

Use of SAS ® PROCS in Survival Analysis

Lida Gharibvand, University of California, Riverside

ABSTRACT

Survival analysis involves the modeling of time-to-event data whereby death or failure is considered an "event". The graphic presentation of survival analysis is a significant tool which facilitates clear understanding of the underlying events. Two SAS ® PROCS, LIFETEST and PHREG can generate some of the survival analysis plots using the ODS graphics option in version 9.1.3. In this paper, we will demonstrate the features of estimated hazard function, survival function, and cumulative martingale residual plots using ODS Graphics. In clinical trials, identifying potential outlier individuals who 'died far too early' or 'lived far too long' as compared to what the fitted model predicts. The cumulative residuals from PROC PHREG are used to investigate functional form error of covariate and validates of the proportion hazard function.

INTRODUCTION TO SURVIVAL ANALYSIS

Survival analysis is the phrase used to describe the analysis of data in the form of times from a well-defined "time origin" until the occurrence of some particular event or "end-point". In medical research, the time origin will often correspond to the recruitment of an individual into an experimental study, such as a clinical trial to compare two or more treatments. This in turn may coincide with the diagnosis of a particular condition, the commencement of a treatment regimen, or the occurrence of some adverse event. If the end point is the death of a patient, the resulting data are literally survival times. However, data of a similar form can be obtained when the end-point is not fatal, such as the relief of pain, or the recurrence of symptoms. In this case, the observations are often referred to as *time to event* data.

The analysis of survival data requires special techniques because the data are almost always incomplete, and familiar parametric assumptions may be unjustifiable. Investigators follow subjects until they reach a pre-specified endpoint (for example, death). However, subjects sometimes withdraw from a study, or the study is completed before the endpoint is reached. In these cases, the survival times (also known as failure times) are *censored*; subjects survived to a certain time beyond which their status is unknown. The uncensored survival times are sometimes referred to as *event* times. Methods for survival analysis must account for both censored and uncensored data. In SAS 9.1, a number of SAS/STAT ® procedures have been modified to use an experimental extension to the Output Delivery System (ODS) that enables them to create statistical graphics as automatically as tables. This facility is referred to as ODS Statistical Graphics (or ODS Graphics for short), and it is invoked when you provide the experimental ODS GRAPHICS statement prior to your procedure statements. Any procedures that use ODS Graphics then create graphics, either by default or when you specify procedure options for requesting specific graphs.

ODS SAS STATISTICAL GRAPHICS

ODS graphics is experimental in SAS 9.1, and is expected to be production in SAS 9.2.1 (SAS Institute, Inc. (2005)). With ODS Graphics, a procedure creates the graphs that are most commonly needed for a particular analysis. Using ODS Graphics eliminates the need to save numerical results in an output data set, manipulate them with a DATA step program, and display them with a graphics procedure. One requirement for using ODS GRAPHICS is that ODS output files must be saved in RTF or HTML format.

SAS PROCS:

There are three SAS procedures available for analyzing survival data: LIFEREG, LIFETEST and PHREG (BPHREG). PROC LIFEREG is a parametric regression procedure for modeling the distribution of survival time with a set of concomitant variables (SAS Institute, Inc. (2007a)). PROC LIFETEST is a nonparametric procedure for estimating the survivor function, comparing the underlying survival curves of two or more samples, and testing the association of survival time with other variables (SAS Institute, Inc. (2007b)). PROC PHREG is a semi-parametric procedure that fits the Cox proportional hazards model (SAS Institute, Inc. (2007c)). PROC BPHREG is an experimental upgrade to PHREG procedure that can be used to fit Bayesian Cox proportional hazards model (SAS Institute, Inc. (2007d)).

SURVIVAL DATA USED

The Mayo liver disease example of Lin, Wei, and Ying (1993) is used here to demonstrate the features. The data consist of 418 patients with primary Biliary cirrhosis (PBC), among which 161 had died as of the date of data listing. The data set contains the following variables:

Time follow-up time in years
 Censor event indicator with value 1 for death time and value 0 for censored time
 Age age in years from birth to study registration
 Alb serum albumin level in gm/dl
 Bili serum Bilirubin level in mg/dl
 Edema Edema with value 0 for presence of no Edema, Edema with value 0.5 for untreated or successfully treated, and Edema with value 1 for unsuccessfully treated Edema
 Prottime prothrombin time in seconds

PROC LIFEREG

PROC LIFEREG is a parametric regression procedure for modeling the distribution of survival or failure time data that can be right-, left-, or interval-censored with a set of concomitant variables. The models for the response variable consist of a linear effect composed of the covariates and a random disturbance term. The distribution of the random disturbance can be taken from a class of distributions that includes the extreme value, normal, logistic, and, by using a log transformation, the exponential, Weibull, lognormal, loglogistic, and three-parameter gamma distributions. These models are equivalent to accelerated failure time models when the log of the response is the quantity being modeled. The accelerated failure time model assumes a parametric form for the effects of the explanatory variables and usually assumes a parametric form for the underlying survivor function that the effect of covariates on an event time distribution is multiplicative on the event time. The LIFEREG Procedure can be also used to perform a Tobit analysis, a regression model for left-censored data assuming a normally distributed error term. For more information on LIFEREG refer the SAS Institute on-line documentation (SAS Institute, Inc. (2007a))

There is no ODS Graphics feature available in PROC LIFEREG (version 9.1.3). However we can generate the survival probability plot using the PROBLOT option. The following example demonstrates how you can use the LIFEREG procedure to fit a parametric model to failure time data. Consider fitting the survival time of the PBC patients with covariates Bili, log(Prottime), log(Alb), Age and Edema. The log transform, which is often applied to blood chemistry measurements, is deliberately not employed for Bili. It is of interest to assess the functional form of the variable Bili in the failure time data model.

```
ODS RTF FILE='path\filename.rtf' style=statistical;
ODS LISTING CLOSE;
DATA pbc;
GOPTIONS RESET=all COLORS=(Black, RED,BLUE,YELLOW,GREEN,MAGENTA,CYAN)
dev=EMF target=EMF XMAX=7 YMAX=7 HTEXT=14pt FTEXT="<ttf> Arial";
PROC LIFEREG DATA=pbc;
CLASS Edema;
MODEL time*Censor(0)= Age logAlb logProttime Bili Edema;
PROBPLOT Cencolor=red cframe=ligr cfit=blue ppout npintervals=simul;
INSET / cfill = white ctext = blue; RUN;
```

The PROC LIFEREG statement invokes the procedure. The MODEL statement is required and specifies the variables used in the regression part of the model as well as the distribution used for the error, or random, component of the model (The default distribution used is Weibull and this can be changed for a better fit). Only a single MODEL statement can be used with one invocation of the LIFEREG procedure. If multiple MODEL statements are present, only the last model is used. Main effects and interaction terms can be specified in the MODEL statement, similar to the GLM procedure. Initial values can be specified in the MODEL statement or in an INEST= data set. If no initial values are specified, the starting estimates are obtained by ordinary least squares. Categorical variables can be specified in CLASS statement. Survival probability plot can be generated with PROBLOT. Executing the code above will produce the following plot.

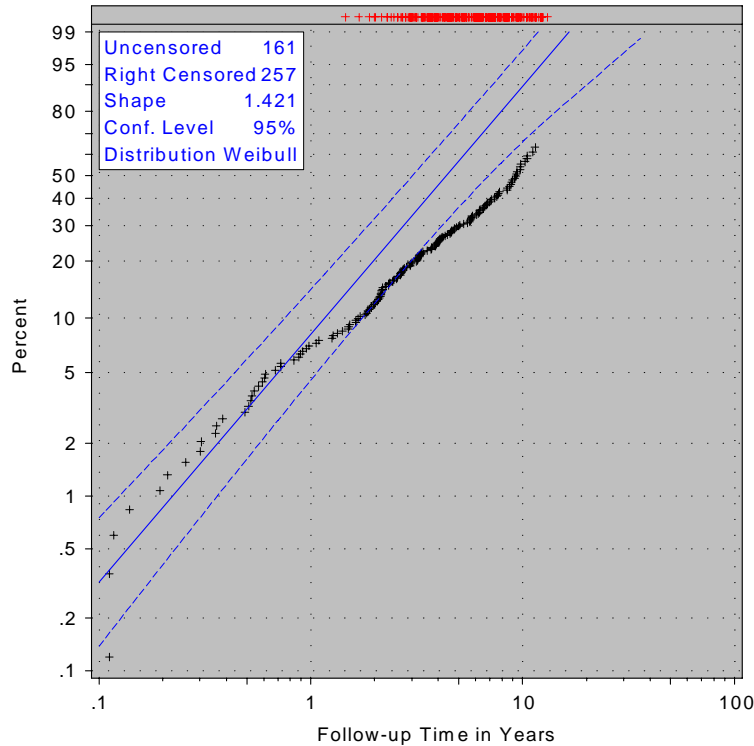


Figure 1: Survival Plot Produced by LIFEREG Procedure

The resulting graphical output is shown in Figure 1. The estimated CDF, a line representing the maximum likelihood fit, and point wise parametric confidence bands are plotted here. The values of right-censored observations are plotted along the top of the graph. Figure 1 assesses the quality of the model fit. As we can see, the model fit is not very satisfactory. We will show how to improve the model fit in this paper.

PROC LIFETEST

PROC LIFETEST is a nonparametric procedure for estimating the survivor function either by the product-limit method (also called the Kaplan-Meier method) or by the life-table method, comparing the underlying survival curves of two or more samples, and testing the association of survival time with other variables. PROC LIFETEST provides non-parametric k-sample tests based on weighted comparisons of the estimated hazard rate of the individual population under the null and alternative hypotheses. This proc also computes rank tests and a likelihood ratio test for testing homogeneity of survival functions across strata. The ODS GRAPHICS features are available in PROC LIFETEST.

```

ODS RTF FILE='path\filename.rtf' style=statistical;
ODS LISTING CLOSE;
ODS GRAPHICS on;
DATA pbc;
PROC LIFETEST DATA=pbc;
TIME Time*Censor(0); SURVIVAL out=Out1 confband=all bandmin=100 bandmax=600
MAXTIME=800 conftype=asinsqrt PLOTS=(stratum, survival, hwb);
STRATA Edema;
TEST Age logAlb logProttime Bili; RUN;
ODS GRAPHICS OFF;
ODS RTF CLOSE;
ODS LISTING;
QUIT;

```

ODS GRAPHICS ON will add statistical graphics to the output. The simplest use of PROC LIFETEST is to request the nonparametric estimates of the survivor function for a sample of survival times. In such a case, only the PROC LIFETEST statement and the TIME statement are required. All statements except the TIME statement are optional, and there is no required order for the statements following the PROC LIFETEST statement. The PROC LIFETEST procedure generates survival graphs from the PLOTS = (s) option. The

legend and long-rank p-value annotations are automatically generated when the STRATA statement is used. To disable the graphics after execution, enter the statement ODS GRAPHICS OFF.

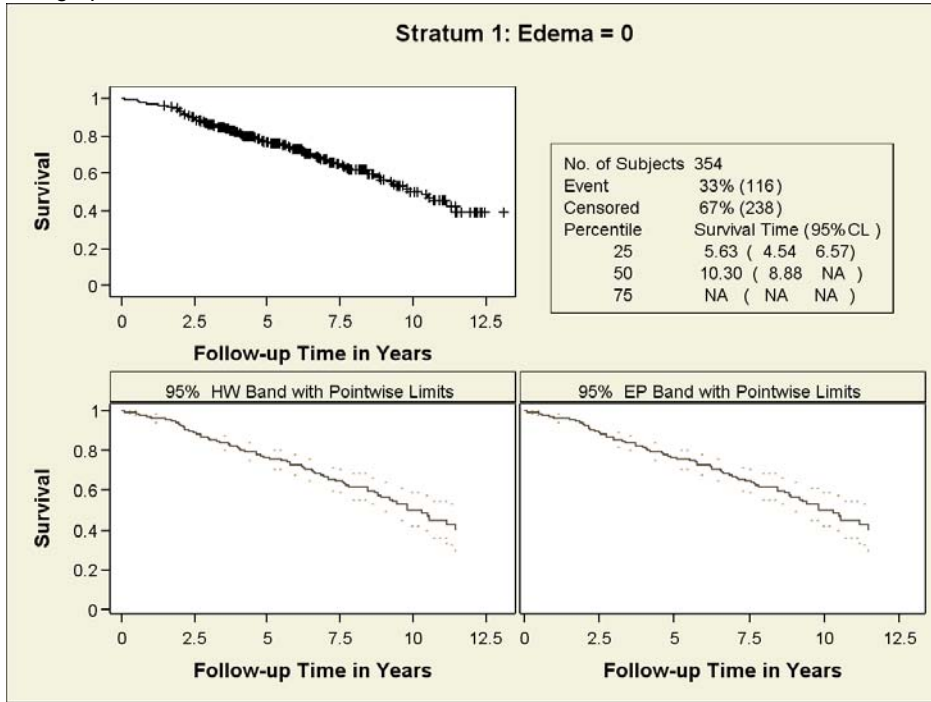


Figure 2: Panel Plot for Patients with Edema=0 (Experimental)

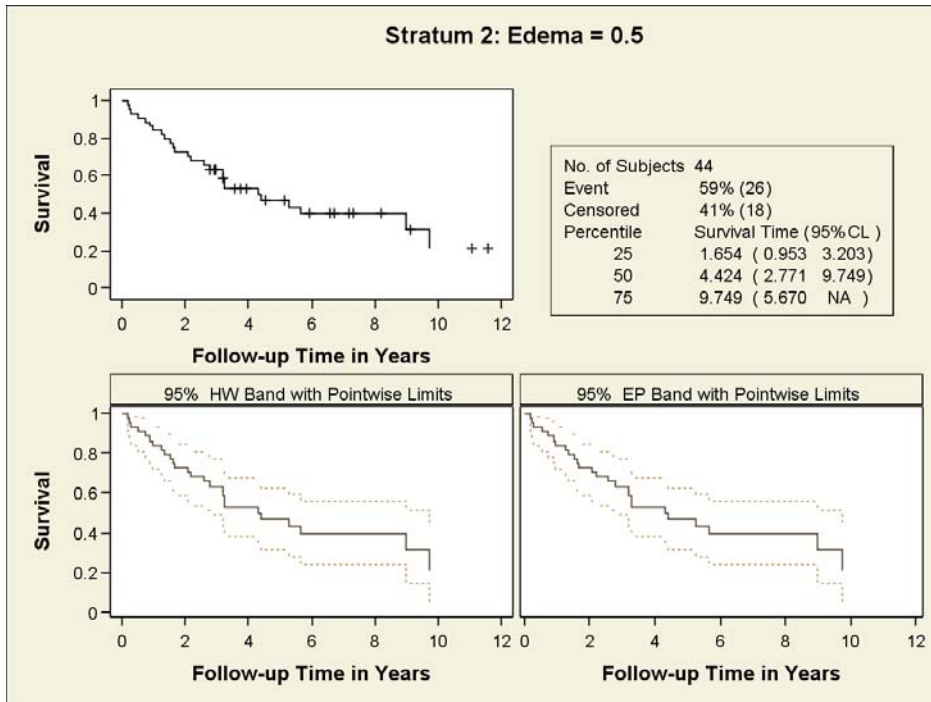


Figure 3: Panel Plot for Patients with Edema=0.5 (Experimental)

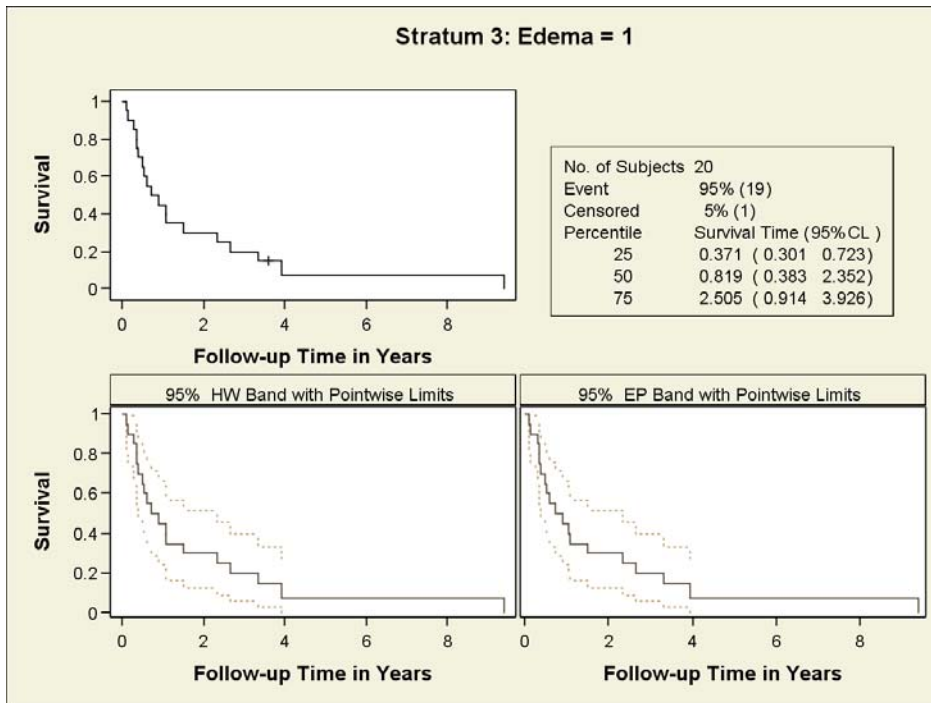


Figure 4: Panel Plot for Patients with Edema=1 (Experimental)

The graphical displays in Figures 2, 3, 4 are requested by specifying the experimental ODS GRAPHICS statement and the experimental PLOTS= option in the SURVIVAL statement. Figure 2 shows Panel Plot for patients with Edema=0 which is the control group. As we can observe, only 33% of patients had events and the median is about 10.30. Figure 4 shows a summary for patients with Edema=1 which is the worse case. In this case the number of events is 95% and the median is 0.819<1.

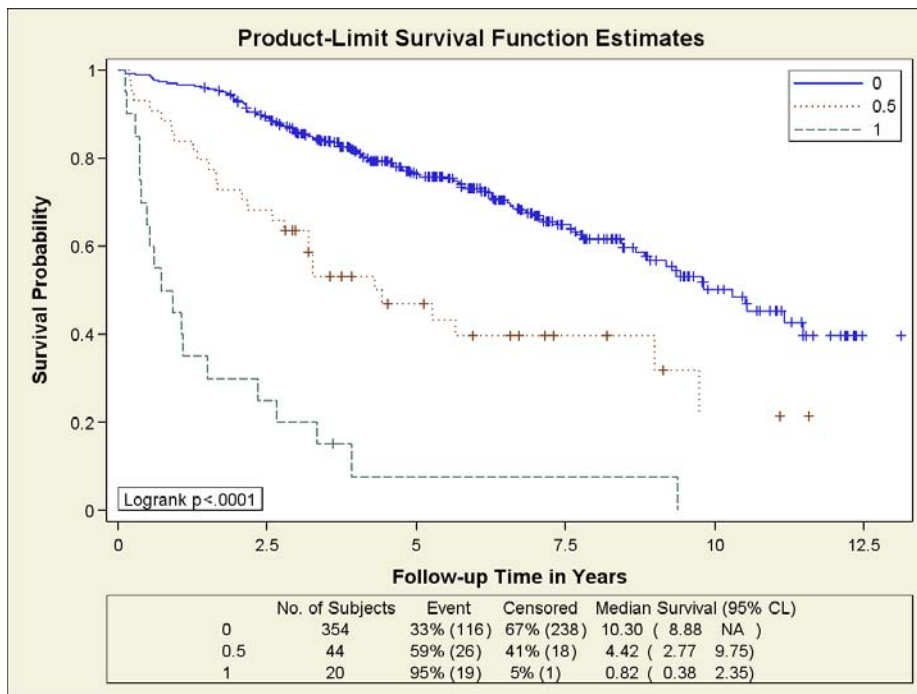


Figure 5: Product-Limit Estimates of Survival of PBC Patients (Experimental)

Figure 5 shows a summary of the estimated survival curves for the three edema groups. Censored observations are plotted as a plus sign. It appears that patients with Edema=0 have the best prognosis while patients with Edema=1 have the worse prognosis, while most of patients showed the event in less than a year. For more information please see SAS ONLINE documentation for SAS/STAT PROC LIFETEST (REF)

PROC PHREG

PROC PHREG is a semi-parametric procedure that fits the Cox proportional hazards model. Cox's semi-parametric model is widely used in survival analysis to model the effect of covariates on hazard rates. Cox's proportional hazards model assumes a parametric form for the effects of the covariates, but it allows an unspecified form for the underlying survivor function. The partial likelihood of Cox model also allows time-dependent covariates. A covariate is time-dependent if its value for any given individual can change over time. The validity of the proportional hazards model can be tested by testing for interaction between time-dependent covariates and the response time.

There is only one ODS GRAPHICS PLOT, cumulative residual plot for detecting model specification error is available in Version 9.1.3. As Lin et al. (1993) suggests, a Cox model for the survival time of the PCB patients with covariates Bili, log(Prottime), log(Albumin), Age and Edema was fitted using the following code.

```
ODS RTF FILE='path\filename.rtf' style=statistical;
ODS LISTING CLOSE;
ODS GRAPHICS on;
DATA pbc;
PROC PHREG DATA=pbc;
MODEL Time*Censor(0)= Bili logAlb logProttime Edema Age;
logProttime=log(Prottime); logAlb=log(Alb);
ASSESS VAR=(Bili) / RESAMPLE; RUN;
ODS GRAPHICS OFF;
ODS RTF CLOSE;
ODS LISTING;
QUIT;
```

The ASSESS statement creates a plot of the cumulative martingale residuals against the values of the covariate Bili, which is specified in the VAR= option. The RESAMPLE option computes the p-value of a Kolmogorov-type supremum test based on a sample of 1,000 simulated residual patterns.

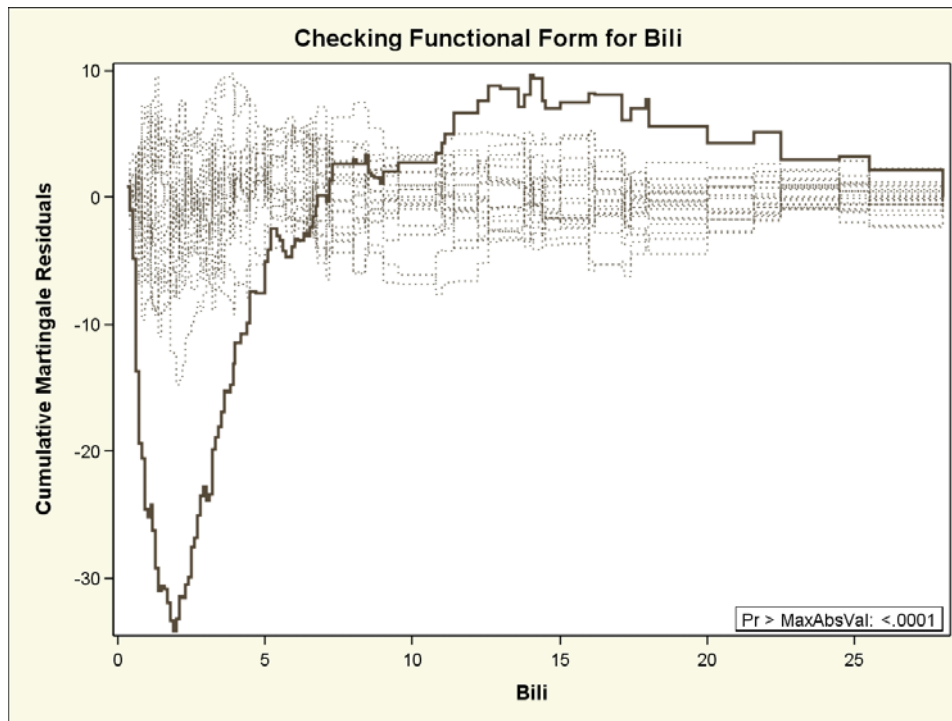


Figure 6: Cumulative Martingale Residuals vs Bili (Experimental)

The plot in Figure 6 displays the observed cumulative martingale residual process for Bili together with 20 simulated realizations from the null distribution. This graphical display is requested by specifying the experimental ODS GRAPHICS statement and the experimental ASSESS statement.

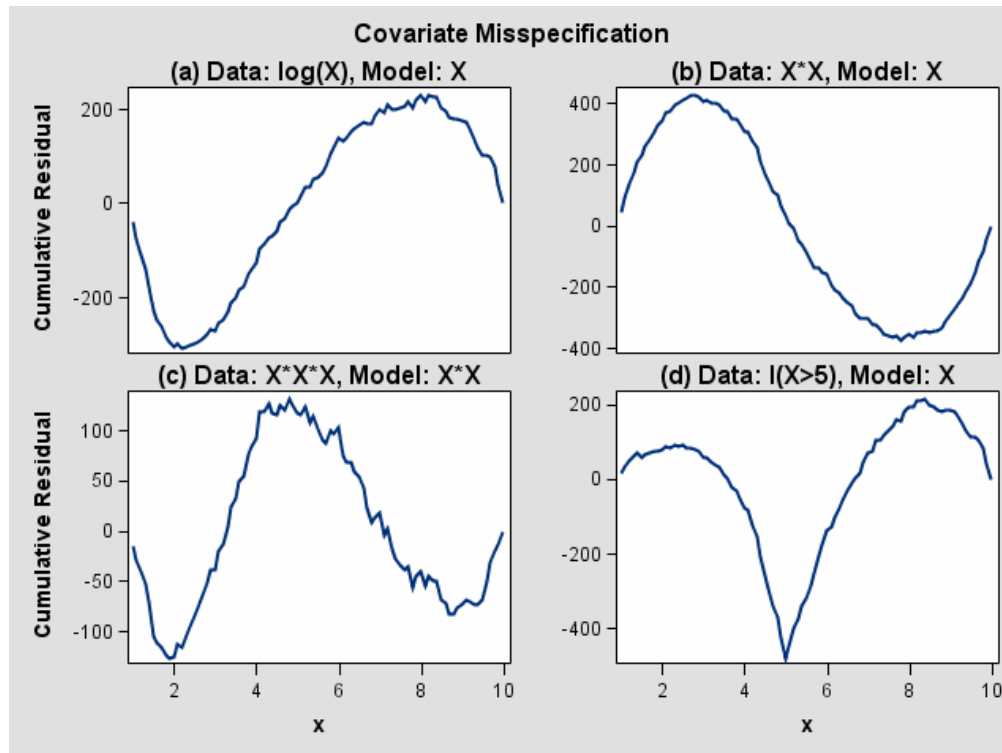


Figure 7: Typical Cumulative Residual Plot Patterns

The curve of observed cumulative martingale residuals in Figure 6 most resembles the behavior of the curve in Figure 7(a), indicating that $\log(\text{Bili})$ might be a more appropriate term in the model than Bili. Next, the analysis of the natural history of the PBC is repeated with $\log(\text{Bilir})$ replacing Bili, and the functional form of $\log(\text{Bilir})$ is assessed (Figure 8). When we compare Figures 6 and 8, we can see that Figure 8 shows a better function form when $\log(\text{Bilir})$ was used. Therefore, the model fit is improved when Bili is replaced by $\log(\text{Bilir})$. Figure 9 displays the results of proportional hazards assumption assessment for $\log(\text{bili})$ and this plot reveals non proportional hazards for $\log(\text{Bili})$ and the proportional hazards assumption appears to be violated for $\log(\text{bili})$. When using PROC PHREG, the experimental ASSESS statement failed to produce the cumulative residual plots in the presence of STRATA variable in Version 9.1.3.

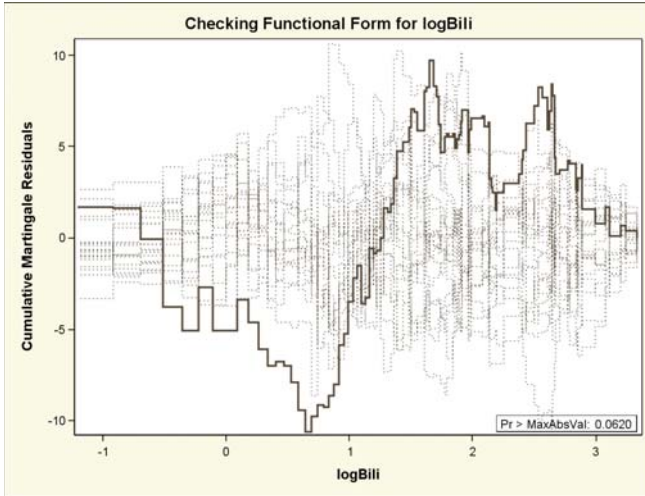


Figure 8: Cumulative Martingale Residuals vs log(Bili)

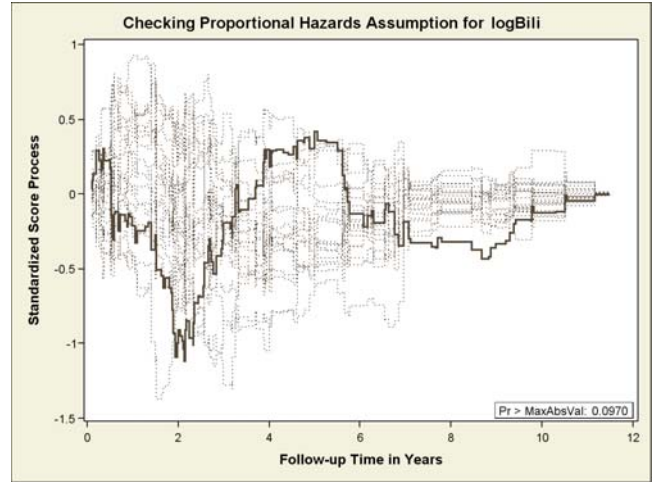


Figure 9: Standardized Score Process for log(Bili)

REFERENCES

Lin, D., Wei, L., and Ying, Z. (1993), "Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals," *Biometrika*, 80, 557 - 572

SAS Institute, Inc. (2005), "TEMPLATE Procedure: Creating ODS Statistical Graphics Output (Experimental)" <http://support.sas.com/rnd/base/topics/statgraph/proctemplate/a002774500.htm>

SAS Institute, Inc. (2007a), SAS/STAT ® User's Guide, SAS OnlineDoc® 9.1.3, Cary, NC: SAS Institute, Inc.

LIFEREG: http://support.sas.com/91doc/getDoc/statug.hlp/lifereg_index.htm

SAS Institute, Inc. (2007b), SAS/STAT ® User's Guide, SAS OnlineDoc® 9.1.3, Cary, NC: SAS Institute, Inc.

LIFETEST: http://support.sas.com/91doc/getDoc/statug.hlp/lifetest_index.htm

SAS Institute, Inc. (2007c), SAS/STAT ® User's Guide, SAS OnlineDoc® 9.1.3, Cary, NC: SAS Institute, Inc.

PHREG: http://support.sas.com/91doc/getDoc/statug.hlp/phreg_index.htm

SAS Institute, Inc. (2007d), SAS/STAT ® User's Guide, SAS OnlineDoc® 9.1.3, Cary, NC: SAS Institute, Inc.

BPHREG: <http://support.sas.com/rnd/app/papers/bayesian.pdf>

Therneau T.M, Grambsch P.M and Fleming. T.R (1990), "Martingale-based residuals for survival models". *Biometrika* 1990 77(1):147-160

CONTACT INFORMATION

This paper is based on a related paper presented at WUSS 2007. Your comments are greatly appreciated and encouraged. Contact the author at:

Lida Gharibvand

University of California, Riverside

Work Phone: (949) 230-5439

Email: lida.gharibvand@email.ucr.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.