ΔAIC	36.35	ΔBIC	31.43

Figure 3: The fit statistics for coping under SPM1L with $J = 9$ based on LBSQP.

Survival Parameter Estimates (Survival Alone)							
Param/Var	**Estimate**	**SE**	**DF**	**T-Value**	**P-Value**	**95% CI**	**Gradient**
icyc	-0.00126	0.1642	1015	-0.01	0.9939	(-0.323, 0.321)	-0.00133
reint	-0.2772	0.1707	1015	-1.62	0.1046	(-0.612, 0.058)	-0.00141
intera	0.2678	0.2337	1015	1.15	0.2520	(-0.191, 0.726)	-0.00211
agegrp	-0.2070	0.1356	1015	-1.53	0.1272	(-0.473, 0.059)	0.000143
nodegrp	0.8926	0.1170	1015	7.63	<.0001	(0.663, 1.122)	-0.00009
er_stat	-0.1982	0.1275	1015	-1.56	0.1203	(-0.448, 0.052)	-0.00221
log λ_1	-4.9829	0.2846	1015	-17.51	<.0001	(-5.541, -4.424)	-0.00031
log λ_2	-2.7823	0.2848	1015	-9.77	<.0001	(-3.341, -2.224)	-0.00051
log λ_3	-2.9729	0.2373	1015	-12.53	<.0001	(-3.439, -2.507)	0.003588
log λ_4	-2.7064	0.2359	1015	-11.47	<.0001	(-3.169, -2.244)	-0.00073
log λ_5	-2.6998	0.2364	1015	-11.42	<.0001	(-3.164, -2.236)	-0.00167
log λ_6	-2.9433	0.2355	1015	-12.50	<.0001	(-3.406, -2.481)	0.001627
log λ_7	-2.8127	0.2387	1015	-11.78	<.0001	(-3.281, -2.344)	-0.00293
log λ_8	-3.0135	0.2367	1015	-12.73	<.0001	(-3.478, -2.549)	0.001908
log λ_9	-3.4253	0.2358	1015	-14.53	<.0001	(-3.888, -2.963)	-0.00119

Figure 4: The estimates of the parameters obtained by (11) with $J = 9$ based on LBSQP.

(SE), the degrees of freedom (DF), the t value, the p value, and the 95% confidence interval (CI) are all shown for each parameter. In addition, the gradient of the negative log-likelihood function is displayed, which can be used to check convergence of `PROC NLMIXED`. A small gradient implies better convergence. The largest absolute value of the gradients shown in Figure 4 is 0.003588, indicating good convergence of `PROC NLMIXED`.

`PROC NLMIXED` produces the table titled "Parameter Estimates" in Figure 5, which consists of three subtables: "Covariance Parameter Estimates", "Longitudinal Parameter Estimates", and "Survival Parameter Estimates". For each parameter in this table, the estimate, SE, DF, t value, p value, 95% CI, and gradient are all provided. The Subtable "Covariance Parameter Estimates" consists of the estimates of the three lower-triangle elements $(\Omega_{00}, \Omega_{10}, \Omega_{11})$ of the

Parameter Estimates							
Param/Var	**Estimate**	**SE**	**DF**	**T-Value**	**P-Value**	**95% CI**	**Gradient**
Covariance Parameter Estimates							
Ω_{00}	0.5350	0.03274	1013	16.34	<.0001	(0.471, 0.599)	-0.08107
Ω_{10}	-0.05658	0.02094	1013	-2.70	0.0070	(-0.098, -0.015)	0.101966
Ω_{11}	0.1725	0.02650	1013	6.51	<.0001	(0.120, 0.224)	-0.02461
σ	0.6149	0.007308	1013	84.14	<.0001	(0.601, 0.629)	0.087386
Longitudinal Parameter Estimates							
Intercept	0.08643	0.07928	1013	1.09	0.2759	(-0.069, 0.242)	0.123621
t	-0.4004	0.02277	1013	-17.58	<.0001	(-0.445, -0.356)	-0.15625
icyc	0.1625	0.07073	1013	2.30	0.0218	(0.024, 0.301)	0.015924
reint	0.1259	0.07007	1013	1.80	0.0726	(-0.012, 0.263)	0.313939
intera	-0.06607	0.09767	1013	-0.68	0.4989	(-0.258, 0.126)	0.002171
agegrp	0.05158	0.05907	1013	0.87	0.3828	(-0.064, 0.167)	-0.13258
nodegrp	0.02780	0.05384	1013	0.52	0.6057	(-0.078, 0.133)	-0.18797
er_stat	0.02232	0.05408	1013	0.41	0.6799	(-0.084, 0.128)	-0.05203
Survival Parameter Estimates							
icyc	-0.06267	0.1695	1013	-0.37	0.7116	(-0.395, 0.270)	0.024537
reint	-0.3355	0.1756	1013	-1.91	0.0563	(-0.680, 0.009)	-0.01849
intera	0.3464	0.2406	1013	1.44	0.1502	(-0.126, 0.818)	0.14234
agegrp	-0.1827	0.1397	1013	-1.31	0.1913	(-0.457, 0.091)	0.049036
nodegrp	0.9522	0.1217	1013	7.82	<.0001	(0.713, 1.191)	0.03694
er_stat	-0.1540	0.1315	1013	-1.17	0.2418	(-0.412, 0.104)	-0.00715
$\log \lambda_1$	-4.9259	0.2888	1013	-17.06	<.0001	(-5.493, -4.359)	0.016501
$\log \lambda_2$	-2.5917	0.2923	1013	-8.87	<.0001	(-3.165, -2.018)	-0.02471
$\log \lambda_3$	-2.7430	0.2482	1013	-11.05	<.0001	(-3.230, -2.256)	-0.0547
$\log \lambda_4$	-2.4318	0.2494	1013	-9.75	<.0001	(-2.921, -1.942)	-0.06882
$\log \lambda_5$	-2.3952	0.2519	1013	-9.51	<.0001	(-2.889, -1.901)	-0.07637
$\log \lambda_6$	-2.6149	0.2529	1013	-10.34	<.0001	(-3.111, -2.119)	-0.08057
$\log \lambda_7$	-2.4741	0.2571	1013	-9.62	<.0001	(-2.979, -1.970)	-0.08093
$\log \lambda_8$	-2.6860	0.2547	1013	-10.54	<.0001	(-3.186, -2.186)	-0.0729
$\log \lambda_9$	-3.1547	0.2510	1013	-12.57	<.0001	(-3.647, -2.662)	-0.05614
β	0.2643	0.06188	1013	4.27	<.0001	(0.143, 0.386)	0.314192

Figure 5: The estimates of the parameters for coping under SPM1L with $J = 9$ based on LBSQP.

Hazard Ratios & λ Estimates		
Parameter	**Estimate**	**95% CI**
HR_icyc	0.9393	(0.674, 1.310)
HR_reint	0.7150	(0.507, 1.009)
HR_intera	1.4140	(0.882, 2.267)
HR_agegrp	0.8330	(0.633, 1.096)
HR_nodegrp	2.5915	(2.041, 3.291)
HR_er_stat	0.8573	(0.662, 1.110)
λ_1	0.007256	(0.004, 0.013)
λ_2	0.07489	(0.042, 0.133)
λ_3	0.06438	(0.040, 0.105)
λ_4	0.08788	(0.054, 0.143)
λ_5	0.09116	(0.056, 0.149)
λ_6	0.07318	(0.045, 0.120)
λ_7	0.08424	(0.051, 0.140)
λ_8	0.06815	(0.041, 0.112)
λ_9	0.04265	(0.026, 0.070)
HR_β	1.3025	(1.154, 1.471)

Figure 6: The hazard ratios and $\boldsymbol{\lambda}$ estimates for coping under SPM1L with $J = 9$ based on LBSQP.

random-effects covariance matrix as well as the estimate of the standard deviation (σ) for the error term. The estimates of the coefficients of the variables from the longitudinal component are shown in Subtable "Longitudinal Parameter Estimates". In Subtable "Survival Parameter Estimates", "Param/Var" lists all the names of the variables as well as the parameters from the survival component. In this subtable, $\log \lambda_1$, $\log \lambda_2$, ..., and $\log \lambda_9$ are the natural logarithms of the piecewise baseline hazards for the three intervals, respectively. The hazard ratios of the covariates and β as well as the estimates of $\boldsymbol{\lambda}$ are shown in Figure 6. From Figure 5, we see that the number of induction cycles has p values of 0.0218 and 0.7116 in the longitudinal and survival submodels, respectively, indicating that the effect of the number of induction courses is significant at the 0.05 level in the longitudinal submodel and is not statistically significant in the survival submodel. The reintroduction of CMF and the interaction between the number of initial cycles and the reintroduction do not have significant effects on QOL or OS based on the large p values. From Figure 5, we also see that (i) the number of positive nodes is highly significant in the survival submodel, implying that a greater number of positive nodes would increase the risk of death, and (ii) a significant coefficient β indicates that coping is highly associated with OS.

From the "Fit Statistics" tables in Figure 3 and Figure 7, we see that coping had the largest values of ΔAIC and ΔBIC and physical well-being had the smallest values of ΔAIC and ΔBIC, which indicate that coping led to the most gain in fitting the OS while physical well-being had the least contribution to the fit of the OS.

Fit Statistics			
Log Likelihood	-8707.14		
AIC	17470.29	BIC	17608.12
AIC_{Long}	15085.29	BIC_{Long}	15144.36
$AIC_{Surv\|Long}$	2385.00	$BIC_{Surv\|Long}$	2463.76
$AIC_{Surv,0}$	2400.95	$BIC_{Surv,0}$	2474.79
ΔAIC	15.96	ΔBIC	11.04

(a) Appetite

Fit Statistics			
Log Likelihood	-8541.07		
AIC	17138.14	BIC	17275.97
AIC_{Long}	14760.26	BIC_{Long}	14819.33
$AIC_{Surv\|Long}$	2377.88	$BIC_{Surv\|Long}$	2456.64
$AIC_{Surv,0}$	2400.95	$BIC_{Surv,0}$	2474.79
ΔAIC	23.08	ΔBIC	18.15

(b) Mood

Fit Statistics			
Log Likelihood	-8767.95		
AIC	17591.90	BIC	17729.73
AIC_{Long}	15197.49	BIC_{Long}	15256.57
$AIC_{Surv\|Long}$	2394.40	$BIC_{Surv\|Long}$	2473.17
$AIC_{Surv,0}$	2400.95	$BIC_{Surv,0}$	2474.79
ΔAIC	6.55	ΔBIC	1.63

(c) Physical well-being

Figure 7: The fit statistics for appetite, mood, and physical well-being under SPM1L with $J = 9$ based on LBSQP.

	$AIC_{Surv\|Long}$			ΔAIC		
J	LBSQP	MBSQP	RBSQP	LBSQP	MBSQP	RBSQP
1	2528.20	2528.20	2528.20	3.70	3.70	3.70
2	2512.19	2512.19	2512.19	8.84	8.84	8.84
3	2449.81	2449.81	2508.67	20.41	20.41	9.65
4	2445.98	2445.98	2445.98	21.52	21.52	21.52
5	2406.02	2447.95	2443.96	29.28	21.55	22.43
6	2407.99	2449.59	2445.61	29.32	21.60	22.47
7	2409.59	2409.59	2447.57	29.41	29.41	22.50
8	2407.34	2407.34	2407.34	30.54	30.54	30.54
9	2364.60	2409.38	2409.22	36.35	30.50	30.64
10	2365.65	2409.90	2411.05	36.54	30.60	30.63

Table 1: $AIC_{Surv|Long}$'s and ΔAIC's for coping under SPM1L with different partition algorithms and different J's.

Since $\Delta AIC = AIC_{Surv,0} - AIC_{Surv|Long}$, both $AIC_{Surv,0}$ and $AIC_{Surv|Long}$ change when J changes. Thus, the "best J" based on ΔAIC may not necessarily lead to the best survival submodel or the best joint model. As an illustration, Table 1 shows the values of $AIC_{Surv|Long}$ and ΔAIC for coping under SPM1L with different partition algorithms and different J's. Note `TMAXI` is set to 2 with `WEIGHT` = 0.5 in all the models. From Table 1, we see that (i) $J = 10$ or $J = 9$ is the "best" choice according to ΔAIC; (ii) $J = 9$ or $J = 8$ is the "best" one based on $AIC_{Surv|Long}$; and (iii) the value of $AIC_{Surv|Long}$ is 2364.60 for LBSQP when

	Coping		Physical well-being	
w	`TMAXI` $= 1$	`TMAXI` $= 2$	`TMAXI` $= 1$	`TMAXI` $= 2$
0	2402.88	2402.95	2402.78	2402.54
0.25	2384.23	2376.58	2399.27	2396.98
0.5	2383.51	2364.60	2399.10	2394.40
0.75	2395.76	2385.60	2401.70	2399.09
1[a]	2402.95	2402.95	2402.82	2402.82

Table 2: $\text{AIC}_{\text{Surv|Long}}$'s for coping and physical well-being under SPM1L with $t_{\max,i}$ adjustments using different w's. [a] No $t_{\max,i}$ adjustment

$J = 9$, which is the smallest one among all the combination of the partition algorithms and J's. Since our goal is to fit the joint model to both the longitudinal data and the survival data, it is more appropriate to use $\text{AIC}_{\text{Surv|Long}}$ to determine the number of intervals for the piecewise constant baseline hazard function.

Table 2 shows the values of $\text{AIC}_{\text{Surv|Long}}$ for coping and physical well-being under SPM1L with $J = 9$ and $t_{\max,i}$ adjustments using different w's. We can see from Table 2 that, among the five values of w, for both $t_{\max,i}$ adjustments, coping and physical well-being have the smallest values of $\text{AIC}_{\text{Surv|Long}}$ at $w = 0.5$. For both coping and physical well-being, `TMAXI` $=$ 2 outperforms `TMAXI` $= 1$ except for $w = 0$ for coping. In addition, both coping and physical well-being have the largest values of $\text{AIC}_{\text{Surv|Long}}$ at $w = 1$, indicating that the full trajectories were not well estimated.

4.2. A simulated data example

We consider a simulated data example, which is similar to the data used in Zhang *et al.* (2014). We generated a simulated data set with $n = 400$ subjects as follows. First, the time points (a_{ij}'s) at which the longitudinal measures were taken were fixed at (0, 21, 42, 63, 84, 105, 126)/30.4375. For each subject, we generated seven binary covariates, $(x_{i1}, \ldots, x_{i7})$, independently from Bernoulli distributions with success probabilities (i.e., $P(x_{ij} = 1)$, $j = 1, \ldots, 7$)) 0.49, 0.92, 0.81, 0.49, 0.38, 0.56, and 0.74, respectively. These proportions were estimated from the data in Zhang *et al.* (2014), corresponding to the covariates treatment, race/ethnicity, gender, age, Karnofsky status, baseline stage of disease, and vitamin supplementation, respectively. Second, we simulated the longitudinal trajectory as

$$\mu_i(a_{ij}) = (\theta_0 + \theta_{0i}) + (\theta_1 + \theta_{1i})a_{ij} + \boldsymbol{\gamma}\boldsymbol{x}_i,$$

where $\theta_0 = 0.62$, $\theta_1 = 0.04$, $\boldsymbol{\gamma} = (-0.11, -0.10, -0.18, -0.06, -0.58, -0.09, 0.10)^\top$, and $(\theta_{0i}, \theta_{1i})^\top \sim N\left(\mathbf{0}, \begin{pmatrix} 0.62 & -0.04 \\ -0.04 & 0.06 \end{pmatrix}\right)$. Finally, we generated the longitudinal data from a $N(\mu_i(a_{ij}), \sigma^2)$ distribution with $\sigma = 0.54$ and t_i^* as

$$t_i^* = -\frac{\log(1-U)}{\lambda \exp\{\beta_1(\theta_0 + \theta_{0i}) + \beta_2(\theta_1 + \theta_{1i}) + \boldsymbol{\alpha}\boldsymbol{x}_i\}},$$

where $\boldsymbol{\alpha} = (-0.36, 0.15, 0.04, -0.003, -0.33, -0.38, -0.07)^\top$, $\beta_1 = 0.26$, $\beta_2 = 1.17$, $\lambda = \exp(-1.67)$, and $U \sim U(0, 1)$. This longitudinal dataset is denoted by D_{Long}. Note that the

values of the parameters were obtained by fitting SPM2L to the data in Zhang *et al.* (2014) in which the longitudinal measure corresponds to pain. The censoring time C_i was generated from an exponential distribution with mean 50, and the right-censoring percentage was about 12%. The failure time and censoring indicator were calculated as $t_i = \min\{t_i^*, C_i\}$ and $\delta_i = 1\{t_i^* \leq C_i\}$, where $1\{A\}$ denotes the indicator function such that $1\{A\} = 1$ if A is true and 0 otherwise. This survival dataset is denoted by D_{Surv}.

We also generated three additional sets of longitudinal data, with longitudinal trajectories simulated from

$$\mu_{\ell i}(a_{ij}) = (\theta_0 + \theta_{0i} + \tau_{\ell 0i}) + (\theta_1 + \theta_{1i} + \tau_{\ell 1i})a_{ij} + \boldsymbol{\gamma}\boldsymbol{x}_i,$$

where (i) $(\tau_{10i}, \tau_{11i})^\top \sim N\left(\mathbf{0}, \begin{pmatrix} 0.1^2 & 0 \\ 0 & 0.1^2 \end{pmatrix}\right)$; (ii) $(\tau_{20i}, \tau_{21i})^\top \sim N\left(\mathbf{0}, \begin{pmatrix} 0.5^2 & 0 \\ 0 & 0.5^2 \end{pmatrix}\right)$; and (iii) $(\tau_{30i}, \tau_{31i})^\top \sim N\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. Then the longitudinal data were generated from a $N(\mu_{\ell i}(a_{ij}), \sigma^2)$ distribution with $\sigma = 0.54$ for $\ell = 1, 2, 3$. These three additional sets of longitudinal data were coupled with the same survival times as in $D_{\text{Long}} + D_{\text{Surv}}$ to form three additional data sets. These resulting data sets are denoted by $D_{\text{Long1}} + D_{\text{Surv}}$, $D_{\text{Long2}} + D_{\text{Surv}}$, and $D_{\text{Long3}} + D_{\text{Surv}}$. This simulation setting is similar to Simulation III in Zhang *et al.* (2014) except for the six additional covariates.

Figure 8 shows the fit statistics for $D_{\text{Long}} + D_{\text{Surv}}$, $D_{\text{Long1}} + D_{\text{Surv}}$, $D_{\text{Long2}} + D_{\text{Surv}}$, and $D_{\text{Long3}} + D_{\text{Surv}}$, respectively, using `JMFit`. We see from Figure 8 that $D_{\text{Long}} + D_{\text{Surv}}$ has the largest values of ΔAIC and ΔBIC.

Fit Statistics			
Log Likelihood	-4051.49		
AIC	8148.99	BIC	8240.79
AIC_{Long}	6104.63	BIC_{Long}	6156.52
$\text{AIC}_{\text{Surv}\|\text{Long}}$	2044.36	$\text{BIC}_{\text{Surv}\|\text{Long}}$	2084.27
$\text{AIC}_{\text{Surv},0}$	2075.71	$\text{BIC}_{\text{Surv},0}$	2107.64
ΔAIC	31.35	ΔBIC	23.37

(a) $D_{\text{Long}} + D_{\text{Surv}}$

Fit Statistics			
Log Likelihood	-4079.49		
AIC	8204.98	BIC	8296.78
AIC_{Long}	6151.87	BIC_{Long}	6203.76
$\text{AIC}_{\text{Surv}\|\text{Long}}$	2053.11	$\text{BIC}_{\text{Surv}\|\text{Long}}$	2093.03
$\text{AIC}_{\text{Surv},0}$	2075.71	$\text{BIC}_{\text{Surv},0}$	2107.64
ΔAIC	22.60	ΔBIC	14.61

(b) $D_{\text{Long1}} + D_{\text{Surv}}$

Fit Statistics			
Log Likelihood	-4399.92		
AIC	8845.84	BIC	8937.64
AIC_{Long}	6778.37	BIC_{Long}	6830.26
$\text{AIC}_{\text{Surv}\|\text{Long}}$	2067.47	$\text{BIC}_{\text{Surv}\|\text{Long}}$	2107.39
$\text{AIC}_{\text{Surv},0}$	2075.71	$\text{BIC}_{\text{Surv},0}$	2107.64
ΔAIC	8.24	ΔBIC	0.25

(c) $D_{\text{Long2}} + D_{\text{Surv}}$

Fit Statistics			
Log Likelihood	-4775.45		
AIC	9596.91	BIC	9688.71
AIC_{Long}	7525.67	BIC_{Long}	7577.56
$\text{AIC}_{\text{Surv}\|\text{Long}}$	2071.23	$\text{BIC}_{\text{Surv}\|\text{Long}}$	2111.15
$\text{AIC}_{\text{Surv},0}$	2075.71	$\text{BIC}_{\text{Surv},0}$	2107.64
ΔAIC	4.47	ΔBIC	-3.51

(d) $D_{\text{Long3}} + D_{\text{Surv}}$

Figure 8: The fit statistics for $D_{\text{Long}} + D_{\text{Surv}}$, $D_{\text{Long1}} + D_{\text{Surv}}$, $D_{\text{Long2}} + D_{\text{Surv}}$, and $D_{\text{Long3}} + D_{\text{Surv}}$ under SPM2L.

The rest of the output for $D_{\text{Long}} + D_{\text{Surv}}$ are provided in Appendix B. The output for $D_{\text{Long}\ell} + D_{\text{Surv}}$ ($\ell = 1, 2, 3$) are omitted here for brevity.

5. Concluding remarks

The `JMFit` SAS macro fits the joint models for longitudinal and survival data. The piecewise exponential constant hazard model is assumed for the baseline hazard function. The time-axis is partitioned into J intervals, which are constructed by four algorithms, namely, ESQP, LBSQP, MBSQP, and RBSQP. `JMFit` allows users to specify the number of intervals J and the partition method. Five models, including SPM1L, SPM1Q, SPM2L, SPM2Q, and TVC, are implemented in `JMFit`. This SAS macro also computes AIC, BIC, ΔAIC, ΔBIC, and the estimates of the parameters in the joint model. The computational time of `JMFit` depends on which of SPM1L, SPM1Q, SPM2L, SPM2Q, or TVC is chosen and how big the dataset is. For the example given in Section 4.1, it took 5 minutes to fit SPM1L with $J = 9$ based on LBSQP for coping on a Dell PC with an Intel i5 processor, 3.30 GHz CPU, and 8 GB of memory. On the same PC, it only took 10 to 11 seconds to fit SPM2L to each of the simulated datasets illustrated in Section 4.2.

The `JMFit` SAS macro provides two versions for SPM1L, SPM1Q, SPM2L, and SPM2Q with the `TS` argument, with `TS = 1` yielding the corresponding two-stage model. Comparing these models with `TS` missing or equal to 0, for the two-stage models, (i) we first fit (1) to the longitudinal data alone and obtain the estimates of $\boldsymbol{\theta}_i$, denoted by $\hat{\boldsymbol{\theta}}_i$; and (ii) we then use $\hat{\boldsymbol{\theta}}_i$ in (2) and (3). The TS models are also substantially different than the TVC model since the TS models do not require the LOCF assumption.

The current version of the `JMFit SAS` macro only fits linear and quadratic models for the longitudinal outcome and the piecewise constant baseline hazard function for the survival submodel. In the joint modeling framework, other dependence structures, such as dependence through the derivatives of the trajectory function or interactions with covariates as well as spline approximations to the baseline hazard could be assumed. In addition, other trajectory functions may be more appropriate to model the time effect on the longitudinal outcomes in certain applications. These additional features could be built in the `JMFit` macro in a future release.

Finally, we note that both the IBCSG dataset and the simulated datasets D_{Surv}, D_{Long}, D_{Long1}, D_{Long2}, and D_{Long3} are available for downloading from the journal website.

Acknowledgments

We would like to thank the Editors-in-Chief, the Editor, and the three anonymous reviewers for their very helpful comments and suggestions, which have led to a much improved version of the paper and the the `JMFit` SAS macro. Dr. M.-H. Chen and Dr. J. G. Ibrahim's research was partially supported by NIH grants #GM 70335 and #P01 CA142538.

References

Akaike H (1973). "Information Theory and an Extension of the Maximum Likelihood Prin-

ciple." In BN Petrov, F Csaki (eds.), *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281. Budapest: Akademiai Kiado.

Chi YY, Ibrahim JG (2006). "Joint Models for Multivariate Longitudinal and Multivariate Survival Data." *Biometrics*, **62**, 432–445. doi:10.1111/j.1541-0420.2005.00448.x.

Crowther MJ (2012). STJM: *Stata Module to Fit Shared Parameter Joint Models of Longitudinal and Survival Data*. URL http://EconPapers.repec.org/RePEc:boc:bocode:s457502.

Crowther MJ, Abrams KR, Lambert PC (2013). "Joint Modeling of Longitudinal and Survival Data." *The Stata Journal*, **13**, 165–184.

Fisher LD, Lin D (1999). "Time-Dependent Covariates in the Cox Proportional-Hazards Regression Model." *Annual Review of Public Health*, **20**, 145–157. doi:10.1146/annurev.publhealth.20.1.145.

Ibrahim JG, Chen MH, Sinha D (2001). *Bayesian Survival Analysis*. Springer-Verlag. doi:10.1007/978-1-4757-3447-8.

Ibrahim JG, Chu H, Chen LM (2010). "Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data." *Journal of Clinical Oncology*, **28**, 2796–2801. doi:10.1200/jco.2009.25.0654.

International Breast Cancer Study Group (1996). "Duration and Reintroduction of Adjuvant Chemotherapy for Node-Positive Premenopausal Breast Cancer Patients." *Journal of Clinical Oncology*, **14**, 1885–1894. doi:10.1200/jco.2005.03.0783.

Philipson P, Sousa I, Diggle P, Williamson P, Kolamunnage-Dona R, Henderson R (2012). ***joineR***: *Joint Modelling of Repeated Measurements and Time-to-Event Data*. R package version 1.0-3, URL https://CRAN.R-project.org/package=joineR.

SAS Institute Inc (2011a). *SAS/IML Software, Version 9.3*. Cary, NC. URL http://www.sas.com/.

SAS Institute Inc (2011b). *SAS/STAT Software, Version 9.3*. Cary, NC. URL http://www.sas.com/.

Proust-Lima C, Philipps V, Liquet B (2016). "Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package **lcmm**." *Journal of Statistical Software*. Forthcoming.

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 (2012). ***JM***: *Joint Modeling of Longitudinal and Survival Data*. R package version 1.1-0, URL https://CRAN.R-project.org/package=JM.

Rizopoulos D (2016). "The R Package **JMbayes** for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC." *Journal of Statistical Software*. Forthcoming.

Schwarz G (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, **6**, 461–464. doi:10.1214/aos/1176344136.

Tsiatis AA, DeGruttola V, Wulfsohn MS (1995). "Modelling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS." *Journal of the American Statistical Association*, **90**, 27–37. doi:10.1080/01621459.1995.10476485.

Zhang D, Chen MH, Ibrahim JG, Boye ME, Wang P, Shen W (2014). "Assessing Model Fit in Joint Models of Longitudinal and Survival Data with Applications to Cancer Clinical Trials." *Statistics in Medicine*, **33**(27), 4715–4733. doi:10.1002/sim.6269.

Zhu H, Ibrahim JG, Chi YY, Tang N (2012). "Bayesian Influence Measures for Joint Models for Longitudinal and Survival Data." *Biometrics*, **68**, 954–964. doi:10.1111/j.1541-0420.2012.01745.x.

A. The macro JMFit

The SAS macro JMFit as well as the five submacros should be stored in a folder named jmfit. Then JMFit can be accessed by including the following lines:

```
filename jmfit "directory of the file JMFit";
%include jmfit(JMFit.sas);
%JMFit(LONG=, SURV=, MODEL=, TS=, TMAXI=, WEIGHT=, NPIECES=, PARTITION=,
  OPTIONS=, INITIAL=, OUTPUT=);
```

Inputs for JMFit

LONG: Data set with the first three columns, SID (subject ID), Y (longitudinal measure), A (time at which Y was taken), and additional columns for covariates (X_{L1}-X_{Lp}), where SID, Y, and A should be arranged in the first, second, and third columns, and X_{L1}, $\cdots$, X_{Lp} should be placed after column 3, which can be enforced in SAS by using the retain command. Note that X_{L1}, ..., X_{Lp} can be time-dependent or baseline covariates. Required.

SURV: Data set with the first three columns, SID, survival time (T), censoring indicator (delta) (1 = death and 0 = censored), and additional columns for covariates (X_{S1}-X_{Sq}), where SID, T, and delta should be arranged in the first, second, and third columns, and X_{S1}, ..., X_{Sq} should be placed after column 3. Required.

MODEL: Model specification. Required. One of

1. SPM1L: Shared Parameter Model 1 with Linear trajectory.
2. SPM1Q: Shared Parameter Model 1 with Quadratic trajectory.
3. SPM2L: Shared Parameter Model 2 with Linear trajectory.
4. SPM2Q: Shared Parameter Model 2 with Quadratic trajectory.
5. TVC: Time-Varying Covariates Model.

TS: Indicates whether to implement the model specified in the MODEL argument or the corresponding two-stage model. If 0 (default), the model specified in the MODEL argument will be fit. If 1, the corresponding two-stage model will be fit instead. It only works for SPM1L, SPM1Q, SPM2L, and SPM2Q.

TMAXI: $t_{\max,i}$ adjustment to the model specified in the MODEL argument. If 0 (default), no $t_{\max,i}$ adjustment will be applied. If 1, the trajectory will become flat after $t^*_{\max,i} = t_{\max,i}+$ WEIGHT$\times\max(t_i - t_{\max,i}, 0)$. If 2, starting at $t^*_{\max,i}$, the trajectory will linearly go down to 0 at the last follow-up survival time. It only works for SPM1L and SPM1Q.

WEIGHT: The proportion ($\in [0, 1]$) of $\max(t_i - t_{\max,i}, 0)$. If 0 (default), the starting point of the modified extrapolation of the trajectory is $t_{\max,i}$. If 1, the trajectory extends to t_i with no $t_{\max,i}$ adjustment. It only works when TMAXI = 1 or TMAXI = 2.

NPIECES: Number of intervals J (≥ 1) for the piecewise constant baseline hazard function. Required.

PARTITION: Algorithm for constructing the partition of the time axis. Required. One of

(i) 1: Equally-Spaced Quantile Partition (ESQP).

(ii) 2: Left Bi-Sectional Quantile Partition (LBSQP).

(iii) 3: Middle Bi-Sectional Quantile Partition (MBSQP).

(iv) 4: Right Bi-Sectional Quantile Partition (RBSQP).

OPTIONS: Allows users to specify options that are available in the PROC NLMIXED statement. For example, OPTIONS = %str(QPOINTS = 5 TECH = CONGRA ABSGCONV = 0.0001) specifies Gaussian quadrature with five quadrature points for approximating the integral of the likelihood over the random effects, the conjugate-gradient optimization, and an absolute gradient convergence criterion of 0.0001.

INITIAL: Allows users to set their own initial values. JMFit will automatically generate the starting values for the model parameters and these initial values will be stored in the data set _initial. Since the order of the parameters is very important when calculating AIC_{Long} and BIC_{Long}, users are recommended to change the initial values in _initial and then rename _initial.

OUTPUT: Name of the output rich text file (RTF). One can also specify the directory in which the file will be put. For example, the output file named "myoutput" will be stored in C:\...\myoutput by %JMFit(..., OUTPUT = C:\...\myoutput); If OUTPUT is not specified, the file will be indexed by the name of Y from LONG and the model's name.

Note 1: (i) The name of the SID variable in LONG should be the same as that of the SID variable in SURV; (ii) A and T should be in the same unit of time (month preferred); (iii) the categorical covariates must be coded as dummy variables; and (iv) the SAS macro allows for any numbers of covariates for both components of the joint model and the covariates for the longitudinal component can be totally different from those for the survival component.

Note 2: (i) "ERROR: Not enough memory to generate code." This might occur if J is too big; (ii) a too long path for the OUTPUT may lead to an error; (iii) the macro is assuming options validvarname = v7; for valid variable names that can be processed in SAS; and (iv) the calculations of AIC_{Long} and BIC_{Long} require the IML Procedure.

Note 3: No missing values are allowed in both data sets.

Output for JMFit

The macro automatically produces an RTF file indexed by the name of Y from LONG and the model's name. The RTF file includes five tables: (i) Number of Subjects; (ii) Fit Statistics; (iii) Survival Parameter Estimates (Survival Alone); (iv) Parameter Estimates; and (v) Hazard Ratios & λ Estimates.

Note: The construction of the Parameter Estimates table is different for each model. For SPM1L, SPM1Q, SPM2L, and SPM2Q, it consists of three subtables: "Covariance Parameter Estimates", "Longitudinal Parameter Estimates", and "Survival Parameter

Estimates"; for the TS model corresponding to SPM1L, SPM1Q, SPM2L, or SPM2Q, there are two tables: "Longitudinal Parameter Estimates (Stage I)" and "Survival Parameter Estimates (Stage II)"; and for the TVC model, there is only one table named "Parameter Estimates".

B. Output for the simulated data $D_{\mathbf{Long}} + D_{\mathbf{Surv}}$ under SPM2L

Number of Subjects	
Subjects_in_long	400
Subjects_in_surv	400
Subjects_Used	400

Figure 9: The number of subjects for $D_{\text{Long}} + D_{\text{Surv}}$.

Survival Parameter Estimates (Survival Alone)							
Param/Var	**Estimate**	**SE**	**DF**	**T-Value**	**P-Value**	**95% CI**	**Gradient**
therapy	-0.3373	0.1081	400	-3.12	0.0019	(-0.550, -0.125)	0.002926
race	-0.00214	0.1811	400	-0.01	0.9906	(-0.358, 0.354)	0.001148
gender	0.05360	0.1272	400	0.42	0.6737	(-0.196, 0.304)	-0.00402
age	0.04130	0.1078	400	0.38	0.7018	(-0.171, 0.253)	0.005914
karnofsky	-0.2666	0.1116	400	-2.39	0.0173	(-0.486, -0.047)	-0.00098
stage	-0.2204	0.1088	400	-2.03	0.0435	(-0.434, -0.006)	-0.00063
bf	-0.03879	0.1262	400	-0.31	0.7587	(-0.287, 0.209)	0.003399
log λ_1	-1.5883	0.2502	400	-6.35	<.0001	(-2.080, -1.096)	0.000383

Figure 10: The estimates of the parameters obtained by fitting the survival data alone for $D_{\text{Long}} + D_{\text{Surv}}$.

Hazard Ratios & λ Estimates		
Parameter	**Estimate**	**95% CI**
HR_therapy	0.7020	(0.563, 0.875)
HR_race	1.0552	(0.732, 1.521)
HR_gender	0.9778	(0.756, 1.264)
HR_age	1.1057	(0.889, 1.375)
HR_karnofsky	0.7499	(0.598, 0.940)
HR_stage	0.8237	(0.661, 1.026)
HR_bf	0.9328	(0.722, 1.205)
λ_1	0.1809	(0.109, 0.301)
HR_β_1	1.2041	(1.046, 1.387)
HR_β_2	3.6966	(2.312, 5.910)

Figure 11: The hazard ratios and $\boldsymbol{\lambda}$ estimates for $D_{\text{Long}} + D_{\text{Surv}}$ under SPM2L.

Parameter Estimates							
Param/Var	**Estimate**	**SE**	**DF**	**T-Value**	**P-Value**	**95% CI**	**Gradient**
Covariance Parameter Estimates							
Ω_{00}	0.6819	0.05753	398	11.85	<.0001	(0.569, 0.795)	0.005795
Ω_{10}	-0.05231	0.01440	398	-3.63	0.0003	(-0.081, -0.024)	0.023612
Ω_{11}	0.06822	0.006320	398	10.79	<.0001	(0.056, 0.081)	-0.0075
σ	0.5252	0.008301	398	63.26	<.0001	(0.509, 0.542)	0.000529
Longitudinal Parameter Estimates							
Intercept	0.6852	0.1916	398	3.58	0.0004	(0.309, 1.062)	-0.0101
longt	0.03229	0.01491	398	2.17	0.0309	(0.003, 0.062)	-0.03416
therapy	-0.01673	0.08539	398	-0.20	0.8447	(-0.185, 0.151)	-0.00516
race	-0.1504	0.1449	398	-1.04	0.3002	(-0.435, 0.135)	-0.01036
gender	-0.1901	0.1019	398	-1.87	0.0628	(-0.390, 0.010)	-0.00528
age	-0.1539	0.08503	398	-1.81	0.0711	(-0.321, 0.013)	-0.02083
karnofsky	-0.5958	0.08773	398	-6.79	<.0001	(-0.768, -0.423)	-0.00274
stage	-0.2010	0.08602	398	-2.34	0.0199	(-0.370, -0.032)	0.010376
bf	-0.07759	0.09840	398	-0.79	0.4309	(-0.271, 0.116)	-0.00665
Survival Parameter Estimates							
therapy	-0.3539	0.1118	398	-3.16	0.0017	(-0.574, -0.134)	-0.01536
race	0.05376	0.1859	398	0.29	0.7726	(-0.312, 0.419)	0.000795
gender	-0.02245	0.1307	398	-0.17	0.8637	(-0.279, 0.234)	0.008286
age	0.1005	0.1109	398	0.91	0.3656	(-0.118, 0.319)	-0.00213
karnofsky	-0.2878	0.1150	398	-2.50	0.0128	(-0.514, -0.062)	0.010698
stage	-0.1940	0.1117	398	-1.74	0.0831	(-0.414, 0.026)	-0.00047
bf	-0.06953	0.1302	398	-0.53	0.5937	(-0.326, 0.186)	0.001858
log λ_1	-1.7098	0.2590	398	-6.60	<.0001	(-2.219, -1.201)	-0.0061
β_1	0.1858	0.07175	398	2.59	0.0100	(0.045, 0.327)	0.003564
β_2	1.3074	0.2387	398	5.48	<.0001	(0.838, 1.777)	0.003804

Figure 12: The estimates of the parameters for $D_{\text{Long}} + D_{\text{Surv}}$ under SPM2L.

Affiliation:

Danjie Zhang
Gilead Sciences, Inc.
333 Lakeside Drive
Foster City, CA 94404, United States of America
E-mail: danjie.zhang@gilead.com

Ming-Hui Chen
Department of Statistics
University of Connecticut
215 Glenbrook Road U-4120
Storrs, CT 06269, United States of America
E-mail: ming-hui.chen@uconn.edu

Joseph G. Ibrahim
Department of Biostatistics
University of North Carolina
McGavran Greenberg Hall CB#7420
Chapel Hill, NC 27599, United States of America
E-mail: ibrahim@bios.unc.edu

Mark E. Boye, Wei Shen
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285, United States of America
E-mail: boyema@lilly.com, shen_wei_x1@lilly.com

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
July 2016, Volume 71, Issue 3 *Submitted:* 2013-05-01
doi:10.18637/jss.v071.i03 *Accepted:* 2015-02-23