

Evaluating Predictive Accuracy of Survival Models with PROC PHREG

Changbin Guo, Ying So, and Woosung Jang, SAS Institute Inc.

Abstract

Model validation is an important step in the model building process because it provides opportunities to assess the reliability of models before their deployment. Predictive accuracy measures the ability of the models to predict future risks, and significant developments have been made in recent years in the evaluation of survival models. SAS/STAT® 14.2 includes updates to the PHREG procedure with a variety of techniques to calculate overall concordance statistics and time-dependent receiver operator characteristic (ROC) curves for right-censored data. This paper describes how to use these criteria to validate and compare fitted survival models and presents examples to illustrate these applications.

Introduction

There has been growing interest in predictive modeling for various applications. A critical task in the model building process is accessing the model's predictive capability systematically. Two important aspects of a prediction model are calibration and discrimination. Calibration refers to the ability of the model to correctly rank the individuals in the sample by risk. Discrimination characterizes the model's ability to correctly classify subjects for their actual outcomes. There are a variety of methodologies to assess the discriminative performance of a prediction model.

The concordance statistic (or C-statistic) is the most commonly used discrimination measure in the context of logistic regression with binary outcomes. The concept underlying concordance is that a subject who experiences a particular outcome has a higher predicted probability of that outcome than a subject who does not experience the outcome. The C-statistic can be calculated as the proportion of pairs of subjects whose observed and predicted outcomes agree (are concordant) among all possible pairs in which one subject experiences the outcome of interest and the other subject does not. The higher the C-statistic, the better the model can discriminate between subjects who experience the outcome of interest and subjects who do not. In the context of survival analysis, various C-statistics have been formulated to deal with right-censored data (Harrell 1986; Uno et al. 2011).

Besides the C-statistic, receiver operator characteristic (ROC) curves and AUC (area under the ROC curve) statistics are also commonly used to assess the discrimination ability of the model with binary outcomes. For survival models with time-to-event outcomes, ROC curves are computed at specific time points. Various definitions and estimators of time-dependent ROC curves and AUC functions have been proposed in the survival setting. Blanche, Latouche, and Viallon (2013) provide a comprehensive survey of the available methods. Time-dependent ROC curves and AUC functions characterize how well the fitted model can distinguish between subjects who experience an event and subjects who do not. Whereas C-statistics provide overall measures of predictive accuracy, time-dependent ROC curves and AUC functions summarize the predictive accuracy at specific times. In practice, it is common to use several time points within the support of the observed event times.

The LOGISTIC procedure in SAS/STAT software fits logistic regression models for binary outcomes and offers a variety of functionality for performing analyses by using C-statistic and ROC curves.

The PHREG procedure in SAS/STAT 14.2 provides a number of state-of-the-art techniques to calculate overall concordance statistics and time-dependent ROC curves and AUC statistics for right-censored data. This new functionality becomes available with the introduction of the CONCORDANCE and ROCOPTIONS options in the PROC PHREG statement as well as the new ROC statement. This paper reviews the existing features in PROC LOGISTIC for C-statistic and ROC curves, presents the new features in PROC PHREG, and illustrates their applications in examples. Key differences between PROC PHREG and PROC LOGISTIC are also examined.

ROC Analysis in PROC LOGISTIC: A Review

Assume that the binary response, E , of an individual or an experimental unit can take one of two possible values, denoted for convenience by 0 and 1. Suppose \mathbf{Z} is a vector of explanatory variables and $\pi = \Pr(E = 1 \mid \mathbf{Z})$ is the response probability to be modeled. The logistic model has the form

$$\text{logit}(\pi) \equiv \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta' \mathbf{Z}$$

where α is the intercept parameter and $\beta = (\beta_1, \dots, \beta_s)'$ is the vector of s slope parameters.

The LOGISTIC procedure fits the model by the method of maximum likelihood.

Receiver Operating Characteristic (ROC) Curves

Receiver operating characteristic (ROC) curves display the discrimination potential of fitted logistic models by evaluating the trade-offs between true positive rate (sensitivity) and false positive rate ($1 - \text{specificity}$). If you have a cutoff value, you can classify the subjects as being positive or negative according to their predicted probabilities. Hence a 2×2 classification table can be constructed to show the relationship between the predicted outcomes and the actual outcomes. Sensitivity is the fraction of positive subjects who are predicted as positive, and specificity is the fraction of negative subjects who are correctly predicted as negative.

The ROC curve plots the sensitivity against one minus the specificity for a series of cutoffs for the fitted probability. The ROC plot is a unit square plot, and the higher the curve rises above the 45-degree line, the more desirable it is. The 45-degree line corresponds to an area under the curve (AUC) of 0.5 and represents where the fractions of true positives and false negatives are equal. PROC LOGISTIC implements a nonparametric test to compare correlated ROC curves that was proposed by DeLong, DeLong, and Clarke-Pearson (1988). You can construct linear contrasts to perform comparisons by using the empirical ROC curves of specified models.

The ROC methodology has become a standard tool for assessing predictive accuracy because it provides a comprehensive evaluation of a fitted model. In practice, it is sometimes more convenient to use the so-called area under the ROC curve (AUC), which summarizes the entire curve. AUC is connected to a variety of well-known statistical measures; the concordance statistic (C-statistic) is the most obvious one, because the two measures are equivalent. As an estimator of the probability of concordance, the C-statistic estimates the concordant probability in a randomly selected pair of subjects. Concordance is the phenomenon that if one subject of the pair experiences the positive outcome and the other does not, the subject with the positive outcome also has a higher predicted probability of the positive outcome than the subject with the negative outcome. This provides the AUC with a convenient interpretation in terms of concordance probability.

Example: Using PROC LOGISTIC for ROC Analysis

This example illustrates the features of the C-statistic and ROC curves in PROC LOGISTIC.

The data set **Liver** consists of data about 418 patients who have primary biliary cirrhosis (PBC), an autoimmune disease of the liver. Each observation represents a patient. The variable **Time** represents the follow-up time in years (the time from registration to liver transplantation, death, or study termination, whichever comes first); the variable **Status** is the censoring indicator (1 for death, 0 for censored); and the explanatory variables are **Age** (age, in years), **Albumin** (serum albumin level, in g/dl), **Bilirubin** (serum bilirubin level, in mg/dl), **Edema** (presence of edema, or swelling), and **Protime** (prothrombin time, in seconds).

```
data Liver;
  input Time Status Age Albumin Bilirubin Edema Protime @@;
  label Time="Follow-Up Time in Years";
  Time= Time / 365.25;
  Outcome = (Time > 5);
  datalines;
400 1 58.7652 2.60 14.5 1.0 12.2 4500 0 56.4463 4.14 1.1 0.0 10.6
1012 1 70.0726 3.48 1.4 0.5 12.0 1925 1 54.7406 2.54 1.8 0.5 10.3
1504 0 38.1054 3.53 3.4 0.0 10.9 2503 1 66.2587 3.98 0.8 0.0 11.0
1832 0 55.5346 4.09 1.0 0.0 9.7 2466 1 53.0568 4.00 0.3 0.0 11.0
2400 1 42.5079 3.08 3.2 0.0 11.0 51 1 70.5599 2.74 12.6 1.0 11.5
3762 1 53.7139 4.16 1.4 0.0 12.0 304 1 59.1376 3.52 3.6 0.0 13.6
3577 0 45.6893 3.85 0.7 0.0 10.6 1217 1 56.2218 2.27 0.8 1.0 11.0
```

```

... more lines ...

989 0 35.0000 3.23 0.7 0.0 10.8 681 1 67.0000 2.96 1.2 0.0 10.9
1103 0 39.0000 3.83 0.9 0.0 11.2 1055 0 57.0000 3.42 1.6 0.0 9.9
691 0 58.0000 3.75 0.8 0.0 10.4 976 0 53.0000 3.29 0.7 0.0 10.6
;

```

The response variable, **Outcome**, is defined according to whether the subject is still at risk at year 5. The following statements use the LOGISTIC procedure to fit the logistic regression model that uses **Bilirubin**, **Age**, and **Edema** as explanatory variables:

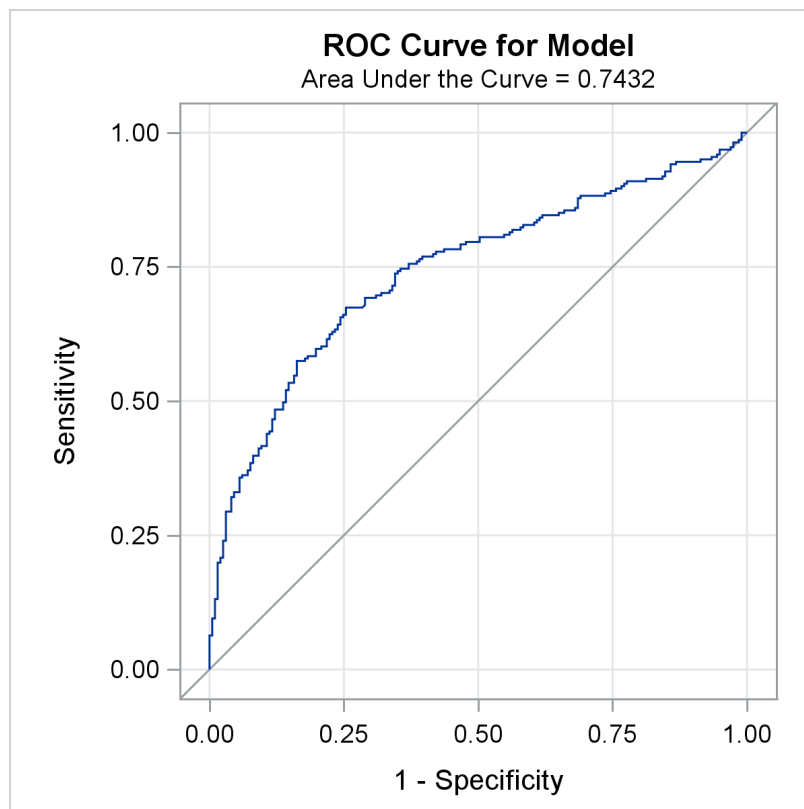
```

ods graphics on;
proc logistic data=Liver plots(only)=roc;
  model Outcome=Bilirubin Age Edema;
run;

```

Figure 1 displays the ROC curve that is produced by the PLOTS= option.

Figure 1 Receiver Operating Characteristic Curve



The AUC (area under the ROC curve) is estimated by the statistic *c* in the "Association of Predicted Probabilities and Observed Responses" table shown in Figure 2. In this example, the AUC is 0.7432.

Figure 2 Association Table

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	74.3	Somers' D	0.486
Percent Discordant	25.7	Gamma	0.486
Percent Tied	0.0	Tau-a	0.243
Pairs	43537	c	0.743

You can use PROC LOGISTIC to compare the ROC curves of different models. Consider three submodels of the previously fitted logistic model, each of which contains two of the three covariates **Bilirubin**, **Age**, and **Edema**. You can assess the difference in the ROC curves among these submodels in PROC LOGISTIC. In the following statements, three ROC statements are specified, one for each submodel:

```
proc logistic data=Liver plots=roc;
  model Outcome=Bilirubin Age Edema / nofit;
  roc 'Bilirubin+Age' Bilirubin Age;
  roc 'Age+Edema' Age Edema;
  roc 'Bilirubin+Edema' Bilirubin Edema;
  roccontrast reference('Age+Edema') / estimate e;
run;
```

The ROCCONTRAST statement implements the nonparametric test of DeLong, DeLong, and Clarke-Pearson (1988) to compare the three ROC curves, the REFERENCE option specifies that the Age+Edema curve is used as the reference curve in the contrast, the E option displays the contrast coefficients, and the ESTIMATE option computes the AUC and tests each comparison.

Figure 3 displays the ROC curves of the three submodels and their AUC statistics.

Figure 3 ROC Curves for the Three Models

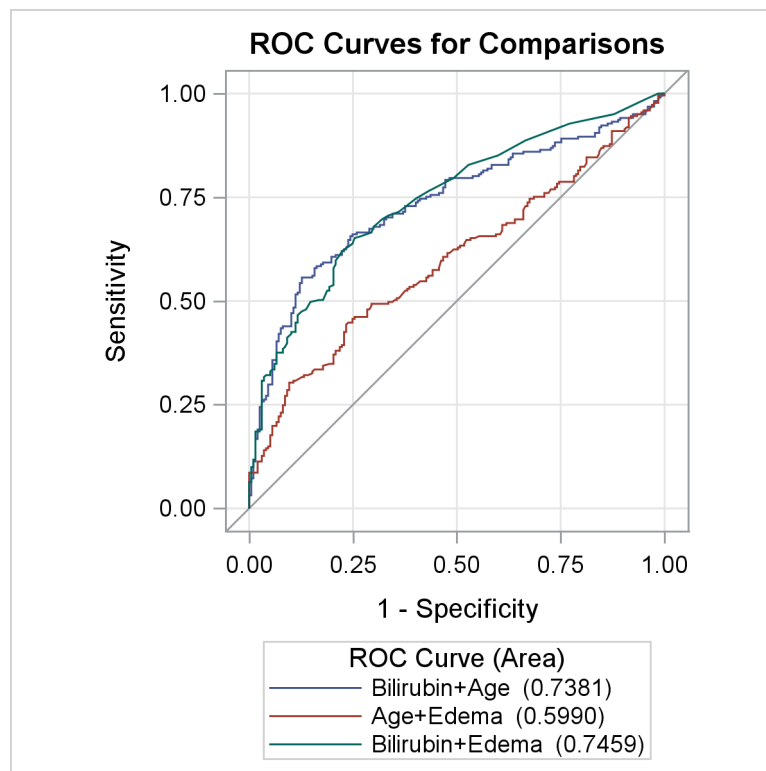


Figure 4 displays the area under the ROC curve along with its standard error and a confidence interval for each model in the comparison. None of the three confidence intervals contain 0.50, indicating that the model-based prediction is significantly different from random guessing, which is represented by the diagonal line in the ROC plots in Figure 3.

Figure 4 ROC Association Table

ROC Association Statistics							
Mann-Whitney							
ROC Model	Area	Standard Error	95% Wald Confidence Limits	Somers' D	Gamma	Tau-a	
Bilirubin+Age	0.7381	0.0245	0.6900 0.7862	0.4762	0.4762	0.2379	
Age+Edema	0.5990	0.0276	0.5449 0.6532	0.1980	0.1983	0.0989	
Bilirubin+Edema	0.7459	0.0238	0.6992 0.7926	0.4918	0.5026	0.2457	

Figure 5 shows that the contrast used 'Age+Edema' as the reference level. This table is produced by specifying the E option in the ROCCONTRAST statement.

Figure 5 ROC Contrast Coefficients

ROC Contrast Coefficients		
ROC Model	Row1	Row2
Bilirubin+Age	1	0
Age+Edema	-1	-1
Bilirubin+Edema	0	1

Figure 6 shows that the 2-degrees-of-freedom test that 'Age+Edema' is different from at least one other test is significant at the 0.05 level.

Figure 6 ROC Test Results (2 Degrees of Freedom)

ROC Contrast Test Results			
Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Age+Edema	2	22.7198	<.0001

Concordance Statistics for Survival Models

In SAS/STAT 14.2, PROC PHREG provides two versions of concordance statistics—Harrell's C-statistic and Uno's C-statistic—for dealing with right-censored data. These new statistics can be considered generalizations of the C-statistic for binary outcomes as implemented in PROC LOGISTIC. This section discusses how these new statistics are formulated and how to request them in PROC PHREG. Their differences from the conventional C-statistic and some computational issues concerning their usage are also discussed.

Two Definitions

The conventional C-statistic is defined for binary outcomes and cannot be directly applied to time-to-event data. By contrast, Harrell's and Uno's versions of the C-statistic are defined specifically for right-censored data. A fundamental difference between Harrell's method and Uno's method is how survival times are ordered in the presence of censoring. Harrell's method offers a straightforward approach by discarding the pairs that have become incomparable because of censoring. This approach, although it is simple, has the shortcoming that the estimates depend on the censoring variable. Uno's method (Uno et al. 2011) is a new formulation that models the censoring distribution and uses it to weight the uncensored observations in the estimation, making the estimates censoring-independent. For a detailed description of these methods, see the section "Estimating Concordance Statistics" in the [Appendix](#).

Syntax in PROC PHREG

The new CONCORDANCE option in the PROC PHREG statement is designed to perform analyses related to overall concordance. The default method is Harrell's C-statistic, which you can request by specifying the CONCORDANCE or CONCORDANCE=HARRELL option in the PROC PHREG statement. To compute the standard error, you can specify the CONCORDANCE=HARRELL(SE) option.

To request Uno's C-statistic, specify the CONCORDANCE=UNO option in the PROC PHREG statement. To compute the standard error, specify the CONCORDANCE=UNO(SE) option. The standard error estimator is based on a perturbation-resampling method proposed by Uno et al. (2011). You can control the perturbation process by specifying the options in [Table 1](#).

Table 1 Options for Perturbation

Option Name	Function	Default Value
ITER=	Specify number of perturbations	100
SEED=	Specify seed for random number generator	Random
DIFF	Compare C-statistics	NA

Differences from Conventional C-Statistic

The conventional C-statistic, as implemented in the LOGISTIC procedure, is for binary outcomes. From this perspective, the C-statistics that are constructed for right-censored data can be considered its generalizations. However, it is worth mentioning that the conventional C-statistic has the alternative expression as the area under the ROC curve, although this property is no longer shared by its survival counterparts. Besides the two types of C-statistics that PROC PHREG provides, there are other definitions of C-statistics for survival outcomes in the literature. Essentially, these C-statistics differ in how they account for censoring.

Speedy Computation

The definitions of C-statistics are based on the notion of paired observations, and so are their interpretations. You can conveniently calculate these C-statistics by looking at all the pairs of subjects in a data set and counting them accordingly. It is worth noting that this simple approach does not scale well, because its computational complexity is on the order of n^2 , meaning that the computational burden grows at a quadratic rate as a function of the number of observations. As a result, the computing time might become unbearably long in some situations. To improve the computational efficiency, PROC PHREG adopts an alternative algorithm based on binary search trees. In this new scheme, the same task can be completed in one pass of the sorted data so that it becomes unnecessary to look at all the pairs. The new algorithm reduces the order of complexity to $n \log(n)$, and experiments with simulated data have demonstrated that the computation time can be reduced from several hours to a few minutes.

Time-Dependent ROC Curves

This section introduces the concept of time-dependent ROC curves for right-censored data and the syntax for their computation as implemented in PROC PHREG.

Concept and Methods

Let T denote the event-time variable, and let Y denote the continuous variable to be assessed. For example, assume that Y is the linear predictor. At time t , a binary outcome can be defined as

$$D_t = I(T \leq t)$$

Suppose c denotes a specific value within the support of Y . The sensitivity (SE) and specificity (SP) can be defined as

$$SE_t(c) = \Pr(Y > c | D_t = 1)$$

$$SP_t(c) = \Pr(Y \leq c | D_t = 0)$$

The ROC curve at time t is defined to be

$$ROC_t(u) = SE_t[1 - SP_t^{-1}(u)]$$

This definition is often referred to in the literature as the “cumulative/dynamic” ROC curve. “Cumulative” means that all events that occur before time t are considered to be “cases.” Information about other types of time-dependent ROC curves is available in the literature—for example, in Heagerty and Zheng (2005).

The AUC statistic at time t is the area under the ROC curve at time t :

$$AUC_t = \int ROC_t(u) du$$

PROC PHREG supports four different techniques for estimating time-dependent ROC curves, as summarized in Table 2.

Table 2 Methods of Estimating Time-Dependent ROC Curves in PROC PHREG

Option Name	Method	Reference
IPCW	Inverse probability of censoring weighting	Uno et al. (2007)
KM	Conditional Kaplan-Meier	Heagerty, Lumley, and Pepe (2000)
NNE	Nearest neighbors	Heagerty, Lumley, and Pepe (2000)
RECURSIVE	Recursive method	Chambless and Diao (2006)

For a detailed description of these methods, see the section “[Estimating Time-Dependent ROC Curves](#)” in the [Appendix](#).

The four methods of estimating the time-dependent ROC curves are constructed in different ways. Both the NNE method and the KM method are constructed by estimating the joint distribution between Y and T (Heagerty, Lumley, and Pepe 2000). The NNE method is set as the default because it produces smooth estimates. In addition to the noninformative censoring that the other methods assume, the IPCW method also assumes that censoring occurs independently of all the covariates. From a computational perspective, the recursive method should be the most efficient, because its calculation is performed on the sorted data and requires only one pass. This feature could be particularly appealing when the data are too big and the time points to be evaluated are numerous. In addition, the recursive approach is implemented based on binary search trees, and its computational complexity is on the order of $n \log(n)$.

Syntax in PROC PHREG

As in PROC LOGISTIC, in the PHREG procedure you use the PLOTS=ROC option to produce time-dependent ROC curves, and you use the ROC statement to specify the models for which concordance and ROC analyses are performed. The ROCOPTIONS option in the PROC PHREG statement, which has been completely redesigned from the ROCOPTIONS option in PROC LOGISTIC, offers various control mechanisms for computing and rendering the ROC curves. [Table 3](#) lists the available suboptions of the ROCOPTIONS option in PROC PHREG and their functions.

Table 3 Suboptions of the ROCOPTIONS Option

Option	Function
AT=	Specifies the list of time points for ROC curves
AUC	Displays the area under the ROC curve
AUCDIFF	Displays AUC differences for each pair of identified models
IAUC	Displays the integrated area under the curve
METHOD= <i>method</i> < (<i>options</i>) >	Specifies the method of calculating ROC curves and AUC statistics
OUTAUC= <i>SAS-data-set</i>	Names the output data set for the AUC plot
OUTROC= <i>SAS-data-set</i>	Names the output data set for the ROC plots

Example: New Features in PROC PHREG

This example uses the **Liver** data from the earlier example to illustrate various uses of concordance statistics and time-dependent ROC curves and AUC statistics in PROC PHREG.

The following statements use the PHREG procedure to fit the Cox regression model that uses **Bilirubin**, **Age**, and **Edema** as explanatory variables:

```
ods graphics on;
proc phreg data=Liver concordance plots=roc rocoptions(at=2 to 10 by 2);
    model Time*Status(0)=Bilirubin Age Edema;
run;
```

The CONCORDANCE option in the PROC PHREG statement displays Harrell’s C-statistic. The PLOTS=ROC option plots the time-dependent ROC curves at time points 2, 4, 6, 8, and 10 years, which are specified in the AT= suboption in the ROCOPTIONS option.

Results of Harrell’s C-statistic are shown in [Figure 7](#).

Figure 7 Harrell’s C-Statistic

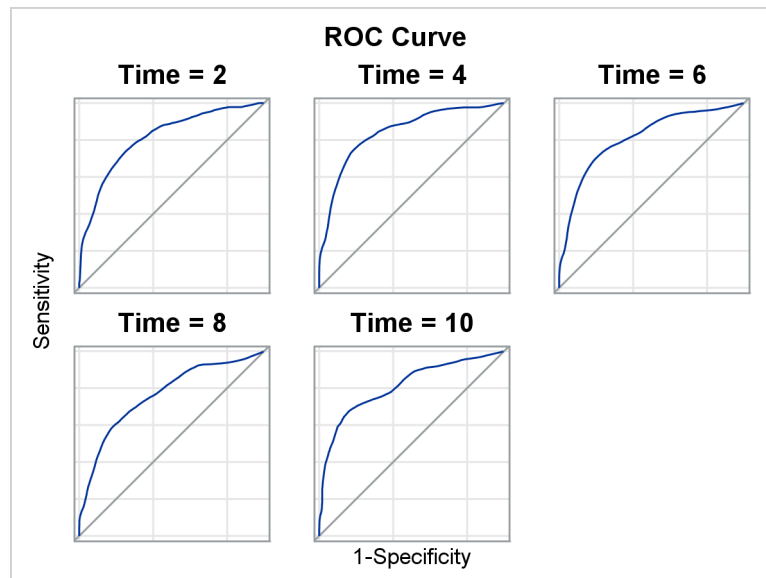
The PHREG Procedure

Harrell's Concordance Statistic					
Comparable Pairs					
Source	Estimate	Concordance	Discordance	Tied in Predictor	Tied in Time
Model	0.7966	34798	8884	2	5

There are 34,798 concordance pairs, 8,884 discordance pairs, 2 pairs that are tied in the linear predictor, and 5 pairs that are tied in the follow-up time. This gives a concordance estimate of 0.7966.

Figure 8 shows the time-dependent ROC curves at the selected years.

Figure 8 ROC Plot at Selected Time Points



By default, these curves are computed by the nearest neighbors technique of Heagerty, Lumley, and Pepe (2000) and are displayed in a panel. It appears that among the five selected years, year 4 has the largest AUC, year 8 has the lowest AUC, and the other years are in between. If you specify the `OVERLAY=INDIVIDUAL` global plot option to display individual plots, each plot also displays the AUC. The following statements make two individual ROC plots, at year 2 and year 6:

```
proc phreg data=Liver plots(overlay=individual)=roc rocoptions(at=2 6);
  model Time*Status(0)=Bilirubin Age Edema;
run;
```

The individual plots of time-dependent ROC curves at year 2 and year 6 are shown in Figure 9 and Figure 10, respectively.

Figure 9 ROC Plots at Year 2

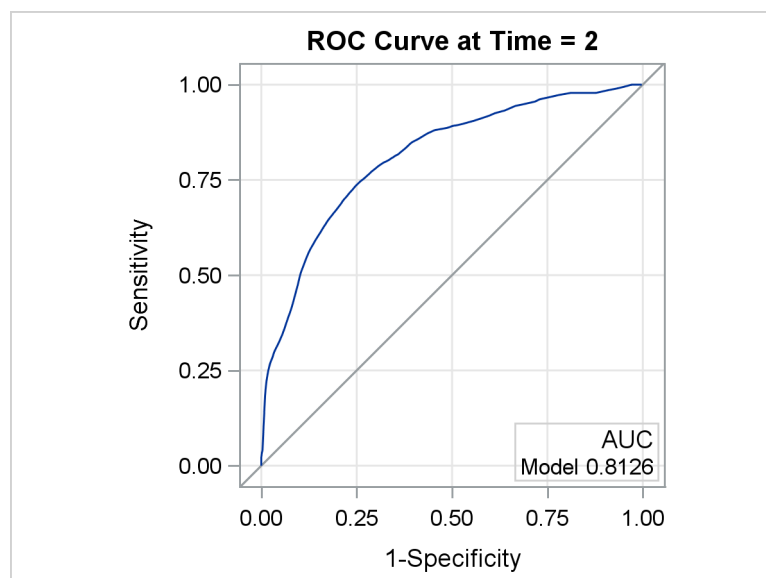
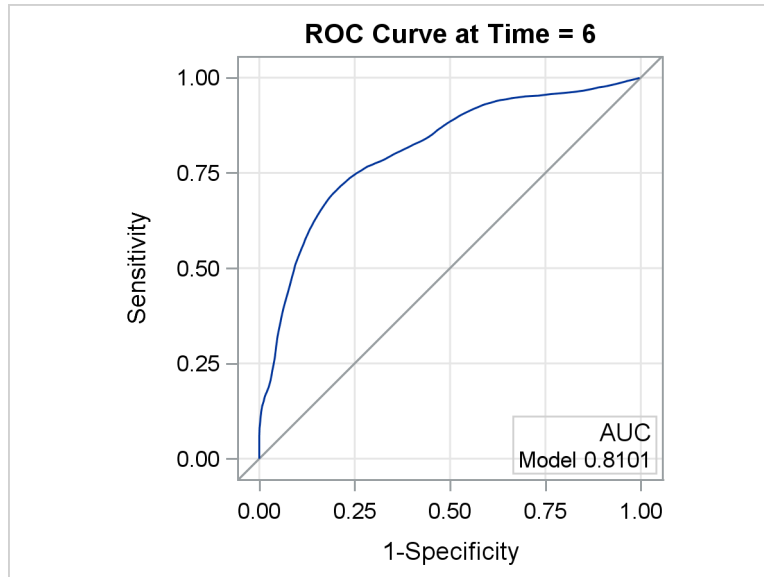


Figure 10 ROC Plots at Year 6



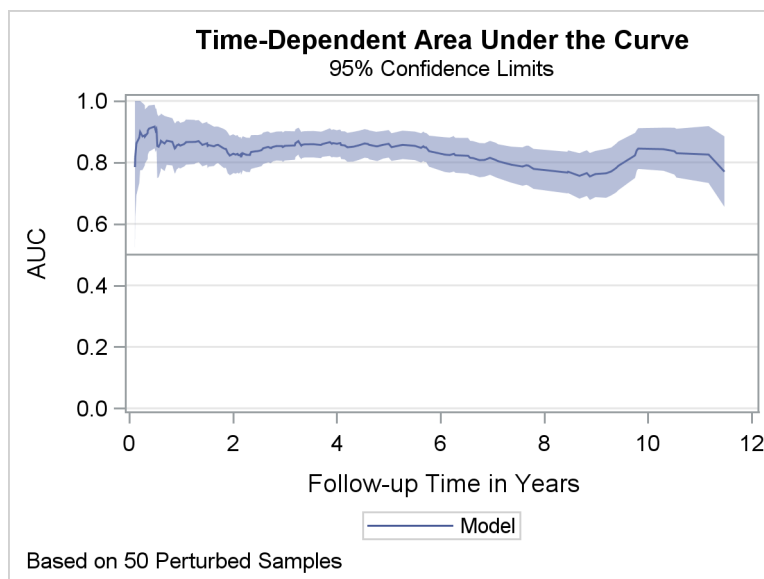
Time-dependent ROC curves change only at the distinct event times. You can examine the AUC at all distinct event times by plotting the curve of the AUC. The following statements plot the curve of the AUC of the fitted model and display the 95% pointwise confidence limits:

```
proc phreg data=Liver plots=auc rocoptions(method=ipcw(cl seed=1234) iauc);  
  model Time*Status(0)=Bilirubin Age Edema;  
run;
```

The PLOTS=AUC option in the PROC PHREG statement plots the AUC curve. The ROCOPTIONS option in the PROC PHREG statement enables you to specify the inverse probability of censoring weighting (IPCW) method to compute the ROC curves, and the CL suboption requests pointwise confidence limits for the AUC curve. The IAUC option computes and displays the integrated AUC over time.

Figure 11 displays the AUC curve and the 95% confidence limits for the fitted model.

Figure 11 AUC Plot with 95% Confidence Limits



It appears that the AUC statistic reaches a high of 0.92 at year 0.21 but mostly hovers around 0.8.

Figure 12 displays the integrated AUC, which averages all available AUC statistics over time.

Figure 12 Integrated AUC

Integrated Time-Dependent Area Under the Curve	
Source	Estimate
Model	0.8284

Consider three submodels of the previously fitted Cox model, each of which contains two of the three covariates **Bilirubin**, **Age**, and **Edema**. You can use the method of Uno et al. (2011) to assess the difference of the concordance probabilities between any two submodels. In the following statements, three ROC statements are specified, one for each submodel:

```
proc phreg data=Liver concordance=uno(diff se seed=1234 iter=100);
  model Time*Status(0)=Bilirubin Age Edema / nofit;
  roc 'Bilirubin+Age' Bilirubin Age;
  roc 'Age+Edema' Age Edema;
  roc 'Bilirubin+Edema' Bilirubin Edema;
run;
```

The DIFF suboption of the CONCORDANCE=UNO option in the PROC PHREG statement calculates all pairwise differences. The SE suboption computes the standard error for each pairwise difference, based on 100 perturbation samples as specified by the ITER= suboption. The seed of the random generator for the perturbation resampling is set to 1234. The NOFIT option is specified in the MODEL statement, because there is no need to fit the specified model.

Figure 13 displays the results of the concordance analysis.

Figure 13 Comparing Uno's Concordance Estimates

The PHREG Procedure

Differences in Uno's Concordance Statistic					
Source	_Source	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Bilirubin+Age	Age+Edema	0.0972	0.0232	17.57	<.0001
Bilirubin+Age	Bilirubin+Edema	-0.0264	0.0231	1.31	0.2529
Age+Edema	Bilirubin+Edema	-0.1236	0.0287	18.51	<.0001

It appears that the two submodels that contain **Bilirubin** have a significantly larger concordance probability than the submodel without **Bilirubin**. In other words, the two submodels that contain **Bilirubin** predict the survival outcomes more accurately than the model without **Bilirubin**.

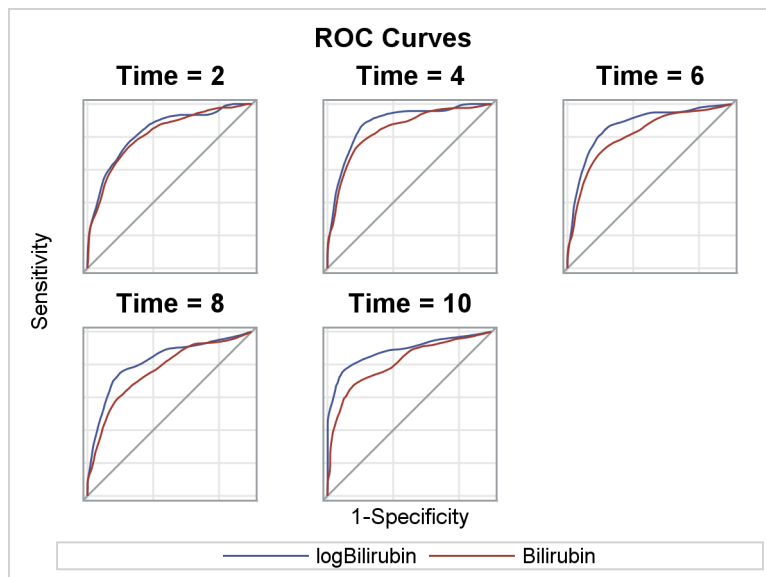
It has been demonstrated elsewhere that the log transform is a much-improved functional form for **Bilirubin** in a Cox regression model (Lin, Wei, and Ying 1993). It is expected that the model that contains **Bilirubin** in the log scale would have more discriminating power than the model that contains **Bilirubin** in the original scale. In the following statements, PROC PHREG is used to fit the model with the log transform for **Bilirubin**:

```
proc phreg data=Liver;
  model Time*Status(0)=logBilirubin Age Edema;
  logBilirubin = log(Bilirubin);
  output out=Liver2 xbeta=Y;
run;
proc phreg data=Liver2 plots=roc rocoptions(at=2 to 10 by 2);
  model Time*Status(0)=Bilirubin Age Edema / roclabel='Bilirubin';
  roc 'logBilirubin' pred=Y;
run;
```

Because the OUTPUT statement is used, the linear predictor variable is saved as the variable **Y** in the output data set **Liver2**. With **Liver2** as the new input data set, PROC PHREG is called to fit the Cox model that contains **Bilirubin** in the original scale. The linear predictor variable **Y** is specified in the ROC statement as the PRED= option value.

Figure 14 displays the ROC curves of the two competing models.

Figure 14 ROC Plots to Evaluate the Log Transform for Bilirubin



The ROC curve for the model with the log scale for **Bilirubin** essentially lies above that of its counterpart with the original scale for all the selected time points. This leads to the conclusion that the log transform for **Bilirubin** improves the predictive power of the model.

Syntax Differences between PROC PHREG and PROC LOGISTIC

For the most part, the syntax in the PHREG procedure for producing concordance statistics and time-dependent ROC curves and AUC statistics is similar to the syntax for ROC analysis in the LOGISTIC procedure. But there are some major differences.

New CONCORDANCE Option

In PROC LOGISTIC, the C-statistic is uniquely defined and is automatically displayed in the printed output. To request C-statistics in PROC PHREG, you must specify the CONCORDANCE option in the PROC PHREG statement. Furthermore, you specify the keyword HARRELL or UNO to request the two different C-statistics.

Redesigned ROCOPTIONS Option

The ROCOPTIONS option in both PROC PHREG and PROC LOGISTIC controls various aspects of computing and rendering the ROC curves. To accommodate the new reality in the survival setting, the ROCOPTIONS option in PROC PHREG has been completely redesigned. For comparison purposes, Table 4 lists some key suboptions of the ROCOPTIONS option in PROC LOGISTIC and their functions.

Table 4 Key Suboptions of ROCOPTIONS Option in PROC LOGISTIC

Option	Function
ALPHA= <i>number</i>	Sets the significance level for the confidence limits
CROSSVALIDATE X	Uses cross validation to compute the ROC
OUT= <i>SAS-data-set-name</i>	Names the output data set for the ROC plots
WEIGHTED	Uses frequency \times weight to compute the ROC

Enhanced ROC Statement

The ROC statement is available in both PROC LOGISTIC and PROC PHREG. You can use it to specify a model to be used in the concordance analysis or ROC analysis. There is no limit to the number of ROC statements that you can use.

In PROC PHREG, you can use one of the following specifications:

effect-list

specifies a list of effects that have previously been specified in the MODEL statement.

PRED=*variable*

specifies a *variable* in the input data set. The *variable* does not have to be specified in the MODEL statement.

SOURCE=*item-store-name*

specifies a one- or two-level name of an existing item store.

The PRED= and *effect-list* options are also supported in PROC LOGISTIC, but the SOURCE= option is not. As an enhancement, the SOURCE= option enables you to import a model that is produced outside PROC PHREG; for example, you can fit a parametric survival model by using PROC LIFEREG, save the model in an item store, and then specify the item-store name in the SOURCE= option in an ROC statement to produce an ROC analysis of the imported model.

Decommissioned ROCCONTRAST Statement

PROC LOGISTIC provides the ROCCONTRAST statement for comparing different ROC curves. Such a setup is no longer used in PROC PHREG because the C-statistic is no longer equal to the area under the ROC curve. To compare Uno's C-statistic between two prediction models, you can use the DIFF option.

Summary

This paper introduces concordance statistics (C-statistics) and time-dependent ROC curves and AUC statistics for assessing the predictive accuracy of survival models that are available in the PHREG procedure in SAS/STAT 14.2. These new methods can be considered generalizations from the conventional methods for binary outcomes that have been available in PROC LOGISTIC. Because the goal of applying these criteria remains the same in the survival setting, you can use your experience with PROC LOGISTIC to perform relevant analyses.

To apply the new methods successfully, it is important to understand how they relate to the conventional methods and to be cautious about the differences. In the conventional setting with binary outcomes, the C-statistic is equivalent to the area under the ROC curve and thus serves as an overall summary of the curve. This relationship breaks down in the survival setting because the ROC curves have been generalized to be time-dependent and the area under the curve becomes time-specific. Also, different versions of the C-statistic, such as Harrell's C-statistic and Uno's C-statistic, exist because the overall concordance probability can be formulated in different ways to deal with censoring. For time-dependent ROC curves, PROC PHREG currently supports only the cumulative/dynamic type and provides multiple estimation methods. [Table 5](#) summarizes these new methods.

Table 5 Summary of New Methods for Evaluating Predictive Accuracy

Method	Option	Standard Error	Limitation
C-statistic	HARRELL (Default)	Yes	Estimate is censoring-dependent
	UNO	Yes	Assumes censoring distribution is estimable
Time-dependent ROC curve	NNE (Default)	No	SPAN= option value is needed
	KM	No	Monotonicity might be violated for small sample
	RECURSIVE	No	Underdefined at tied events
	IPCW	Yes	Assumes censoring distribution is estimable

Appendix

Estimating Concordance Statistics

For the i th individual ($1 \leq i \leq n$) in a sample, let X_i , Δ_i , and \mathbf{Z}_i be the observed time, event indicator (1 for death, 0 for censored), and covariate vector, respectively.

For the i th individual, the Cox proportional hazards model in PROC PHREG assumes the hazard function to be

$$\lambda_i(t) = \lambda(t; \mathbf{Z}_i) = \lambda_0(t) \exp(\mathbf{Z}_i' \boldsymbol{\beta})$$

where $\lambda_0(t)$ is an arbitrary and unspecified baseline hazard function and $\boldsymbol{\beta}$ is the vector of true regression parameters that is associated with the explanatory variables. Let $\hat{\boldsymbol{\beta}}$ denote the maximum partial likelihood estimates of $\boldsymbol{\beta}$. The estimated linear predictor for the i th individual is $Y_i = \hat{\boldsymbol{\beta}}' \mathbf{Z}_i$.

- **Harrell's Concordance**

Harrell (1986) defines the concordance probability as

$$C_H = \Pr(\boldsymbol{\beta}' \mathbf{Z}_1 > \boldsymbol{\beta}' \mathbf{Z}_2 | T_1 < T_2, T_1 < \min(D_1, D_2))$$

where \mathbf{Z}_1 and \mathbf{Z}_2 are the covariate vectors for a pair of subjects, the survival times are denoted as T_1 and T_2 , and the censoring times are denoted as D_1 and D_2 .

Assuming that there are no ties in the event times and the predictor scores, C_H can be estimated as

$$\hat{C}_H = \frac{\sum_{i \neq j} \Delta_i I(X_i < X_j) I(\hat{\boldsymbol{\beta}}' \mathbf{Z}_i > \hat{\boldsymbol{\beta}}' \mathbf{Z}_j)}{\sum_{i \neq j} \Delta_i I(X_i < X_j)}$$

When there are ties in the predictor scores, the preceding calculation can be adjusted to be

$$\hat{C}_H = \frac{\sum_{i \neq j} \Delta_i I(X_i < X_j) [I(\hat{\boldsymbol{\beta}}' \mathbf{Z}_i > \hat{\boldsymbol{\beta}}' \mathbf{Z}_j) + 0.5 I(\hat{\boldsymbol{\beta}}' \mathbf{Z}_i = \hat{\boldsymbol{\beta}}' \mathbf{Z}_j)]}{\sum_{i \neq j} \Delta_i I(X_i < X_j)}$$

For standard errors of \hat{C}_H , PROC PHREG uses the estimator derived based on the delta method (Kang et al. 2015).

- **Uno's Concordance**

Uno et al. (2011) propose the following method of estimating the concordance probability:

$$C_U = \Pr(\boldsymbol{\beta}' \mathbf{Z}_1 > \boldsymbol{\beta}' \mathbf{Z}_2 | T_1 < T_2)$$

If τ is a specified time point within the support of the censoring variable, Uno et al. (2011) also define a truncated version of the concordance probability as

$$C_U = \Pr(\boldsymbol{\beta}' \mathbf{Z}_1 > \boldsymbol{\beta}' \mathbf{Z}_2 | T_1 < T_2, T_1 < \tau)$$

You can specify a τ value in the TAU= option in the PROC PHREG statement. If the TAU= option is not specified, then there is no truncation and the τ value is taken as the largest event time.

Let $\hat{G}(t)$ be the Kaplan-Meier estimate of the censoring distribution (assuming no covariates). C_U is consistently estimated by

$$\hat{C}_U = \frac{\sum_{i=1}^n \sum_{j=1}^n \Delta_i \hat{G}(X_i^-)^{-2} I(X_i < X_j, X_i < \tau) [I(\hat{\boldsymbol{\beta}}' \mathbf{Z}_i > \hat{\boldsymbol{\beta}}' \mathbf{Z}_j) + 0.5 * I(\hat{\boldsymbol{\beta}}' \mathbf{Z}_i = \hat{\boldsymbol{\beta}}' \mathbf{Z}_j)]}{\sum_{i=1}^n \sum_{j=1}^n \Delta_i \hat{G}(X_i^-)^{-2} I(X_i < X_j, X_i < \tau)}$$

Estimating Time-Dependent ROC Curves

• Inverse Probability of Censoring Weighting Approach

Let $\hat{G}(t)$ be the Kaplan-Meier estimate of the censoring distribution (assuming no covariates). Assuming that the censoring distribution is independent of the failure time distribution, the sensitivity (SE) and specificity (SP) under a specific threshold value c can be consistently estimated as

$$\widehat{SE}_t(c) = \frac{\sum_{i=1}^n \Delta_i I(\hat{\beta}' \mathbf{Z}_i > c, X_i \leq t) / \hat{G}(X_i)}{\sum_{i=1}^n \Delta_i I(X_i \leq t) / \hat{G}(X_i)}$$

$$\widehat{SP}_t(c) = \frac{\sum_{i=1}^n I(\hat{\beta}' \mathbf{Z}_i \leq c, X_i > t)}{\sum_{i=1}^n I(X_i > t)}$$

$ROC_t(c)$ can be estimated by substituting these estimated sensitivities and specificities. The estimated AUC_t is calculated by using the trapezoidal rule to integrate the estimated $ROC_t(c)$ curve.

Uno et al. (2007) propose estimating the standard errors of the AUC_t estimator by using the perturbation-resampling method. Suppose $\hat{\sigma}^2$ is the sample variance based on M realizations of the perturbed AUC_t . The $100(1 - \alpha)\%$ confidence limits for AUC_t are $\widehat{AUC}_t \pm z_{\alpha/2} \hat{\sigma}$, where \widehat{AUC}_t is the estimated AUC_t and $z_{\alpha/2}$ is the upper $100\alpha/2$ percentile of the standard normal distribution.

• Conditional Kaplan-Meier Approach

Using Bayes' theorem, you can write the sensitivity and specificity, respectively, as

$$SE_t(c) = \Pr(Y > c | D_t = 1) = \frac{[1 - S(t|Y > c)] \Pr(Y > c)}{1 - S(t)}$$

$$SP_t(c) = \Pr(Y \leq c | D_t = 0) = \frac{S(t|Y \leq c) \Pr(Y \leq c)}{S(t)}$$

where $S(\cdot)$ is the survivor function and $S(\cdot|Y > c)$ is the conditional survivor function for $Y > c$.

Heagerty, Lumley, and Pepe (2000) use the Kaplan-Meier method to estimate the survivor function $S(\cdot)$ and the conditional survivor function $S(\cdot|Y > c)$. The latter was estimated using subjects where the condition $Y > c$ is met. The sensitivity and specificity are estimated, respectively, as

$$\widehat{SE}_t(c) = \frac{[1 - \hat{S}_{KM}(t|Y > c)] [1 - \hat{F}_Y(c)]}{1 - \hat{S}_{KM}(t)}$$

$$\widehat{SP}_t(c) = \frac{\hat{S}_{KM}(t|Y \leq c) \hat{F}_Y(c)}{\hat{S}_{KM}(t)}$$

where $\hat{S}_{KM}(\cdot)$ is the Kaplan-Meier estimator and $\hat{F}_Y(c) = \sum_i I(Y_i \leq c)/n$.

• Nearest Neighbors Approach

Following Akritas (1994), the bivariate survival function, $S(c, t) = \Pr(Y > c, T > t)$, can be estimated as

$$\hat{S}_{b_n}(c, t) = \frac{1}{n} \sum_i \hat{S}_{b_n}(t|Y = Y_i) I(Y_i > c)$$

where $\hat{S}_{b_n}(t|Y = Y_i)$ is a smoothed estimate of the conditional survival function. Define the weighted Kaplan-Meier estimator as

$$\hat{S}_{b_n}(t|Y = Y_i) = \prod_{s \in \{X_i: i=1, \dots, n, \Delta_i=1\}, s \leq t} \left[1 - \frac{\sum_j K_{b_n}(Y_i, Y_j) I(X_i = s) \Delta_i}{\sum_j K_{b_n}(Y_i, Y_j) I(X_i = s)} \right]$$

where $K_{b_n}(Y_i, Y_j)$ is a kernel function that depends on the parameter b_n . Akritas (1994) uses the nearest neighbor kernel, $K_{b_n}(Y_i, Y_j) = I\{-b_n < \hat{F}_Y(Y_i) - \hat{F}_Y(Y_j) < b_n\}$, where $0 < 2b_n < 1$; this effectively selects the nearest $2b_n$ proportion of observations in the neighborhood. The default value for b_n is 0.05. You can specify a different value by using the SPAN= suboption in METHOD=NNE in the ROCOPTIONS option in the PHREG statement.

The sensitivity and specificity can then be estimated, respectively, as

$$\widehat{SE}_t(c) = \frac{1 - \hat{F}_Y(c) - \hat{S}_{b_n}(c, t)}{1 - \hat{S}_{b_n}(t)}$$

$$\widehat{SP}_t(c) = 1 - \frac{\hat{S}_{b_n}(c, t)}{\hat{S}_{b_n}(t)}$$

where $\hat{S}_{b_n}(t) = \hat{S}_{b_n}(-\infty, t)$. For more information, see Heagerty, Lumley, and Pepe (2000).

• Recursive Approach

Chambless and Diao (2006) propose estimating time-dependent ROC curves by using a recursive approach akin to the Kaplan-Meier method. Let $t_1 < t_2 < \dots < t_M$ be the distinct event times in the data. The area under the curve at time t_m , $1 \leq t_m \leq M$, can be derived as

$$AUC_{t_m} = \frac{\sum_{k=1}^m \gamma_k \lambda(t_k) (1 - \lambda(t_k)) S(t_{k-1}) - \sum_{k=1}^m \tau_k \lambda(t_k) (1 - S(t_{k-1})) S(t_{k-1})}{S(t_m) (1 - S(t_m))}$$

where $S(\cdot)$ is the survivor function, $\lambda(\cdot)$ is the hazard function, $t_0 = 0$, $\tau_0 = 0$, and

$$\tau_k = \Pr(\beta' \mathbf{Z}_i > \beta' \mathbf{Z}_j | X_i = t_k, \Delta_i = 1, X_j > t_k)$$

$$\gamma_k = \Pr(\beta' \mathbf{Z}_i > \beta' \mathbf{Z}_j | X_i = t_{k-1}, \Delta_i = 1, X_j = t_k, \Delta_j = 1)$$

In a recursive fashion, the sensitivity and specificity at time t_m can be shown, respectively, to be

$$SE_{t_m}(c) = \sum_{k=1}^m \rho_k(c) \lambda(t_k) S(t_{k-1}) / [1 - S(t_m)]$$

$$SP_{t_m}(c) = \frac{\Pr(\beta' \mathbf{Z}_i \leq c) - \sum_{k=1}^m [1 - \rho_k(c)] \lambda(t_k) S(t_{k-1})}{S(t_m)}$$

where $\rho_k(c) = \Pr(\beta' \mathbf{Z}_i > c | X_i = t_k, \Delta_i = 1)$.

Define \mathcal{R}_k to be the risk set at time t_k , and let r_k be the number of subjects in \mathcal{R}_k . Let $\mathbf{Z}_{(k)}$ be the covariate vector for the subject whose event time is t_k . The unknown parameters τ_k , γ_k , and $\rho_k(c)$ can be estimated as

$$\hat{\tau}_k = \frac{1}{k-1} \sum_{i=1}^k I(\hat{\beta}' \mathbf{Z}_{(i)} > \hat{\beta}' \mathbf{Z}_{(k)})$$

$$\hat{\gamma}_k = \frac{1}{r_k-1} \sum_{j \in \mathcal{R}_k} I(\hat{\beta}' \mathbf{Z}_{(k)} > \hat{\beta}' \mathbf{Z}_j)$$

$$\hat{\rho}_k(c) = I(\hat{\beta}' \mathbf{Z}_{(k)} > c)$$

When there is only one event at each event time, $\lambda(t_k)$ is estimated as $\hat{\lambda}(t_k) = 1/r_k$ and $S(t_k)$ is estimated by the Kaplan-Meier method as $\hat{S}(t_k) = \hat{S}(t_{k-1})[1 - \hat{\lambda}(t_{k-1})]$. In the case of a tie, the order of the events in the calculation is the same as the order of their appearance in the input data set.

REFERENCES

- Akritas, M. G. (1994). "Nearest Neighbor Estimation of a Bivariate Distribution under Random Censoring." *Annals of Statistics* 22:1299–1327.
- Blanche, P., Latouche, A., and Viallon, V. (2013). "Time-Dependent AUC with Right-Censored Data: A Survey." In *Risk Assessment and Evaluation of Predictions*, edited by M.-L. T. Lee, M. Gail, R. Pfeiffer, G. Satten, T. Cai, and A. Gandy, 239–251. New York: Springer.
- Chambless, L. E., and Diao, G. (2006). "Estimation of Time-Dependent Area under the ROC Curve for Long-Term Risk Prediction." *Statistics in Medicine* 25:3474–3486.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics* 44:837–845.

- Harrell, F. E. (1986). "The PHGLM Procedure." In *SUGI Supplemental Library Guide, Version 5 Edition*. Cary, NC: SAS Institute Inc.
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). "Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker." *Biometrics* 56:337–344.
- Heagerty, P. J., and Zheng, Y. (2005). "Survival Model Predictive Accuracy and ROC Curves." *Biometrics* 61:92–105.
- Kang, L., Chen, W., Petrick, N. A., and Gallas, B. D. (2015). "Comparing Two Correlated C Indices with Right-Censored Survival Outcome: A One-Shot Nonparametric Approach." *Statistics in Medicine* 34:685–703.
- Lin, D. Y., Wei, L. J., and Ying, Z. (1993). "Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals." *Biometrika* 80:557–572.
- Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., and Wei, L. J. (2011). "On the C-Statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data." *Statistics in Medicine* 30:1105–1117.
- Uno, H., Cai, T., Tian, L., and Wei, L. J. (2007). "Evaluating Prediction Rules for t-Year Survivors with Censored Regression Models." *Journal of the American Statistical Association* 102:527–537.

ACKNOWLEDGMENTS

The authors thank Ed Huddleston for his valuable editorial assistance in preparing this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors:

Changbin Guo	Ying So	Woosung Jang
SAS Institute Inc.	SAS Institute Inc.	SAS Institute Inc.
SAS Campus Drive	SAS Campus Drive	SAS Campus Drive
Cary, NC 27513	Cary, NC 27513	Cary, NC 27513
changbin.guo@sas.com	ying.so@sas.com	woosung.jang@sas.com

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.