Paper 3175-2019

Joint Analysis of Failure Times and Time-Varying Covariates

Joseph C. Gardiner, Department of Epidemiology and Biostatistics,

Michigan State University, East Lansing, MI 48824

Abstract

Statistical models for analyses of failure times include the proportional hazards model and the accelerated failure time model. These models can be extended to assess the influence of a longitudinally assessed biomarker (a time-varying covariate) on the survival distribution by modeling the hazard function or the scale parameter of a parametric survival distribution. If the biomarker is updated intermittently at a few time points, a straight-forward approach applies the most recent values preceding the failure times. SAS® procedures PHREQ, LIFEREG and SEVERITY can be used for analyses. Joint models for the failure time and biomarker parse their joint distribution into conditionally independent components given random effects. Using the GLIMMIX procedure, the biomarker trajectory is constructed as a linear function of random effects and polynomials or splines of time. When incorporated into the survival model as a time-varying covariate, the joint model, called a shared parameter model is estimated using the NLMIXED procedure. The joint model provides a more complete use of the data on failure times and the longitudinal data on the biomarker. Recent software developments, including a SAS macro, have harnessed SAS procedures to address analyses of shared parameter models. We provide a brief overview of methods and demonstrate their application with previously published biomedical data.

1. Introduction

The Cox proportional hazards (PH) model and the accelerated failure time (AFT) model are two standard approaches to analysis of survival times. Data on sampled units comprise covariates $\mathbf{x}$ assessed at the time origin $t=0$ and a single time to event T that may be observed in the follow up period ending at U, if $T \le U$ or the right censoring time U is observed if $T > U$. The objective is to assess the influence of $\mathbf{x}$ on some aspects of the (conditional) survival distribution of T, $S(t \mid \mathbf{x}) = P[T > t \mid \mathbf{x}]$. In the PH model the hazard function has the form $h(t \mid \mathbf{x}) = h_0(t)\exp(\mathbf{x}'\beta)$ where the baseline hazard $h_0(t)$ is unspecified. In biomedical applications the importance of the parameters β cannot be overemphasized. They have an interpretation as log hazard ratios. For example, if x_1 is a binary indicator for treatment, and all other covariates are held fixed, the hazard ratio $\exp(\beta_1)$ measures the adjusted relative impact of treatment on survival. Estimation of β_1 via optimization of the partial likelihood provides a basis for inference. For specified covariate profiles, we get an estimate the survival function from $S(t \mid \mathbf{x}) = \exp\left(-H_0(t)\exp(\mathbf{x}'\beta)\right)$ where $H_0(t)$ is the cumulative baseline hazard. PROC PHREG provides for comprehensive analyses of the PH model.

The AFT model structures $\log T = \mathbf{x}'\beta + \sigma\varepsilon$ where the random ε has a specified parametric distribution independent of $\mathbf{x}$ and $\sigma > 0$ is a scale parameter. The AFT class includes the lognormal, log-logistic and Weibull distributions. The Weibull distribution is the only continuous distribution

that is in both AFT and PH classes. Its survival and hazard functions are $S(t \mid \mathbf{x}) = \exp\left(-(t / \theta(\mathbf{x}))^{\gamma}\right)$, $h(t \mid \mathbf{x}) = \gamma t^{\gamma-1} \exp(-\gamma \mathbf{x}'\beta)$, where the scale is modeled as $\log \theta(\mathbf{x}) = \mathbf{x}'\beta$ and $\gamma = \sigma^{-1}$ is the shape parameter. Scale parameter (SP) models have the form $S(t \mid \mathbf{x}) = S_0\left((t / \theta(\mathbf{x}))^{\gamma}\right)$ where S_0 is a survival distribution. A prototype is the Burr distribution that specifies $S_0(t) = \left(1 + \alpha^{-1} t\right)^{-\alpha}$ where $\alpha > 0$ is a shape parameter. Special cases are the log-logistic $(\alpha = 1)$ and Pareto $(\gamma = 1)$ distributions, and the Weibull, $S_0(t) = \exp(-t)$ is obtained as a limiting case as $\alpha \to \infty$.

PROC LIFEREG is dedicated to the analysis of the AFT model from left, right or interval censored data. For SP models, PROC SEVERITY offers many options for defining survival distributions by calling subroutines written in PROC FCMP. Data on survival times may have combinations that are left, right or interval censored and left or right truncated.

Time-varying covariates

With time-varying covariates (TVC) $\mathbf{x}(t)$, the foregoing needs modification. The hazard function is $h(t \mid \mathbf{x}(t)) = \lim_{\Delta t \downarrow 0} P[t \le T < t + \Delta t \mid T \ge t, \mathbf{x}(t)] / \Delta t$ with the heuristic interpretation that $h(t \mid \mathbf{x}(t))\Delta t$ is approximately the probability of the event occurring in $[t, t + \Delta t)$, conditional on the covariate history up to t and being at risk of the event. Studies where the TVC are piecewise constant can be handled by construction of the likelihood for $(T \wedge U, \delta)$ given $\mathbf{x} = \{\mathbf{x}(t_m), m = 0, \ldots, M-1\}$ where $0 = t_0 < t_1 < \ldots < t_{M-1} < T \wedge U \equiv t_M$ are the observation times of the covariates, and $\delta = [T \le U]$. The maintained assumptions are: conditional independence of (T, U), given $\mathbf{x}$ and strict exogeneity on the conditional distribution, $D(T \mid T \ge t_{m-1}, \mathbf{x}) = D(T \mid T \ge t_{m-1}, \mathbf{x}(t_{m-1})), m = 1, \ldots, M$ (Wooldridge, 2010). With $\mathbf{x}(t)$ constant $(= \mathbf{x}(t_{m-1}))$ on $[t_{m-1}, t_m)$, the contribution to the likelihood by a censored observation U (at t_M) is $\prod_{m=1}^{M} P[T > t_m \mid T \ge t_{m-1}, \mathbf{x}(t_{m-1})] = \exp\left(-\sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} h(u \mid \mathbf{x}(t_{m-1}))du\right)$. An observed event time T (at t_M) contributes $h(t_M \mid \mathbf{x}(t_{M-1}) \times P[T > t_M \mid T \ge t_{M-1}, \mathbf{x}(t_{M-1})]$. The log-likelihood for a single datum can be written compactly as $\delta \log\left(h(t_M \mid \mathbf{x}(t_{M-1}))\right) - \sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} h(u \mid \mathbf{x}(t_{m-1}))du$.

From a random sample of observations $\{(T_i \wedge U_i, \delta_i, \mathbf{x}_i(t_{m_i})) : 0 \le m_i \le M_i - 1, 1 \le i \le n\}$ a data set can be constructed with multiple records for each subject that stacks vertically the records at the measurement times. The number of records may vary by subject, and it is not necessary that all TVCs change at the measurement times in the subject. The last record has information on whether or not t_{M_i} is a failure or a censoring time. The log-likelihood for the sample is

[1] $$\sum_{i=1}^{n} \left(\delta_i \log\left(h(t_{M_i} \mid \mathbf{x}_i(t_{M_i - 1}))\right) - \sum_{m_i=1}^{M_i} \int_{t_{m_i - 1}}^{t_{m_i}} h(u \mid \mathbf{x}_i(t_{m_i - 1}))du \right).$$

Section 2 focuses on implementing [1] with some examples on specifying the hazard function. The choices are a fully specified parametric distribution (e.g., Weibull, lognormal, log-logistic), or a PH model with the baseline hazard expressed as a piecewise constant function.

Joint Models

In joint modelling, a separate process models the TVC data $\{\mathbf{x}_i(t_{m_i}) : 0 \le m_i \le M_i - 1\}$. We focus on an univariate continuous covariate with observations for the *i*-th subject denoted by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iM_i})'$ at the measurement at times $\{t_{ij} : 1 \le j \le M_i\}$. A linear mixed model is posited for $E(\mathbf{Y}_i \mid \mathbf{x}_i, \mathbf{b}_i)$ with fixed covariates $\mathbf{x}_i$ and random effects $\mathbf{b}_i$. It permits modeling the time dynamics of the underlying process $Y_i(t)$ where $Y(t_{ij}) = Y_{ij}$, usually with low order polynomials of *t*. PROC GLIMMIX is ideal for this purpose. Next, some features of the conditional mean $\mu_i(t, \mathbf{b}_i) = E\left(Y_i(t) \mid \mathbf{x}_i, \mathbf{b}_i\right)$ are transferred into the conditional hazard function of T_i, $h(t \mid \mathbf{x}_i, \mathbf{b}_i) = h_0(t) \exp\left(\mathbf{x}_i'\beta + f(\mu_i(t, \mathbf{b}_i), \lambda)\right)$ where f is a scalar function. The parameters of the joint model are (β, λ) and parameters in the specification of the baseline function $h_0(t)$, e.g., piecewise constant or Weibull form $h_0(t) = \gamma t^{\gamma - 1}$. The log-likelihood [1] is replaced by a joint log-likelihood for $(\log T_i, \mathbf{Y}_i)$ conditional on $(\mathbf{x}_i, \mathbf{b}_i)$, and the marginal distribution obtained by integrating with respect to random effects distribution. Section 4 addresses a series of joint models. PROC NLMIXED is the computational engine.

2. Application

The data set used for illustration comes from a study of patients who had primary biliary cirrhosis (now called primary biliary cholangitis), an autoimmune disease of the liver (Murtaugh *et al*, 1994). Patients (n=312) were randomized to D-penicillamine or placebo. Over the follow-up period of nearly 13 years there were 140 deaths. A small number of patients (n=29) received a transplant. The event of interest is death or transplant. Time to event is in years from date of randomization. Age at baseline, gender and treatment group are fixed covariates. There are several biomarkers with follow-up measurements. Assessments of serum bilirubin (in mg/ml) vary from 1 to 16 with about 59% of patients having at least 5 records. L_BILI is serum bilirubin log transformed. An indicator for liver enlargement, hepatomegaly (HEPATOM) is also time-dependent. We consider only the baseline value HEPATOM0.

The data file PBCSEQ2 has multiple records for patients (ID) with time variables t0, t1 that identify the interval [t0, t1) in which L_BILI is assumed constant. START and STOP in days are converted to years in t0, t1. EVENT labels patient status at t1, with LAST signaling the last record.

Data for ID=3

Obs	id	start	stop	t0	t1	event	last	drug	age	sex	hepatom	bili
12	3	0	176	0.00	0.48	0	.	1	70.07	0	0	1.40
13	3	176	364	0.48	1.00	0	.	1	70.07	0	1	1.10
14	3	364	743	1.00	2.03	0	.	1	70.07	0	0	1.50
15	3	743	1012	2.03	2.77	2	1	1	70.07	0	0	1.80

Although formats could be gainfully applied, to save ourselves some consternation indicators are included in PBCSEQ2 as well has time variables RCTIME and LTTIME.

```
data pbcseq2;
set g.pbcseq2;
by id;
retain hepatom0;
if first.id then hepatom0=hepatom;
if t0>0 then lttime=t0;
if event=0 then rctime=t1;
else rctime=.;
drug_PENCL=(drug=1);
sex_female=(sex=1);

proc format;
value affirm 0='no' 1='yes';
value  sex   0='male' 1='female' ;
value status 0='alive' 1='transplant' 2='dead';
value  drug  1='D-pencl' 0='Placebo';
run;
```

Parametric models

Consider fitting parametric distributions using the log-likelihood [1]. The key statements in PROC SEVERITY are DIST, LOSS and SCALEMODEL. The analysis variable is t1. The right-censored and left-truncated options are used to exploit the formation of [1]. Covariate values for each input record are supplied through

[2] $\log\theta(\mathbf{x}) = \beta_0 + \beta_1 drug\_pencl + \beta_2 sex\_female + \beta_3 age + \beta_4 hepatom0 + \beta_5 L\_BILI$

```
proc severity data=pbcseq2 vardef=n outest=est covout;
dist burr logn llogistic weibull;
loss t1/rightcensored=rctime lefttruncated=lttime;
scalemodel drug_pencl sex_female age hepatom0 L_BILI;
nloptions gconv=0 tech=quanew singular=1.0e-9 covfuzz=1.0e-12;
run;
```

Table 1: Estimates from Weibull, Lognormal and Log-logistic scale parameter models

	Weibull			Lognormal			Log-logistic		
Parameter	**Estimate**	**StdErr**	**p-value**	**Estimate**	**StdErr**	**p-value**	**Estimate**	**StdErr**	**p-value**
drug_PENCL	0.0454	0.1178	0.7000	–0.0462	0.1474	0.7542	–0.0359	0.1432	0.8022
sex_female	–0.0719	0.1631	0.6593	0.0533	0.2111	0.8005	–0.0870	0.2175	0.6893
age	–0.0316	0.0059	<.0001	–0.0335	0.0069	<.0001	–0.0331	0.0071	<.0001
hepatom0	–0.4041	0.1289	0.0017	–0.2804	0.1546	0.0697	–0.3453	0.1498	0.0212
L_BILI	–1.0227	0.0863	<.0001	–1.0279	0.0809	<.0001	–0.9063	0.0773	<.0001
Intercept	5.5960	.	.	4.9213			5.0126		
Gamma	1.3321	0.0867	<.0001	1.0277	0.0582	<.0001	1.9425	0.1294	<.0001
–2 LOGL	786.12			852.20			858.64		
–2 $LOGL_0$*	1110.34			1120.91			1111.01		

* –2 LOGL for model omitting L_BILI.

We make a few remarks at this juncture.

(1) The fitted Burr distribution (not shown) gave almost identical estimates to that of the Weibull. The shape α was very large that made the Weibull a good approximation to the Burr: as $\alpha\to\infty$, $S(t)=\left(1+\alpha^{-1}(t/\theta)^{\gamma}\right)^{-\alpha}\to\exp\left(-(t/\theta)^{\gamma}\right)$. SEVERITY has its lexicon for shape and intercept parameters. Here, we use β_0 for intercept and γ for shape. The lognormal survival distribution is $S(t)=\Phi\left(-\log\left(t/\theta\right)^{\gamma}\right)$.

(2) Although the log-logistic distribution is not predefined in SEVERITY, its distribution and density functions can be programmed in PROC FCMP, and then called within SEVERITY (Gardiner, 2014).

(3) The covariance matrix **G** of the model parameters is obtained from the OUTEST=EST and COVOUT options. VARDEF=N prevents the multiplier $N/(N-q)$ being applied to **G** where q is the number of model parameters. In testing the significance of effects, SEVERITY uses the t-distribution with $N-q$ degrees of freedom. Our input file has N=1945 records for n=312 subjects.

(4) The Weibull hazard $h(t\,|\,\mathbf{x})=\gamma t^{\gamma-1}\exp(-\gamma\mathbf{x}'\boldsymbol{\beta})$ with TVC, $x_5\equiv L\_BILI$ provides an interpretation for β_5 in [2] from the hazard ratio (HR) for one unit increase in L_BILI at time t. Keeping all other covariates "fixed", $HR\equiv h(t\,|\,x_5+1)/h(t\,|\,x_5)=\exp(-\gamma\beta_5)$. With estimates, $\hat{\gamma}$ = 1.3321, $\hat{\beta}_5$ = –1.0227 we obtain $\widehat{HR}=3.91$. A 95% confidence interval (CI) for HR is constructed using the approximation of the variance $Var(\hat{\gamma}\hat{\beta}_5)=\hat{\beta}_5^2Var(\hat{\gamma})+\hat{\gamma}^2Var(\hat{\beta}_5)+2\hat{\gamma}\hat{\beta}_5Cov(\hat{\gamma},\hat{\beta}_5)$. Assuming asymptotic normality of the MLE, a routine calculation gives the 95% CI, (3.32, 4.60).

Piecewise constant hazard function

Let $h(t\,|\,\mathbf{x})=h_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})$ with a flexible form for the baseline hazard, $h_0(t)=\sum_{j=1}^{J}\lambda_j[a_{j-1}\le t<a_j]$ where $0=a_0<a_1<\ldots<a_J=\infty$ partitions the time axis into J intervals. The constants $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_J)$ and regression coefficients $\boldsymbol{\beta}$ constitute the parameters of the model. Parametrization in terms of log-hazards $\alpha_j=\log\lambda_j$ is preferred. The corresponding survival distribution is piecewise exponential. PHREG provides a Bayes analysis of the posterior distributions of ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$). Our objective here is the obtain the MLE of ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$) from the log-likelihood [1]. Fortunately, PHREG produces the MLE as starting values for the full Bayes analysis.

By default J=8 intervals partition the time axis, with cut-points to get approximately the same number of events in each interval—169/8 ≈ 21. The partition does not depend on covariates. The cut-points are saved in the PARTITION output (Table 2). The BAYES statement asks for only a single Monte-Carlo iteration (NMC=1) and the default PIECEWISE option is sufficient. The default burn-in NBI=2000, default priors for parameters are maintained, but all plots are switched off. The piecewise exponential model produces the MLE in Table 3.

```
ods output partition=partition;
proc phreg data=pbcseq2;
class drug(ref='Placebo') sex(ref='male')
      hepatom0(ref='no')/param=ref;
format drug drug. hepatom0 affirm. sex sex.;
model (t0,t1)*event(0)=Drug_pencl sex_female age L_BILI;
bayes nmc=1 plots=none piecewise=loghazard;
run;
```

Table 2: Cut-points for time axis in the piecewise exponential model

cut1	cut2	cut3	cut4	cut5	cut6	cut7	cut8
0	1.008926	2.221827	2.951484	3.957672	5.098018	6.492991	8.462928

Table 3: Estimates from the Piecewise Exponential model (PROC PHREG)

Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**95% Confidence Limits**	
Alpha1	–7.1506	0.6173	–8.3605	–5.9406
Alpha2	–7.2728	0.6079	–8.4642	–6.0814
Alpha3	–6.6224	0.6069	–7.8119	–5.4329
Alpha4	–6.6326	0.5841	–7.7774	–5.4877
Alpha5	–6.5534	0.5841	–7.6983	–5.4085
Alpha6	–6.6137	0.5945	–7.7790	–5.4484
Alpha7	–6.2153	0.5693	–7.3310	–5.0996
Alpha8	–6.2700	0.5656	–7.3786	–5.1613
Drug (D_pencl)	–0.0618	0.1575	–0.3706	0.2469
Sex (female)	0.0815	0.2176	–0.3450	0.5080
age	0.0424	0.00778	0.0272	0.0577
hepatom0 (yes)	0.5589	0.1770	0.2120	0.9059
L_BILI	1.3696	0.0841	1.2047	1.5345

The adjusted HR for one-unit increase in L_BILI is 3.93, (95% CI, 3.34, 4.64). In joint modeling of the failure time and L_BILI, we will return to the piecewise exponential model for initial parameter values. The models described thus far all concern the conditional distribution $D(T \wedge U, \delta \,|\, \mathbf{x})$ with fixed binary covariates DRUG, SEX, HEPATOM0, and one continuous covariate AGE. Serum bilirubin, logged (L_BILI) is a continuous TVC. From the log-likelihood construction [1] we estimated the Weibull, lognormal, log-logistic and the piecewise exponential models (Table 1 and Table 3). Overall, all models show a significant effect (on survival) of L_BILI, HEPATOM0, and AGE, but non-significance of DRUG and SEX, at least at the 10% level. Of course L_BILI is the most convincing of these effects. In proportional hazards models--- Weibull, piecewise exponential, the effect of L_BILI can be summarized by the hazard ratio. It is assuring that the results are practically the same. However, it must be emphasized that these results for L_BILI are not results of joint modelling. They can be biased and often are appreciably so (Ye and Yu, 2014). We now turn to true joint models.

3. Joint Models: Preliminary sub-models for failure time and marker

The failure time T is assumed to have a piecewise exponential distribution. We begin with the construction of the log-likelihood to replace [1]. Each interval $[t_{m-1},t_m)$ is split by the cut-points $\{a_j : 0 \le j \le J\}$. For example, if $t_1 \in [a_{j-1},a_j)$ the first interval $[0,t_1)$ is split by $\{0,a_1,a_2 \ldots,a_{j-1},t_1\}$. Multiple records with variable names TSTART, TSTOP are created for the j intervals ending with the interval $[a_{j-1},t_1)$. The next interval $[t_1,t_2)$ is split by $\{t_1,a_j,\ldots,a_{k-1},t_2\}$, if $t_2 \in [a_{k-1},a_k), k > j$ or, not split if $t_2 \in [a_{j-1},a_j)$. To generalize, define

$\Delta_{jm} = (t_m - \max(t_{m-1},a_{j-1}))[a_{j-1} \le t_m < a_j] + (a_j - \max(t_{m-1},a_{j-1}))[t_m \ge a_j]$, provided $\max(t_{m-1},a_{j-1}) \le a_j,t_m$.

Next, replace in [1] for a single subject, $\int_{t_{m-1}}^{t_m} h(u \mid \mathbf{x}(t_{m-1}))du = \sum_{j=1}^{J} \exp(\alpha_j + \mathbf{x}'(t_{m-1})\beta)\Delta_{jm}$. At the last record, $[t_{M-1},t_M)$ we have $\log\left(h(t_M \mid \mathbf{x}(t_{M-1})\right) = \sum_{j=1}^{J} \alpha_j \Delta_{jM} + \mathbf{x}'(t_{M-1})\beta$. Once again we have applied the covariate value at the beginning of the interval $[t_{m-1},t_m)$. Therefore the contribution to the log-likelihood of a single subject with multiple records on $[t_{m-1},t_m), m = 1,\ldots,M$ is

[3] $\delta\left[\sum_{j=1}^{J} \alpha_j \Delta_{jM} + \mathbf{x}'(t_{M-1})\beta\right] - \sum_{m=1}^{M}\sum_{j=1}^{J} \exp(\alpha_j + \mathbf{x}'(t_{m-1})\beta)\Delta_{jm}$.

The piecewise exponential model was estimated using PHREG by calling the BAYES option (Tables 2 and 3). Instead, we can use LIFEREG with an expanded data to reflect the log-likelihood [3]. The duration variables Δ_{jm} (RISKTIME) are exponentially distributed. We also add indicators alpha1,…,alpha8 for the intervals of constant hazard.

Expanded records for ID=3

The original records are Obs=20, 21, 22, 24. Two additional records are created by cut-points 1.0089 and 2.2218. Last observation is a failure time. alpha4 to alpha8 are zero.

Obs	t0	t1	tstart	tstop	risktime	alpha1	alpha2	alpha3	L_BILI	event
20	0.0000	0.4819	0.0000	0.4819	0.4819	1	0	0	0.3365	0
21	0.4819	0.9966	0.4819	0.9966	0.5147	1	0	0	0.0953	0
22	0.9966	2.0343	0.9966	1.0089	0.0123	1	0	0	0.4055	0
23	0.9966	2.0343	1.0089	2.0343	1.0254	0	1	0	0.4055	0
24	2.0343	2.7708	2.0343	2.2218	0.1875	0	1	0	0.5878	0
25	2.0343	2.7708	2.2218	2.7708	0.5490	0	0	1	0.5878	2

```
ods output parameterestimates=parms_pe;
proc lifereg data=TVC_PE2;
model risktime*event(0)=alpha1-alpha7 drug_pencl sex_female age
hepatom0 L_BILI/dist=exponential;
run;
```

Table 4: Estimates from the Piecewise Exponential model (from PROC LIFEREG)

Maximum Likelihood Parameter Estimates						
Parameter	**Estimate**	**Standard Error**	**95% Confidence Limits**		**Chi–Square**	**p-value**
Intercept	6.2701	0.5656	5.1615	7.3788	122.87	<.0001
alpha1	0.8806	0.3160	0.2613	1.4999	7.77	0.0053
alpha2	1.0028	0.3172	0.3810	1.6246	9.99	0.0016
alpha3	0.3524	0.3153	–0.2656	0.9705	1.25	0.2637
alpha4	0.3626	0.3120	–0.2489	0.9741	1.35	0.2451
alpha5	0.2835	0.3114	–0.3268	0.8938	0.83	0.3627
alpha6	0.3437	0.3112	–0.2663	0.9537	1.22	0.2695
alpha7	–0.0547	0.3096	–0.6614	0.5521	0.03	0.8598
drug_PENCL	0.0618	0.1575	–0.2469	0.3706	0.15	0.6946
sex_female	–0.0815	0.2176	–0.5080	0.3449	0.14	0.7079
age	–0.0424	0.0078	–0.0577	–0.0272	29.73	<.0001
hepatom0	–0.5590	0.1770	–0.9060	–0.2120	9.97	0.0016
L_BILI	–1.3696	0.0841	–1.5346	–1.2047	264.93	<.0001

–2 LOGL=783.64

Table 3 and Table 4 are for the same model. LIFEREG parameterizes the piecewise exponential hazard as $h(t \mid \mathbf{x}) = \exp(-(\alpha_j + \mathbf{x}'\beta))$, $t \in [a_{j-1}, a_j)$ with intercept. The indicators alpha1 to alpha7 are contrasts with the intercept alpha8. We will use the parameter estimates saved in **`parms_pe`** to inform starting values in the joint model.

Linear mixed model for the marker L_BILI

Let $Y(t)$ denote the continuous serum bilirubin (logged) at time *t*. Observations for the *i*-th subject are $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iM_i})'$ at assessment times $\{t_{ij} : 1 \le j \le M_i\}$. Henceforth $Y(t_{ij}) = Y_{ij}$.

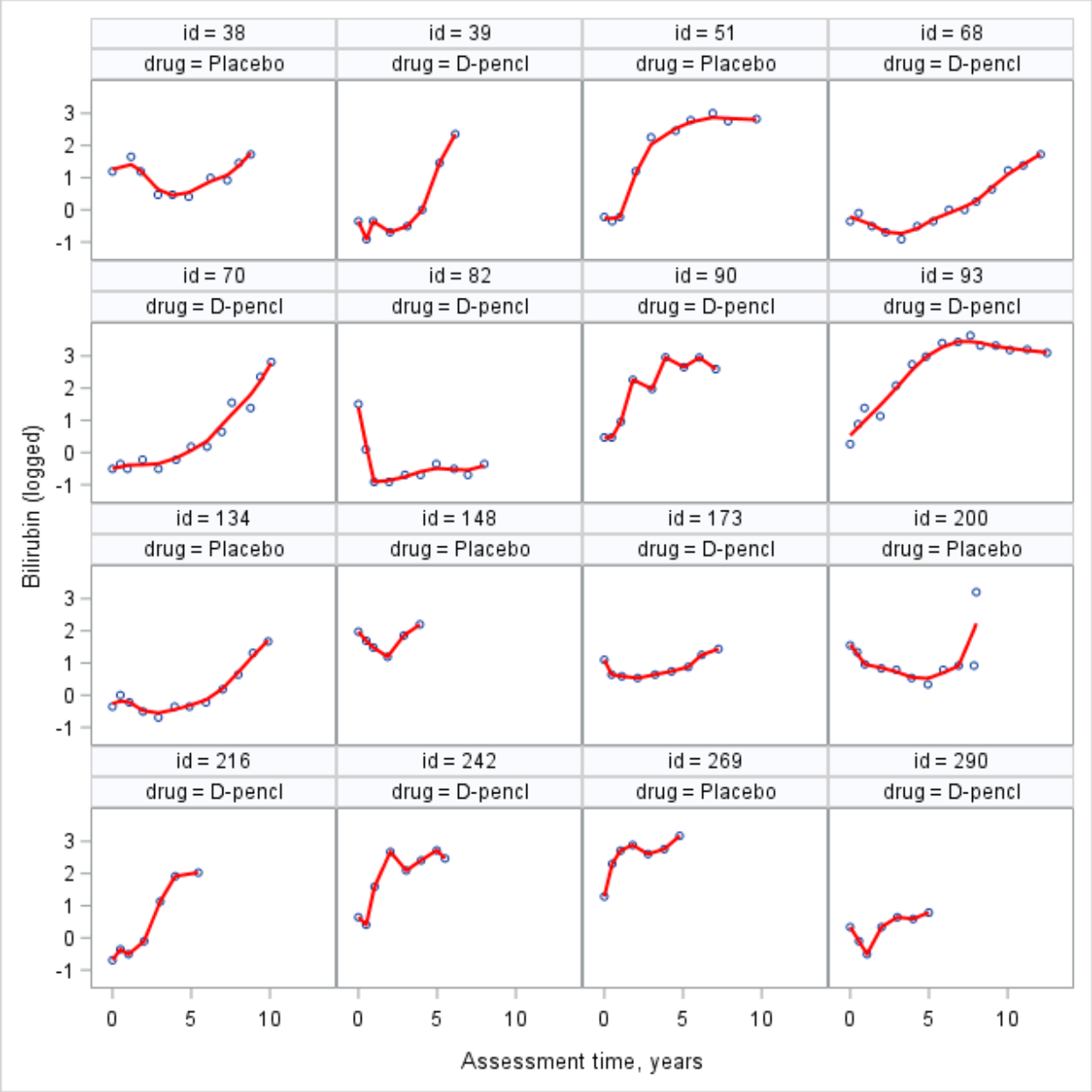
Plotting the longitudinal profiles for a sample of subjects might suggest how to structure the time component. Figure 1 is similar to figure 1.1 in Rizopoulos (2012). (Loess smoother could be different). We might consider linear and quadratic terms for t_{ij}. Other empirical data sets may support more complex structures such as splines (Crowther *et al*, 2012). We consider the model:

$$[4] \quad \begin{aligned} E(Y_{ij} \mid x_{i1}, \mathbf{b}_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 t_{ij} + \beta_3 x_{i1} t_{ij} + \beta_4 t_{ij}^2 + \beta_5 x_{i1} t_{ij}^2 + b_{i0} + b_{i1} t_{ij} + b_{i2} t_{ij}^2 \\ Var(Y_{ij} \mid x_{i1}, \mathbf{b}_i) &= \sigma^2 \end{aligned}$$

where $x_{i1} \equiv$ drug_pencl and $\mathbf{b}_i = (b_{i0}, b_{i1}, b_{i2})' \sim NORMAL(\mathbf{0}, \mathbf{G})$.

From [4] in the placebo group, $E(Y_{ij} \mid x_{i1} = 0) = \beta_0 + \beta_2 t_{ij} + \beta_4 t_{ij}^2$, and in the D-pencillamine treated group $E(Y_{ij} \mid x_{i1} = 1) = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) t_{ij} + (\beta_4 + \beta_5) t_{ij}^2$.

Figure 1: Profiles of serum bilirubin (logged) in selected subjects



The figure is generated by

```
ods graphics /height=6 in width=6 in;
proc sgpanel data=pbcseq2 noautolegend;
where id in (38 39 51 68 70 82 90 93 134 148 173 200 216 242 269 290);
panelby id DRUG/columns=4 rows=4;
loess x=t0 y=L_BILI/ smooth=.4 lineattrs=(color=red);
colaxis label='Assessment time, years';
rowaxis label='Bilirubin (logged)';
format drug drug.;
run;
```

Assuming $\{Y_{ij} : 1 \le j \le n_i\}$ are independent and normally distributed given $\mathbf{b}_i$, the mixed model [4] is estimated by maximum likelihood. The unstructured covariance matrix **G** is parameterized by the Cholesky decomposition, **type=Chol.** If $\mathbf{G} = \mathbf{CC}'$ where **C** is lower triangular, then $\mathbf{C} = [s_{11}, s_{21}\ s_{22}, s_{31}\ s_{32}\ s_{33}]$.

```
ods output parameterestimates=parms_gl covparms=covparms;
proc glimmix data=pbcseq2 noclprint gradient method=quad(qpoints=1);
class id;
model L_BILI=drug_pencl|t0|t0 /s dist=normal;
random intercept t0|t0 /subject=id type=Chol;
nloptions gconv=0;
run;
```

Estimated parameter estimates in Table 5 appear to support most profiles in Figure 1. We could use COVTEST statements to see if a simpler structure for **G** is warranted.

Table 5. Estimates of fixed effects in the Linear Mixed model

Effect	Estimate	Standard Error	DF	t Value	p-value	Gradient
Intercept	0.5854	0.08245	310	7.10	<.0001	–0.00001
drug_PENCL	–0.1434	0.1159	1087	–1.24	0.2160	–9.55E–6
t0	0.1626	0.03171	284	5.13	<.0001	0.000035
drug_PENCL*t0	0.007118	0.04370	1087	0.16	0.8706	0.000027
t0*t0	0.001444	0.003497	258	0.41	0.6799	–0.00027
drug_PENCL*t0*t0	–0.00231	0.004445	1087	–0.52	0.6026	0.000031

–2 LOGL=2864.84

Our mixed model is a way station to building the joint model. The syntax is more than that needed. The two output data sets **parms_gl** and **covparms** save the estimates that will be used as initial values for parameters in the joint model.

Construction of the joint likelihood

The expanded data set used to fit the piecewise exponential model (Table 4) is appended with relabeled variables t0_L, t1_L, Y together with the previous t0, t1, L_BILI. All other fixed covariates DRUG_PENCL, SEX_FEMALE, AGE, HEPATOM0 are included. The original data set PBCSEQ2 is a multiple record file of 1945 records for 312 patients. It is expanded in JOINT_1 to 3412 records to accommodate the log-likelihood [3] of the piecewise exponential model. The mixed model [4] uses 1945 records that are extracted by IND=1.

Expanded data records for ID=3 in data set JOINT_1

Obs	t0	t1	tstart	tstop	risktime	L_BILI	alpha1	alpha2	alpha3	t0_L	t1_L	Y	IND
20	0.0000	0.4819	0.0000	0.4819	0.4819	0.3365	1	0	0	0.0000	0.4819	0.3365	1
21	0.4819	0.9966	0.4819	0.9966	0.5147	0.0953	1	0	0	0.4819	0.9966	0.0953	1
22	0.9966	2.0343	0.9966	1.0089	0.0123	0.4055	1	0	0	0.9966	2.0343	0.4055	1
23	0.9966	2.0343	1.0089	2.0343	1.0254	0.4055	0	1	0	.	.	.	0
24	2.0343	2.7708	2.0343	2.2218	0.1875	0.5878	0	1	0	2.0343	2.7708	0.5878	1
25	2.0343	2.7708	2.2218	2.7708	0.5490	0.5878	0	0	1	.	.	.	0

For starting values of parameters assemble a data set **PARMS_ALL** from **parms_pe, parms_gl** and **covparms**. See Tables 4 and 5. The covariance parameters are the residual variance σ^2 and from the lower triangular matrix **C**. Not all of them are needed or appropriate in joint models. For example, a12 below will not be used.

Parameters in the piecewise exponential model

Obs	pname	Estimate	parameter
1	Intercept	6.2701	a0
2	alpha1	0.8806	a1
3	alpha2	1.0028	a2
4	alpha3	0.3524	a3
5	alpha4	0.3626	a4
6	alpha5	0.2835	a5
7	alpha6	0.3437	a6
8	alpha7	–0.0547	a7
9	drug_PENCL	0.0618	a8
10	sex_female	–0.0815	a9
11	age	–0.0424	a10
12	hepatom0	–0.5590	a11
13	L_BILI	–1.3696	a12

Parameters in the linear mixed model

Obs	pname	Estimate	parameter
14	Intercept	0.5854	b0
15	drug_PENCL	–0.1434	b1
16	t0	0.1626	b2
17	t0*t0	0.0014	b3
18	drug_PENCL*t0	0.0071	b4
19	drug_PENCL*t0*t0	–0.0023	b5
20	CHOL(1,1)	0.9961	s11
21	CHOL(2,1)	0.0519	s21
22	CHOL(2,2)	0.2977	s22
23	CHOL(3,1)	0.0003	s31
24	CHOL(3,2)	–0.0224	s32
25	CHOL(3,3)	0.0109	s33
26	Residual	0.0931	sigsq

4. Joint models for failure time and longitudinal marker

We now proceed to estimating a series of joint models with NLMIXED. Broadly, they involve passing some aspect of the conditional mean function for the marker L_BILI

[5.0] $$\mu_i(t) = (\beta_1 + \beta_4 t + \beta_5 t^2)x_{i1} + (\beta_0 + b_{i0}) + (\beta_2 + b_{i1})t + (\beta_3 + b_{i2})t^2$$

with random effects $\mathbf{b}_i = (b_{i0}, b_{i1}, b_{i2})$ into the individual hazard function. For example, Rizopoulos (2012) considers

[5.1] $$h_i(t \mid x_{i1}, x_{i2}, \mathbf{b}_i) = \exp\left(-\{\alpha_j + c_0 + c_1 x_{i1} + c_2 x_{i2} + (\lambda_1 + \lambda_2 x_{i2})\mu_i(t)\}\right), t \in [a_{j-1}, a_j)$$

where x_{i1}=DRUG_PENCL and x_{i2} = HEPATOM0. This is our **Model 1.** The intercept c_0 and $\{\alpha_j, 1 \le j \le 7\}$, parameterize the piecewise constant baseline hazard as in Table 4. Parameters λ_1, λ_2 are called association parameters. The log-likelihood [3] is no longer tenable because from [1] the cumulative hazard terms are

[5.2] $\int_{t_{m_i-1}}^{t_{m_i}} h_i(t \mid x_{i1}, x_{i2}, \mathbf{b}_i))dt = \exp\left(-(\alpha_j + c_0 + c_1 x_{i1} + c_2 x_{i2})\right) \int_{t_{m_i-1}}^{t_{m_i}} \exp\left(-(\lambda_1 + \lambda_2 x_{i2})\mu_i(t)\right)dt$,

with some caveats. Our multiple record data file JOINT_1 may further split $[t_{m_i-1}, t_{m_i})$ by the cut-points a_j. For example, for ID=3 we see that 2 additional records (IND=0) are added to the original 4 records (IND=1), making the integral over 6 sub-intervals [TSTART, TSTOP). In general, evaluation of the integral in [5.2] would need numerical methods. Gauss-Kronrod quadrature is the most cited (Garcia-Hernandez and Rizopoulos, 2018, Crowther *et al*, 2012).

We will proceed here with a simple expedient knowing well that it would introduce some bias. Heckman and Singer (1986) caution that the bias could be appreciable in some empirical applications. The expedient evaluates the integral in [5.2] at the beginning (TSTART) of the sub-intervals, and therefore approximates the integral as a crude Riemann-sum over sub-intervals. But this will vary by individual. The JMFit SAS macro (Zhang *et al*, 2016) that fits several shared parameter models employs 200 sub-intervals. The %JM SAS macro (Garcia-Hernandez and Rizopoulos, 2018) is the most versatile and comprehensive. It has options for fitting models with other baseline hazards (e.g., Weibull, splines, fractional polynomials (Royston and Parmar (2002)) and generalized linear mixed models for the marker process. It calls Gauss-Kronrod quadrature for numerical integration.

NLMIXED assembles the log-likelihood from $D\left(Y_i(t) \mid \mathbf{b}_i\right) \equiv NORMAL(\mu_i(t), \sigma^2)$ and $D(T_i \wedge U_i, \delta_i \mid \mathbf{b}_i)$ specified via [5.1] and [5.2]. The integration with respect to random effects is by the default Gauss-Hermite quadrature, **`method=gauss.`** Optimization of the log-likelihood with respect to model parameters is governed by **`tech=newrap.`** Starting values of parameters are extracted from **`PARMS_ALL(where=(parameter not in ("a9" "a10" "a12" )));`** An educated guess for (**`lam1, lam2`**) completes the initial values specification.

Fortitude and patience, together with aid from an arsenal of options in NLMIXED often lead to satisfactory convergence and plausible results. Applying the parameter estimates from a previous run as starting values for the next invocation, and tweaking the quadrature points (**`qpoints=`**) and gradient convergence (**`gconv=`**) options support confidence in the final results. Also, look for convergence with small gradients. We have asked for empirical standard errors (**`empirical`**) probably adding slightly to the overall computational effort. The execution time for the entire program was 1 hr, 27 mins, on a desktop 4 core Xeon processor. Table 6 shows the results.

From [5.1] for an unit increase in the marker at time *t*, the log hazard-ratio is $(-\lambda_1)$ in subjects without hepatomegaly $(x_{i2} = 0)$ and $(-\lambda_1 - \lambda_2)$ in subjects with hepatomegaly $(x_{i2} = 1)$. Estimates and standard errors are obtained from two **`estimate`** statements, and then transformed to hazard ratios in a simple data step. Confidence limits are based on the t-distribution (DF=309).

The results from NLMIXED are comparable to those shown in Rizopoulos (2012) via their R-program JM. For HEP=0, HR=3.2 (95% CI: 2.5, 4.1), and for HEP=1, HR=4.0 (95% CI: 3.1, 5.2).

	Hazard Ratio (HR) and 95% CL		
Label	**HR**	**LCL**	**UCL**
HAZARD RATIO marker @ HEP=0	3.21	2.30	4.47
HAZARD RATIO marker @ HEP=1	4.23	3.22	5.58

LCL=95% lower CL; UCL=95% upper CL. HEP=hepatomegaly

Table 6: Estimates from the Joint model (Model 1)

Parameter Estimates							
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**p-value**	**95% Confidence Limits**	
lam1	–1.1654	0.1691	309	–6.89	<.0001	–1.4981	–0.8327
lam2	–0.2781	0.2157	309	–1.29	0.1983	–0.7025	0.1463
a0	4.1419	0.3710	309	11.16	<.0001	3.4118	4.8720
a1	0.4182	0.3547	309	1.18	0.2393	–0.2797	1.1161
a2	0.6755	0.3565	309	1.89	0.0591	–0.02600	1.3770
a3	0.1825	0.3566	309	0.51	0.6092	–0.5192	0.8842
a4	0.2523	0.3465	309	0.73	0.4670	–0.4294	0.9341
a5	0.1814	0.3476	309	0.52	0.6021	–0.5025	0.8654
a6	0.1443	0.3474	309	0.42	0.6782	–0.5393	0.8278
a7	–0.08040	0.3478	309	–0.23	0.8173	–0.7648	0.6040
a8 drug	–0.08110	0.1748	309	–0.46	0.6429	–0.4250	0.2628
a11 hepatom0	–0.1774	0.4221	309	–0.42	0.6746	–1.0081	0.6533
b0	0.5846	0.08943	309	6.54	<.0001	0.4086	0.7606
b1 drug	–0.1423	0.1161	309	–1.23	0.2210	–0.3707	0.08604
b2 time	0.1650	0.03459	309	4.77	<.0001	0.09691	0.2330
b3 time^2	0.003314	0.004060	309	0.82	0.4151	–0.00468	0.01130
b4 drug×time	0.001619	0.04391	309	0.04	0.9706	–0.08478	0.08802
b5 drug×time^2	–0.00204	0.004398	309	–0.46	0.6433	–0.01069	0.006616
s11	0.9958	0.03712	309	26.82	<.0001	0.9227	1.0688
s21	0.05478	0.02598	309	2.11	0.0358	0.003660	0.1059
s22	0.3034	0.02165	309	14.01	<.0001	0.2608	0.3460
s31	0.001345	0.003566	309	0.38	0.7062	–0.00567	0.008362
s32	–0.02206	0.002899	309	–7.61	<.0001	–0.02777	–0.01636
s33	0.01174	0.002144	309	5.48	<.0001	0.007525	0.01596
sigsq	0.09188	0.007029	309	13.07	<.0001	0.07805	0.1057

–2 LOG L=3709.0

Syntax for Model 1

```
ods output additionalestimates=estimates
           parameterestimates=parms_NL5;
proc nlmixed data=JOINT_1 method=gauss tech=newrap gconv=0
     qpoints=7 empirical;
parms/data=g.parms_NL4; /*parameter update from a previous call*/
dummy=1;
```

```
/*----SURVIVAL----*/
xb=b0+b1*DRUG_pencl+b2*tstart+b3*tstart*tstart+b4*drug_pencl*tstart
     +b5*drug_pencl*tstart*tstart
          +Z0+Z1*tstart+Z2*tstart*tstart;
xa=a0+a1*alpha1+a2*alpha2+a3*alpha3+a4*alpha4+a5*alpha5+a6*alpha6
     +a7*alpha7+a8*drug_PENCL+a11*hepatom0
     +lam1*xb+lam2*xb*hepatom0;
fail=(event ~=0);
LLIK_S= -fail*xa-risktime*exp(-xa);

/*----LONGITUDINAL MIXED MODEL---*/
PI=constant("PI");
xb_L=b0+b1*drug_pencl+b2*t0_L+b3*t0_L*t0_L+b4*drug_pencl*t0_L
       +b5*drug_pencl*t0_L*t0_L
       +Z0+Z1*t0_L+Z2*t0_L*t0_L;
resid=(Y-xb_L);

if IND=0 then LLIK_L=0;
else
 LLIK_L=-.5*log(2*PI)-.5*resid**2/sigsq-.5*log(SIGsq);

g11=s11**2;
g21=s21*s11; g22=s21**2+s22**2;
g31=s31*s11; g32=s31*s21+s32*s22; g33=s31**2+s32**2+s33**2;

estimate "HAZARD RATIO marker @ HEP=0"  lam1;
estimate "HAZARD RATIO marker @ HEP=1"  lam1+lam2;

model dummy~general(LLIK_S+LLIK_L);
random z0 z1 z2 ~normal([0,0,0], [g11, g21, g22, g31, g32, g33])
     subject=id;
run;
```

Model 2 (Shared parameter): The survival model is piecewise exponential with hazard function

$$h_i(t \mid x_{i1}, x_{i2}, \mathbf{b}_i) = \exp\left(-(\alpha_j + c_0 + c_1 x_{i1} + c_2 x_{i2} + \lambda_0 b_{i0} + \lambda_1 b_{i1})\right), t \in [a_{j-1}, a_j).$$

Notice that there are no additional time-functions in the linear predictor. Among the models fit by the JMFit SAS macro (Zhang *et al*, 2016), this is called SPM2L. To fit this model, modify the syntax in the linear predictor **xa** for the random effects by **lam0*z0+lam1*z1.** With the same requests made for Model 1, and beginning the computations at initial values suggested by Model 1, satisfactory convergence was achieved in 1hr, 11mins. Results are in Table 7 (columns 1-3).

Model 3 (Shared parameter with linear trajectory): The survival model is piecewise exponential with hazard function

$$h_i(t \mid x_{i1}, x_{i2}, \mathbf{b}_i) = \exp\left(-\{\alpha_j + c_0 + c_1 x_{i1} + c_2 x_{i2} + \lambda(b_{i0} + b_{i1} t)\}\right), t \in [a_{j-1}, a_j).$$

This model is called SPM1L in JMFit. The time function in the cumulative hazard can be integrated in closed form. Results in Table 7 (columns 4-6).

Model 4 (Shared parameter with linear trajectory, but intercept and slope in marker model): The survival model is the same as Model 3, but the longitudinal marker model is reduced to linear variation only—Table 7 (columns 7-9).

The association parameters in model 2 are significant pointing to a correlation between $Y_i(t)$ and $\log T_i$. In models 3 and 4 the hazard function (conditional on $\mathbf{b}_i$) has the form of the Gompertz distribution (Collett, 2015). Model 4 is a restriction of model 3 that involves fewer covariance parameters. The 3 DF test for comparing the two models is significant.

Table 7: Estimates from the Joint models (Models 2, 3, and 4)

	MODEL 2			MODEL 3			MODEL 4		
Parms	**Estimate**	**Stderr**	**p-value**	**Estimate**	**Stderr**	**p-value**	**Estimate**	**Stderr**	**p-value**
lam	.	.	.	–0.7012	0.1804	0.0001	–1.2294	0.09505	<.0001
lam1	–3.1127	0.4200	<.0001	.	.	.	.	.	.
lam0	–1.2355	0.1170	<.0001	.	.	.	.	.	.
a0	1.7102	0.3402	<.0001	2.1759	0.5872	0.0002	1.3530	0.4101	0.0011
a1	2.8669	0.4619	<.0001	1.4342	0.6236	0.0221	2.5615	0.4342	<.0001
a2	2.5378	0.4216	<.0001	1.5490	0.6249	0.0137	2.6428	0.4368	<.0001
a3	1.5314	0.3865	<.0001	0.8332	0.6194	0.1796	1.8670	0.4231	<.0001
a4	1.3599	0.3665	0.0002	0.9130	0.5952	0.1261	1.7958	0.4134	<.0001
a5	1.0146	0.3556	0.0046	0.7644	0.5754	0.1850	1.5140	0.3983	0.0002
a6	0.6900	0.3462	0.0471	0.5752	0.5523	0.2985	1.1832	0.3858	0.0024
a7	0.2355	0.3368	0.4849	0.1810	0.4744	0.7031	0.5650	0.3626	0.1202
a8 drug	0.1217	0.2542	0.6326	0.0650	0.2217	0.7695	0.03608	0.2572	0.8885
a11 hepatom0	–0.7284	0.1881	0.0001	–1.0375	0.2275	<.0001	–0.6747	0.1922	0.0005
b0	0.5800	0.08952	<.0001	0.5721	0.08934	<.0001	0.5439	0.08809	<.0001
b1 drug	–0.1440	0.1162	0.2161	–0.1332	0.1162	0.2526	–0.1118	0.1159	0.3356
b2 time	0.1916	0.03553	<.0001	0.1732	0.03601	<.0001	0.1843	0.01906	<.0001
b3 time²	–0.0009	0.00394	0.8285	–0.00044	0.003935	0.9102	.	.	.
b4 drug× time	0.00396	0.04508	0.9300	0.02130	0.05101	0.6765	0.00708	0.02428	0.7708
b5 drug× time²	–0.00213	0.00445	0.6332	–0.00349	0.005010	0.4860	.	.	.
s11	0.9964	0.03721	<.0001	0.9890	0.03718	<.0001	0.9973	0.03717	<.0001
s21	0.07698	0.02642	0.0038	0.08056	0.02468	0.0012	0.07849	0.01433	<.0001
s22	0.3102	0.02274	<.0001	0.2820	0.02730	<.0001	–0.1654	0.01519	<.0001
s31	–0.00132	0.00357	0.7127	–0.00337	0.003182	0.2906	.	.	.
s32	–0.02289	0.00313	<.0001	–0.01948	0.004220	<.0001	.	.	.
s33	0.01001	0.00156	<.0001	0.009400	0.001613	<.0001	.	.	.
sigsq	0.09323	0.00723	<.0001	0.09689	0.008246	<.0001	0.1203	0.00988	<.0001
–2 log L	3759			3830			3918		

6. Bayesian Analyses of Joint Models

Adopting a Bayesian framework, Guo and Carlin (2004) discuss a series of shared parameter models. The marker process has two random effects $\mathbf{b}_i = (b_{i0}, b_{i1})$, for intercept and slope. For the survival component their model XI specifies the hazard function, $h(t \mid \mathbf{x}_i, \mathbf{b}_i) = h_0(t)\exp(\mathbf{x}_i'\beta + \lambda_0 b_{i0} + \lambda_1 b_{i1})$ with Weibull baseline, $h_0(t) = \gamma t^{\gamma-1}$ and shape parameter γ>0. A more complex model replaces the random component with $\lambda_0 b_{i0} + \lambda_1 b_{i1} + \lambda_2(b_{i0} + b_{i1}t)$. The empirical application is a study of survival in AIDS patients who were treated by antiretroviral drugs, ddI vs ddC (Abrams *et al*, 1994). The longitudinal marker is the square-root of the CD4 cell count assessed at baseline and subsequently at 2, 6, 12 and 18 months. This data set has been extensively adopted for illustration. See Garcia-Hernandez and Rizopoulos (2018), Littell *et al*, (2006) Ye and Yu (2014) for different specifications of $h_0(t)$ and the random effects component. Ibrahim *et al*, (2001) has a discussion of the fundamentals of joint modeling in the Bayesian context.

As note previously, one challenging complexity in fitting joint models is the need to carry out an integration of the hazard function with respect to *t*. Crowther *et al*, (2012) and Royston and Parmer (2002) describe advantages of modeling the cumulative hazard $H(t \mid \mathbf{x}_i, \mathbf{b}_i)$ so that a proportional cumulative hazards on $\log H(t \mid \mathbf{x}_i, \mathbf{b}_i)$ can have flexible functions for $s = \log(t)$, including the natural cubic spline (NCS), basis splines and fractional polynomials. Gauss-Kronrod quadrature is the conduit for numerical integration wherever it is needed.

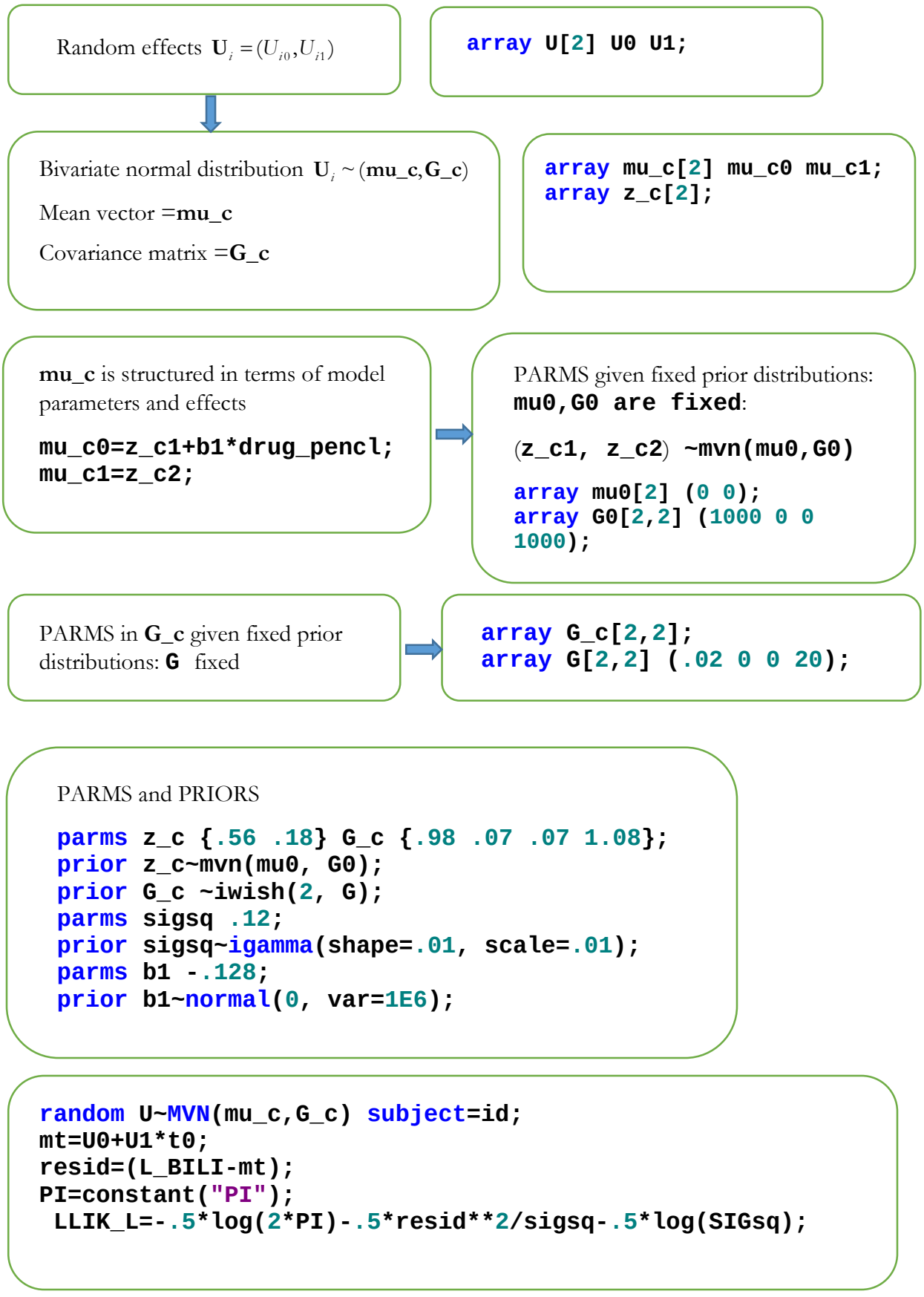
In the Bayes analysis, PHREG offers the piecewise constant hazard and the NCS for proportional hazards models (SAS/STAT Analytics 15.1). The joint model that we consider next, is far less ambitious in its scope. Consider model 4 of the previous section specifying marker process $\mu_i(t) = \beta_1 x_{i1} + (\beta_0 + b_{i0}) + (\beta_2 + b_{i1})t \equiv U_{i0} + U_{i1}t$ with hierarchically centered random effects (U_{i0}, U_{i1}) and x_{i1} is the binary indicator for drug_pencl. The hazard function has a Weibull baseline $h_i(t \mid x_{i1}, \mathbf{b}_i) = \gamma t^{\gamma-1}\exp\left(-\gamma(c_0 + c_2 x_{i1} + \lambda\mu_i(t)\right)$.

The PROC MCMC syntax for the marker component is entirely analogous to what is needed to fit a normal likelihood based linear mixed model. Figure 2 summarizes the features.

The log likelihood for the survival component is composed of $\delta_i \log h_i(T_i \mid x_{i1}, \mathbf{b}_i)$ and $-\int_{t_{m-1}}^{t_m} h_i(t \mid x_{i1}, \mathbf{b}_i)dt$ where $\log h_i(t \mid x_{i1}, \mathbf{b}_i) = \log(\gamma) + \gamma\left(\log t - (c_0 + c_2 x_{i1} + \lambda(U_{i0} + U_{i1})t\right) - \log t$. The integral contains $\int_{t_{m-1}}^{t_m} \gamma t^{\gamma-1}\exp(-\gamma\lambda U_{i1}t)dt$. Whether or not re-parameterization could lead to efficiencies in computations by MCMC is unclear. Direct programing of the log-likelihood is simple. The integration is carried out via the QUAD CALL in PROC MCMC. A subroutine in PROC FCMP must be invoked first, before initiating PROC MCMC.

```
proc fcmp outlib=sasuser.funcs.QUAD;
subroutine intfun(t, ft, c1, C);
outargs ft;
ft=c1*(t**(c1-1))*exp(-C*t);
endsub;run;
```

Figure 2: Longitudinal component for Bayes analysis



Survival component for Bayes analysis

The parameters are labeled $(c0, c1, c2)$ for the intercept, shape γ (Weibull) and regression coefficient of drug_pencl, respectively. The association parameter is λ --**lam.**

```
parms c0 2.3 c1 2.1 c2 .05;
prior c0 c2 ~Normal(0, var=1E6);
prior c1 ~gamma(shape=1E4, iscale=1E4);
parms lam -3;
prior lam ~Normal(0, var=1E4);

xa=c0+c2*drug_pencl+lam*U0;
fail=(event ~=0);
LP=log(t1)-xa;
C=c1*lam*U1;

LLIK_S1=log(c1)+c1*LP-log(t1)-C*t1; LLIK_S1=fail*LLIK_S1;
if C=0 then do;
     if t0>0 then LLIK_s2=-exp(-c1*xa)*(t1**c1-t0**c1);
          else LLIK_s2=-exp(-c1*xa)*(t1**c1);
          end;
else do;
 call quad("IntFun",integral, t0, t1, c1, C);
     LLIK_S2=-exp(-c1*xa)*integral;
     end;

LLIK_S=LLIK_s1+LLIK_s2;
model general(LLIK_L+LLIK_S);
run;
```

The last two statements combine the longitudinal and survival components of the joint model.

Parameters in Joint model					
Block	**Parameter**	**Array Index**	**Sampling Method**	**Initial Value**	**Prior Distribution**
1	z_c1		N–Metropolis	0.5600	MVNormal(mu0, G0)
	z_c2			0.1800	
2	G_c1	[1,1]	Conjugate	0.9800	iWishart(2, G)
	G_c2	[1,2]		0.0700	
	G_c3	[2,1]		0.0700	
	G_c4	[2,2]		1.0800	
3	sigsq		N–Metropolis	0.1200	igamma(shape=.01, scale=.01)
4	b1		N–Metropolis	–0.1280	normal(0, var=1E6)
5	c0		N–Metropolis	2.3000	normal(0, var=1E6)
	c1			2.1000	gamma(shape=1E4, iscale=1E4)
	c2			0.0500	normal(0, var=1E6)
6	lam		N–Metropolis	–3.0000	normal(0, var=1E4)

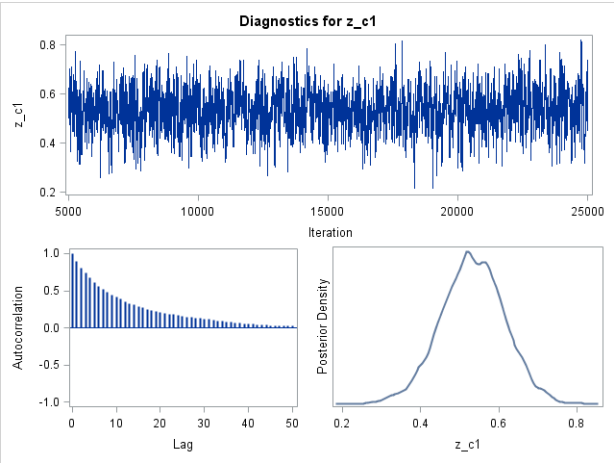
Results from PROC MCMC are summarized in Table 8. A burn-in of 5000 samples was discarded before 20000 Markov chain samples were generated. All other options are at their default values.

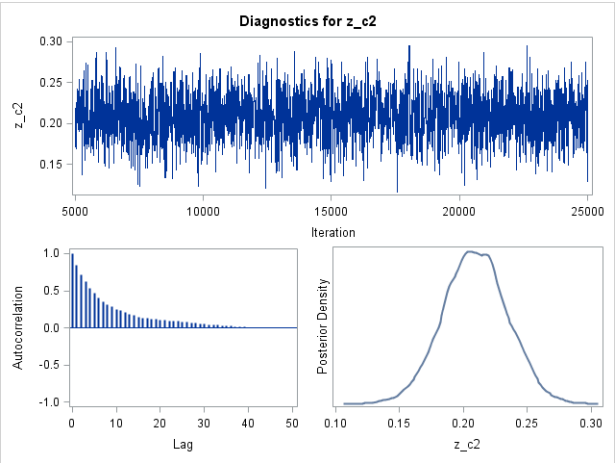
```
proc mcmc data=pbcseq2 seed=22819 nmc=20000 nbi=5000;
```

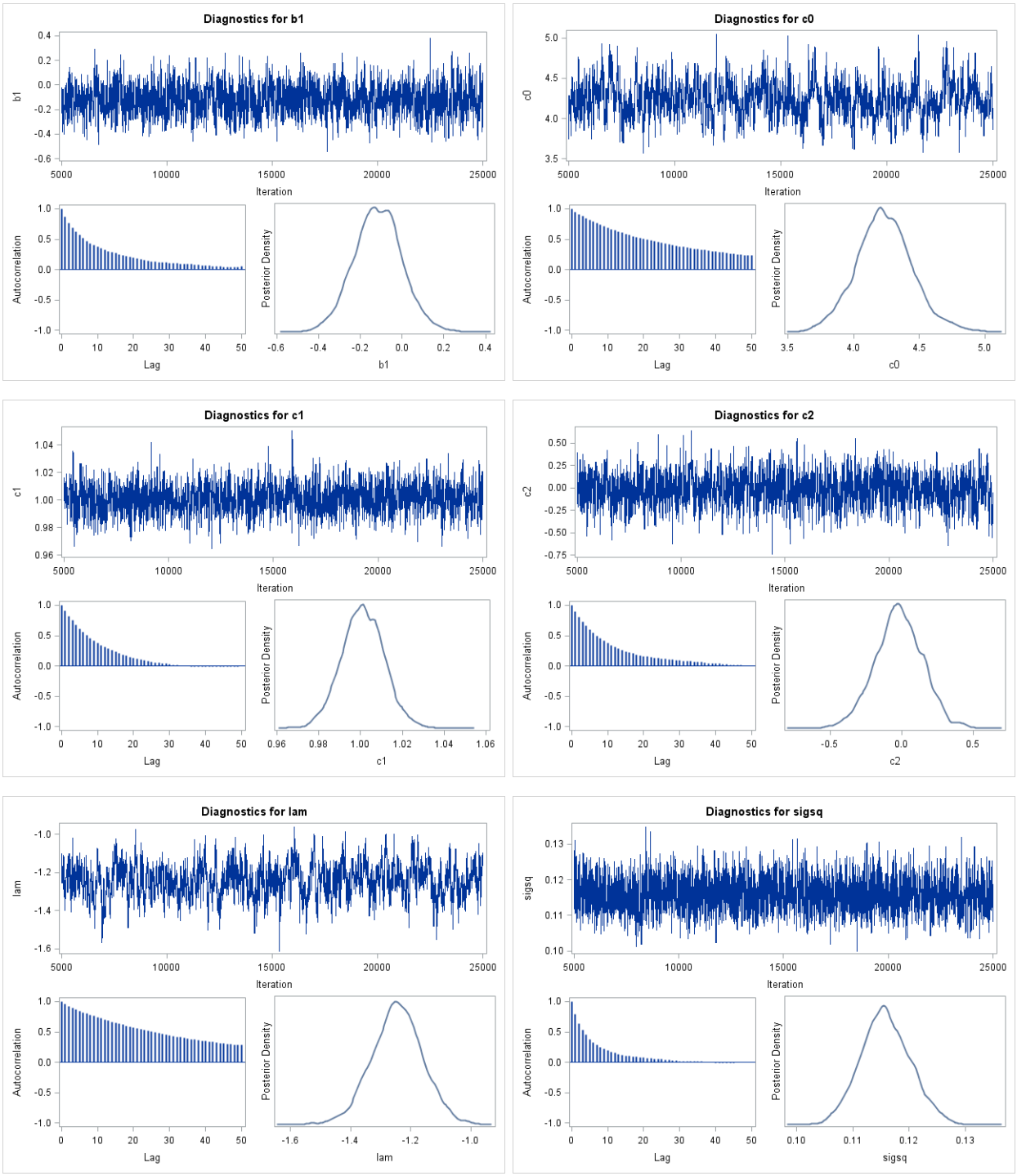
Table 8: Posterior Summaries and 95% HPD Intervals from Joint model

Posterior Summaries and Intervals					
Parameter	**N**	**Mean**	**Standard Deviation**	**95% HPD Interval**	
z_c1	20000	0.5330	0.0813	0.3684	0.6863
z_c2	20000	0.2092	0.0253	0.1595	0.2579
G_c1	20000	0.9956	0.0854	0.8372	1.1686
G_c2	20000	0.0861	0.0278	0.0310	0.1395
G_c3	20000	0.0861	0.0278	0.0310	0.1395
G_c4	20000	0.1505	0.0150	0.1227	0.1808
sigsq	20000	0.1158	0.00439	0.1076	0.1248
b1 drug	20000	–0.1161	0.1141	–0.3328	0.1190
c0	20000	4.2355	0.2133	3.7733	4.6322
c1 shape	20000	1.0011	0.0102	0.9803	1.0203
c2 drug	20000	–0.0199	0.1670	–0.3385	0.3039
lam	20000	–1.2452	0.0847	–1.4059	–1.0749

Execution time for the PROC MCMC call was long, over 2.5 hrs cpu time. It could be that our parameterization was not ideal and a simpler evaluation of the survival likelihood could a made with one QUAD call. Trace, autocorrelation and density (TAD) plots are shown next for parameters except G_c1 to G_c4—which did exceeding well because of our choice of starting values and priors.







7. Concluding Remarks

There has been considerable new development in the area of joint modeling of longitudinal markers and survival outcomes. Extensions of the basic model of single repeated measures marker, modelled by a linear mixed model and a single time-to-event, include competing risks for the latter and multivariate longitudinal outcomes for the former, as well as adoption of the Bayesian framework

(Hickey *et al*, 2018; Li and Luo, 2019; Lu, 2017). Generalized mixed models for the marker allow for repeated binary, binomial and count outcomes. In PROC GLIMMIX we have a versatile tool for this part of the analysis. For the survival part, modelling the hazard function posited in proportional hazards form pervades a majority of methods. However, flexible modelling of the baseline component as piecewise constant function, restricted cubic splines, or fractional-polynomials, permit a fairly broad catalog of models. The thorny part is the inclusion of functions of the marker process into the linear predictor of the hazard function. Shared random effects models are easier to handle primarily because the computational task is far less demanding than when incorporating a function of the marker that has covariates, random effects and time trajectories. Nevertheless, software programs in R, Stata and SAS have "softened" the burden. In this article we show how the suite of SAS procedures, LIFEREG, SEVERITY, GLIMMIX, MCMC with NLMIXED providing the "glue" can be gainfully applied to address some joint modelling problems.

Several reviews highlight the breadth of theoretical developments and have sought to demonstrate real-world applications (Sudell *et al*, 2018; Hickey *et al*, 2016; Asar *et al*, 2015; Ye and Yu, 2014; Tsiatis and Davidian, 2004). A forthcoming book chapter promises to be a comprehensive review of the state of the art (Papageorgiou *et al*, 2019).

REFERENCES

Abrams DI, Goldman AI, Launer C, et al. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human-immunodeficiency-virus infection. *New England Journal of Medicine.* 1994;330(10):657-662.

Asar O, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International Journal of Epidemiology.* 2015;44(1):334-344.

Collett D. *Modelling Survival Data for Medical Research, Third Edition.* Boca Raton, FL: CRC Press; 2015.

Crowther MJ, Abrams KR, Lambert PC. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine.* 2012;31(30):4456-4471.

Garcia-Hernandez A, Rizopoulos D. %JM: A SAS Macro to Fit Jointly Generalized Mixed Models for Longitudinal Data and Time-to-Event Responses. *Journal of Statistical Software.* 2018;84(12):1-29.

Gardiner JC, Luo Z, Tang X, Ramamoorthi RV. Fitting heavy-tailed distributions to health care data by parametric and bayesian methods. *Journal of Statistical Theory and Practice.* 2014;8(4):619-652.

Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. *American Statistician.* 2004;58(1):16-24.

Heckman JJ, Singer B. Econometric Analysis of Longitudinal Data. In: Griliches Z, ed. *Handbook of Econometrics.* Vol 3. Handbooks in Economics series, book 2 Amsterdam; Oxford and Tokyo: North-Holland; distributed in the U.S. and Canada by Elsevier Science New York; 1986:1689-1763.

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology.* 2016;16.

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society Series A.* 2018;181(4):1105-1123.

Ibrahim JG, Chen M-H, Sinha D. *Bayesian Survival Analysis.* New York: Springer-Verlag; 2001.

Li K, Luo S. Bayesian functional joint models for multivariate longitudinal and time-to-event data. *Computational Statistics & Data Analysis.* 2019;129:14-29.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. *SAS for Mixed Models Analyses, Second Edition.* Cary, NC.: SAS Institute Inc; 2006.

Lu T. Bayesian semiparametric mixed-effects joint models for analysis of longitudinal-competing risks data with skew distribution. *Statistics and Its Interface.* 2017;10(3):441-450.

Murtaugh PA, Dickson ER, Vandam GM, et al. Primary biliary-cirrhosis - prediction of short-term survival based on repeated patient visits. *Hepatology.* 1994;20(1):126-134.

Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and its Applications.* Vol 6. 2019.

Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R.* Boca-Raton, FL. CRC-Press 2012.

Royston P, Parmar MKB. 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.* 2002;21(15):2175-2197.

Sudell M, Kolamunnage-Dona R, Tudur-Smith C. Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis (vol 16, 168, 2016). *BMC Medical Research Methodology.* 2018;18.

Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica.* 2004;14(3):809-834.

Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data, Second Edition.* Cambridge, MA: MIT Press; 2010.

Ye W, Yu M. Joint Models of Longitudinal and Survival Data. In: Klein JP, VanHouwelingen HC, Ibrahim JG, Scheike TH, eds. *Handbook of Survival Analysis.* Boca Raton, FL: CRC Press; 2014:523-547.

Zhang D, Chen M-H, Ibrahim JG, Boye ME, Shen W. JMFit: A SAS Macro for Joint Models of Longitudinal and Survival Data. *Journal of Statistical Software.* 2016;71(3):1-24.

CONTACT INFORMATION

We welcome your comments and questions. Please contact

Joseph C. Gardiner
Department of Epidemiology and Biostatistics
College of Human Medicine
Michigan State University
East Lansing, MI 48824
jgardiner@epi.msu.edu