Regression Analysis of Duration and Severity Data –New Capabilities with SAS® Software

Joseph C. Gardiner

Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI

ABSTRACT

Duration and severity data arise in several fields including biostatistics, demography, economics, engineering and sociology. SAS® procedures LIFETEST, LIFEREG and PHREG are the workhorse for analysis of time to event data in applications in biostatistics. Similar methods apply to the magnitude or severity of a random event, where the outcome might be right, left or interval censored and/or, right or left truncated. All combinations of types of censoring and truncation could be present in the data set. Regression models such as the accelerated failure time model, the Cox model and the non-homogeneous Poisson model have extensions to address time-varying covariates in the analysis of clustered outcomes, multivariate outcomes of mixed types, and recurrent events. We present an overview of new capabilities that are available in the procedures QLIM, QUANTLIFE, RELIABILITY and SEVERITY with examples illustrating their application using empirical data sets drawn from easily accessible sources.

INTRODUCTION

In its simplest form survival data comprise a time to a terminal event T measured from a time origin $t=0$, time-invariant covariates $\mathbf{x}$ and an indicator δ that labels whether or not T was actually observed in the sampling unit by the end of the observational period U. If $T > U$ the survival time is right censored at U and we define $\delta = 0$; otherwise, when T is observed, $T \le U$ and we set $\delta = 1$. The two traditional survival models are the Cox proportional hazards model for the hazard function h of T given by $h(t \mid \mathbf{x}) = h_0(t) \exp(\mathbf{x}'\beta)$ where the baseline hazard h_0 is left unspecified, and the accelerated failure time (AFT) model given by $\log T = \mathbf{x}'\beta + \sigma\varepsilon$ where the random variable ε has a fully specified parametric distribution independent of $\mathbf{x}$.

In the AFT model, the parameters are related to the conditional mean $E(T \mid \mathbf{x})$ or to the conditional p-th percentile, $t_p(\mathbf{x})$, $0 < p < 1$. For example, the Weibull distribution for T (that is, ε is extreme value), has survival function $S(t \mid \mathbf{x}) = \exp(-(t / \theta(\mathbf{x}))^\gamma)$. With the parameterization $\log \theta(\mathbf{x}) = \mathbf{x}'\beta$ we have $E(T \mid \mathbf{x}) = \exp(\mathbf{x}'\beta)\Gamma(\sigma + 1)$ and $t_p(\mathbf{x}) = \exp(\mathbf{x}'\beta)\big(-\log(1 - p)\big)^\sigma$ where $\sigma = \gamma^{-1}$. In the lognormal distribution for T (that is, ε is standard normal) we have $S(t \mid \mathbf{x}) = \Phi\big(-\log(t / \theta(\mathbf{x}))^{1/\sigma}\big)$, $\log \theta(\mathbf{x}) = \mathbf{x}'\beta$,

$E(T \mid \mathbf{x}) = \exp(\mathbf{x}'\beta + \frac{1}{2}\sigma^2)$ and $t_p(\mathbf{x}) = \exp\{\mathbf{x}'\beta + \sigma\Phi^{-1}(p)\}$. The Weibull is the prototype distribution with the proportional hazard (PH) form $h(t \mid \mathbf{x}) = \gamma t^{\gamma-1}\exp(-\gamma\mathbf{x}'\beta)$. In both examples the parameterization is through the scale $\theta(\mathbf{x})$.

The Burr distribution is an example of a mixed proportional hazards model in which conditional on $(\mathbf{x}, \nu)$ the hazard has the form $h(t \mid \mathbf{x}, \nu) = \nu h_0(t)\exp(\mathbf{x}'\beta)$ where $h_0(t)$ is the Weibull hazard and the frailty ν is Gamma distributed with shape α and mean 1. With some re-parameterization the survival function is $S(t \mid \mathbf{x}) = \left(1 + (t / \theta(\mathbf{x}))^\gamma\right)^{-\alpha}$ and $E(T \mid \mathbf{x}) = \exp(\mathbf{x}'\beta)\Gamma(1+\sigma)\Gamma(\alpha-\sigma) / \Gamma(\alpha)$, $\sigma < \alpha$, $t_p(\mathbf{x}) = \exp(\mathbf{x}'\beta)\left\{(1-p)^{-1/\alpha} - 1\right\}^\sigma$ where $\sigma = \gamma^{-1}$. Two special cases of the Burr are the Pareto $(\gamma = 1)$ and the log-logistic $(\alpha = 1)$ (Gardiner $et\ al$, 2014).

Time-varying covariates

Extending these models to accommodate time-varying covariates (TVC) greatly expands their reach in applications. In this paper we demonstrate through examples how this can be achieved with the procedures PHREG and SEVERITY. Obviously there are some restrictions. The TVC should be piecewise constant. In practice, in addition to the event data (T, δ) a single TVC $x(t)$ would have observed values at times $0 = t_0 < t_1 < \ldots < t_{M-1} < T \wedge U \equiv t_M$. The points $\{t_m : m = 0, \ldots, M\}$ are now fixed. The initial value $x(0)$ remains constant in the interval $[0, t_1)$, and at t_1 it is updated to $x(t_1)$ which remains constant in $[t_1, t_2)$ and so on. In the last interval the value $x(t_{M-1})$ applies until end of follow up at the event time T or censoring time U whichever is observed. The data set describing these data will have M records for the subject (M might vary by subject), with variables for the left and right endpoints of the intervals, covariate, and event status at the right endpoint. For each intermediate record for $[t_{m-1}, t_m)$, T is considered left-truncated at t_{m-1} and right-censored at t_m (with indicator $\delta_m = 0$), except for the last record where the event status recorded in δ will become operational, that is $\delta_M = \delta$. The same data processing can be applied to examples with several TVCs under consideration, with each record capturing data on updated values of at least one covariate.

Parametric models

With the notation that $\mathbf{x}(t)$ includes the covariate history through time t, the definition of the hazard function is formally $h(t \mid \mathbf{x}(t)) = \lim_{\Delta t \downarrow 0} P[t \le T < t + \Delta t \mid T \ge t, \mathbf{x}(t)] / \Delta t$. The observed data on a subject is $(T, \delta, \mathbf{x})$ where $\mathbf{x} = \{\mathbf{x}(t_m), m = 0, \ldots, M-1\}$. Assume conditional independence of (T, U), given $\mathbf{x}$ and strict exogeneity of $\mathbf{x}$. Denoting a conditional distribution by $D(\cdot \mid \cdot)$ the strict exogeneity assumption is $D(T \mid T \ge t_{m-1}, \mathbf{x}) = D(T \mid T \ge t_{m-1}, \mathbf{x}(t_{m-1})), m = 1, \ldots, M$ (Wooldridge, 2010). The contribution to the log-likelihood of a generic datum $(T, \delta, \mathbf{x})$ is given by

$\sum_{m=1}^{M} \log\left(P[T > t_m \mid T \ge t_{m-1}, \mathbf{x}(t_{m-1})]\right) + \delta_M \log\left(h(t_M \mid \mathbf{x}(t_{M-1}))\right)$. The term in the summation is $-\int_{t_{m-1}}^{t_m} h(u \mid \mathbf{x}(t_{m-1})) du$. For observations from a random sample, there is a second summation over subjects with appropriate subscripting, $\{T_i, \delta_i, \mathbf{x}_i(t_{im}), m = 0, \ldots, M_i - 1\}$. We make three comments.

(1) Knowing the hazard function in parametric form does in principle provide an avenue for estimation of parameters via maximum likelihood. In our application to follow, we use PROC SEVERITY to estimate the Weibull and lognormal AFT models, and the Burr and its special cases, Pareto and log-logistic.

(2) Assuming proportional hazards, $h(t \mid \mathbf{x}(t)) = h_0(t) \exp(\mathbf{x}'(t)\beta)$ the log-likelihood simplifies to

$$\sum_{m=1}^{M} \left(-\int_{t_{m-1}}^{t_m} h_0(u) du\right) \exp(\mathbf{x}'(t_{m-1})\beta) + \delta_M \left(\mathbf{x}'(t_{M-1})\beta + \log h_0(t_M)\right).$$

For the Weibull, $h_0(t) = \gamma t^{\gamma-1}$. Thus we obtain a log-likelihood parameterized by (β, γ). Again, PROC SEVERITY can be used for estimation.

(3) Assume h_0 to be piecewise constant: $h_0(t) = \lambda_j, t \in [a_{j-1}, a_j)$ $j = 1, \ldots, J$ with a fixed partition of the time axis, $0 = a_0 < a_1 < \ldots < a_{J-1} < a_J = +\infty$. The distribution of T is called the piecewise exponential. There is an advantage of using the parameterization $\alpha_j = \log(\lambda_j), j = 1, \ldots, J$ because of the lack of log concavity of the log-likelihood function in λ_j when $[a_{j-1}, a_j)$ does not capture observed event times. Maximum likelihood estimation provides estimates for $\beta, \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$. The BAYES option in PROC PHREG is a trick to obtain the maximum likelihood estimates because they are the default starting parameter values for the MCMC computations for the posterior distribution of $(\beta, \boldsymbol{\alpha})$. The entire Bayes machination can be suppressed.

Semi-parametric models

The investment in structuring the data set in multiple record formats for TVCs pays off very well when estimating the Cox proportional hazards model $h(t \mid \mathbf{x}(t)) = h_0(t) \exp(\mathbf{x}'(t)\beta)$. The counting process style input in PROC PHREG exploits precisely this format. The baseline hazard h_0 is unspecified. The aforementioned likelihood derivation must be replaced by the partial likelihood given by $\prod_{i=1}^{n} \prod_{t>0} \left\{ \dfrac{Y_i(t) \exp(\beta' \mathbf{x}_i(t))}{\sum_{k=1}^{n} Y_k(t) \exp(\beta' \mathbf{x}_k(t))} \right\}^{\Delta N_i(t)}$ where $N_i(t) = [T_i \le t, \delta_i = 1]$, $Y_i(t) = [T_i \ge t, U_i \ge t]$ are respectively, the observed event indicator and at-risk indicator: In the i-th subject $N_i(t) = 1$ if the event has occurred by time t; otherwise $N_i(t) = 0$. Also $Y_i(t) = 1$ if the i-th subject is at risk of the event at time $t-$; otherwise $Y_i(t) = 0$.

APPLICATION

The Panel Study of Income Dynamics (PSID) which began in 1968 is the longest running longitudinal household survey in the world. Data were collected at each wave on socioeconomic factors including health, education, marriage and detailed information on childbearing. For our illustration we use a data set assembled by Lillard and Panis (1996, 2003) of 3371 respondents on the duration (DUR) of their first marriage with a modification by Rabe-Hesketh and Skrondal (2012) that describes DUR as a continuous variable measured in years from the start of marriage to divorce or end of follow up. Approximately 31% of the marriages ended in divorce. Censoring may occur due to death of a spouse, or if the couple were still married at their last interview. Covariates for our analysis are the time-fixed factors at marriage: race of husband/wife (RACE), husband's education (HISEDUC), and the difference in age between husband and wife (AGEDIFF). The categories are reflected in the following formats:

```
proc format;

value agediff low-<-10='She older'
              -10-<10, 10='Within 10 years'
              10<-high='He older';
value educ low-<12='<12 yrs' 12-<16='12-<16 yrs' 16-high='>=16 yrs';
value race 1='Both White' 2='Both Black' 3='Black/White'  4='White/Black';
run;
```

During the marriage the number of children ($x(t)$=NUMKIDS) born is incorporated in the data set DIVORCE_LG as multiple records per couple identified by ID. The intervals [TSTART, TSTOP) describes the dynamics of change in NUMKIDS. Initially, at TSTART=0 we have NUMKIDS=0. Subsequently, at the birth of a child at TSTOP—time elapsed since marriage in years, NUMKIDS increments by 1.

Table 1: Duration of marriage and child birth data for three couples

ID	DUR	TSTART	TSTOP	flag	rctime	lttime	numkids	divorce
11	0.7670	0.0000	0.7670	0	0.7670	.	0	0
11	32.5120	0.7670	32.5120	0	32.5120	0.7670	1	0
11	34.9430	32.5120	34.9430	1	34.9430	32.5120	2	0
13	2.5850	0.0000	2.5850	0	2.5850	.	0	0
13	2.8702	2.5850	2.8702	1	.	2.5850	1	1
15	18.1840	0.0000	18.1840	1	.	.	0	1

In table 1, DUR=TSTOP is the current duration of marriage. Couple ID=11 had two children. Their first child was born at 0.767 years, their second child at 32.512 years and their marriage was either intact or a spouse had demised (DIVORCE=0) at last follow up at 34.943 years. Couple ID=13 had one child at 2.585 years and divorced (DIVORCE=1) at 2.8702 years. The marriage of couple ID=15 who had no children lasted 18.184 years, ending in divorce. FLAG is an indicator for last record. RCTIME and LTTIME will be described later. The semi-parametric analysis is carried out in PHREG with the following syntax.

4

```
ods output hazardratios=HRS;
proc phreg data=c.divorce_lg;
class agediff(ref='Within 10 years')  race(ref='Both White')
hiseduc(ref='12-<16 yrs')/param=ref;
model (tstart, tstop)*divorce(0)= race agediff hiseduc numkids;
format race race. agediff agediff. hiseduc educ.;
hazardratio numkids/cl=wald;
hazardratio race/diff=ref cl=wald;
hazardratio agediff/diff=ref cl=wald;
hazardratio hiseduc/diff=ref cl=wald;
run;
```

Parameter estimates from maximum partial likelihood estimation are in Table 2.

Table 2: Analysis of maximum likelihood estimates

Parameter	Class	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Type 3
race	Black/White	1	0.45057	0.10630	17.9657	<.0001	<.0001
race	Both Black	1	0.15300	0.10019	2.3322	0.1267	
race	White/Black	1	0.33244	0.09939	11.1872	0.0008	
agediff	He older	1	−0.27856	0.21665	1.6532	0.1985	0.0716
agediff	She older	1	0.48829	0.26127	3.4927	0.0616	
hiseduc	<12 yrs	1	−0.30615	0.06824	20.1275	<.0001	<.0001
hiseduc	>=16 yrs	1	−0.27797	0.10568	6.9186	0.0085	
numkids		1	−0.11018	0.02208	24.8956	<.0001	<.0001

The default Wald tests for each of the four effects in the model are displayed in column "Type3". Age differences do not appear to matter overall (p=.072, df=2) on the hazard for divorce. The negative coefficient for numkids indicates that the hazard for divorce is lower in couples with more children. We cannot extrapolate this to a causal effect because the assumption of exogeneity of the TVC $x(t)$ is implausible. Lillard (1993) comment on couples who face economic and marital problems might delay having children. The hazard for divorce influences the path of $x(t)$. A more involved analysis would entail a multi-state model for child-bearing with forward flowing states $0 \rightarrow 1 \rightarrow 2 \rightarrow \ldots$ with exit to the terminal state, divorce (Hougaard, 2000).

The HAZARDRATIO statement in PHREG is used to generate a table of hazard ratios and 95% Wald confidence intervals. The default results are in Table 3 and displayed graphically in Figure 1.

```
proc sgplot data=HRS;
scatter y=description x=hazardratio/xerrorlower=Waldlower
xerrorupper=Waldupper markerattrs=(symbol=DiamondFilled size=8);
refline 1 / axis=x;
xaxis label="Hazard Ratio " min=0;
yaxis label="Comparisons";
title "Hazard Ratio and Wald 95% CI ";
run;
```

Table 3: Hazard ratios and 95% confidence intervals

Obs	Description	HazardRatio	WaldLower	WaldUpper
1	numkids Unit=1	0.896	0.858	0.935
2	race Black/White vs Both White	1.569	1.274	1.933
3	race Both Black vs Both White	1.165	0.958	1.418
4	race White/Black vs Both White	1.394	1.148	1.694
5	agediff He older vs Within 10 years	0.757	0.495	1.157
6	agediff She older vs Within 10 years	1.630	0.976	2.719
7	hiseduc <12 yrs vs 12-<16 yrs	0.736	0.644	0.842
8	hiseduc >=16 yrs vs 12-<16 yrs	0.757	0.616	0.932

Figure 1: Display of the hazard ratios of 95% confidence intervals
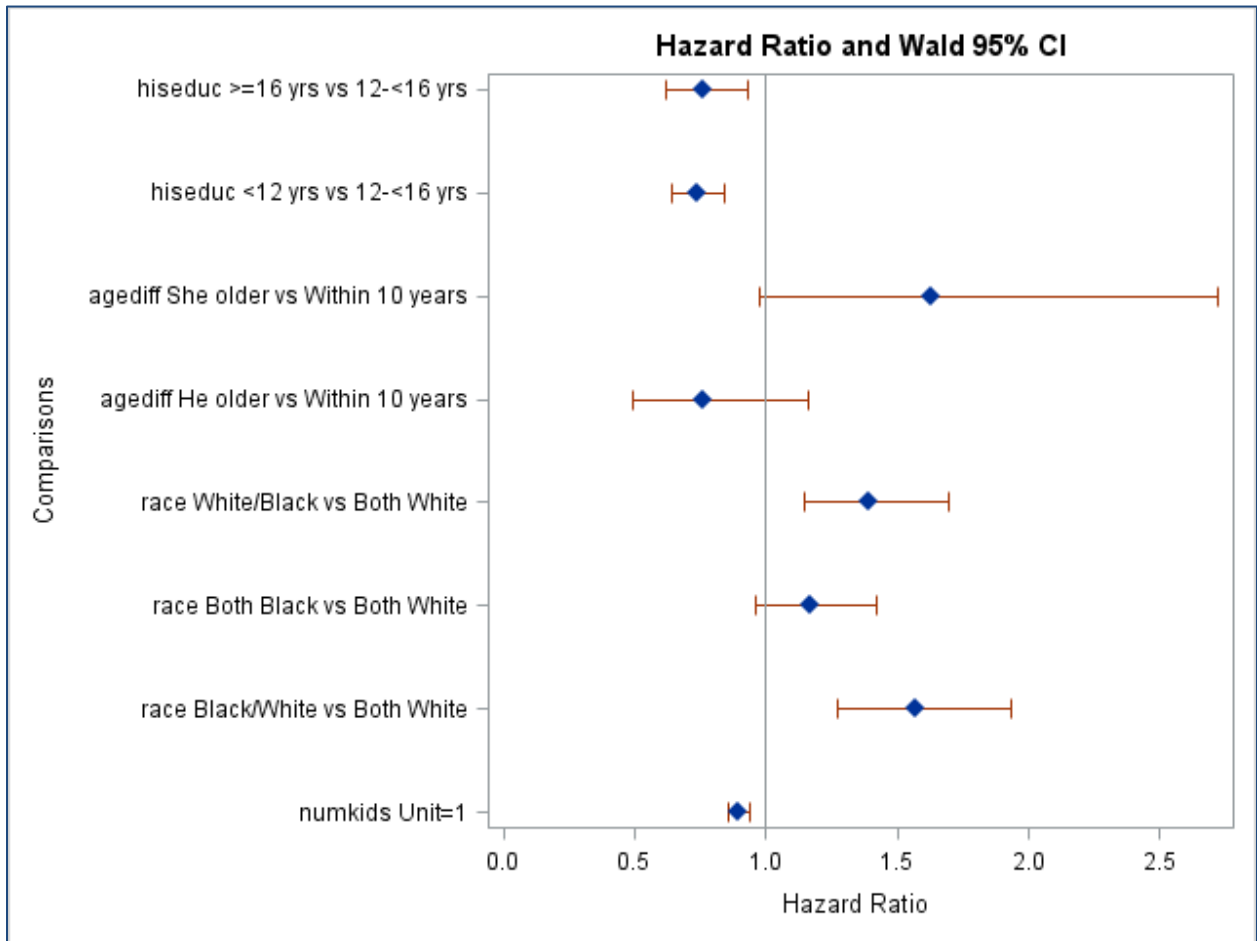
Parametric models

We will use PROC SEVERITY to estimate Burr, Weibull and lognormal models. For each record we must define the right censoring time (RCTIME) and left-truncation time (LTTIME). If these variables have missing values it means that censoring and or truncation are not present. Note that SEVERITY will not permit the value 0 for LTTIME in the initial record of each couple. The value 0 is operationally equivalent to a missing value for left-truncated times. See Table 1.

SEVERITY currently does not support a CLASS statement and therefore to keep the same parameterization of the semiparametric model we must create indicator variables for race, agediff and hiseduc. They are shown in the SCALEMODEL statement which structures the scale $\theta(\mathbf{x})$ as $\log\theta(\mathbf{x}) = \mathbf{x}'\beta$ with an intercept. In principle SEVERITY can fit any parametric distribution that has a scale parameter. There are 10 pre-defined distributions, including the Burr, Weibull and lognormal. However, with assistance from the FCMP procedure one only needs to define the cumulative distribution and density functions, and then invoke the procedure within SEVERITY. See Gardiner *et al.* (2014) for details on fitting the Coxian distribution which is mixture of Erlang distributions.

The following syntax fits five distributions declared in the DIST statement. Log-logistic was programmed in the FCMP procedure. The LOSS statement specifies TSTOP as the analysis variable. In table 1 we see that it is the current duration of the marriage (DUR) which may be left-truncated and /or right censored, indicated by the options after the slash/.

```
proc severity data=divorce_lg print=all outcdf=cdf;
dist burr weibull logn pareto llogistic;
loss TSTOP/rightcensored=rctime lefttruncated=lttime;
scalemodel race_BW race_BB race_WB  heolder sheolder hedropout
hecollege numkids;
run;
```

One interesting feature is the fit statistics generated for the parametric distributions. The fitted cumulative survival distribution function (CDF) $\hat{S}(t)$ can be defined in several ways. The full average is the mixture $\hat{S}(t) = n^{-1}\sum_{i=1}^{n} S(t \mid \hat{\theta}(\mathbf{x}_i), \hat{\omega})$ where $\hat{\theta}(\mathbf{x}_i) = \exp(\mathbf{x}_i'\hat{\beta})$ and $\hat{\omega}$ denotes estimates of all parameters that are not associated with covariates (e.g., intercept, shape parameters). This can be potentially a very time-consuming computation if there are several continuous covariates. With the current application we had no concerns. There are <200 configurations of the covariates. Another choice of $\hat{S}(t)$ is the calculation at the mean $\bar{\theta} = n^{-1}\sum_{i=1}^{n} \hat{\theta}(\mathbf{x}_i)$. SEVERITY allows for other choices with the DFMIXTURE option in the SCALEMODEL statement. Fit criteria (AIC, BIC, KS, CVM) compare the CDF to the empirical survival distribution function (EDF). The OUTCDF= option saves the results that can be used for plotting.

Voluminous output is created by the above syntax. Only a small portion has been corralled in the Tables 4 and 5. From the 'smaller the better criterion' the Burr distribution fits the best. Because $\alpha < \sigma$, it does not have a finite mean.

Table 4: Fit statistics for selected parametric distributions

Dist	−2 Log Likelihood		AIC		AICC		BIC		KS		AD		CvM	
Burr	10245	*	10267	*	10267	*	10344	*	8.00757	*	1447		496.42439	*
Weibull	10391		10411		10411		10482		23.20500		735.22624	*	499.23054	
Logn	10287		10307		10307		10378		15.65472		961.08157		497.23254	
Pareto	10350		10370		10370		10441		14.06586		1040		496.74218	
LLogistic	10333		10353		10353		10424		19.06309		826.44920		497.95849	
Note: The asterisk (*) marks the best model according to each column's criterion.														

Table 5: Maximum likelihood estimates for selected parametric distributions

Parameter	Class	Burr Parameter Estimate	Standard Error*	Weibull Parameter Estimate	Standard Error*	Lognormal Parameter Estimate	Standard Error*
race	**Black/White**	−0.30385	0.11680	−0.56380	0.10964	−0.45325	0.12069
race	**Both Black**	0.07365	0.10282	−0.22831	0.10366	−0.06992	0.10902
race	**White/Black**	−0.25721	0.10625	−0.43332	0.10226	−0.36453	0.10976
agediff	**He older**	−0.10821	0.21169	0.34951	0.22467	0.14928	0.21046
agediff	**She older**	−0.57429	0.29731	−0.52924	0.27080	−0.58613	0.28968
hiseduc	**<12 yrs**	0.19915	0.07475	0.39222	0.07009	0.30029	0.07410
hiseduc	**>=16 yrs**	0.32622	0.10340	0.27430	0.10978	0.32558	0.11198
numkids		0.20475	0.02418	0.13281	0.02040	0.16431	0.02444
intercept		1.54030		3.86197		3.53125	
sigma		0.53913		1.03598		1.59407	
alpha		0.17037					

*Default computed by procedure. See below.

In all three models numkids is significant. Percentiles of the duration of marriage are obtained directly from the formula for $t_p(\mathbf{x})$. Keeping all other variables fixed, for the lognormal the effect of one additional child at a timepoint is exp(0.16431)=1.18, that is, an 18% increase in the any percentile of duration. For the Burr the effect is larger, about a 23% increase. We again note that the strict exogeneity of the TVC $\mathbf{x}(t)$: $D(T \,|\, T \geq t_{m-1}, \mathbf{x}) = D(T \,|\, T \geq t_{m-1}, \mathbf{x}(t_{m-1})), m = 1, \ldots, M$, is used in the derivation of the log-likelihood. This assumption is perhaps untenable with $x(t)$=NUMKIDS(t).

If a confidence interval (CI) for the median duration for marriage is desired at a specified covariate value, we might start with a CI for $\log t_p(\mathbf{x})$. The covariance matrix $\mathbf{V}$ of $(\hat{\omega}, \hat{\beta})$ can be saved in a data set created with the COVOUT option in the SEVERITY statement.. The procedure makes a degrees-of-freedom adjustment. If $\mathbf{G} = -\dfrac{\partial^2 LogL}{\partial \beta \partial \beta'}$ denotes the Hessian matrix, then $\mathbf{V} = \dfrac{N}{d}\mathbf{G}^{-1}$ where d=N−p by default, p=#parameters. The option VARDEF=N that makes d=N is appropriate here because N=# records in the data set is much larger than the sample size of couples (=3371).

The piecewise constant hazard model is fitted with PHREG invoking the BAYES statement. A Bayesian analysis is not the intent, but the starting values of all parameters for the Gibbs sampler are the maximum likelihood estimates of β, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ in the hazard $h(t \mid \mathbf{x}) = \exp(\alpha_j + \mathbf{x}'\beta)$, $t \in [a_{j-1}, a_j)$, $j = 1, \ldots, J$. By default the number of intervals $J=8$, and can be set with the NINTERVALS option. The cut-points are selected to have approximately equal number of events in the intervals. In this example there are 1032 marriages that ended in divorce, so each interval contains 129 divorces. The INTERVAL option allows one to change the cut-points. For an example of options and details of using the BAYES statement see Gardiner (2010). Accepting implicit defaults, a piecewise constant hazard model with the same covariates as in table 2 is obtained by

```
proc phreg data=c.divorce_lg;
class agediff(ref='Within 10 years') race(ref='Both White')
hiseduc(ref='12-<16 yrs')/param=ref;
model (tstart, tstop)*divorce(0)= race agediff hiseduc numkids;
format race race. agediff agediff. hiseduc educ.;
bayes nbi=1000 nmc=2000 seed=20314 coeffprior=uniform plots=none
PIECEWISE=loghazard(prior=uniform ninterval=8);
run;
```

With some annotation of the default output the results of ML estimation are in Table 6.

Table 6: Maximum Likelihood Estimates Piecewise Constant Hazard Model

Parameter	Cut-point /Class	Estimate	Standard Error	95% Confidence Limits	
Alpha1	2.505443	–4.0793	0.0957	–4.2668	–3.8917
Alpha2	4.397765	–3.6483	0.0965	–3.8375	–3.4592
Alpha3	6.242479	–3.4899	0.0977	–3.6814	–3.2984
Alpha4	8.372772	–3.4885	0.0991	–3.6827	–3.2943
Alpha5	10.92927	–3.5192	0.1004	–3.7160	–3.3225
Alpha6	15.56494	–3.9031	0.1019	–4.1029	–3.7033
Alpha7	21.58546	–3.8336	0.1038	–4.0371	–3.6301
Alpha8	Infty	–4.4567	0.1068	–4.6661	–4.2474
race	Black/White	0.4672	0.1064	0.2587	0.6757
race	Both Black	0.1659	0.1002	–0.0305	0.3623
race	White/Black	0.3489	0.0994	0.1540	0.5437
agediff	He older	–0.3068	0.2166	–0.7314	0.1179
agediff	She older	0.4855	0.2613	–0.0265	0.9976
hiseduc	<12 yrs	–0.3187	0.0683	–0.4524	–0.1849
hiseduc	>=16 yrs	–0.2743	0.1057	–0.4814	–0.0671
numkids		–0.1108	0.0218	–0.1534	–0.0681

RECURRENT EVENTS

Recurrent events refer to events of the same type that can have multiple occurrences during a follow-up period. There are many examples such as tumor recurrence, hospital readmissions, asthma attacks, breakdowns in repairable systems, etc. In the observational period $[0, \tau]$, occurrences at times $T_1 < T_2 < T_3 < \dots$ produce the cumulative number of occurrences to time t, $N(t) = \sum_{k=1}^{\infty} [T_k \leq t]$ and the gap times between occurrences $W_1 = T_1$, $W_2 = T_2 - T_1, \dots, W_k = T_k - T_{k-1}, \dots$. Changes in the process $\mathbf{N} = \{N(t) : t \geq 0\}$ could be studied through its *intensity* $\lambda(t) = \lim_{\Delta t \to 0} P[\Delta N(t) = 1 | \mathfrak{I}_{t-}] / \Delta t$ where it is assumed that the jump $\Delta N(t)$ at time t in the process $\mathbf{N}$ is at most 1. The past history $\mathfrak{I}_{t-}$ captures all that is known about the process and TVCs just before t. Cook and Lawless (2007) construct a likelihood for the events occurring at times $\{T_k\}$ in $[0, \tau]$ but the theory does not readily give the mean $\mu(t) = E(N(t))$, variance $\sigma^2(t) = Var(N(t))$ or the distribution of event counts in an interval. When $\mathfrak{I}_{t-}$ excludes the past history of the process itself, but may contain exogenous TVCs, it is customary to replace the intensity by the *rate of occurrence function* $\rho(t)$ and use results from Poisson processes to make inference on $\rho(t)$ or $\mu(t)$.

Non-homogeneous Poisson process

The prototypical process $\mathbf{N}$ is the non-homogeneous Poisson process (NHPP) in which the independent increments property assures that $\rho(t)$ does not depend on the past history of $\mathbf{N}$. The *mean cumulative function* (MCF) is $\mu(t) = \int_0^t \rho(u) du$ and $P[N(t) = k] = \exp(-\mu(t)) \frac{\{\mu(t)\}^k}{k!}$, $k = 0, 1, \dots$. In the NHPP gap times $\{W_k : k \geq 1\}$ are not independent, $P[W_k > w | T_{k-1} = t] = \exp\left(-\int_t^{t+w} \rho(s) ds\right)$. In the case of a homogeneous Poisson process (HPP) where $\rho(t) = \rho$ is constant, the gap times are independent with exponential distribution $P[W_k > w] = \exp(-\rho w)$.

Time-invariant covariates $\mathbf{x}$ are included in $\rho(t | \mathbf{x}) = \rho_0(t) \exp(\mathbf{x}'\beta)$ where $\rho_0(t)$ is parametrically specified: (1) HPP $\rho_0(t) = 1$, (2) NHPP(LOG) $\rho_0(t) = \exp(\gamma t)$, (3) NHPP(PROP) $\rho_0(t) = \gamma t^{\gamma-1}$, and (4) NHPP(POW) $\rho(t | \mathbf{x}) = \gamma t^{\gamma-1} \eta^{-\gamma}$, (5) NHPP(CA) $\rho(t | \mathbf{x}) = \gamma t^{\gamma-1} \eta$, where $\eta = \mathbf{x}'\beta$.

Estimation in the NHPP uses for the i-th subject the observed event times $T_{i1} < T_{i2} < T_{i3} < \dots < T_{in_i}$, and the end of observation τ_i to construct the likelihood $\prod_{j=1}^{n_i} \{\rho(T_{ij} | \mathbf{x}_i)\} \exp(-\mu(\tau_i | \mathbf{x}_i))$. Note that the period $(T_{in_i}, \tau_i]$, contributes the term $\exp\left(-\int_{T_{in_i}}^{\tau_i} \rho(s | \mathbf{x}_i) ds\right)$ to the likelihood, unless $\tau_i = T_{in_i}$.

Therefore when constructing a data set with recurrences for analysis this end of observation information must be included. Estimation of parameters is via maximum likelihood.

APPLICATION

A data set of multiple incidences of mammary tumors in 48 female rats is constructed from a carcinogenicity experiment (Cook and Lawless, 2007; Hougaard, 2000). The study began with 76 animals that were given a carcinogen and treated with retinyl acetate for 60 days. At day 60, 48 surviving animals free of tumors were randomized to continued treatment (TRT=1, n= 23) with retinoid prophylaxis or control (TRT=0, n=25). Tumors appearance was recorded intermittently ending on day 182. The control group has animals with multiple tumors detected on the same day.

Table 7 shows the layout of the data. ID=3 had tumors at day 63 and 68. STATUS=1 for these two records, and a third record is added to the period (63, 182] with STATUS=0. For ID=32 in addition to single events at day 66, 77 and 98, there were three tumors at day 102. They are given separate records. ID=48 has tumors at day 84, 134 and 182. An additional record is included for the end of observation although the interval (TSTART, TSTOP] is null. Note that for each animal their last record has STATUS=0 (censored). The variables t0, t1 reset the time origin to day=60.

Table 7: Tumor recurrence times in rats

ID	TRT	days	status	tstart	tstop	t0	t1	lttime	rctime
3	1	63	1	0	63	0	3	.	.
3	1	68	1	63	68	3	8	3	.
3	1	182	0	68	182	8	122	8	122
32	0	66	1	0	66	0	6	.	.
32	0	77	1	66	77	6	17	6	.
32	0	98	1	77	98	17	38	17	.
32	0	102	1	98	102	38	42	38	.
32	0	102	1	102	102	42	42	42	.
32	0	102	1	102	102	42	42	42	.
32	0	182	0	102	182	42	122	42	122
48	0	84	1	0	84	0	24	.	.
48	0	134	1	84	134	24	74	24	.
48	0	182	1	134	182	74	122	74	.
48	0	182	0	182	182	122	122	122	122

Use PROC RELIABILITY to estimate the NHPP for tumor recurrences with rate function $\rho(t \mid \mathbf{x}) = \gamma t^{\gamma-1} \exp(\beta_0 + \beta_1(TRT = 0))$.

```
ods output modobstats=obstats;
proc reliability data=mammary_all;
unitid id;
class trt;
distribution NHPP(prop);
model t1*status(0)=TRT/hpptest obstats;
estimate "mean ratio" TRT -1 1/cl exp;
format trt trt.;
run;
```

```
proc format;
value trt 0='CNT' 1='TRT';
run;
```

The **unitid** statement is critical to identifying records of the same subject (animal). The censoring indicator **status(0)** is required to identify the end of observation time.

Table 8: NHPP–Proportional intensity parameter estimates

Parameter		Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits	
				Lower	Upper
Intercept		–4.0302	0.3684	–4.7522	–3.3082
TRT	CNT	0.8097	0.1520	0.5118	1.1076
TRT	TRT	0.0000	0.0000	0.0000	0.0000
Shape		1.0420	0.0719	0.9102	1.1929
–Log L	893.1958				

The estimate of the mean (rate) ratio, treatment vs control $\exp(-\beta_1)$ is 0.455 (95% CI: 0.330, 0.599) calculated from table 8 or from the ESTIMATE statement. The HPP option produces a likelihood ratio test of $H_0 : \gamma = 1$, which is not significant (p=0.554). To fit the HPP distribution use **distribution HPP.**

Observation-wise computations of the estimated MCF $\mu(t \,|\, \mathbf{x}) = t^\gamma \exp(\mathbf{x}'\beta)$ under the model are saved in the modobstats=OBSTATS data set. Standard errors and 95% CIs are computed from $Var(\hat{\mu}(t \,|\, \mathbf{x})) = \hat{\mu}^2(t \,|\, \mathbf{x}) \times Var\{\log(\hat{\mu}(t \,|\, \mathbf{x}))\}$ and under asymptotic normality of the MLE $(\hat{\gamma}, \hat{\beta})$. The MCFs at 122 days are: TRT= 2.652 (95% CI: 2.064, 3.409), CNT=5.960 (95% CI: 5.076, 6.998) showing a wide divergence between TRT and CNT in the mean count of tumors. The MCFs may be plotted using the MCFPLOT statement as described in the next section.

Nonparametric Analyses

In addition to the model-based MCF, a nonparametric estimate of the MCF and related statistics are obtained from the MCFPLOT statement. To keep the nonparametric analysis separate from the NHHP analyses use

```
ods output mcfest=mcfest mcfDest=mcfdiff;
proc reliability data=mammary_all;
unitid id;
class trt;
mcfplot t1*status(0)=trt/logintervals variance=lawless overlay;
mcfplot t1*status(0)=trt/ mcfdiff;
format trt trt.;
run;
```

The first mcfplot statement assures separate calculations for the two groups TRT and CNT. The MCF given by $\hat{\mu}(t) = \int_0^t \frac{dN(s)}{Y(s)}$ where $N(t) = \sum_{i=1}^n N_i(t)$ and $Y(t) = \sum_{i=1}^n Y_i(t)$ denote respectively, the observed counts of events and number at risk at t within group. Among variance estimators of $\hat{\mu}(t)$ are the default (VARIANCE=LAWLESS) which is robust to departures from the Poisson assumptions, and VARIANCE=POISSON. They are given by

$$Var_L(\hat{\mu}(t)) = \sum_{i=1}^n \left\{ \int_0^t \frac{Y_i(u)}{Y(u)} \left(dN_i(u) - \frac{dN(u)}{Y(u)} \right) \right\}^2 , \ Var_P(\hat{\mu}(t)) = \int_0^t \frac{dN(s)}{Y^2(s)} .$$

The option **logintervals** requests computation of the 95% CI based on $\log(\hat{\mu}(t))$. The default plot includes values of $\hat{\mu}(t)$, standard errors and pointwise CIs at each event time. Due to ties in tumor recurrence times, there is some overprinting. Using PROC SGPLOT with the calculations saved in the ODS mcfest data set, a custom graph may be produced. See Figure 2. Note the change in names of the variables TRT and t1 to GROUP and AGE.

```
data mcfest2;
set mcfest;
by group age;
if last.age;
run;

proc sgplot data=mcfest2 noautolegend;
step x=age y=mcf/group=group name='group';
band x=age lower=lower upper=upper/group=group type=step
transparency=.7;
label age='Days from randomization';
keylegend 'group' /title='GROUP' location=inside position=bottomright;
run;
```

Displaying the parametric NHPP analysis with nonparametric point estimates requires only the FIT option **mcfplot t1*status(0)=trt/fit=model;** and of course the previous distribution and model statements (Figure 3).

The second mcfplot statement generates estimates for the difference between CNT and TRT, $\hat{\mu}_C(t) - \hat{\mu}_T(t)$ together with standard errors and pointwise 95% CIs and a default plot (Figure 4). Because of ties in event times, the saved mcfdiff data set is used to construct the plot at the latest time point within the tied sets (Figure 5).

PROC RELIABILITY performs a log-rank type test of equality of the rate functions $H_0 : \rho_1(t) = \rho_0(t)$ against the alternative $H_A : \rho_1(t) = \rho_0(t)\exp(\beta)$ where the baseline rate $\rho_0(t)$ (CNT) is unspecified. Because the alternative makes $\rho_1(t) \le \rho_0(t)$ or $\rho_1(t) \ge \rho_0(t)$ for all t, the test is appropriate if the two MCFs are not expected to across. This is supported by the figures. The test is

based on the statistic $U(\tau) = \int_0^\tau w(u)\{d\mu_1(u) - d\mu_0(u)\}$ where $w(u) = \dfrac{a(u)Y_1(u)Y_0(u)}{Y_\bullet(u)}$, $Y_\bullet(u) = Y_0(u) + Y_1(u)$

and subscripts refer the two groups. Tests are produced for constant and linear weight $a(u)$.

Table 9: Tests for equality of mean cumulative functions

Weight Function	Statistic	Variance	Chi Square	DF	Pr > Chi Square
Constant	39.625000	104.532426	15.020608	1	0.0001
Linear	12.669414	14.070846	11.407562	1	0.0007

Figure 2: MCF and 95% CI (nonparametric)   **Figure 3: MCF and 95% CI (NHPP)**
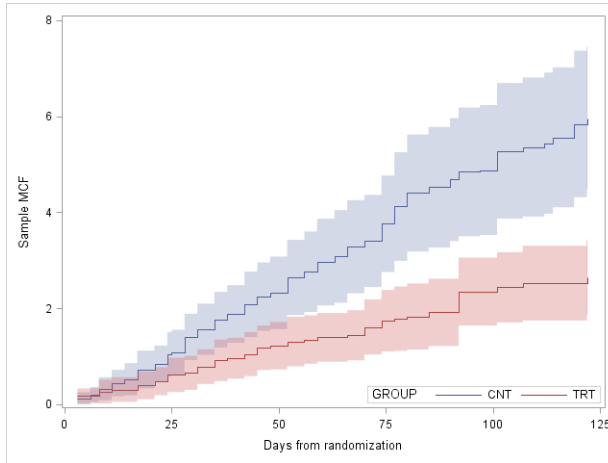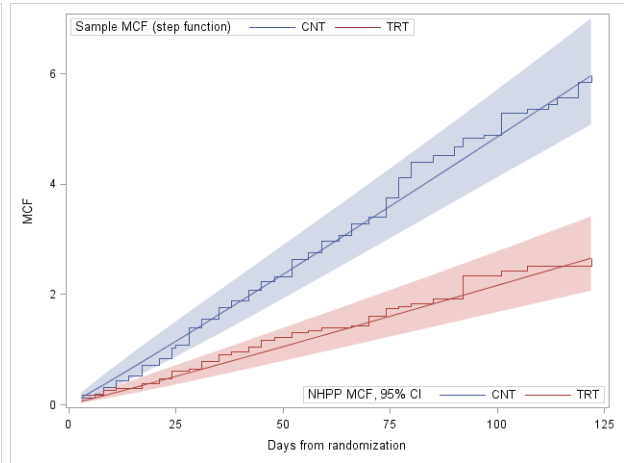
Figure 4: Default plot for difference in MCFs   **Figure 5: Plot for difference in MCFs**
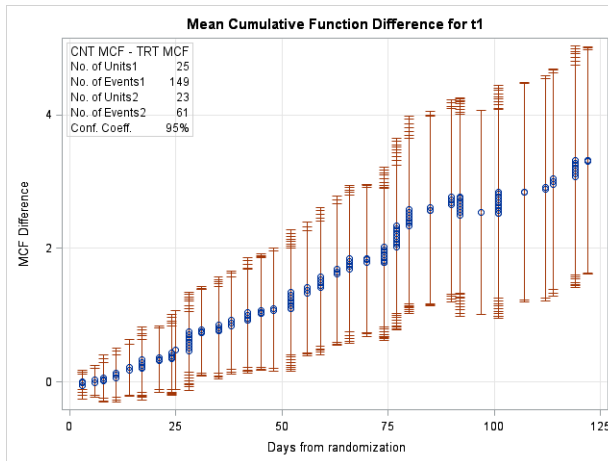
14
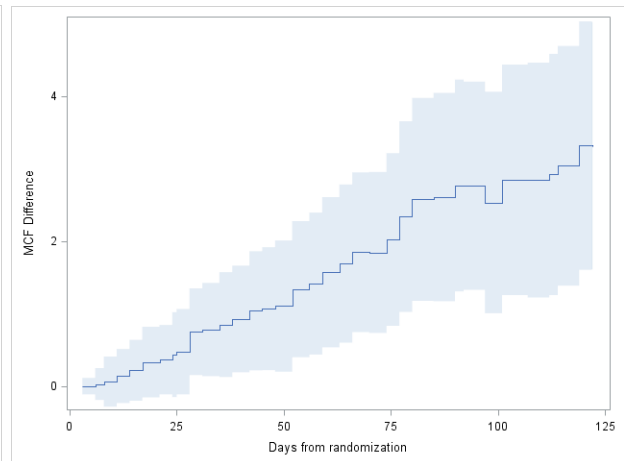
Analyses using PROC SEVERITY

We return to the NHPP and revisit the construction of the likelihood of the event times $T_{i1} < T_{i2} < T_{i3} < \ldots < T_{in_i}$ in the observation interval $[0, \tau_i]$. The conditional distribution of the gap time $W_{ij} = T_{ij} - T_{ij-1}$ given T_{ij-1} is $P[W_{ij} > w \,|\, T_{ij-1} = t] = \exp\left(-\int_t^{t+w} \rho(s)ds\right)$. For the event occurring at T_{ij} the contribution to the likelihood is $\exp\left(-\int_{T_{ij-1}}^{T_{ij}} \rho(s)ds\right)\rho(T_{ij})$. The period $(T_{in_i}, \tau_i]$ with no event contributes $\exp\left(-\int_{T_{in_i}}^{\tau_i} \rho(s \,|\, \mathbf{x}_i)ds\right)$ unless $\tau_i = T_{in_i}$, i.e., the end of observation coincides with an event time. The likelihood function for the i-th unit used by PROC RELIABILITY for NHPP analyses is a product of these terms which neatly simplifies to $\prod_{j=1}^{n_i}\{\rho(T_{ij} \,|\, \mathbf{x}_i)\}\exp\left(-\mu(\tau_i \,|\, \mathbf{x}_i)\right)$.

Go back to the interval $(T_{ij-1}, T_{ij}]$ and use the property that the event occurring at T_{ij} is left-truncated at T_{ij-1}. For the interval $(T_{in_i}, \tau_i]$, we have left truncation at T_{in_i} and perhaps right censoring at τ_i. This is precisely the situation that is handled by PROC SEVERITY where the analysis variable is a positive response (here T_{ij}) that might be both left truncated and right censored. This construction exploits structure of the likelihood (the T_{ij}'s are dependent). In Table 7 we created the left truncation time LTIME and right censoring time RCTIME. To specify left truncation, SEVERITY requires that (1) LTTIME > 0, (2) t1 > LTTIME. A missing value for LTTIME or RCTIME indicates the absence of truncation or censoring. If RCTIME=t1 then the observation is right censored. Due to ties we have several records with t1=LTTIME as well as for some end of observation records.

In order to keep all records and mimic our previous NHPP analysis we the null intervals (t0, t1] are altered very slightly t0=t0+**.0001**; t1=t1+**.00015**. See also Johnston and So (2003).

```
proc severity data=mammary_sev print=all vardef=n;
dist weibull;
loss t1/lefttruncated=lttime rightcensored=rctime;
scalemodel trt;
run;
```

Table 10: Analysis of gap times

Parameter Estimates				
Parameter	**Estimate**	**Standard Error**	**t Value**	**p-value**
Theta	21.99552	3.12289	7.04	<.0001
Gamma	1.04196	0.07190	14.49	<.0001
TRT	0.77708	0.15542	5.00	<.0001
−Log L	893.1958			

Tables 8 and 10 have differences in parameterization. The hazard is $h(t \,|\, \mathbf{x}) = \gamma t^{\gamma-1} \exp(-\gamma \mathbf{x}'\beta)$. Hence the group effect (TRT vs CNT) is $-\gamma\beta_1 = -0.8097$. Note that the **vardef=n** option is advisable to get the appropriate standard errors.

Analyses using PROC PHREG

The counting process formulation of multiple intensity models allows analysis of the proportional rates (means) model. Exogenous TVCs are allowed in the intensity $\lambda(t \mid \mathfrak{I}_{t-}) = \rho_0(t)\exp(\mathbf{x}'(t)\beta)$ but the MCF $\mu(t) = \int_0^t \rho_0(s)\exp(\mathbf{x}'(u)\beta)ds$ depends on the covariate path in [0, t]. With fixed covariates $\mathbf{x}_0$ we would get $\mu(t \mid \mathbf{x}_0) = \left(\int_0^t \rho_0(s)ds\right)\exp(\mathbf{x}_0'\beta) = \mu_0(t)\exp(\mathbf{x}_0'\beta)$. We used the same modified data set `mammary_sev` to break ties in tumor recurrence times to ensure that intervals (t0, t1) are non-null.

```
data covar;
input trt @@;
format trt trt.;
datalines;
1 0
;
run;

proc phreg data=mammary_sev covsandwich(aggregate)
           plots(overlay cl)=MCF;
id id;
class TRT(ref='CNT')/param=ref;
model (t0, t1)*status(0)= TRT/risklimits;
baseline covariates=covar out=stats_MCF MCF=_ALL_/rowid=TRT;
format trt trt.;
run;
```

The mean (hazard) ratio (TRT vs CNT) is 0.448 (95% CI: 0.305, 0.656), similar to the results from the NHPP analysis. Instead of the default MCF plot, the output data set `stats_MCF` was used to produce Figure 5.

Figure 5: Mean Cumulative Function and 95% CI

16
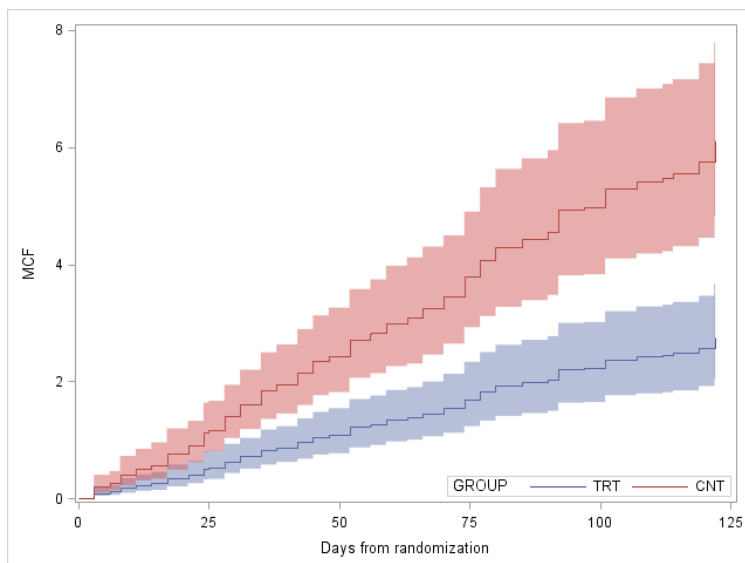
DISCUSSION

There are some common features in the two applications discussed in this paper. In our first application we demonstrated how PROC PHREG and PROC SEVERITY can be used to analyze a single duration outcome T with time-varying covariates (TVC) $\mathbf{x} = \{\mathbf{x}(t) : t \geq 0\}$. The TVC is observed and may change only at discrete time points $\{t_m : m = 0, \ldots, M\}$. The second application concerns recurrent events, multiple occurrences of the same event type, with the focus on the cumulative event count $N(t)$ and the occurrence times $\{T_k : k \geq 1\}$ but the covariate $\mathbf{x}$ was fixed. In both examples, right censoring signals end of observation. We showed how PROC RELIABILITY and PROC SEVERITY can be used for analysis of recurrence data. In both applications a structural feature of left truncation and right censoring that arises from likelihood construction permits the use of PROC SEVERITY in analyses. PROC QLIM can address right censored or truncated data, not both in the same application, and assumes normality of the response. PROC LIFEREG is the workhorse for analysis of the AFT model handling all types of censoring (left, right, or interval, and combinations thereof) within the same application. However, it does not permit truncation.

The handling of TVC can be problematic depending upon their type. Endogenous covariates that affect the evolution of the underlying event process are more difficult to deal with theoretically (Aalen *et al*, 2008). In our first application the single duration was the time from marriage to divorce. Approximately 31% of the 3371 marriages in our data set ended in divorce. Right censoring of the duration was due to death of a spouse, or if the couple were still married at their last interview. The counting process style of data input was applied to update the TVC, $x(t)$=NUMKIDS at birth of children during the marriage. With the tenuous assumption that this covariate was strictly exogenous, the semi-parametric Cox regression model was applied to assess the impact of covariates on the hazard function. As described by Lillard (1993) endogeneity of $x(t)$ could be addressed through a joint model with hazard functions for marriage dissolution and for child bearing. Another analytic strategy is to consider a state transition model for $x(t)$ with forward transitions $0 \rightarrow 1 \rightarrow 2 \rightarrow \ldots$, with exit to the terminal state (divorce) from any state (Hougaard, 2000). Such multistate modelling can be dealt with in PROC PHREG (Gardiner *et al*, 2008, Andersen and Keiding, 2002).

The parametric likelihood for $(T, \mathbf{x})$ illuminates how the exogeneity assumption enters into its construction. In addition we see how left truncation and right censoring appear in the likelihood terms. For example, for time points t_1, t_2 with $t_1 < t_2$,

$$P[T > t_2 \mid \mathbf{x}(t_m), 0 \leq m \leq M] = P[T > t_2 \mid T > t_1, \mathbf{x}(t_m), 0 \leq m \leq M] P[T > t_1 \mid \mathbf{x}(t_m), 0 \leq m \leq M]$$

The first term on the right hand side is $P[T > t_2 \mid T > t_1, \mathbf{x}(t_1)]$ which incorporates the exogeneity and shows the left truncation at t_1. Both PROC PHREG and PROC SEVERITY can handle left truncation and right censoring. In the latter it is made explicit through the rightcensored= and lefttruncated= options in the LOSS statement. In the former it is implicit in the counting process style input.

In our second application on analysis of recurrent events, we focus on the intensity function $\lambda(t) = \lim_{\Delta t \to 0} P[\Delta N(t) = 1 \mid \mathfrak{I}_{t-}] / \Delta t$ for event occurrence. Different modelling strategies are needed depending on what is contained in the model for $\lambda(t)$ from the past history $\mathfrak{I}_{t-}$. If the past influences the intensity only though exogenous covariates, we can again use PROC PHREG for analysis of the proportional means (rates) models $\lambda(t \mid \mathbf{x}) = \rho_0(t) \exp(\mathbf{x}'(t)\beta)$. With fixed covariates $\mathbf{x}_0$ the MCF $\mu(t \mid \mathbf{x}_0) = \left(\int_0^t \rho_0(s) ds \right) \exp(\mathbf{x}_0'\beta) = \mu_0(t) \exp(\mathbf{x}_0'\beta)$ is estimated.

The same nonparametric generalized Aalen estimator $\mu(t \mid \mathbf{x}_0)$ is obtained in PROC RELIABILITY for discrete $\mathbf{x}_0$. The application to tumor recurrences has a single covariate at two levels, TRT and CNT. In this case a nonparametric test for equality of the two MCFs is produced. The strength of PROC RELIABILITY is in its capabilities with the NHPP in which case the baseline $\rho_0(t)$ is parametrically specified. There are options to specify $\rho_0(t)$.

We must mention that TVCs are also allowed in the NHPP in a manner analogous to the description made previously for covariate values $\{\mathbf{x}(t_m) : m = 0, \ldots, M\}$ at the event times. The contribution to the likelihood for the event at T_{ij} is $\left\{ \rho(T_{ij} \mid \mathbf{x}(T_{ij-1})) \right\} \exp\left(-\int_{T_{ij-1}}^{T_{ij}} \rho(s \mid \mathbf{x}(T_{ij-1})) ds \right)$ and for no event in $(T_{in_i}, \tau_i]$ is $\exp\left(-\int_{T_{in_i}}^{\tau_i} \rho(s \mid \mathbf{x}(T_{in_i})) ds \right)$. Likelihood construction exploits the independent increments property of the Poisson process.

Situations where the intensity $\lambda(t)$ depends on the past history of the event process $\mathbf{N} = \{N(t) : t \geq 0\}$ are not directly handled by the software. For example the negative binomial model with fixed covariates has intensity $\lambda(t \mid \mathbf{x}) = \rho(t \mid \mathbf{x}) \left(\dfrac{1 + \phi N(t-)}{1 + \phi \mu(t \mid \mathbf{x})} \right)$ with say, $\rho(t \mid \mathbf{x}) = \rho_0(t) \exp(\mathbf{x}'\beta)$. It can be derived from a random effects Poisson model. Cook and Lawless (2007) provide parametric and semi-parametric analyses. See also Kalbfleisch and Prentice (2002) for an example of a recurrent events model for the intensity that captures some features of the past of the process $\mathbf{N}$.

REFERENCES

Aalen O, Borgan O, Gjessing H. *Survival and Event History Analysis.* New York, NY: Springer-Verlag; 2008.
Andersen PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research.* 2002;11(2):91-115.
Cook RJ, Lawless JF. *The Statistical Analysis of Recurrent Events.* New York, NY: Springer-Verlag; 2007.
Gardiner JC, Luo Z, Ramamoorthi RV. Fitting Heavy-tailed Distributions to Healthcare Data by Parametric and Bayesian Methods. *Journal of Statistical Theory and Practice.* 2014. In press.

Gardiner JC. Survival Analysis: Overview of Parametric, Nonparametric and Semiparametric approaches and New Developments, Paper 252-2010. SAS Global Forum 2010; Seattle, WA.

Gardiner JC, Luo Z, Liu L. Analysis of Multiple Failure Times Using SAS Software. In: Khattree R, Naik D, eds. *Computational Methods in Biomedical Research*. Boca Raton, FL: Chapman & Hall/CRC; 2008.

Hougaard P. *Analysis of Multivariate Survival Data*. New York, NY: Springer-Verlag; 2000.

Johnston GJ, So Y. Analysis of data from recurrent events. Paper 273-28. SAS User's Group International, 2003.

Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data, 2nd Edition*. Hoboken, NJ: John Wiley & Sons; 2002.

Lawless JF. *Statistical Models and Methods for Lifetime Data, 2nd Edition*. Hoboken, NJ: John Wiley & Sons; 2003

Lillard LA. Simultaneous-equations for hazards - marriage duration and fertility timing. *Journal of Econometrics*. 1993;56(1-2):189-217.

Lillard LA, Panis CWA. Marital status and mortality: The role of health. *Demography*. 1996;33(3):313-327.

Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modelling using Stata, Volume II: Categorical Responses, Counts, and Survival, 3rd Edition*. College Station, TX: Stata Press; 2012.

Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data, 2nd Edition*. Cambridge, MA: MIT Press; 2010.

ACKNOWLEDGMENTS

CONTACT INFORMATION

We welcome your comments and questions. Please contact

Joseph C. Gardiner
Division of Biostatistics
Department of Epidemiology and Biostatistics
B629 West Fee Hall
Michigan State University
East Lansing, MI 48824
jgardiner@epi.msu.edu