



## **%JM: A SAS Macro to Fit Jointly Generalized Mixed Models for Longitudinal Data and Time-to-Event Responses**

**Alberto Garcia-Hernandez**  
Astellas Pharma Global Development

**Dimitris Rizopoulos**  
Erasmus University Medical Center

---

### **Abstract**

In clinical research subjects are usually observed during a period of time. Primary and secondary endpoints are often either responses measured longitudinally over time or the time at which an event of interest occurs. Joint modeling is increasingly being used for multiple purposes such as to adjust the analysis of the longitudinal response for informative dropout mechanisms.

In this paper we present %JM, a SAS macro that fits jointly generalized mixed models for longitudinal data and proportional hazards models for time-to-event responses. The macro fits normal, binary, binomial and Poisson longitudinal responses and allows choosing among a range of options to fit the trajectories of the longitudinal response over time: a linear function, splines, natural cubic splines and  $B$ -splines. For the time-to-event response, that might be right-censored, the macro fits parametric, stratified or not, proportional hazards models with the following baseline risk functions: exponential, Weibull, piecewise exponential and the generalizations of the Weibull and the Gompertz models based on splines.

%JM offers several options to connect the longitudinal model and the time-to-event model: current-value-dependent and slope-dependent shared parameters, lagging effects, cumulative effects, random effects coefficients as shared parameters and interaction effects.

*Keywords:* SAS macro, joint modeling, time-to-event, longitudinal, generalized mixed models, shared parameter models, survival data.

---

## **1. Introduction**

The joint modeling of longitudinal and time-to-event data has received significant attention over the past years. The two contexts where these models have been proven most useful

are: (a) to evaluate longitudinal response over time and analyze this response jointly with the time to study discontinuation in order to adjust for informative dropout mechanisms; (b) to evaluate the effect of a surrogate marker measured over time into an event such as death avoiding the bias that occurs when Andersen and Gill's extended Cox model is used with endogenous biomarkers measured with error (Prentice 1982). For further background on this family of models we recommend the reviews of Tsiatis and Davidian (2004) and Wu, Liu, Yi, and Huang (2012) and the book of Rizopoulos (2012b).

One of the difficulties that joint models have encountered for a broad use in clinical research is the lack of software and the intensive use of computational resources. Fortunately this has recently changed and several packages and procedures are already available. Three R (R Core Team 2018) packages; **JM** (Rizopoulos 2010), **joineR** (Williamson, Kolamunnage-Dona, Philipson, and Marson 2008; Philipson, Diggle, Sousa, Kolamunnage-Dona, Williamson, and Henderson 2018) and **lcmm** (Proust-Lima, Philipps, Diakite, and Liqueur 2017a; Proust-Lima, Philipps, and Liqueur 2017b); and one Stata (Stata Corp 2014) module, **stjm** (Crowther 2012), fit these models using maximum likelihood whereas the R package **JMBayes** (Rizopoulos 2016, 2017) produces Markov chain Monte Carlo simulations to approach this problem from a Bayesian perspective. All these software packages are limited to normal longitudinal data with the exception of the **JMBayes** R package that allows user-defined likelihood functions for the longitudinal data.

In SAS (SAS Institute Inc. 2004), Guo and Carlin (2004) proposed using the NL MIXED procedure although they did not address the challenge implicit in the estimation of the survival function of the time-to-event model that, in general, does not have a closed form. More recently, Zhang, Chen, Ibrahim, Boye, Wang, and Shen (2014) have written several SAS macros based on the NL MIXED procedure that use the Riemann integral to compute the cumulative hazard function. This work is, though, limited to normally-distributed longitudinal responses that change over time following either a linear or a quadratic function, time-to-event responses with a piecewise constant baseline hazard, and the two most popular association structures, namely a trajectory current-value-dependent shared parameter and the use of the random effects coefficients as shared parameters.

This paper presents **%JM** (available at <http://www.jm-macro.com/>), a SAS macro that fits jointly generalized mixed models for longitudinal data and parametric proportional hazards models for time-to-event responses. The macro accepts normal, binary, binomial and Poisson longitudinal responses and offers several options to fit the trajectories of this response over time: a linear function, splines, natural cubic splines and  $B$ -splines. For the time-to-event response, the macro includes the following baseline risk functions: exponential, Weibull, piecewise exponential and the generalizations of the Weibull and the Gompertz models based on splines. **%JM** has multiple options to connect the longitudinal model and the time-to-event model: current-value-dependent and slope-dependent shared parameters, lagging effects, cumulative effects, random effects coefficients as shared parameters and interaction effects.

This paper is structured as follows. Section 2 formulates the joint models fitted by the macro. Section 3 presents the macro. Section 4 illustrates its use with two examples, and Section 5 concludes the paper. A full description of all macro parameters is given in Appendix A.

## 2. Joint model specification

This paper focuses on a group of joint models where the association structure between longitudinal and time-to-event outcomes is modeled via non-observed latent variables (Rizopoulos and Lesaffre 2014). In particular, we will use normally-distributed random effects  $b_i$  as the latent variables that associate both responses and given these, the longitudinal responses  $y_i$  and the time-to-event responses  $T_i$  are assumed to be independent. The probability of the observations for a subject  $i$   $\{T_i, \delta_i, y_i\}$ , conditioned to the random effects  $b_i$ , can thus be decomposed as follows

$$p(T_i, \delta_i, y_i | b_i; \theta) = p(T_i, \delta_i | b_i; \theta)p(y_i | b_i; \theta),$$

where  $T_i$  is the observed time-to-event or time-to-censoring,  $\delta_i$  is the indicator of event or censoring,  $y_i = \{y_{ij}; 1 \leq j \leq n_i\}$  the collection of observations in the longitudinal response for subject  $i$  and  $\theta$  the vector of parameters that includes parameters for the time-to-event response and for the longitudinal model.

### 2.1. The longitudinal submodel

For the longitudinal response, we will fit generalized random-effects ( $G$ -sided) mixed-effects models. The random effects  $b_i$  are assumed to be normally distributed with mean zero and a variance-covariance matrix  $G$ . Given the random effects, the observations for the longitudinal response on the same subject are assumed to be independent.

$$p\{y_i | b_i; \theta_y\} = \prod_j p\{y_i(t_{ij}) | b_i; \theta_y\},$$

where  $\theta_y$  refers the vector of parameters for the longitudinal model. This generalized random-effects linear model for longitudinal data can be formulated as follows:

$$g\{m_i(t)\} = g[E\{y_i(t) | b_i\}] = X_i(t)(\beta_t + b_i) + Z_i\beta_b,$$

where  $m_i(t)$  denotes the expected value for subject  $i$  at time  $t$  on the longitudinal response  $y_i(t)$ ,  $g(\cdot)$  is a link function,  $X_i(t)$  is the design matrix of fixed and random effects that model the trajectories of the longitudinal response over time, associated with the coefficients  $\beta_t$  and  $b_i$  respectively, and  $Z_i$  and  $\beta_b$  are the design matrix and the coefficients associated with the baseline covariates.

#### *Random intercepts and slopes*

In the context of random-effects linear models several options exist to adequately model the trajectories of the longitudinal response over time. A simple random intercepts and slopes model consists of a fixed-effect intercept  $\beta_0$ , a fixed-effect slope  $\beta_1$ , a random-effect intercept  $b_{i0}$ , a random-effect slope  $b_{i1}$  and a vector of coefficients  $\beta_b$  associated with a set of baseline covariates, as shown below

$$g\{m_i(t)\} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + Z_i\beta_b. \quad (1)$$

We need to estimate the variances of the random effects  $b_{i0}$  and  $b_{i1}$  as well as their covariance unless it is assumed to be 0. Occasionally we might need to fit a distinct trajectory for certain

subgroups. For instance, a separate slope for subjects in active treatment ( $TRT = 1$ ), with respect to those in placebo ( $TRT = 0$ ), would require the second term of Equation 1 to be expanded to  $(\beta_1 + \beta_2 TRT_i + b_{i1})t$ .

### Random splines

A simple straight line may not be adequate to model the trajectories of the longitudinal response over time. The time-dependent component of the random-effects longitudinal model, that we have denoted above as  $X_i(t)(\beta_t + b_i)$ , can be modeled using splines. A spline is a smooth function, that we will denote as  $s(t; \beta_t, b_i)$ , that is often constructed as a linear combination of truncated power functions. A spline of degree  $n$  and  $N$  internal knots can be expressed as

$$s(t; \beta_t, b_i) = \sum_{j=0}^{N+n} (\beta_j + b_{ij})B_j(t),$$

where  $B_j(t)$  are the basis functions, and  $\beta_j$  and  $b_{ij}$  their associated fixed and random-effects coefficients respectively. Splines without knots are standard polynomials. For instance, a spline of degree 3 and no knots is a function that includes a linear term, a quadratic term and a cubic term.

Natural cubic splines are cubic splines resulting after setting the second and third derivatives to zero at the boundary knots and therefore fitting a straight line outside these. Four (two fixed and two random) less effects are required with respect to the complete cubic spline with the same number of internal knots. The basis functions  $B_0(t)$  to  $B_{N+1}(t)$  can be calculated, for example, using Equation 2 (Hastie, Tibshirani, and Friedman 2009):

$$B_0(t) = 1; \quad B_1(t) = t; \quad B_{j+1}(t) = d_j(t) - d_N(t), \quad (2)$$

where  $N$  is the number of internal knots,  $d_j(t) = (t - v_j)_+^3 / (v_{N+1} - v_j)^3$ ,  $v_0$  is the lower boundary knot,  $v_1$  to  $v_N$  are the internal knots and  $v_{N+1}$  refers to the upper boundary knot.

Splines can also be expressed as a linear combination of Bézier basis functions through the generalization of the Bézier curve.  $B$ -splines are maximally differentiable bases with a number of desirable properties (Farouki and Rajan 1987; Farouki and Goodman 1996) that, in general, lead to better convergence as compared to the bases defined using truncated power functions. Bézier basis functions can be produced using the Boor recurrence relation (Racine 2018).

## 2.2. The time-to-event submodel

For the time-to-event response we will fit parametric proportional hazards models in which the baseline hazard function needs to be specified. The variable of interest  $T^*$  is the time to the event that might not be observed for all subjects. The observed variables are the time to event or censoring  $T_i$  and the indicator of event  $\delta_i$ . The general expression of this model is given by

$$h_i(t) = h_0(t) \exp(\gamma^\top w_i),$$

where  $h_0(t)$  is the baseline hazard rate that must be specified,  $w_i$  denotes the vector of covariates and  $\gamma$  the corresponding vector of regression coefficients.

*Exponential*

The simplest model for the time-to-event response assumes that the baseline risk is constant over time:

$$h_0(t) = \lambda.$$

*Weibull*

The Weibull distribution adds a shape parameter to allow the hazard risk to increase or decrease over time using a power function of time.

$$h_0(t) = \frac{k}{\lambda^k} t^{k-1}.$$

*Piecewise exponential*

In standard survival analysis it is customary to leave  $h_0(\cdot)$  unspecified as in the Cox model. In the joint modeling framework such an approach may underestimate the standard errors of the parameter estimates (Hsieh, Tseng, and Wang 2006), and it is more adequate to define this function. A standard approach is the piecewise-constant hazard model with a baseline hazard function defined as follows:

$$h_0(t) = \sum_{q=1}^{Q+1} \xi_q I(v_{q-1} < t \leq v_q),$$

where  $v_0 = 0, v_1, \dots, v_Q$  are  $Q$  internal knots that split the range of times into  $Q+1$  intervals,  $v_{Q+1}$  is a value higher than the maximum observed time and  $\xi_q$  is the hazard at the interval defined from  $v_{q-1}$  to  $v_q$ .

A piecewise exponential model can get as close to the unspecified Cox model as desired by increasing the number of knots. However, too many knots might not be recommended and, in order to avoid the underestimation of the standard errors mentioned before, it is advisable to maintain at least 10–20 events per parameter including the parameters involved in estimating the longitudinal response (Harrell 2001).

*Flexible generalization of the Gompertz and the Weibull models*

The Gompertz and the Weibull models can be formulated in terms of the baseline log-hazard function as  $\log\{h_0(t)\} = \kappa_0 + \kappa_1 t$  and  $\log\{h_0(t)\} = \kappa_0 + \kappa_1 \log(t)$  respectively. A flexible generalization of these models can be constructed as  $\log\{h_0(t)\} = s(x, \kappa)$  where  $s(x, \kappa)$  is a spline function of  $x$ , with  $x = t$  and  $x = \log(t)$  for the generalizations of the Gompertz and Weibull models respectively, and  $\kappa$  the vector of coefficients of the spline function. For example, a generalization of the Weibull model using natural cubic splines has the following formulation

$$\log\{h_0(t)\} = \sum_{j=0}^{N+1} \kappa_j B_j(x),$$

where  $x = \log(t)$ ,  $N$  is the number of internal knots and  $B_0(x)$  to  $B_{N+1}(x)$  are the basis functions defined in Section 2.1.

### The Royston-Parmar model

The Weibull model can also be formulated with respect to the logarithm of the cumulative hazard function,  $\log\{H_0(t)\} = \log[\int_0^t h_0(t)dt]$ , as a monotonic linear equation  $\log\{H_0(t)\} = \kappa_0 + k_1 \log(t)$  with  $\kappa_1 > 0$ . Royston and Parmar (2002) proposed a flexible generalization of the Weibull model defined as  $\log\{H_0(t)\} = \sum_{n=0}^{N+1} \kappa_n B_n(x)$  where  $B_0(x)$  to  $B_{N+1}(x)$  are the basis functions of a monotonic natural cubic spline with  $N$  internal knots on  $x = \log(t)$ . Taking into account that  $h(t) = dH(t)/dt$ , the log-hazard function of this model can be formulated in terms of the basis functions  $B_0(x)$  to  $B_{N+1}(x)$  and their first derivatives  $B'_0(x)$  to  $B'_{N+1}(x)$  as

$$\log\{h_0(t)\} = -\log(t) + \sum_{n=0}^{N+1} \kappa_n B_n(x) + \log\left\{ \sum_{n=0}^{N+1} \kappa_n B'_n(x) \right\},$$

where  $x = \log(t)$ .

### Stratification

It might not be reasonable to assume that the whole sample has the same baseline hazard function, that is, the proportional hazard assumption between subgroups might not hold. In these situations it is convenient to treat those subgroups as strata thus a different baseline hazard is fitted for each subgroup, as given by

$$h_{ik}(t) = h_{0k}(t) \exp(\gamma^\top w_i),$$

where  $h_{0k}(t)$  is the baseline risk of stratum  $k$ . This model requires  $Q \times K$  parameters to define all baseline hazard functions where  $Q$  is the number of parameters required to estimate the hazard function of each stratum and  $K$  is the number of strata.

### Exogenous time-dependent covariates

One of the motivations for the development of joint models of longitudinal and time-to-event responses was to find an alternative to the extended Cox model, also known as the Andersen-Gill model, when the time-dependent covariate  $y_i(t)$  is not exogenous. Endogenous time-dependent covariates are thus included in this model as the dependent longitudinal responses  $y_i(t)$ . However there could be situations where, in addition to the endogenous time-dependent response  $y_i(t)$ , it is of interest to include an exogenous time-dependent covariate  $w_i(t)$  following the principles of the extended Cox model. In this case, the general expression of the time-to-event model changes to

$$h_i(t) = h_0(t) \exp(\gamma^\top w_i(t)),$$

where  $w_i(t)$  is a covariate vector that contains both baseline and exogenous time-dependent covariates.

## 2.3. The joint model

The most standard approach to connect the longitudinal and the time-to-event submodels in this framework is to assume that the risk of event at a given time  $t$  depends on the estimated

value of the longitudinal response at that time (Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997).

$$h_i(t) = h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\}. \quad (3)$$

For the case that the baseline risk  $h_0(t)$  is exponential, it should be noted that, once a shared-parameter  $\alpha$  is included between the subject-level estimate of the longitudinal response  $m_i(t)$  and the time-to-event response  $T_i$ , the resulting hazard  $h_i(t)$  for subject  $i$ , given in Equation 3, is not constant anymore but changes as the longitudinal response changes over time.

#### *Other parameterizations*

The above model assumes that the risk of event at time  $t$  is related to the expected value of the longitudinal response at exactly this time point  $t$ . This might not always be the most appropriate parameterization and thus other types of association have been suggested in the literature (Rizopoulos and Ghosh 2011; Rizopoulos 2012b). For each patient the longitudinal model fits a trajectory in time and it may be the case that it is not the value but the rate of increase (or decrease) of the longitudinal response that is associated with the risk of event.

$$h_i(t) = h_0(t) \exp\{\gamma^\top w_i + \alpha m'_i(t)\}.$$

A cumulative association assumes that the risk of event does not depend on the expected value of the longitudinal response at that time point but on the cumulative effect.

$$h_i(t) = h_0(t) \exp\left\{\gamma^\top w_i + \alpha \int_0^t m_i(s) ds\right\}.$$

Occasionally there is a latency period between the value of the longitudinal response and the appearance of an increased risk. These situations can be modeled fixing a time-lagged effect model.

$$h_i(t) = h_0(t) \exp[\gamma^\top w_i + \alpha m_i\{\max(0, t - c)\}],$$

where  $c$  specifies the time lag. It is also possible to connect the longitudinal and the time-to-event models through the individual coefficients of the random effects.

$$h_i(t) = h_0(t) \exp(\gamma^\top w_i + \alpha^\top b_i), \quad (4)$$

where  $\alpha$  is the vector of coefficients associated with the vector of random effects  $b_i$ . For example, using a random intercepts and slopes model for the longitudinal response, the last term of Equation 4 would be  $\alpha_1 b_{i0} + \alpha_2 b_{i1}$ . From a computational point of view, the main advantage of this parameterization is that the association between both responses is time-independent what leads to an analytical form of the cumulative hazard function.

#### *Interaction effects*

It might not be correct to assume that the effect of the longitudinal response on the event risk is the same for all subjects and thus shared-parameter interactions need to be considered. A standard parameterization with interactions would be formulated as given by

$$h_i(t) = h_0(t) \exp[\gamma^\top w_{i1} + \alpha^\top \{w_{i2} \cdot m_i(t)\}], \quad (5)$$

where  $w_{i1}$  denotes the vector of covariates,  $\gamma$  its corresponding vector of coefficients,  $w_{i2}$  the vector of interaction terms, and  $\alpha$  its corresponding vector of interaction coefficients. The first term of  $w_{i2}$ , is usually taken as 1 so  $\alpha_1$  represents the association between the longitudinal and the time-to-event responses when all other interaction terms are zero.

#### *Association term: mean versus linear predictor*

For non-normal longitudinal responses, the link function is usually distinct from the identity function leading to two options to link the longitudinal and the time-to-event models: (a) a linear relationship between the log-hazard function and the expected response of the longitudinal response  $m_i(t)$  as shown in Equation 3; and (b) the log-hazard function linearly related to the linear predictor  $g\{m_i(t)\} = X_i(t)\beta + Z_i(t)b_i$  as illustrated in Equation 6.

$$h_i(t) = h_0(t) \exp\{\gamma^\top w_i + \alpha g\{m_i(t)\}\} = h_0(t) \exp[\gamma^\top w_i + \alpha\{X_i(t)(\beta_t + b_i) + Z_i\beta_b\}], \quad (6)$$

where  $m_i(t)$  is the estimated value for longitudinal response  $y_i(t)$ ,  $g(\cdot)$  is the link function,  $X_i(t)$  is the design matrix of fixed and random effects that model the trajectories of the longitudinal response over time, associated with the coefficients  $\beta_t$  and  $b_i$  respectively, and  $Z_i$  and  $\beta_b$  are the design matrix and the coefficients associated with the baseline covariates.

### 3. The %JM macro

This section will introduce the main functionalities of the %JM macro. In addition, Appendix A contains the complete list of macro parameters.

#### 3.1. The conditional log-likelihood function

The NLMIXED procedure allows to manually define the individual components of the conditional log-likelihood function, that are different for the longitudinal and the time-to-event responses, through the use of the `ll` statement. The probability of an observation  $y_i(t_{ij})$  from the longitudinal response, conditional to the random effects parameters, always has a closed form as given below, for example, for the normal distribution.

$$\log\{p\{y_i(t_{ij}) \mid b_i; \theta_y\}\} = -\frac{1}{2} \log(2\pi) - \log(\sigma) - \frac{1}{2} \left\{ \frac{y_i(t_{ij}) - m_i(t_{ij})}{\sigma} \right\}^2, \quad (7)$$

where  $\theta_y$  denotes the vector of parameters of the longitudinal response model decomposed into the subvectors  $\theta_y = (\beta_t, \beta_b, \sigma^2)$ . In particular, the %JM macro does not estimate the residual variance,  $\sigma^2$ , but its log-transformed square root,  $\log(\sigma)$ .

On the other hand, the probability function of the time-to-event response, conditional to the random effects parameters, can be expressed as

$$\log\{p(T_i, \delta_i \mid b_i; \theta)\} = \delta_i \log\{h_i(T_i \mid b_i; \theta_y, \theta_T)\} - \int_0^{T_i} h_i(s \mid b_i; \theta_y, \theta_T) ds, \quad (8)$$

where  $\theta = (\theta_y, \theta_T)$  and  $\theta_T = (\kappa, \gamma, \alpha)$  is the vector of parameters of the time-to-event model including the association parameters  $\alpha$ . In order to provide the conditional probability function of the time-to-event response to the `ll` statement of the NLMIXED procedure, the macro

needs to calculate the integral involved in the second term of Equation 8. Since this integral may not have a closed form, depending on the model chosen for both the longitudinal and the time-to-event responses, the %JM macro approximates it using the 15-points Gauss-Kronrod rule as given by

$$\int_0^{T_i} h_i(s | b_i; \theta) ds \approx \frac{T_i}{2} \sum_{k=1}^{15} w_k h_i(t_{ik} | b_i; \theta),$$

where  $t_{ik}$  are the 15 Kronrod-rule nodes for the integral from  $-1$  to  $1$  re-scaled into the  $0$  to  $T_i$  interval, and  $w_k$  are the 15 Kronrod-rule weights for the integral from  $-1$  to  $1$ . The NLMIXED procedure allows defining arrays in the programming statements to code the conditional probability function. This functionality is useful to apply the 15-points Gauss-Kronrod rule for the definition of the conditional likelihood function of the time-to-event response.

### 3.2. Approximation of the marginal log-likelihood function

Under the assumptions of this model, the marginal (joint) log-likelihood of the observations on subject  $i$  can be expressed as follows:

$$\log p(T_i, \delta_i, y_i; \theta_y, \theta_T) = \log \int p(T_i, \delta_i | b_i; \theta_y, \theta_T) \left[ \prod_j p\{y_i(t_{ij}) | b_i; \theta_y\} \right] p(b_i; \theta_b) db_i,$$

where  $\theta_b$  denotes the unique parameters of the random-effects covariance matrix  $G$ .

The NLMIXED procedure approximates the marginal likelihood, the integral of the conditional-likelihood function over the random effects, using adaptive or non-adaptive Gauss-Hermite quadrature. By default, the NLMIXED procedure uses adaptive Gauss-Hermite quadrature, and selects the number of quadrature points adaptively by evaluating the log-likelihood function at the initial parameters until two successive evaluations have a relative difference less than 0.0001. The NLMIXEDOptions macro parameter allows personalization of the approximation method with the options available in the NLMIXED procedure. For example, `NLMIXEDOptions = noad noadscale qpoints=100` would lead to a non-adaptive Gauss-Hermite quadrature with 100 quadrature points.

### 3.3. Optimization routines

The %JM macro uses the optimization techniques available in the NLMIXED procedure. If the NLMIXEDOptions macro parameter is not provided by the user, the default dual Quasi-Newton algorithm is used. Table 1 shows the execution times required to fit the joint model of Section 4.1 with some of the optimization routines available in the NLMIXED procedure.

### 3.4. Initial parameters

The %JM macro fits first the separate longitudinal and time-to-event models using either the MIXED procedure or the GLIMMIX procedure for normal and non-normal longitudinal responses respectively and the LIFEREG procedure for the time-to-event responses. The estimates obtained from these disjoint models, together with zeros for the association parameters  $\alpha$ , are provided as initial parameters to the NLMIXED procedure to fit the joint model.

Alternatively, users can provide initial parameters, that have been previously stored in a dataset, using the InitialParameters macro parameter. Appendix B.2 specifies the structure expected for the dataset of initial parameters. Sometimes it is useful to fit a first joint

Optimization technique	NLMIXEDOptions	Time	
		(minutes:seconds)	-2LL
Double-dogleg	tech = dbldog	3:44	8628.4
Quasi-Newton	tech = quanew	5:00	8628.4
Newton-Raphson with search	tech = newrap	6:54	8628.4
Newton-Raphson with ridging	tech = nrridg	5:04	8628.4
Trust region	tech = trureg	4:36	8628.4

Table 1: Example 1, execution times.

model, save the parameter estimates in a dataset using the `OutputParameters` macro parameter, and then use these estimates as initial parameters in subsequent calls to the macro.

### 3.5. Setup of the longitudinal model

For the longitudinal submodel, the `%JM` macro fits either normal, binary, binomial or Poisson responses by specifying the distribution in the `LongiType` macro parameter. The user might change the link function, from the canonical link that is used by default, by specifying `LongiLink = identity | logit | log | probit`.

The longitudinal model can include one or more baseline covariates (`LongiCovariates = <var list>`) and there are several options to model the trajectories of the longitudinal response over time (`LongiTimeModel = linear | spline | naturalcubic | bspline`). For each of these models it is possible to specify variables that interact with the trajectory (`LongiTimeInter = <var list>`). If one wishes to estimate not only a distinct trajectory shape but also a distinct intercept for a particular group, the corresponding 0–1 variable needs to be added as a baseline covariate as well.

For splines,  $B$ -splines and natural cubic splines users can specify the number of desired internal knots using the `LongiNknots` parameter. With respect to the location of the knots, if not given by the user, the macro places the knots using an equally-spaced-quantile partition of the longitudinal-response observation times. Splines and  $B$ -splines also require specifying the degree thus for example `LongiDegree = 3` would fit a cubic polynomial.

In addition,  $B$ -splines and natural cubic splines, use both a lower and an upper boundary knot although the role of these is slightly different. The  $B$ -splines basis functions  $B_{i,n}(t)$ , generated using the Boor recurrence relation (Racine 2018), fit a horizontal line outside the boundary knots thus these bases might not be adequate to predict data beyond them. On the other hand, natural cubic splines fit a straight, but not necessarily horizontal, line outside the boundary knots what makes this model often an attractive option to have prudent predictions for points beyond the observed time span. If not given by the user, using the `LongiLowerKnot` and `LongiUpperKnot` parameters, zero and the maximum observed time of the longitudinal responses are used as lower and upper boundary knots respectively.

Table 2 shows the possible structures available for the random effects variance-covariance matrix  $G$ .

### 3.6. Setup of the time-to-event model

The `%JM` macro allows using the popular exponential, Weibull and piecewise exponential dis-

Structure name	Macro parameter	$G$ matrix
Unstructured	<code>LongiGMatrix = un</code>	$\begin{bmatrix} \sigma_0^2 & \sigma_{10} & \sigma_{20} \\ \sigma_{10} & \sigma_1^2 & \sigma_{21} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix}$
Variance components	<code>LongiGMatrix = vc</code>	$\begin{bmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix}$
Ante-dependence	<code>LongiGMatrix = ante(1)</code>	$\begin{bmatrix} \sigma_0^2 & \rho_{10}\sigma_0\sigma_1 & \rho_{10}\rho_{21}\sigma_0\sigma_2 \\ \rho_{10}\sigma_0\sigma_1 & \sigma_1^2 & \rho_{21}\sigma_1\sigma_2 \\ \rho_{10}\rho_{21}\sigma_0\sigma_2 & \rho_{21}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

Table 2: Random-effects variance-covariance matrices.

tributions for the baseline hazard (`EventModel = exponential | weibull | piecewise`) as well as the generalizations of the Gompertz model using splines,  $B$ -splines or natural cubic splines (`EventModel = gspline | gbspline | gnaturalcubic`), the equivalent generalizations of the Weibull model (`EventModel = wspline | wspline | wnaturalcubic`) and the analogous models parameterized over the log-cumulative hazard function using the Royston and Parmar approach (`EventModel = rpspline | rpbspline | rpnaturalcubic`).

Baseline covariates are given through the `EventCovariates` parameter whereas exogenous time-dependent covariates can be included using a vertical data structure as explained in Appendix B.1.

For the piecewise exponential model and the models using splines, the macro allows the user to fix the number of knots (`EventNknots = <n>`) and/or the position of the knots (`Eventknot1 = <time>, ...`). If not specified by the user, six knots will be used that split the time period into an equally-spaced-quantile partition of observed, maybe log-transformed, times. In order to calculate the partition, the user might choose either to use the default option, all censored or non-censored event times or to use only actual (non-censored) events (`AdditionalOptions = eventtimes`).

In addition, it is possible to add a stratification variable (`EventStrata = <var name>`), that allows the macro to be used either for situations where the proportional hazards assumption does not hold or to fit time-to-events models with competing events.

### 3.7. Setup of the association structure

The parameterizations to associate the longitudinal model with the time-to-event model are selected setting the `SharedParam` parameter to one or several of the following options: `current_value`, `slope`, `cumulative` or `coefficients`. Some variables can modify the effect of the longitudinal response on the event risk model, as depicted in Equation 5; this can be modeled including these variables in the `SharedParamInt` parameter.

When a random coefficients parameterization is chosen (`SharedParam = coefficients`), the

user needs to specify the random effects to be included as shared parameters by specifying, for example, `SharedCoefficients = bi0 bi1`. Lagging in the association between the longitudinal and the time-to-event response is fixed by giving a positive value to the `SharedLagging` parameter. Lagging can be used together with the current-value, the slope or the cumulative parameterization thus, for example, a time-lagged slope-dependent model would result in Equation 9.

$$h_i(t) = h_0(t) \exp[\gamma^\top w_i + \alpha m'_i\{\max(0, t - c)\}]. \quad (9)$$

For the cumulative association a linear relationship between the log-hazard function and the cumulative linear predictor,  $\int_0^t g\{m_i(s)\}ds$ , is assumed. The macro approximates this cumulative value as illustrated below for a longitudinal model using splines:

$$\int_0^{T_i} g\{m_i(t)\}dt = \sum_{j=0}^{N+n} (\beta_j + b_{ij})C_j(T_i) + Z_i\gamma T_i,$$

where  $C_j(x) = \int_0^x B_j(t)dt$ ,  $B_j(t)$  are the spline basis functions and  $Z_i$  and  $\gamma$  are the design matrix and the coefficients associated with the baseline covariates. The cumulative basis functions  $C_j(x)$  are approximated using the Gauss-Kronrod rule for the observed times  $T_j$  and using the trapezoidal rule for the Gauss-Kronrod rule nodes  $t_{ik}$  required to approximate the cumulative hazard.

Several parameterizations, separated between spaces in the `SharedParam` parameter, can be combined. For example, the option `SharedParam = current_value slope` would lead to the following model

$$h_i(t) = h_0(t) \exp\{\gamma^\top w_i + \alpha_1 m_i(t) + \alpha_2 m'_i(t)\}.$$

Including multiple parameterizations should be done cautiously. It is possible, for instance, to include both a current-value-dependent shared parameter  $\alpha_1$  and a cumulative-dependent shared parameter  $\alpha_2$  in the same model. However such a parameterization may have high collinearity.

### 3.8. Macro installation

The following six macros need to be compiled, e.g., by using the `%include` statement:

```
%include '&location./calculateknotspartition.sas';
%include '&location./spline.sas';
%include '&location./ncspline.sas';
%include '&location./bspline.sas';
%include '&location./kronrodrule15p.sas';
%include '&location./jm.sas';
```

A convenient solution to avoid the `%include` statements in each program is to include these macros in a location referred by the SAS `autocall` facility.

## 4. Examples

### 4.1. An example with a normally-distributed longitudinal response

The fitting of a joint model will be illustrated with the dataset from a study to compare two antiretroviral drugs (Abrams *et al.* 1994). The square-root-transformed CD4 cell count, assumed normally distributed, is fitted using a random intercepts and slopes model with gender as baseline covariate and a separate slope for the subjects with DDI treatment.

$$m_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1} + \beta_2 \text{DDI}_i)t + \beta_3 \text{FEMALE}_i,$$

where  $\beta_0$  to  $\beta_3$  are the coefficients associated with fixed effects and  $b_{i0}$  and  $b_{i1}$  are random effects. The matrix of variances-covariances associated with the random effects was assumed unstructured so three parameters have to be estimated: the variances of  $b_{i0}$  and  $b_{i1}$  and the covariance between them.

The time to death is fitted using a Royston-Parmar model, using a natural cubic spline with 3 knots as smoothing function, and both gender and DDI as baseline covariates. Both models are associated through both the estimated current-value and the slope of the longitudinal response as given by

$$h_i(t) = h_0(t) \exp\{\gamma_1 \text{DDI}_i + \gamma_2 \text{FEMALE}_i + \alpha_1 m_i(t) + \alpha_2 m'_i(t)\}.$$

This model is fitted using the following code:

```
%JM(Data = mydata.aids, SubjectVar = patient, LongiTimeModel = linear,
  LongiVar = cd4, LongiTimevar = obstime, LongiTimeInter = ddi,
  LongiCovariates = female, LongiGMatrix = un, EventTimeVar = time,
  EventVar = death, EventVal = 1, EventModel = rpnaturalcubic,
  EventCovariates = ddi female, EventNknots = 3, NLMIXEDOptions = gconv=0,
  OutputPredictions = mi, OutputPredictData = mydata.aids_mi,
  SharedParam = current_value slope, AdditionalOptions = calculateexectime);
```

Some input parameters (`LongiType = normal`, `InitialParameters = disjoint`) are not specified because the default values of the macro are used. The omission of the `method` and `technique` options in the `NLMIXEDOptions` macro parameter implies that the default integration method and optimization technique, adaptive Gaussian quadrature and dual quasi-Newton respectively, are used. The inclusion of `gconv=0` in this macro parameter is convenient to avoid the premature stopping of the optimization algorithm due to a relative change in the gradient being very small without the gradient value itself being sufficiently small (Kieran, Tao, and Gibbs 2012). The input parameter `OutputPredictions` is used to request the NLMIXED procedure to save a dataset with subjects' predicted longitudinal responses  $m_i(t)$ , using the empirical Bayes estimates for the random effects. For this model, the macro provides the following output:

Joint Model Summary:

Longitudinal Response Distribution: Normal. Link function: identity.

Longitudinal Process: Random Intercepts and Slopes Model

Event Process: (Natural-cubic-splines-based) Royston-Parmar model

baseline risk function.  
 Parameterization: current\_value slope.  
 Event Process: 467 subjects, 188 events.

The NL MIXED Procedure

Specifications

Data Set	WORK._JMDATA	
Dependent Variable		JMResponse
Distribution for Dependent Variable		General
Random Effects		bi0 bi1
Distribution for Random Effects		Normal
Subject Variable		patient
Optimization Technique		Dual Quasi-Newton
Integration Method		Adaptive Gaussian Quadrature

Dimensions

Observations Used	1872
Observations Not Used	0
Total Observations	1872
Subjects	467
Max Obs Per Subject	6
Parameters	17
Quadrature Points	1

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	3	4357.95401	36.53343	1056.432	-4452.23
.	.	.	.	.	.
68	134	4314.219	5.28E-11	0.000372	-0.00002

NOTE: FCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	8628.4
AIC (smaller is better)	8662.4
AICC (smaller is better)	8662.8
BIC (smaller is better)	8732.9

## Additional Estimates

Label	Estimate	Standard Error ...
Longitudinal Process: Intercept	7.1637	0.2322 ...
Longitudinal Process: Slope	-0.2247	0.02241 ...
Longitudinal Process: DDI by slope	-0.00143	0.03057 ...
Longitudinal Process: FEMALE	0.6356	0.7474 ...
Longitudinal Process: Log SD Residual	0.5593	0.02714 ...
Longitudinal Process: Log SD Random Intercept	1.5178	0.03582 ...
Longitudinal Process: Log SD Random Slope	-1.6497	0.09648 ...
Longitudinal Process: Covariance Intercept by Slope	0.06422	0.07178 ...
Event Process: Spline parameter 0	-5.1115	0.7846 ...
Event Process: Spline parameter 1	1.4639	0.3083 ...
Event Process: Spline parameter 2	0.002638	0.05402 ...
Event Process: Spline parameter 3	-0.4785	1.0466 ...
Event Process: Spline parameter 4	1.0682	2.5062 ...
Event Process: DDI	0.3938	0.2688 ...
Event Process: FEMALE	0.5655	0.3802 ...
Association (current-value)	-0.3721	0.05389 ...
Association (slope-dependent)	-5.7478	1.4133 ...

Execution Time (hours:minutes:seconds): 0:05:00.

The joint model finds a strong association of both the square root CD4 cell count and its slope with the risk of death. A unit decrease in the marker corresponds to a  $\exp(0.3721) = 1.45$ -fold increase in the risk of death while a unit decrease in the slope is associated with a  $\exp(5.7478) = 313.50$ -fold increase in this risk.

The NLMIXED procedure selected one quadrature point, resulting in the Laplace approximation. Alternatively, end users can force the use of a specific number of quadrature points by adding, for example, `qpoints=5` to the `NLMIXEDOptions` macro parameter. Figure 1 illustrates, for a group of parameters, the estimates obtained depending on the number of quadrature points used to approximate the marginal log-likelihood function.

The macro does not directly produce graphs but allows saving estimates and predictions in datasets so users can create the desired additional output. The replication scripts produce two additional plots: Figure 2 shows the predicted longitudinal response versus time of patients 1 to 20; Figure 3 provides four diagnostic plots.

## 4.2. An example with a binary longitudinal response

In this section, the fitting of a joint model with non-normal longitudinal responses will be illustrated with the dataset of Section 4.1 on the binary longitudinal response  $\pi_i$  defined as having the CD4 cell count higher or equal to 30 cells per cubic millimeter.

$$\log\left\{\frac{\pi_i(t)}{1 - \pi_i(t)}\right\} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1} + \beta_2 \text{DDI}_i)t + \beta_3 \text{FEMALE}_i. \quad (10)$$

The time to death is fitted using the same Royston-Parmar model used in Section 4.1. Both

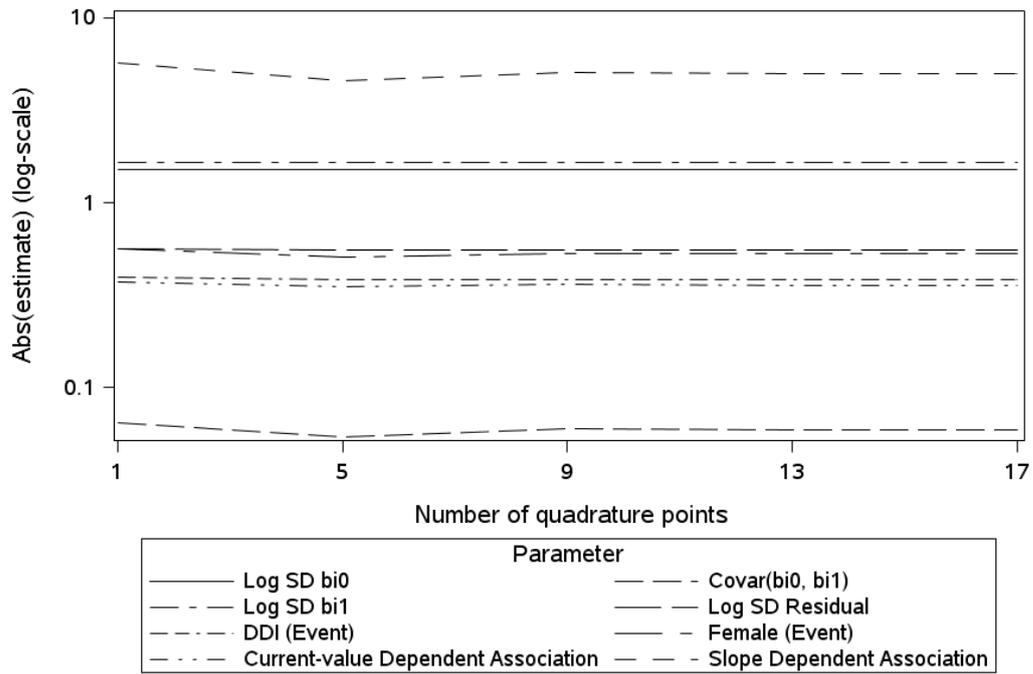


Figure 1: Example 1, joint model estimates versus number of quadrature points.

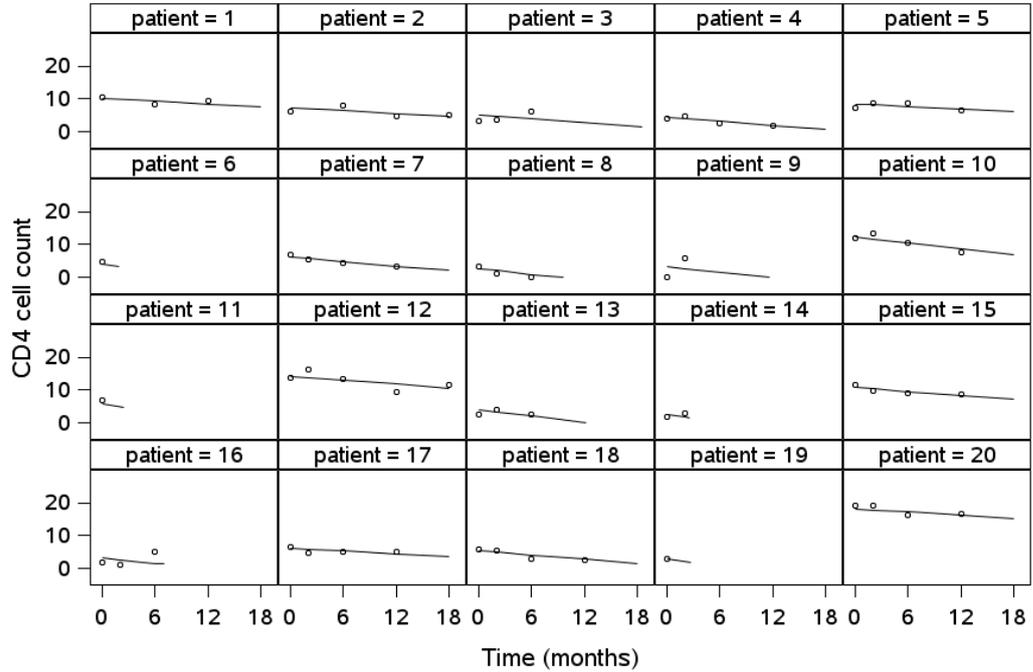


Figure 2: Example 1, joint model empirical Bayes estimates versus time of 20 subjects.

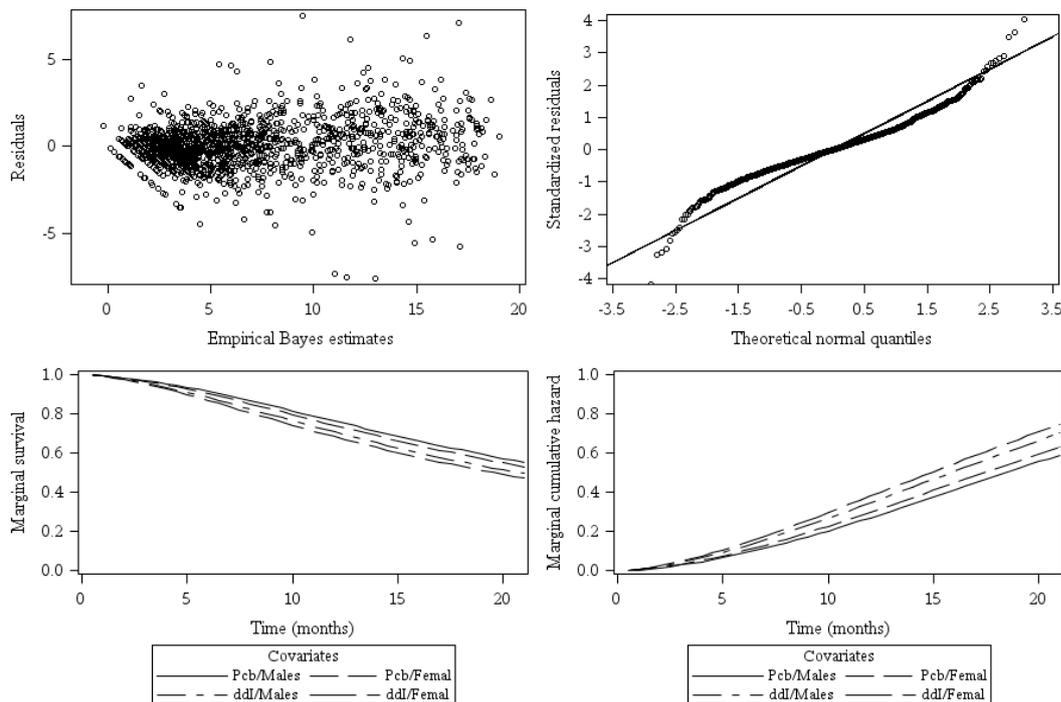


Figure 3: Example 1, diagnostic plots.

submodels are associated using the random effects coefficients as shared parameters.

$$h_i(t) = h_0(t) \exp(\gamma_1 \text{DDI}_i + \gamma_2 \text{FEMALE}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1}). \quad (11)$$

For non-normal data, the %JM macro uses first the GLIMMIX procedure to fit the longitudinal model, in this case a random-effects logistic regression, and takes the estimates as initial parameters for the joint model. The following code shows how this longitudinal model is fitted with the GLIMMIX procedure:

```
proc glimmix data = mydata.aids;
  class patient;
  model cd4_a30 (Event = "1") = obstime obstime*ddi female
    / distribution = binary link = logit;
  random int obstime/ sub = patient type = un;
run;
```

By default, the GLIMMIX procedure uses pseudo-likelihood (PL) estimation methods based on linearization. The PL methods can produce biased estimates in situations where the number of observations per subject is small, particularly, for binary data (Breslow and Lin 1995; Lin and Breslow 1996; Flom, McMahon, and Pouget 2006; Pinheiro and Chao 2006). In line with this, we can set the LongiModelOptions macro parameter to `method=quad` to avoid the PL approach and use the adaptive Gauss-Hermite approximation in the submodel used to estimate the initial parameters of the longitudinal response. This joint model is therefore fitted using the following code:

```
%JM(Data = mydata.aids, SubjectVar = patient, LongiType = binary,
    LongiTimeModel = linear, LongiVar = cd4, LongiTimevar = obstime,
    LongiTimeInter = ddi, LongiCovariates = female, LongiGMatrix = un,
    LongiModelOptions = method=quad, EventTimeVar = time, EventVar = death,
    EventVal = 1, EventModel = rpnaturalcubic, EventCovariates = ddi female,
    EventNknots = 3, NLMIXEDOptions = gconv=0, OutputPredictions = mi,
    OutputPredictData = mydata.aids_mi, SharedParam = coefficients,
    SharedCoefficients = bi0 bi1, AdditionalOptions = calculateexetime);
```

The macro provides the following output:

Joint Model Summary:

Longitudinal Response Distribution: Binary. Link function: logit.

Longitudinal Process: Random Intercepts and Slopes Model

Event Process: (Natural-cubic-splines-based) Royston-Parmar model  
baseline risk function.

Parameterization: coefficients.

Event Process: 467 subjects, 188 events.

Dimensions

Observations Used	1872
Observations Not Used	0
Total Observations	1872
Subjects	467
Max Obs Per Subject	6
Parameters	16
Quadrature Points	11

(Part of the output has been omitted)

The NLMIXED Procedure

Additional Estimates

Label	Estimate	Standard Error ...
Longitudinal Process: Intercept	0.5921	0.2896 ...
Longitudinal Process: Slope	-0.2676	0.06017 ...
Longitudinal Process: DDI by slope	0.01252	0.06276 ...
Longitudinal Process: FEMALE	0.09023	0.9430 ...
Longitudinal Process: Log SD Random Intercept	1.5272	0.1215 ...
Longitudinal Process: Log SD Random Slope	-1.5249	0.4717 ...
Longitudinal Process: Covariance Intercept by Slope	0.4700	0.4203 ...

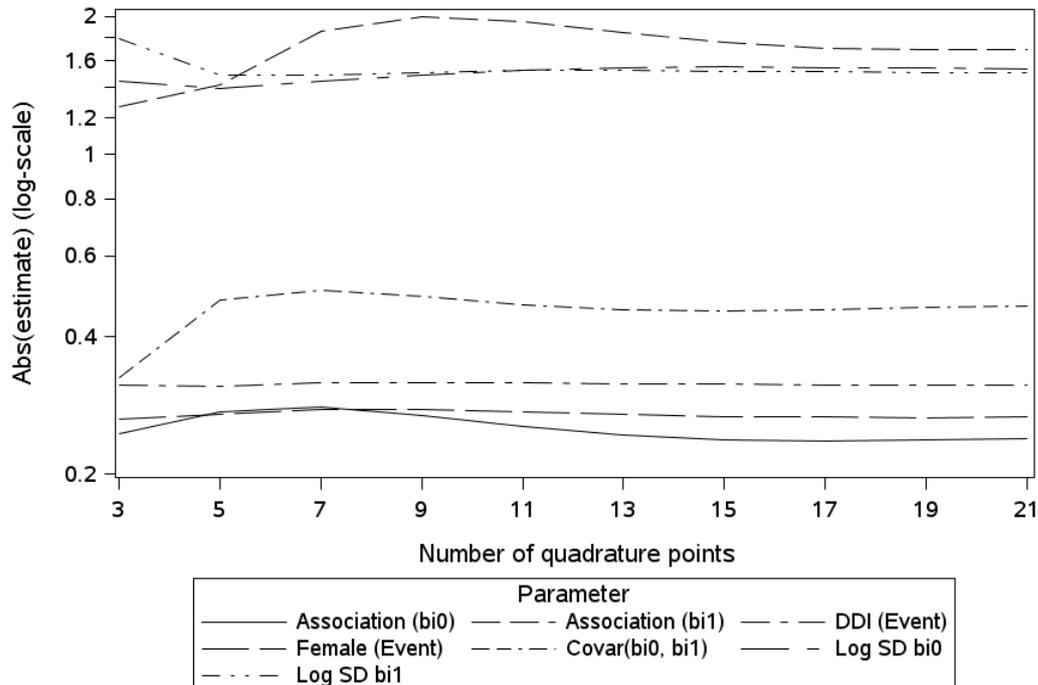


Figure 4: Example 2, joint model estimates versus number of quadrature points.

```

Event Process: Spline parameter 0          -5.2979    0.5002 ...
Event Process: Spline parameter 1          1.4566    0.2993 ...
Event Process: Spline parameter 2          0.02640   0.05466 ...
Event Process: Spline parameter 3         -0.6294    1.0858 ...
Event Process: Spline parameter 4          1.3301    2.6382 ...
Event Process: DDI                          0.3170    0.1693 ...
Event Process: FEMALE                       0.2733    0.3106 ...
Association (BI0-dependent)                 -0.2542    0.1082 ...
Association (BI1-dependent)                 1.9533    2.7321 ...

```

Execution Time (hours:minutes:seconds): 1:29:26.

The NL MIXED procedure used eleven quadrature points to approximate the marginal log-likelihood function using the adaptive Gauss-Hermite method requiring an execution time of about 90 minutes. This result is in line with previous evidence showing that the Gauss-Hermite technique requires a large number of quadrature points in the random-effects logistic regression model (Lesaffre and Spiessens 2001). In this context, it is advisable to check the sensitivity of the results to changes in the number of quadrature points, thus the model is successively fitted with different numbers of quadrature points, by setting the NL MIXED Options macro parameter to `qpoints=<n>`. Figure 4 shows for a selected subset the absolute values of the parameter estimates in dependence of the number of quadrature points.

## 5. Conclusions

The joint modeling of longitudinal responses and time to event responses is useful in situations when we suspect that the two processes may be associated, such as the survival analysis with time-dependent covariates or the analysis of longitudinal responses with informative dropouts. In this context the methods based on a joint likelihood formulation are known to yield precise inferences, but be computationally demanding (Tsiatis and Davidian 2004).

The literature and software available for this class of models is still limited. Some solutions are available in R and Stata although their focus is mostly on normally-distributed longitudinal responses (Rizopoulos 2010; Philipson *et al.* 2018; Proust-Lima *et al.* 2017a; Crowther 2012; Rizopoulos 2017).

This paper illustrated the possibilities of the SAS NL MIXED procedure to fit this group of models. This procedure allows defining the conditional log-likelihood function, which is different for the longitudinal response and the time-to-event response, and uses either adaptive or non-adaptive Gauss-Hermite quadrature to approximate the integral of the conditional log-likelihood function over the random effects.

We presented %JM, a SAS macro that uses the NL MIXED procedure to fit jointly generalized mixed models for longitudinal data and proportional hazards models for time-to-event responses. The macro fits normal, binary, binomial and Poisson longitudinal responses and provides several options to fit non-linear trajectories of the longitudinal response over time. For the time-to-event response, that might be right-censored, the macro fits parametric, stratified or not, proportional hazards models allowing for a broad range of risk functions. %JM offers several options to connect the longitudinal and the time-to-event models.

Some of the limitations of %JM are inherited from the NL MIXED procedure. The approximation of the marginal log-likelihood function is likely to require longer execution times than the pseudo-adaptive approach proposed by Rizopoulos (2012a) for these models, available in the JM R package. One of the strengths of this macro is that it is not limited to normal longitudinal responses, however the modeling of non-normal longitudinal responses is expected to be more computationally intensive (Lesaffre and Spiessens 2001). Macro users are encouraged to routinely check the dependence of the results on the number of quadrature points.

## References

- Abrams DI, Goldman AI, Launer C, Korvick JA, Neaton JD, Crane LR, Grodesky M, Wakefield S, Muth K, Kornegay S, *et al.* (1994). “Comparative Trial of Didanosine and Zalcitabine in Patients with Human Immunodeficiency Virus Infection Who are Intolerant of or Have Failed Zidovudine Therapy.” *New England Journal of Medicine*, **330**(10), 657–662. doi:10.1056/nejm199403103301001.
- Breslow NE, Lin X (1995). “Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion.” *Biometrika*, **82**(1), 81–91. doi:10.2307/2337629.
- Crowther MJ (2012). *STJM11: Stata Module to Fit Shared Parameter Joint Models of Longitudinal and Survival Data*. Stata module version 12.1, URL <http://EconPapers.RePEc.org/software/bocbocode/s457502.htm>.

- Farouki R, Goodman T (1996). “On the Optimal Stability of the Bernstein Basis.” *Mathematics of Computation of the American Mathematical Society*, **65**(216), 1553–1566. doi:[10.1090/s0025-5718-96-00759-4](https://doi.org/10.1090/s0025-5718-96-00759-4).
- Farouki RT, Rajan VT (1987). “On the Numerical Condition of Polynomials in Bernstein Form.” *Computer Aided Geometric Design*, **4**(3), 191–216. doi:[10.1016/0167-8396\(87\)90012-4](https://doi.org/10.1016/0167-8396(87)90012-4).
- Faucett CL, Thomas DC (1996). “Simultaneously Modelling 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](https://doi.org/10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1).
- Flom PL, McMahon JM, Pouget ER (2006). “Using PROC NLMIXED and PROC GLIMMIX to Analyze Dyadic Data with Binary Outcomes.” In *Proceedings of 19th Annual Conference of the Northeast SAS Users Group*, 179–2007. URL <http://www2.sas.com/proceedings/forum2007/179-2007.pdf>.
- Guo X, Carlin BP (2004). “Separate and Joint Modelling of Longitudinal and Event Time Data Using Standard Computer Packages.” *The American Statistician*, **58**(1), 16–24. doi:[10.1198/0003130042854](https://doi.org/10.1198/0003130042854).
- Harrell FE (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag. doi:[10.1007/978-3-319-19425-7](https://doi.org/10.1007/978-3-319-19425-7).
- Hastie T, Tibshirani R, Friedman J (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd edition. Springer-Verlag, New York. doi:[10.1007/978-0-387-84858-7](https://doi.org/10.1007/978-0-387-84858-7).
- Hsieh F, Tseng YK, Wang JL (2006). “Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited.” *Biometrics*, **62**(4), 1037–1043. doi:[10.1111/j.1541-0420.2006.00570.x](https://doi.org/10.1111/j.1541-0420.2006.00570.x).
- Kiernan K, Tao J, Gibbs P (2012). “Tips and Strategies for Mixed Modeling with SAS/STAT Procedures.” In *SAS Global Forum*, 332–2012. URL <http://support.sas.com/resources/papers/proceedings12/332-2012.pdf>.
- Lesaffre E, Spiessens B (2001). “On the Effect of the Number of Quadrature Points in a Logistic Random Effects Model: An Example.” *Journal of the Royal Statistical Society C*, **50**(3), 325–335. doi:[10.1111/1467-9876.00237](https://doi.org/10.1111/1467-9876.00237).
- Lin X, Breslow NE (1996). “Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion.” *Journal of the American Statistical Association*, **91**(435), 1007–1016. doi:[10.1080/01621459.1996.10476971](https://doi.org/10.1080/01621459.1996.10476971).
- Philipson P, Diggle P, Sousa I, Kolamunnage-Dona R, Williamson P, Henderson R (2018). **joineR**: *Joint Modelling of Repeated Measurements and Time-to-Event Data*. R package version 1.2-3, URL <https://CRAN.R-project.org/package=joineR>.
- Pinheiro JC, Chao EC (2006). “Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models.” *Journal of Computational and Graphical Statistics*, **15**(1), 58–81. doi:[10.1198/106186006x96962](https://doi.org/10.1198/106186006x96962).

- Prentice RL (1982). “Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model.” *Biometrika*, **69**(2), 331–342. doi:10.1093/biomet/69.2.331.
- Proust-Lima C, Philipps V, Diakite A, Lique B (2017a). *lcmm: An R Package for Estimation of Latent Class Mixed Models and Joint Latent Class Models*. R package version 1.7.8, URL <https://CRAN.R-project.org/package=lcmm>.
- Proust-Lima C, Philipps V, Lique B (2017b). “Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package **lcmm**.” *Journal of Statistical Software*, **78**(2), 1–56. doi:10.18637/jss.v078.i02.
- Racine J (2018). *A Primer on Regression Splines*. Vignette available in the R package **crs**, URL <https://CRAN.R-project.org/package=crs>.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 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*. Chapman & Hall/CRC. doi:10.1201/b12208.
- Rizopoulos D (2016). “The R Package **JMbayes** for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC.” *Journal of Statistical Software*, **72**(7), 1–45. doi:10.18637/jss.v072.i07.
- Rizopoulos D (2017). *JMbayes: Joint Modeling of Longitudinal and Time-to-Event Data under a Bayesian Approach*. R package version 0.8-70, URL <https://CRAN.R-project.org/package=JMbayes/>.
- Rizopoulos D, Ghosh P (2011). “A Bayesian Semiparametric Multivariate Joint Model for Multiple Longitudinal Outcomes and a Time-to-Event.” *Statistics in Medicine*, **30**(12), 1366–1380. doi:10.1002/sim.4205.
- Rizopoulos D, Lesaffre E (2014). “Introduction to the Special Issue on Joint Modelling Techniques.” *Statistical Methods in Medical Research*, **23**(1), 3–10. doi:10.1177/0962280212445800.
- Royston P, Parmar MK (2002). “Flexible Parametric Proportional-Hazards and Proportional-Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects.” *Statistics in Medicine*, **21**(15), 2175–2197. doi:10.1002/sim.1203.
- SAS Institute Inc (2004). *SAS/STAT Software, Version 9.1*. Cary. URL [http://support.sas.com/documentation/onlinedoc/91pdf/index\\_91.html](http://support.sas.com/documentation/onlinedoc/91pdf/index_91.html).

- Stata Corp (2014). *Stata Data Analysis Statistical Software: Release 14*. College Station, Texas. URL <http://www.stata.com/stata14/>.
- Tsiatis AA, Davidian M (2004). “Joint Modeling of Longitudinal and Time-to-Event Data: An Overview.” *Statistica Sinica*, **14**(3), 809–834.
- Williamson P, Kolamunnage-Dona R, Philipson P, Marson AG (2008). “Joint Modelling of Longitudinal and Competing Risks Data.” *Statistics in Medicine*, **27**(30), 6426–6438. doi:10.1002/sim.3451.
- Wu L, Liu W, Yi GY, Huang Y (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. Article ID 640153.
- Wulfsohn MS, Tsiatis AA (1997). “A Joint Model for Survival and Longitudinal Data Measured with Error.” *Biometrics*, **53**(1), 330–339. doi:10.2307/2533118.
- 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.

## A. Macro parameters

The %JM macro needs to be called using keyword parameters. Tables 3, 4 and 5 provide the description of all %JM macro parameters.

## B. Input and output datasets

### B.1. Input dataset

Table 6 illustrates the standard structure for the input dataset with separate columns for the longitudinal and the time-to-event responses. In addition, the macro allows using a vertical input dataset, with separate rows for the longitudinal and the time-to-event responses, by specifying `datastructure = vertical` in the macro call.

A vertical structure requires a variable called `JMDist` that should equal `LONGI` for the lon-

Macro parameter	Description
<code>Data = &lt;dataset&gt;</code>	Name of the input dataset. Required.
<code>DataStructure = standard   vertical</code>	Structure of the input dataset. Standard: input data with separate columns for longitudinal and time-to-event responses (default); vertical: input data structured in separate rows for longitudinal and time-to-event responses. See further details in Appendix B.1.
<code>Where = &lt;condition&gt;</code>	Condition to select a subset of rows.
<code>SubjectVar = &lt;var name&gt;</code>	Name of the variable that identifies the subject. Required.
<code>InitialParameters = disjoint   &lt;dataset&gt;</code>	Option to estimate initial parameters. <code>disjoint</code> : uses parameters obtained from the disjoint longitudinal and time-to-event submodels (default); <code>&lt;dataset&gt;</code> : uses a dataset where a set of initial parameters has been stored, e.g., obtained from a previous run to the macro.
<code>OutputParameters = &lt;library.dataset&gt;</code>	Dataset to store the estimated model parameters. By default, the estimated parameters are saved in <code>work._jmoutputparameters</code> . See further details on the structure of this dataset in Appendix B.2.
<code>Outputpredictions = &lt;list of model terms&gt;</code>	Requests the macro to create datasets with predictions. Available terms are: <code>XiBeta</code> , <code>XiBeta_Zibi</code> , <code>mi log_hi</code> and <code>cum_hi</code> . It is possible to specify a transformation of these such as <code>exp(-cum_hi)</code> . When the term depends not only on fixed effects but also on random effects, the empirical Bayes estimates of these are used.
<code>OutputpredictData = &lt;list of dataset names&gt;</code>	List of names of the (libraries and) datasets used to save the predictions.
<code>ListingFile = 'file name'</code>	File to send the output by the <code>ods listing</code> statement.

Table 3: %JM macro parameters (Part 1 of 3).

Macro parameter	Description
LongiType = normal   binary   binomial   poisson	Distribution of the longitudinal response $y_i(t)$ . <code>normal</code> is used by default.
LongiLink = identity   logit   log   probit	Link function. The canonical link is used by default.
LongiVar = <var name>	Name of the variable with the longitudinal response. Required.
LongiEventCat = <category>	For binary longitudinal responses only, category of interest. Default value is 1.
LongiTrialsVar = <var name>	For binomial longitudinal responses only, name of the variable with number of trials.
LongiOffset = <var name>	Name of the variable to be included in the linear model with a coefficient fixed to 1.
longiTimeVar = <var name>	Name of the variable with the time at which the longitudinal response has been observed. Required.
LongiTimeModel = linear   spline   bspline   naturalcubic	Approach to fit the trajectories of the longitudinal response versus time. Required.
LongiLowerKnot = <time> LongiUpperKnot = <time>	Lower and upper boundary knots for the modeling of the longitudinal response using $B$ -splines or natural cubic splines. For further details see Sections 2.1 and 3.5.
LongiNKnots = <positive integer>	Number of internal knots for the modeling of the longitudinal response using splines, $B$ -splines or natural cubic splines. No internal knots are assumed by default.
LongiKnot1 = <time>, ..., LongiKnot9 = <time>	Internal knots for the modeling of the longitudinal response using splines. For further details see Sections 2.1 and 3.5.
LongiCovariates = <var list>	List of baseline covariates of the longitudinal response model.
LongiTimeInter = <var name>	Name of the variable that interacts with the longitudinal trajectory.
LongiGMatrix = vc   un   ante(1)	Structure of the $G$ matrix. <code>vc</code> : variance components (default); <code>un</code> : unstructured; <code>ante(1)</code> : ante-dependence.
LongiGMatrixParam = varcovar   chol   varcorr	Parameterization of the $G$ matrix. <code>varcovar</code> : parameters for variances and covariances (default); <code>chol</code> : parameters for the Cholesky decomposition; <code>varcorr</code> : parameters for variances and correlations.
LongiSmallLogSD = <value>	Value to use as initial value for the log-transformed standard deviation of random effects where the disjoint linear model has estimated 0. Default value is $-10$ .
LongiModelOptions = <options>	Options in the MIXED (or GLIMMIX) statement of the MIXED (or GLIMMIX) procedure used to estimate the initial parameters. For normal responses, by default, <code>method = ml</code> .
LongiGLINLOptions = <options>	For non-normal responses, options in the <code>nloptions</code> statement of the GLIMMIX procedure used to estimate the initial parameters.

Table 4: %JM macro parameters (Part 2 of 3).

Macro parameter	Description
EventTimeVar = <var name>	Name of the variable with the time at which the event or the censoring occurs. Required.
EventVar = <var name>	Name of the variable with the indicator of event or censoring. Required.
EventVal = <category>	Category that indicates event of interest. All other categories will be assumed censored observations. Required.
EventModel = exponential weibull   piecewise   gspline gbspline   gnaturalcubic wspline   wbspline wnaturalcubic   rpspline rpbspline   rpnaturalcubic	Distribution of the baseline hazard in the time-to-event model. Required.
EventTimeStart = <var name>	For data in vertical structure, name of the variable with the start of the interval.
EventStrata = <var name>	For stratified models, name of the variable to identify the strata.
EventNKnots = <positive integer>	Number of internal knots for the modeling of the time-to-event responses using a piecewise exponential model or splines. Default value is 6.
EventKnot1 = <time>, ..., EventKnot9 = <time>	Internal knots for the modeling of the time-to-event responses using a piecewise exponential model or splines.
EventCovariates = <var list>	List of baseline covariates of the time-to-event model.
SharedParam = current_value   slope   cumulative   coefficients	Parameterization to associate the longitudinal model with the time-to-event model. Multiple choices are possible. The current-value parameterization is used by default. For further details see Sections 2.3 and 3.7.
SharedLagging = <positive real>	Lagging in the association between the longitudinal and the time-to-event response. Default value is 0.
SharedParamInt = <var list>	List of variables that interact with the effect of the longitudinal response on the event risk model.
SharedCoefficients = bi0   bi1   bi2   bi3   ...	For the random-effects coefficients parameterization, random effects used as shared parameters. Required if SharedParam = coefficients.
SharedLongiTerm = mean   linear_predictor   coefficients	Option to link the longitudinal model and the time-to-event model. See Section 2.3 for further details.
NLMIXEDOptions = <options>	Options in the NLMIXED statement of the NLMIXED procedure used to fit the joint model.
AdditionalOptions = <option>	Requests the macro additional options. See further details in Appendix C.
SmallTimeDif = <time>	Amount of time small enough to adequately estimate the slope of $m_i(t)$ using the central difference approximation. 0.001 is used by default.

Table 5: %JM macro parameters (Part 3 of 3).

Patient	DDI	Obstime	CD4	Death	Time
1	0	0	10.7	0	16.97
1	0	6	8.4	0	16.97
1	0	12	9.4	0	16.97
1	0	18	.	0	16.97

Table 6: Standard input data structure.

Patient	JMDist	DDI	Obstime	CD4	Start	Stop	Weight	Death
1	LONGI	0	0	10.7				
1	LONGI	0	6	8.4				
1	LONGI	0	12	9.4				
1	LONGI	0	18	.				
1	EVENT	0	.	.	0	5	75.1	0
1	EVENT	0	.	.	5	10	72.8	0
1	EVENT	0	.	.	10	16.97	71.6	0

Table 7: Vertical input data structure.

gitudinal responses and `EVENT` for the time-to-event responses as shown in Table 7. Using a vertical structure is convenient either to include time-dependent covariates in the model or to increase the precision of the approximation of the survival function. With the data of Table 7, the macro would apply the Gauss-Kronrod rule separately for each of the three time intervals shown.

## B.2. Parameter estimates and predictions datasets

In addition to the standard output displayed in the examples of Sections 4.1 and 4.2, the parameter estimates are stored in the dataset specified by the `OutputParameters` macro parameter. If not given by the user, the estimates are saved in `work._jmoutputparameters`. Table 8 shows the parameters dataset created in the example of Section 4.1. This dataset includes not only the information displayed in the standard output but also, for each rescaled parameter, the gradient of the negative log-likelihood.

In addition, the `OutputPredictions` macro parameter can be used to request the `NLMIXED` procedure to save one or more datasets with subjects' predictions as illustrated in Section 4.1. This dataset has the usual format produced by the `PREDICT` statement.

## C. Additional options

The following options can be requested by adding one or more of the keywords below to the `AdditionalOptions` macro parameter:

- `calculateexetime`: calculates and displays the macro execution time.
- `competing`: informs the macro that the variable given in the `EventStrata` parameter defines competing events. With this option, the macro expects a vertical data structure where rows have been included for each of the competing time-to-event responses.

Parameter	Estimate	StandardError	DF	tValue	Probt	...
L_INTERCEPT	7.1637	0.2322	465	30.85	< .0001	...
L_TIME_B1	-0.2247	0.0224	465	10.03	< .0001	...
L_TIME_B1_BY_DDI	-0.0014	0.0306	465	-0.05	0.9628	...
L_FEMALE	0.6356	0.7474	465	0.85	0.3955	...
L_LOGSD_INTERCEPT	1.5178	0.0358	465	42.37	< .0001	...
L_COV_B0_B1	0.0642	0.0718	465	-0.89	0.3714	...
L_LOGSD_TIME_B1	-1.6497	0.0965	465	17.10	< .0001	...
L_LOGSD_RESIDUAL	0.5593	0.0271	465	20.61	< .0001	...
E_DDI	0.3938	0.2688	465	1.47	0.1435	...
E_FEMALE	0.5655	0.3802	465	1.49	0.1376	...
E_TIME_B0	-5.1115	0.7846	465	6.51	< .0001	...
E_TIME_B1	1.4639	0.3083	465	4.75	< .0001	...
E_TIME_B2	0.0026	0.0540	465	0.05	0.9611	...
E_TIME_B3	-0.4785	1.0466	465	-0.46	0.6477	...
E_TIME_B4	1.0682	2.5062	465	0.43	0.6701	...
ASSOCT_TD	-0.3721	0.0539	465	-6.90	< .0001	...
ASSOCT_SL	-5.7478	1.4133	465	-4.07	< .0001	...

Table 8: Output parameter estimates dataset.

- **eventtimes**: uses only actual (non-censored) events to calculate the time partition for the time-to-event model. Applicable for the piecewise exponential model and the time-to-event models using splines.
- **fitstatistics**: creates a dataset with the fit statistics. The dataset will be placed in the same location and with the same name as the parameters dataset, but adding the suffix `_fit`.
- **onlyevent**: fits only the time-to-event model.
- **onlylongi**: fits only the longitudinal model.
- **noint**: removes the intercept from the longitudinal model.
- **nomacrowarn**: disables all macro warnings.
- **skipmacroheader**: skips the header produced by the macro before the execution of the NL MIXED procedure.
- **untouchedlisting**: disables all macro `ods listing` statements.

**Affiliation:**

Alberto Garcia-Hernandez  
Global Data Science  
Astellas Pharma Global Development Europe  
Sylviusweg 62, 2333 BE Leiden, The Netherlands  
E-mail: [alberto.garciahernandez@astellas.com](mailto:alberto.garciahernandez@astellas.com)

Dimitris Rizopoulos  
Department of Biostatistics  
Erasmus University Medical Center  
PO Box 2040, 3000 CA Rotterdam, The Netherlands  
E-mail: [d.rizopoulos@erasmusmc.nl](mailto:d.rizopoulos@erasmusmc.nl)  
URL: <http://eur.academia.edu/DimitrisRizopoulos>