Paper 2831-2018

Evaluating the accuracy of clinical prediction models for binary and survival outcomes

Joseph C. Gardiner, Department of Epidemiology and Biostatistics,

Michigan State University, East Lansing, MI 48824

Abstract

Clinical prediction models employ regression-based methods to elucidate potential predictors of outcomes. For a binary outcome, the LOGISTIC and HPLOGISTIC procedures offer options for model development, testing and validation. Several fit statistics can be used to gauge the predictive accuracy of a model as well as for comparisons between competing models. In the appropriate context these statistics include sensitivity, specificity, positive and negative predictive values, the receiver operating characteristic (ROC) curve and concordance indices. A similar development for survival models faces many challenges, including the feature of time-dependent outcome and censoring in accrued survival data. New options in the PHREG procedure permit calculation of some of the aforementioned fit statistics. We discuss their interpretation and illustrate their application with empirical data sets.

1. Introduction

Clinical prediction models are ubiquitous in medicine (Steyerberg, 2009). Prediction models have at their core a diagnostic marker based on contemporaneous and antecedent variables that can predict the likelihood of a future random event. Khorana *et al* (2008) develop a prediction rule for the risk of venous thromboembolism (VTE) in cancer patients on active chemotherapy. A single-index is constructed from baseline clinical and laboratory assessments that could help identify patients at high risk of VTE so that prophylaxis could be initiated. Gardiner *et al* (2016) report on the incidence of hospital-acquired pressure ulcers in a population-based retrospective cohort. The objective was to assess the relative importance of patient variables such as gender, race, age, comorbidities, body mass index in identifying patients who might be at higher risk of acquiring a pressure ulcer during their hospital stay.

From studies in which the event outcome and its potential correlates are available, a series of prognosis models can be evaluated with respect to their discriminative ability. Their common structure models the presence of disease $(D=1)$ from potential predictor variables **x** through a logistic regression model, $P[D=1|\mathbf{x}]=\left(1+\exp(-\mathbf{x}'\beta)\right)^{-1}$. From the distribution of the single index $M=\mathbf{x}'\beta$, the *sensitivity* $P[M>c|D=1]$ and *specificity* $P[M\le c|D=0]$ can be compared by varying the cut-off value c. Sensitivity is called the *true positive ratio* $TP(c)$ and 1 minus specificity, the *false positive ratio* $FP(c)$. The receiver operating characteristic (ROC) curve is a display of the points $(FP(c),TP(c))$ for varying cut-offs that could aid in a judicious choice of c. The area under the ROC curve (AUC) is a summary statistic called the c-statistic. It has an interpretation as the probability that the marker for a randomly selected subject from the diseased population $(D=1)$ is greater than the marker for a randomly selected subject from the non-diseased population $(D=0)$. Since these statistics depend on **x**, one can assess competing logistic models with different covariate sets in their ability to provide a relatively simpler discrimination between true positive and false positive ratios. For example, a c-statistic above 0.75 is considered excellent. Sub-models with fewer covariates may be compared with respect to their c-statistics.

For a binary outcome, the LOGISTIC procedure is the workhorse for estimating the logistic model. Options for ROC analyses of competing models and statistical comparisons between them are available. Both LOGISITC and HPLOGISTIC provide a scheme for development, testing and validation of the prediction model.

Survival outcomes

We begin with a time-to-event T measured from origin $t = 0$ and a marker M measured at baseline. Higher values of the marker are indicative of worse prognosis for the event. Potential predictor variables $\mathbf{x}$ of the survival distribution $S(t \mid \mathbf{x}) = P[T > t \mid \mathbf{x}]$ can be assessed using the semiparametric Cox proportional hazards model (PHM) $h(t \mid \mathbf{x}) = h_0(t)\exp(\mathbf{x}'\beta)$ with baseline hazard function h_0 and the single index $M = \mathbf{x}'\beta$. An alternative model is the parametric accelerated failure time model (AFT), $\log T = \mathbf{x}'\beta + \sigma\varepsilon$ with a specified distribution on ε and scale parameter σ (>0). For example, the extreme-value distribution for ε corresponds to Weibull survival $S(t \mid \mathbf{x}) = \exp\left(-\{t / \theta(\mathbf{x})\}^{1/\sigma}\right)$ where $\log\theta(\mathbf{x}) = \mathbf{x}'\beta$. In the AFT context we use $M = -\mathbf{x}'\beta$ to maintain the convention that higher values of the marker indicate poorer survival outcome.

The next step is to define sensitivity and specificity of the diagnostic marker. Several definitions have been proposed (Heagerty and Zhang, 2005, Pepe *et al*, 2008). PHREG adopts the *cumulative/dynamic* definition:

$$TP(c,t) = Sensitivity(c,t) = P[M > c \mid T \le t], \text{ and } Specificity(c,t) = P[M \le c \mid T > t] = 1 - FP(c,t). \quad (1)$$

Interpreting T as the time to disease onset, the definitions are similar to those for a binary outcome, by taking $D = [T \le t]$. With t fixed, the true positive ratio $TP(c,t)$ and false positive ratio $FP(c,t)$ are monotone decreasing functions of the cut-off c, but not necessarily strictly decreasing. As probabilities they are bounded on [0, 1]. It is desirable for estimators of these quantities to have the same properties.

<u>*ROC curve and AUC*</u>

Given $TP(c,t)$ and $FP(c,t)$, the points of the ROC curve are $ROC(t) = \{(FP(c,t), TP(c,t)) : c \in \Re\}$. Formally, define $\mathrm{ROC}(p, t) = TP\{[FP(p,t)]^{-1}, t\}$ where $[FP(p,t)]^{-1} = \inf\{c : FP(c,t) \le p\}$, $p \in [0,1]\}$. For each t, $p \to \mathrm{ROC}(p, t)$ is a function on [0, 1]. The area under the ROC is $AUC(t) = \int_0^1 ROC(p,t)dp$.

From independent pairs $(M_1, T_1), (M_2, T_2)$ we may interpret $AUC(t)$ as

$$AUC(t) = P[M_1 > M_2 \mid T_1 \le t < T_2] \quad (2)$$

At time t, given that the event has occurred in subject 1, but has not yet occurred in subject 2, $AUC(t)$ is the probability that the marker M_1 in subject 1 is greater than the marker M_2 in subject 2. It measures the ability of the markers to correctly order the event status at time t.

<u>*Integrated AUC*</u>

An 'average' $AUC(t)$ for all t is obtained as $E\left(AUC(T)\right) = \int_0^\infty AUC(t)(-dS(t))$

2. Estimation

With survival data we must address incomplete observation of the survival outcome, that is, *T* might be (right) censored. Let $\{(T_i^*, \delta_i, \mathbf{x}_i) : 1 \le i \le n\}$ denote the observations from a random sample with observed time $T_i^* = \min(T_i, C_i)$, event time T_i, censoring time C_i and (right) censoring indicator $\delta_i = [T_i \le C_i]$. Thus $\delta_i = 1$ if T_i^* is an event time, and $\delta_i = 0$ otherwise. The covariates $\mathbf{x}_i$ are time-invariant and used to define the marker $M_i = \mathbf{x}_i'\beta$. Estimation of β is via maximum partial likelihood in the Cox PHM. Minimally, we assume (T_i, C_i) are conditionally independent given $\mathbf{x}_i$, but in much of what follows the censoring distribution is assumed not to depend on $\mathbf{x}_i$. In several applications this assumption may not be tenable (Blanche *et al*, 2013a). Following historical development of ROC analyses, PHREG offers four methods of estimation of the quantities $TP(c,t)$, $FP(c,t)$, the ROC curve, *ROC*(*t*) and *AUC*(*t*).

Conditional Kaplan –Meier (METHOD=KM)

Start with the definitions (1) to obtain

$$TP(c,t) = P[M > c \mid T \le t] = \frac{P[T \le t \mid M > c]P[M > c]}{P[T \le t]} = \frac{1 - S(t \mid M > c)}{1 - S(t)}(1 - F_M(c))$$

$$FP(c,t) = P[M > c \mid T > t] = \frac{P[T > t \mid M > c]P[M > c]}{P[T > t]} = \frac{S(t \mid M > c)}{S(t)}(1 - F_M(c)),$$

in terms of the survival distribution $S(t) = P[T > t]$, the conditional survival distribution $S(t \mid M > c)$ and the distribution of the marker $F_M(c) = P[M \le c]$. Estimation is simple plug-in with the Kaplan-Meier (KM) estimator of $S(t)$ from all data, for $S(t \mid M > c)$ using only the subsample, and $F_M(c)$ estimated by its empirical cumulative distribution function. This is essentially done in Lu and Liu (2006).

The KM estimators of $S(t \mid M > c)$ are based on different subsamples due to varying cut-off *c*. The resulting estimators of $TP(c,t)$ and $FP(c,t)$ need not be monotone in *c* and could take values outside [0, 1]. The estimated *ROC*(*t*) could also be non-monotone and take values outside the unit square. Heagerty *et al* (2000) give a numerical example to illustrate this issue. It is not necessarily seen only with small to moderate size samples.

Inverse-probability of censoring weighted (IPCW) *estimator* (METHOD=IPCW)

From definitions (1), $TP(c,t) = \frac{P[T \le t, M > c]}{P[T \le t]}$ and $FP(c,t) = \frac{P[T > t, M > c]}{P[T > t]}$. Use plug-in empirical estimates for the numerator and denominator of $TP(c,t)$, but weighted by the inverse probability of not being censored at time T_i^*, which is estimated by $\hat{S}_C(T_i^* -)$ where $\hat{S}_C(.)$ is the KM estimator of the censoring distribution, assumed independent of covariates.

Use the data $\{(T_i^*, \delta_i) : 1 \le i \le n\}$ with LIFETEST regarding censoring times as the response to obtain $\hat{S}_C(.)$. The explicit formulae are (Blanche *et al,* 2013a):

$$\widehat{TP}(c,t) = \frac{\sum_{i=1}^{n} [T_i^* \le t, M_i > c]\delta_i / \hat{S}_C(T_i^* -)}{\sum_{i=1}^{n} [T_i^* \le t]\delta_i / \hat{S}_C(T_i^* -)} \text{ and } \widehat{FP}(c,t) = \frac{\sum_{i=1}^{n} [T_i^* > t, M_i > c]}{\sum_{i=1}^{n} [T_i^* > t]}. \quad (3)$$

These IPCW estimators are monotone and bounded on [0. 1]. The weighting scheme was proposed by Uno *et al* (2007) motivated by $P[C \ge T \mid T, M] = S_C(T-)$ which results in $E\left\{\frac{\delta[T \le t, M > c]}{S_C(T-)}\right\} = P[T \le t, M > c]$. We obtain consistency and asymptotic normality under standard assumptions (Uno *et al*, 2007, Hung and Chiang, 2010a). Assume that the censoring distribution does not depend on the marker *M*. This is sufficient to demonstrate the convergence in probability, $\widehat{TP}(c,t) \to TP(c,t)$, $\widehat{FP}(c,t) \to FP(c,t)$. From (3) we also get an estimator $\widehat{AUC}(t)$ of $AUC(t)$. Hung and Chiang (2010b) and Blanche *et al* (2013a) give an explicit expression suggested from U-statistics:

$$\widehat{AUC}(t) = \frac{n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} [T_i^* \le t < T_j^*, M_i > M_j] \frac{\delta_i}{\hat{S}_C(T_i^* -)\hat{S}_C(t)}}{\hat{S}_T(t)(1 - \hat{S}_T(t))} \quad (4)$$

where $\hat{S}_T(.)$ estimates the event distribution, and the interpretation of $0/0$ as 0. For inference, we also need estimates of standard errors. PHREG implements a sophisticated perturbation-resampling method to compute standard error of the estimator of $AUC(t)$. The aforementioned papers mention the bootstrap to get estimates of standard errors. In the current version of PHREG (SAS/STAT® version 9.4, Analytics 14.2), the IPCW method is the best developed. Enhancements are planned in upcoming releases. Two other estimation methods offered by PHREG will be discussed later. The seminal article by the developers Guo, So and Jang (2017) is highly recommended.

3. Application and Illustration

For illustration we use a data set on the survival experience of 256 end-stage renal disease (ESRD) patients from the ADEMEX study (Vonesh, 2012, Paniagua *et al*, 2002). Patients were randomized to either high dose peritoneal dialysis (TRT=1) or standard dose (TRT=0). Covariates at baseline are patient age in years, gender, and diabetic status. Baseline and updated values of the glomerular filtration rate (ml/min), serum albumin (g/dL) and normalized nitrogen appearance (g/kg/day) were assessed. For our purposes only baseline values of these variables will be used. Survival time (ITTtime) is in months from randomization and ITTdeath is the censoring indicator, value 1 for death and 0 for censoring. Censoring could be for any one of the following reasons: (a) patient received kidney transplant, (b) return of kidney function, (c) true loss to follow up, or (d) reached study termination date. We consider a PHM with all seven of the aforementioned covariates. We request an ROC estimation based on the IPCW estimators of sensitivity and specificity (METHOD=IPCW). The marker is $M_i = \mathbf{x}_i' \hat{\beta}$ estimated from the PHM for each patient.

Apply the formats:

```
proc format;
 value trt 0='Control' 1='Treated';
 value sex 0='Male' 1='Female';
 value affirm 0='no' 1='yes';
run;

proc phreg data=survival_ph plots=roc
     rocoptions(method=ipcw at= 6 to 24 by 6 outroc=rocdata);
class trt (ref='Control') sex(ref='Male')
diabetic(ref='no')/param=ref;
model ITTtime*ITTdeath(0)=Trt Age Sex Diabetic Albumin0 nPNA0 GFR0;
format trt trt. sex sex. diabetic affirm.;
output out=stats_ph xbeta=xbeta;
run;
```

The data set STATS_PH has the values of the marker XBETA. The marker is practically continuous with no tied values. Cut-offs c for construction of the ROC are from this support set. Summary statistics for XBETA are

Analysis Variable : xbeta Linear Predictor							
N	Mean	Std Dev	Median	25th Pctl	75th Pctl	Minimum	Maximum
256	–1.8874	1.1337	–1.7475	–2.5544	–1.0783	–5.6830	1.0384

From the formula (3), ROCDATA saves the computation of the estimates of sensitivity (_Sensitivity_= $\widehat{TP}(c,t)$) and specificity (_Specificity_= 1− $\widehat{FP}(c,t)$) at the distinct marker values (_Cutoff_). Because of the request (**at= 6 to 24 by 6**) we have 4 sets of calculations for each of the requested times t. ROCDATA has 1024=256×4 records. The name ITTtime is applied to t. Although default ROC curves, $ROC(t)$ are produced by the **plots=roc** request, the data set ROCDATA has the requisite information for customized plotting. Points on the ROC curve are in descending order of the cutoff values going from left to right. If interested in calculating the estimate of $AUC(t)$ from ROCDATA, we must add to the calculation the area of the trapezoid at the extreme right, that is $\frac{1}{2}(1-\widehat{FP}(c_{\min},t))(1+\widehat{TP}(c_{\min},t))$ where $c_{\min}$ is the minimum marker value.

Figure 1 is obtained from the plot request **plots(overlay=individual)=roc(tick).**

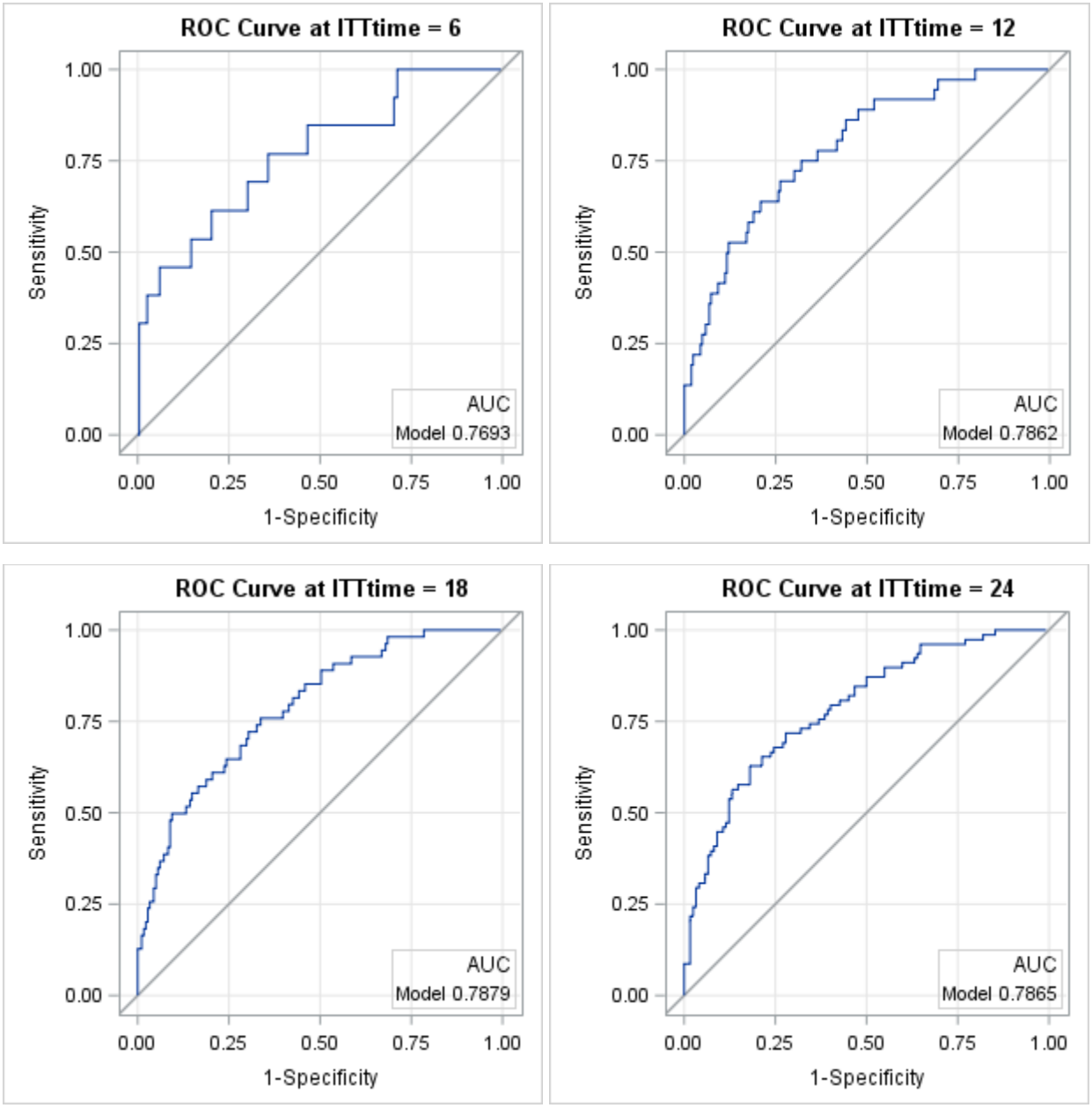
For each t, the estimated ROC curve $ROC(t)$ is a step function. Sensitivity values change at event times.

Estimation of $AUC(t)$

Although ROCDATA has the information to compute the estimate of $AUC(t)$ at the requested time points, it is far simpler to apply the options in PHREG statement: our PHM is the same.

```
proc phreg data=survival_ph plots=auc rocoptions(method=ipcw iauc
outauc=aucdata);
```

Figure 1: ROC(*t*) plots at *t* =6, 12, 18 and 24 months from the IPCW method



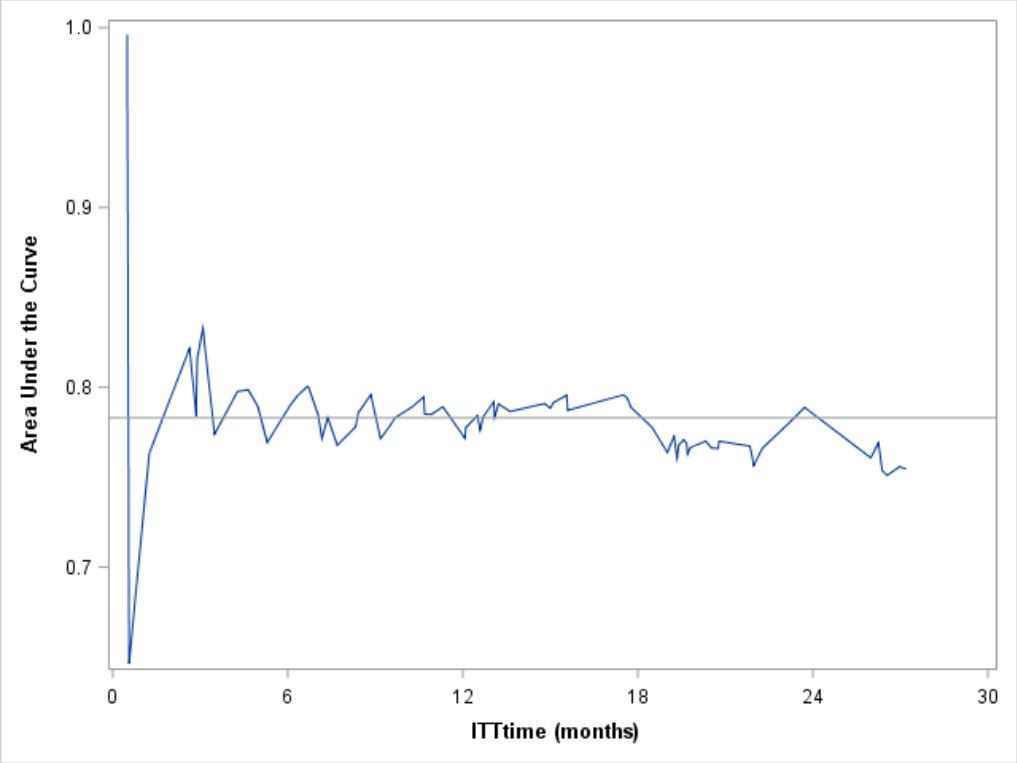
Note that the AT= option should not be used when OUTAUC= is requested. Estimates of *AUC* (*t*) (_AUC_) are computed at all event times (79 distinct times for 84 events) and saved in AUCDATA. Summary statistics are:

Analysis Variable : _AUC_ Area Under the Curve							
N	**Mean**	**Std Dev**	**Median**	**25th Pctl**	**75th Pctl**	**Minimum**	**Maximum**
79	0.7813	0.0325	0.7830	0.7680	0.7893	0.6462	0.996

The calculation is simply the sum of the trapezoidal areas under *ROC*(*t*). Formula (4) is not needed. A default plot is produced by the **`plots=auc`** request, or use AUCDATA to create a custom plot.

Figure 2 and the previous table show that $AUC(t)$ is approximately constant at the median 0.7830 (shown as a reference line) for times from 6 to 18 months.

Figure 2: Estimates of $AUC(t)$ at all event times from the IPCW method



A summary statistic $E\big(AUC(T)\big)$ is called the *integrated* (*time-dependent*) *area under the curve* (IAUC). It is estimated by a weighted sum of $\widehat{AUC}(t)$ with weights $w(t) = \hat{S}_T(t-) - \hat{S}_T(t)$ from the KM estimator of the event time distribution. The option **`iauc`** does the computation as a sum $\sum_t \widehat{AUC}(t)w(t)$ over the distinct event times. The estimate is 0.7795. The range of integration can be restricted to $(0, \tau)$ by the TAU= option.

Confidence intervals AUC

The IPCW method computes standard errors of the estimates of $AUC(t)$ by a sophisticated perturbation-resampling method. Let $\{\psi_i : 1 \le i \le n\}$ be independent and identically distributed (IID) exponential variates with mean=1. Modify formulae (3) as:

$$\widehat{TP}(c,t) = \frac{\sum_{i=1}^{n}[T_i^* \le t, M_i^* > c]\delta_i\psi_i / S_C^*(T_i^*-)}{\sum_{i=1}^{n}[T_i^* \le t]\delta_i\psi_i / S_C^*(T_i^*-)} \quad \text{and} \quad \widehat{FP}(c,t) = \frac{\sum_{i=1}^{n}[T_i^* > t, M_i^* > c]\psi_i}{\sum_{i=1}^{n}[T_i^* > t]\psi_i}$$

where $M_i^* = \mathbf{x}_i'\beta^*$, β^* and $S_C^*(.)$ are modified versions of $\hat{\beta}$ and $\hat{S}_C(.)$. See PHREG documentation for details. We get the standard error $\hat{\sigma}(t)$ of $\widehat{AUC}(t)$ based on a specified number (ITER=) of perturbed

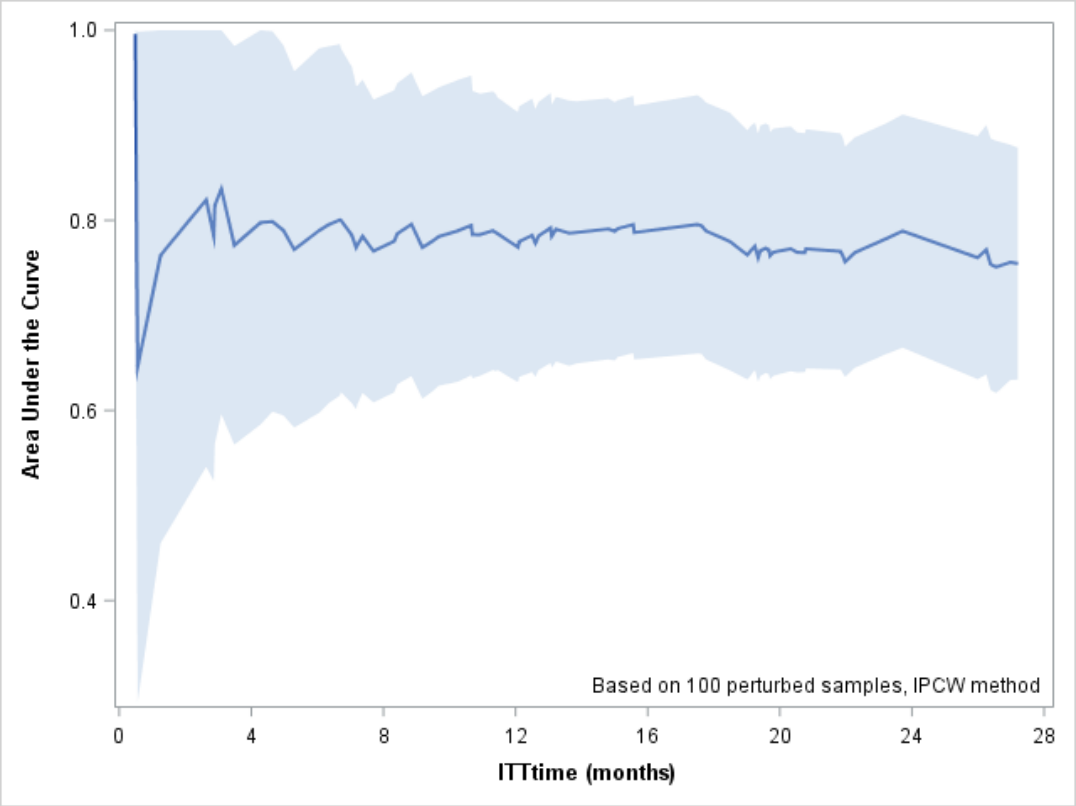
samples and subsequently a $100(1-\alpha)\%$ confidence interval for $AUC(t)$ computed as $\left(\widehat{AUC}(t)-z_{1-\frac{1}{2}\alpha}\hat{\sigma}(t), \widehat{AUC}(t)+z_{1-\frac{1}{2}\alpha}\hat{\sigma}(t)\right)$. Here $z_{1-\frac{1}{2}\alpha}$ denotes the $100(1-\frac{1}{2}\alpha)$-th percentile of the standard normal distribution. The following options save the calculation and produce a plot with the 95% confidence limits (pointwise). The defaults are ALPHA=.05 and ITER=50.

```
proc phreg data=survival_ph plots=auc rocoptions(method=ipcw(cl
iter=100 seed=13118) outauc=aucdata);
```

The **plots=auc** request generates an AUC plot similar to Figure 3 which is recreated from the AUCDATA set.

```
proc sgplot data=aucdata noautolegend;
series x=ITTtime y=_AUC_/lineattrs=(thickness=2);
band x=ITTtime lower=_lowerAUC_  upper=_upperAUC_/transparency=.5;
xaxis values=(0 to 28 by 4) label='ITTtime (months)'
labelattrs=(weight=bold);
yaxis labelattrs=(weight=bold);
inset "Based on 100 perturbed samples, IPCW method"
/position=bottomright;
run;
```

Figure 3: Estimates of $AUC(t)$ with 95% confidence intervals at all event times



Comparison of proportional hazards models

Our previous discussion employed a PHM with 7 covariates (predictors). Adjusted hazard ratios and 95% Wald confidence intervals are obtained via HAZARDRATIO statements (not shown) and corresponding p-values from Type3 Wald tests. For example, a 0.25-unit increase in baseline albumin is associated with a reduction in the risk of death by 29%, HR=0.71, 95% CI: 0.60, 0.84. Using the ASSESS statement, supremum tests for the functional form of the continuous covariates, and tests of the proportional hazards assumption on all covariates indicted no substantive violations of assumptions.

Table 1: Hazard ratios and 95% confidence intervals from the PHM

Description	Hazard Ratio	95% Lower CL	95% Upper CL	p-value
TRT, Treated vs Control	0.690	0.443	1.075	.1013
SEX, Female vs Male	1.191	0.750	1.892	.4588
DIABETIC, yes vs no	2.039	1.140	3.647	.0163
AGE, Unit=10	1.296	1.013	1.660	.0394
ALBUMIN0, Unit=0.5	0.708	0.595	0.843	.0001
nPNA0, Unit=0.25	0.608	0.459	0.807	.0006
GFR0, Unit=2	1.092	0.921	1.294	.3111

Consider a PHM with the predictors **diabetic age Albumin0 nPNA0 GFR0.** Call it our Full Model. Now consider sub-models with one or more predictors removed. Multiple ROC statements allow the fitting of the sub-models, a simple comparison of AUC differences and a graphic displaying all the ROC curves. The format for **`diabetic`** is not applied, due to a glitch in the software version used here.

```
proc phreg data=survival_ph plots(overlay=individual)=roc(tick)
rocoptions(method=ipcw(cl iter=50 seed=20918) at=12 aucdiff);
model ITTtime*ITTdeath(0)=diabetic age Albumin0 nPNA0 GFR0
/roclabel='Full Model';
roc "diabetic age Albumin0 nPNA0" diabetic age Albumin0 nPNA0;
roc "diabetic age Albumin0" diabetic age Albumin0;
roc "diabetic Albumin0 nPNA0" diabetic Albumin0 nPNA0;
roc "diabetic Albumin0 nPNA0 GFR0" diabetic Albumin0 nPNA0 GFR0;
run;
```

A busy Figure 4 shows ROC curves at t=12 for the 5 models. The plot options were used to insert the AUC values into the plot. The option **`aucdiff`** generates a table of differences in AUC for each pair of models. Models receive a label from their respective roc statement, and the label for the full model is supplied in the model statement via ROCLABEL. To suppress comparisons with the full model use the NOFIT option. Amongst comparisons to the Full model, a 4-variable sub-model without GFR0 would be sufficient. We noticed in Table 1 that GFR0 was not a significant predictor. This sub-model also has the highest AUC and does well in comparison with the other three sub-models. However, when a 95% CI for each AUC difference is computed by the UNO method (Table 2), we find that none are significant because the 95% CI straddles the zero value. Figure 4 says just about the same.

Figure 4: ROC curves at t =12 months for Full and Sub-models (IPCW method)

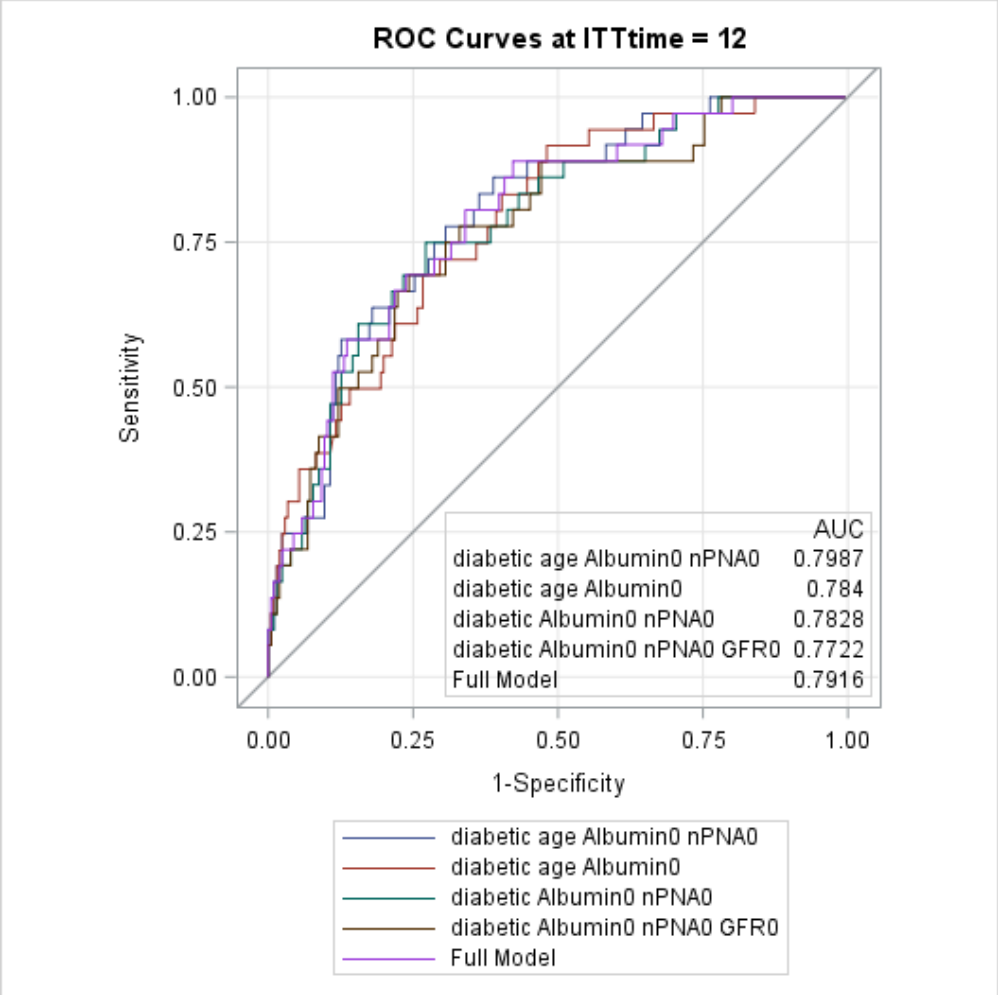


Table 2: Estimate at t =12 of AUC difference and 95% confidence limits for model pairs

PairID	Pair	AUC Estimate	Lower 95% CL	Upper 95% CL
1	diabetic age Albumin0 nPNA0 - diabetic age Albumin0	0.0147	–0.0279	0.0572
2	diabetic age Albumin0 nPNA0 - diabetic Albumin0 nPNA0	0.0158	–0.0507	0.0824
3	diabetic age Albumin0 nPNA0 - diabetic Albumin0 nPNA0 GFR0	0.0264	–0.0303	0.0831
4	diabetic age Albumin0 nPNA0 - Full Model	0.0071	–0.0153	0.0294
5	diabetic age Albumin0 - diabetic Albumin0 nPNA0	0.0012	–0.0679	0.0702
6	diabetic age Albumin0 - diabetic Albumin0 nPNA0 GFR0	0.0118	–0.0507	0.0743
7	diabetic age Albumin0 - Full Model	–0.0076	–0.0582	0.0430
8	diabetic Albumin0 nPNA0 - diabetic Albumin0 nPNA0 GFR0	0.0106	–0.0206	0.0418
9	diabetic Albumin0 nPNA0 - Full Model	–0.0088	–0.0754	0.0579
10	diabetic Albumin0 nPNA0 GFR0 - Full Model	–0.0194	–0.0707	0.0319

ROC analysis with marker data from an external source

As seen in formulae (3) for $\widehat{TP}(c,t)$ and $\widehat{FP}(c,t)$ we need only information on survival data $\{(T_i^*,\delta_i):1\le i\le n\}$ and the marker $\{M_i:1\le i\le n\}$. The latter can be generated from another source, for example an accelerated failure time (AFT) model fitted in LIFEREG or a scale parameter model fitted in SEVERITY. The AFT is a parametric model with the form $\log(T)=\mathbf{x}'\beta+\sigma\varepsilon$ where ε has a specified parametric distribution. When ε has the logistic distribution we get the log-logistic survival distribution $S(t\,|\,\mathbf{x})=\left(1+\{t/\theta(\mathbf{x})\}^{1/\sigma}\right)^{-1}$; when ε has the extreme-value distribution we get the Weibull survival distribution $S(t\,|\,\mathbf{x})=\exp\left((-t/\theta(\mathbf{x}))^{1/\sigma}\right)$ where $\sigma>0$ is a shape parameter and $\log\theta(\mathbf{x})=\mathbf{x}'\beta$ (Gardiner *et al*, 2014). The Weibull has the proportional hazards property but the log-logistic does not. In AFT models a higher value of the marker $M=\mathbf{x}'\beta$ is indicative of better survival. Therefore, we must flip the sign of the marker to keep in line with our previous discussion of the Cox PHM.

The following syntax fits the log-logistic model with all 7 covariates. Change the DIST option to **dist=Weibull** to fit the Weibull.

```
proc lifereg data=survival_ph;
class trt sex diabetic;
format trt trt. sex sex. diabetic affirm.;
model ITTtime*ITTdeath(0)=Trt Age Sex Diabetic Albumin0 nPNA0
GFR0/dist=llogistic;
output out=stats_lr xbeta=xbeta_l;
run;
```

The Weibull shows close similarity with our previous PHM (Table 3).

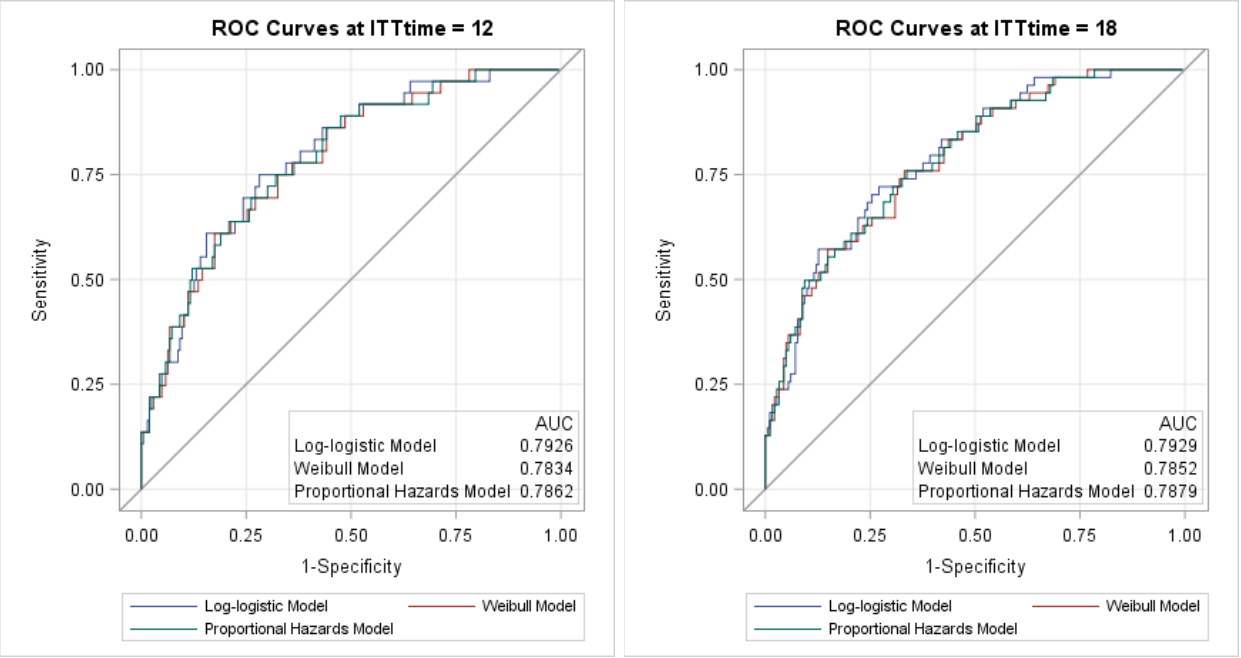
Table 3: Type III Analysis of Effects

		LOG-LOGISTIC MODEL		WEIBULL MODEL		Cox PH Model	
Effect	**DF**	**Wald Chi-Square**	**p-value**	**Wald Chi-Square**	**p-value**	**Wald Chi-Square**	**p-value**
Trt	1	1.8151	0.1779	3.1390	0.0764	2.6854	0.1013
Age	1	3.5791	0.0585	3.7040	0.0543	4.2430	0.0394
Sex	1	0.3320	0.5645	0.3487	0.5549	0.5489	0.4588
Diabetic	1	4.8499	0.0276	6.3937	0.0115	5.7667	0.0163
Albumin0	1	16.5919	<.0001	13.4807	0.0002	15.0446	0.0001
nPNA0	1	9.4132	0.0022	11.3234	0.0008	11.9177	0.0006
GFR0	1	0.2750	0.6000	0.9932	0.3190	1.0261	0.3111

Create the marker data for log-logistic and Weibull models in the same data set. This may be done by (1) fit log-logistic and save XBETA_l, (2) use the output file to fit Weibull and save XBETA_w, (3) flip sign to get marker_l and marker_w in the data set **stats_lr3**. The now familiar syntax produces the ROC curves at two time points for the three models (Figure 5).

```
proc phreg data=stats_lr3 plots(overlay=individual)=roc(tick)
rocoptions(method=ipcw at=12,18);
class trt (ref='Control') sex(ref='Male') diabetic(ref='no')
/param=ref;
format trt trt. sex sex. diabetic affirm.;
model ITTtime*ITTdeath(0)=Trt Age Sex Diabetic Albumin0 nPNA0
GFR0/roclabel='Proportional Hazards Model';
roc "Log-logistic Model" pred=marker_l;
roc "Weibull Model" pred=marker_w;
run;
```

Figure 5: ROC curves at t=12, 18 months for Log-logistic, Weibull and Cox PH models



Checking independence of the censoring distribution and marker

The IPCW method assumes that the censoring distribution is not dependent on the marker *M*. An informal check of this assumption can be made by plotting the KM estimates $\hat{S}_C(t)$ in strata defined by the values of *M*. First, fit a PHM and save the marker data. Second, create categories for the distribution of *M*. We use tertiles to create 3 strata. Third, obtain KM estimates from LIFETEST with a STRATA statement.

The PHM fitted is

```
proc phreg data=survival_ph;
class diabetic(ref='no')/param=ref;
model ITTtime*ITTdeath(0)= Diabetic Age Albumin0 nPNA0;
format diabetic affirm.;
output out=stats_ph xbeta=xbeta;
run;
```

Tertiles of the marker are obtained from UNIVARIATE:

```
proc univariate data=stats_ph;
var xbeta;
output out=xb_pctl pctlpts=( 33 66) pctlpre=P_;
run;
```

Issue a format to create 3 tertiles categories T1, T2, T3:

```
proc format;
value trtl low-<-2.26350='T1' -2.26350-< -1.23428 ='T2'
                -1.23428-high='T3';
run;
```

Obtain KM estimates for the event time distribution and separately for the censoring time distribution:

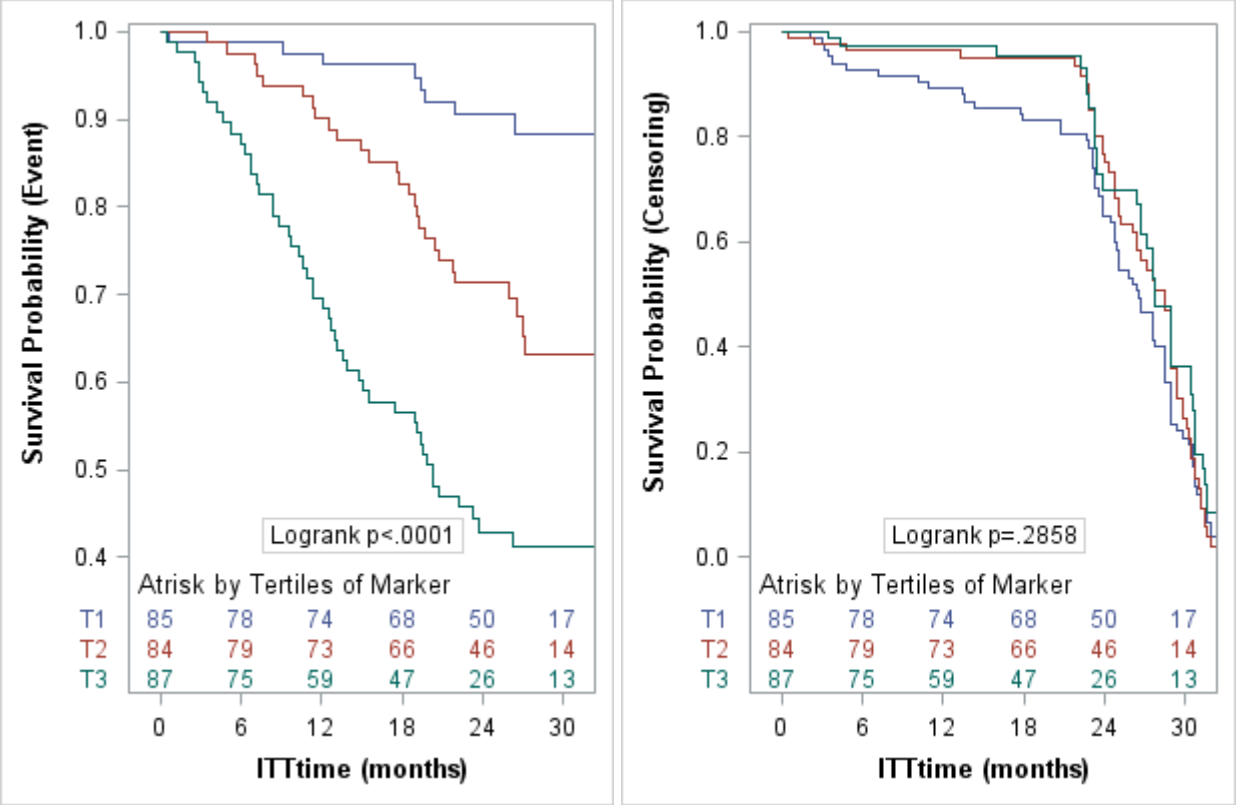
```
ods output survivalplot=surv_T; /*surv_C*/
proc lifetest data=stats_ph
     plots=survival(atrisk=(0 to 30 by 6) atrisktickonly nocensor
test);
     strata xbeta/test=logrank;
format xbeta trtl.;
time ITTtime*ITTdeath(0); *ITTtime*ITTdeath(1);
label ITTtime='ITTtime (months)';
run;
```

For purposes of plotting, two ODS output data sets are created. Two invocations of SGPLOT will create an enhanced plot of the event time distribution and censoring distribution (Matange, 2016). The syntax for the latter is:

```
ods graphics on/height=4.2in width=3.2in;
proc sgplot data=surv_C noautolegend;
step x=time y=survival /group=stratum lineattrs=(thickness=1);
xaxistable atrisk/x=tatrisk class=stratum location=inside
                          colorgroup=stratum valueattrs=(size=9)
                          title=" Atrisk by Tertiles of Marker";
xaxis values=(0 to 30 by 6) labelattrs=(weight=bold);
yaxis label="Survival Probability (Censoring)"
labelattrs=(weight=bold);
inset "Logrank p=.2858" / border textattrs=(size=9) position=s;
run;
```

The plots are in Figure 6. To show detail, the vertical scale is different for the two plots. We expect a significant association of the marker with the event time (left panel), and fortunately we find no association of the marker with the censoring time (right panel)—at least from the informal investigation made here. Assessments are based on the log-rank test.

Figure 6: KM estimates of event time and censoring distributions by marker categories



When the censoring distribution is dependent on the marker, some arguments of the IPCW method for consistency of estimators break down. Blanche *et al* (2013a) offer a modification of the IPCW method called the conditional IPCW (CIPCW) which replaces the observation weights in formulae (3) for $\widehat{TP}(c,t)$ by $\delta_i \left\{\hat{S}_C(T_i^* - \mid M_i)\right\}^{-1}$ and introduces a weight $\left\{\hat{S}_C(t \mid M_i)\right\}^{-1}$ in $\widehat{FP}(c,t)$. They suggest a Cox PMH or any other model be applied to estimate $S_C(t \mid M)$ or preferably from the bivariate distribution $P[C > t, M > c]$. Via simulation studies, the CIPCW estimators show robustness to dependency of the censoring distribution on the marker and perform well in comparison to the nearest neighbor method which is the default method in PHREG.

Nearest Neighbor Method (METHOD=NNE)

PHREG offers METHOD=NNE for the nearest neighbor approach (Heagerty *et al*, 2000) based on a bivariate survival distribution for (*M*, *T*) introduced by Akritas (1994). The NNE method is not currently fully developed: we can get the relevant estimates for ROC analysis, but standard errors are not available for the estimator of $AUC(t)$. Consistency and asymptotic normality have been established, but the difficulty lies in estimating the asymptotic variance. Resampling methods could be used (Hung and Chiang, 2010a, 2011).

Using the data $\{(T_i^*, \delta_i, \mathbf{x}_i) : 1 \le i \le n\}$ with the marker $M_i = \mathbf{x}_i'\hat{\beta}$, estimate the bivariate survival distribution of $S(c,t) = P[M > c, T > t]$ by $\hat{S}_{b_n}(c,t) = n^{-1}\sum_{i=1}^{n} \hat{S}_{b_n}(t \mid M = M_i)[M_i > c]$. The conditional survival distribution is estimated by a kernel-smoothed weighted KM-estimator,

$$\hat{S}_{b_n}(t \mid M = M_i) = \prod_{s \le t}\left[1 - \frac{\sum_j K_{b_n}(M_i, M_j)[T_j^* = s]\delta_j}{\sum_j K_{b_n}(M_i, M_j)[T_j^* \ge s]}\right]$$

where $K(M_i, M_j) = [\,|F_M(M_i) - F_M(M_j)| < b_n\,]$, $0 < b_n < ½$ and $F_M(.)$ is the cumulative distribution function (CDF) of the marker M. Ties amongst observed times are allowed (just as in KM). At the marker value M_i, $\sum_j K_{b_n}(M_i, M_j)[T_j^* = s]\delta_j$ counts events at time $T_j^* = s$ whose associated marker value M_j is to close M_i; also $\sum_j K_{b_n}(M_i, M_j)[T_j^* \ge s]$ is the risk set at time s, but restricted to individuals whose maker value is close to M_i. By default bandwidth b_n =0.05, so that 10% of nearest neighbors are used. Use option SPAN= to change setting. A suggested choice is $b_n = O(n^{-1/3})$.

True Positive Ratio (*Sensitivity*) *and False Positive Ratio* (1−*Specificity*)

Directly from their definitions we get the corresponding estimators

$$\widehat{TP}(c,t) = \frac{1 - F_M(c) - \hat{S}_{b_n}(c,t)}{1 - \hat{S}_{b_n}(-\infty, t)},\ \widehat{FP}(c,t) = \frac{\hat{S}_{b_n}(c,t)}{\hat{S}_{b_n}(-\infty, t)}.$$

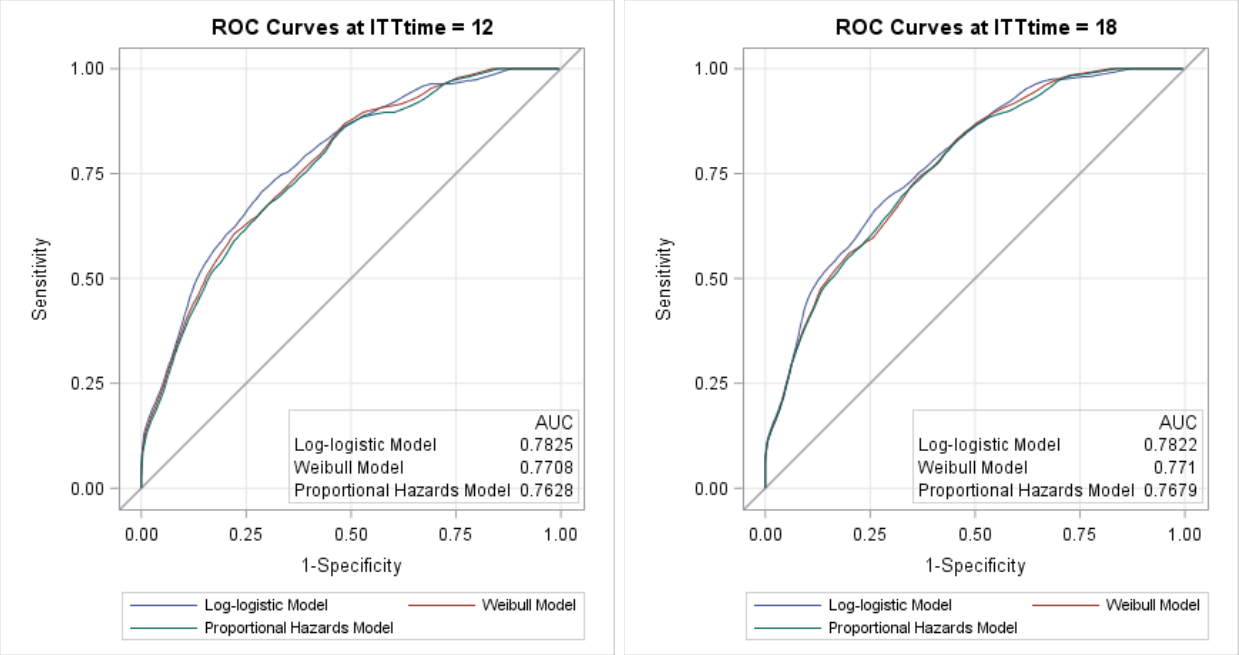
These estimators are monotone in c and bounded on [0.1].

Consider the same models described in the previous section for the log-logistic, Weibull, and Cox PHM with covariates **`Trt Age Sex Diabetic Albumin0 nPNA0 GFR0`**. ROC analysis using the NNE method requires only one change in the **`rocoptions.`**

```
proc phreg data=stats_lr3 plots(overlay=individual)=roc(tick)
rocoptions(method=NNE at=12, 18 outroc=rocdata);
class trt (ref='Control') sex(ref='Male')
diabetic(ref='no')/param=ref;
format trt trt. sex sex. diabetic affirm.;
model ITTtime*ITTdeath(0)=Trt Age Sex Diabetic Albumin0 nPNA0
GFR0/roclabel='Proportional Hazards Model';
roc "Log-logistic Model" pred=marker_l;
roc "Weibull Model" pred=marker_w;
run;
```

Figure 7 plots the ROC curves. Notice that the NNE method produces smoother curves than the IPCW method. The AUC can be computed by the trapezoidal rule from the output data set ROCDATA.

Figure 7: ROC curves at t=12, 18 by the NNE method



Recursive Method (METHOD=RECURSIVE)

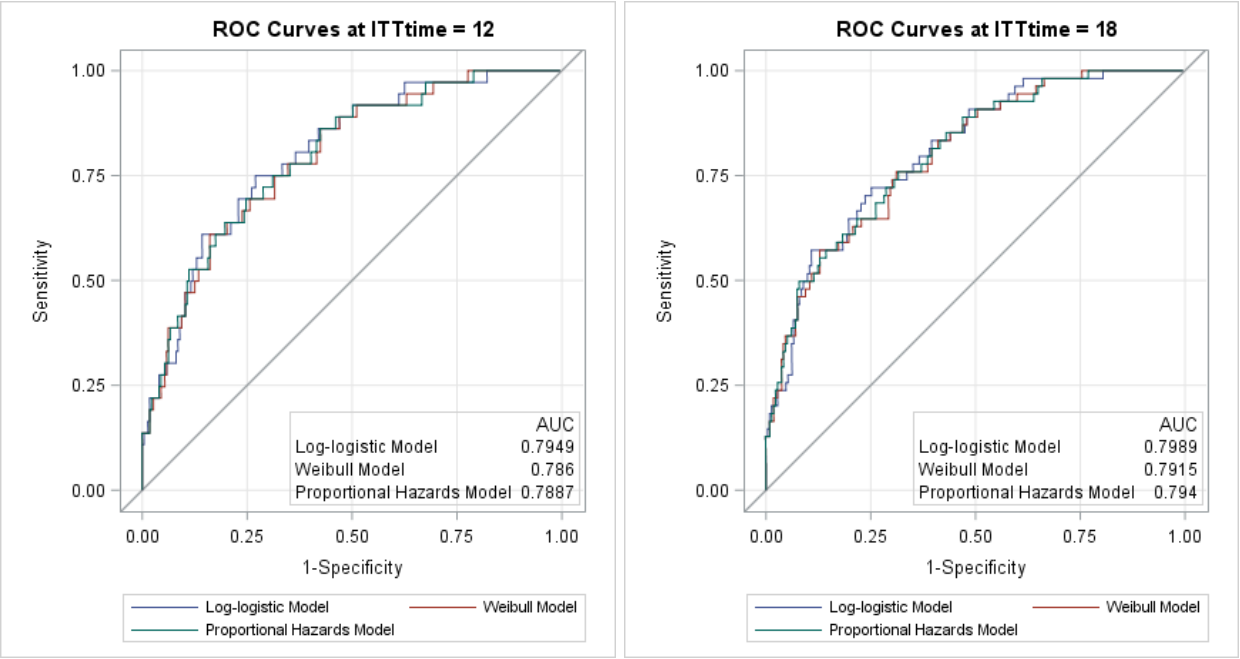
This method proposed by Chambless and Diao (2006) for estimation of $TP(c,t), FP(c,t), AUC(t)$ follows a recursive computation using the ordered distinct survival times up to t. Explicit formulae are available for the three estimators. Unfortunately, the estimator of $FP(c,t)$ need not be monotone in c or bounded in [0,1]. The $TP(c,t)$ estimator is however, monotone and bounded. The censoring distribution is assumed not to depend on the marker.

Let $0 = t_0 < t_1 < t_2 < \ldots < t_K$ denote the distinct event times. At time t_k, let d_k =#events and r_k =#at risk. For the set D_k of subjects with events at t_k, let $\rho_k(c) = \#\{i \in D_k : M_i > c\} = \sum_{i=1}^{n} \delta_i [M_i > c, T_i^* = t_k]$. The estimators of true positive and false positive ratios at $t_m, m = 1, \ldots, K$

$$\widehat{TP}(c,t_m) = \frac{\sum_{k=1}^{m} \rho_k(c)\hat{S}(t_{k-1})/r_k}{1-\hat{S}(t_m)}, \quad \widehat{FP}(c,t_m) = \frac{1-F_M(c)-\sum_{k=1}^{m} \rho_k(c)\hat{S}(t_{k-1})/r_k}{\hat{S}(t_m)}$$

where $\hat{S}$ is the KM estimator. We get estimates of the marker $M_i = \mathbf{x}_i'\hat{\beta}$ from a survival model. Note that $\widehat{TP}(c,t_m) \le 1$ follows from $\rho_k(c) \le d_k$ and $\hat{S}(t_{k-1}) - \hat{S}(t_k) = \hat{S}(t_{k-1})d_k / r_k$. We cannot guarantee that $\widehat{FP}(c,t_m)$ is monotone or bounded on [0, 1] although $\widehat{FP}(-\infty,t_m) = 1$ and $\widehat{FP}(+\infty,t_m) = 0$. Figure 8 shows ROC plots that are very similar to the NNE method (Figure 7) and IPCW method (Figure 5).

Figure 8: ROC curves at t=12, 18 by the RECURSIVE method



Concordance Analysis

For a binary outcome the area under the ROC curve is the c-statistic. If F and G denote the CDFs of the marginal distributions of the markers (X,Y) from the diseased $(D_X=1)$ and non-diseased populations $(D_Y=0)$, respectively, then $P[X>Y]=\int_{-\infty}^{\infty}(1-F(u))dG(u)$. Evaluation of the integral in terms of $TP(c),FP(c)$ shows that the integral is precisely $AUC=\int_0^1 TP\left(FP^{-1}(u)\right)du$ (Vexler, 2016).

In general, for independent pairs $(M_1,T_1),(M_2,T_2)$ from survival data the *concordance index* is defined as $C_U=P[M_1>M_2\,|\,T_1<T_2]$ (Heagerty *et al*, 2005), Uno *et al*, 2011). For continuous distributions, IID random variables, $P[T_1<T_2]=½$ and $C_U=2P[M_1>M_2,T_1<T_2]=P[(M_1-M_2)(T_1-T_2)<0]$.

Kendall's Tau (K) is defined as the difference of probability of 'concordance' and of probability of 'discordance', $K=P[(M_1-M_2)(T_1-T_2)>0]-P[(M_1-M_2)(T_1-T_2)<0]=2P[(M_1-M_2)(T_1-T_2)>0]-1$ for continuous distributions (Nelson, 2006). For what we call 'concordance' in the survival context, $C_U=½(K+1)$.

With survival data we must address censoring which makes it impossible to order event times beyond the last follow up time. Uno defines a truncated version which we use in the sequel:
$C_U(\tau)=P[M_1>M_2\,|\,T_1<T_2,T_1<\tau]$ where τ is a value in the support of the censoring distribution. To allow for ties in the marker define (Gerds *et al,* 2013)
$C_U(\tau)=P[M_1>M_2\,|\,T_1<T_2,T_1<\tau]+½P[M_1=M_2\,|\,T_1<T_2,T_1<\tau]$ and a similar modification for the area under the curve, $AUC(t)=P[M_1>M_2\,|\,T_1\le t<T_2]+½P[M_1=M_2\,|\,T_1\le t<T_2]$.

The concordance index $C_U(\tau)$ quantifies the ability of the marker to order the events times up to τ whereas $AUC(t)$ quantifies the ability of the marker to order the event status at time *t*. Both measures can be used to inform discriminative performance of prediction models. An interesting discussion on the merits of these measures has been initiated by Blanche *et al* (2016).

Uno's method: The censoring distribution is assumed not to depend on the marker. An estimator of $C_U(\tau)$ is obtained as

$$\hat{C}_U(\tau)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}[T_i^*<T_j^*, T_i^*<\tau]\left([M_i>M_j]+\tfrac{1}{2}[M_i=M_j]\right)w_i}{\sum_{i=1}^{n}\sum_{j=1}^{n}[T_i^*<T_j^*, T_i^*<\tau]w_i} \text{ where } w_i=\frac{\delta_i}{\hat{S}_C(T_i^*-)\hat{S}_C(T_i^*)} .$$

Because $\hat{S}_C(t)$ changes only at censoring times, and the calculation is made at event times the distinction in $\hat{S}_C(T_i^*-)$ and $\hat{S}_C(T_i^*)$ will apply only when the event time is tied with a censoring time. We also want $\hat{S}_C(T_i^*)>0$. It is assured by $\hat{S}_C(\tau)>0$, if τ is in the support of the censoring distribution. If not specified by the option TAU=, the maximum event time is used. Asymptotic theory ensures the convergence $\sqrt{n}\left(\hat{C}_U-C_U\right)\rightarrow NORMAL(0,\sigma^2)$ where σ^2 can be estimated by the perturbation-resampling method similar to that applied to get confidence intervals for $AUC(t)$. Table 4 assembles the results from several calls to PHREG. (Value-list in the option TAU= is not currently available).

```
ods output auc=auc concordance=concordance;
proc phreg data=survival_ph concordance=UNO(SE iter=100 seed=22018)
tau=12 /* 18 24 28 */
rocoptions(method=ipcw(cl iter=100 seed=22018) auc at=12 18 24
27.2039);
class trt(ref='Control')sex(ref='Male')diabetic(ref='no')
/param=ref;
model ITTtime*ITTdeath(0)=Trt Age Sex Diabetic Albumin0 nPNA0 GFR0;
format trt trt. sex sex. diabetic affirm.;
run;
```

Table 4: Estimates and 95% confidence intervals for the Concordance Index and AUC

	Concordance Index				Area Under the Curve at Tau			
Tau	Estimate	Standard Error	95% LCL	95% UCL	Estimate	Standard Error	95% LCL	95% UCL
12	0.7729	0.0370	0.7005	0.8454	0.7862	0.0374	0.7129	0.8594
18	0.7665	0.0343	0.6993	0.8337	0.7879	0.0361	0.7171	0.8586
24	0.7466	0.0311	0.6857	0.8076	0.7865	0.0321	0.7237	0.8493
28*	0.7324	0.0326	0.6684	0.7964	0.7561	0.0393	0.6791	0.8331

*Tau=28 is proxy for the maximum event time 27.204 months.

The similarity of the results for the two measures is assuring, but this could be a characteristic of this data set. Just as in Table 2 for the AUC(*t*) we can carry out comparisons based on the concordance index for sub-models of the main PHM. The syntax is analogous: we will get a table of estimated differences in the concordance index, the standard error and p-value for testing a null difference for each of the 10 individual pairs of the five models (results not shown).

```
proc phreg data=survival_ph concordance=uno(diff SE iter=100
seed=22018) tau=12;
model ITTtime*ITTdeath(0)=diabetic age Albumin0 nPNA0
GFR0/roclabel='Full Model';
roc "diabetic age Albumin0 nPNA0" diabetic age Albumin0 nPNA0;
roc "diabetic age Albumin0" diabetic age Albumin0;
roc "diabetic Albumin0 nPNA0" diabetic Albumin0 nPNA0;
roc "diabetic Albumin0 nPNA0 GFR0" diabetic Albumin0 nPNA0 GFR0;
run;
```

Harrell's c-index (C_H) is estimated by the option **concordance=Harrell(SE)**, with standard errors but no comparisons are made between models. Harrell's c-index is defined as the proportion of useable subject pairs in which the event times and markers are concordant (Harrell *et al,* 1996; Harrell, 2015). Formally, defined as $C_H = P[M_1 > M_2 \mid T_1 < T_2, T_1 < \min(\tilde{C}_1, \tilde{C}_2)]$ where $\tilde{C}_1, \tilde{C}_2$ are the censoring times. An useable pair (T_i^*, T_j^*) comprises of either (i) distinct values with the lower value being an event say $\delta_i = 1$ and $T_i^* < T_j^*$, or (ii) tied values $T_i^* = T_j^*$ with exactly one being an event, i.e. $(\delta_i, \delta_j) = (1,0)$ or $(\delta_i, \delta_j) = (0,1)$. Higher marker values are associated with shorter 'survival'. A pair is concordant (or discordant) if the marker is larger (or smaller) for the event, than for the comparator. An estimator is $\hat{C}_H = (n_c + \frac{1}{2} n_m) / (n_c + n_d + n_m)$ where $n_c = \#$ concordant pairs, $n_d = \#$ discordant pairs $n_m = \#$ tied in marker pairs. This expression is entirely analogous to the calculation of c-statistic for a prediction model from logistic regression for a binary outcome in LOGISTIC or HPLOGISTIC.

Concluding Remarks

Of the four methods available in PHREG for estimating $TP(c,t), FP(c,t)$ the IPCW method is best developed. Estimators are monotone in the cutoff c, and bounded in [0,1]. For sub-models of the main PHM, we can use ROC statements to make comparisons between models based on $AUC(t)$, and obtain confidence intervals for the differences. However, the method assumes that censoring does not depend on the marker. The KM method may lead to estimators of $TP(c,t), FP(c,t)$ that are not-monotone in c or bounded in [0,1]. With the RECURSIVE method we can guarantee the monotonicity and boundedness for the estimator of $TP(c,t)$ only. The NNE method is based on the joint survival distribution of (M,T) and produces estimators of $TP(c,t), FP(c,t)$ with the monotonicity and boundedness properties. The censoring distribution may depend on the marker. Finally, when censored observations are absent in our survival data, the IPCW, KM and RECURSIVE methods lead to the usual empirical estimators based on definition (1). Robustness to marker dependent censoring is a desirable property in applications to observational studies. With administrative censoring in trial data, the censoring distribution is plausibly independent of covariates, but may not be so with loss to follow up or withdrawals from study.

Throughout, we adopted the cumulative/dynamic definition of $TP(c,t), FP(c,t)$ as implemented in PHREG. See Heagerty and Zheng (2005) for two other definitions, and a recent review by Kamarudin *et al* (2017). Although the focus is on a single endpoint in the time-to-failure analysis, information on the type or cause of failure leads to competing risks analysis (Beyersmann and Scheike, 2014, Gardiner, 2016). PHREG offers two methods for competing risks analysis, one based on modelling the cause-

specific hazard functions (Andersen *et al*, 2002, Anderson and Keiding, 2012) and another based on modelling the sub-distribution hazards (Fine and Gray, 1999). Blanche *et al* (2013b) provide an approach to ROC analysis in competing risks models. Enhancements to PHREG will likely offer options to extend the reach of the current methodology.

References

Akritas MG. Nearest-neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics.* 1994; 22(3):1299-1327.

Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. *Statistical Methods Medical Research.* 2002;11(2):203-215.

Andersen PK, Keiding N. Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine.* 2012; 31(11-12):1074-1088.

Beyersmann J, Scheike TH. Classical Regression Models for Competing Risks. In: Klein JP, VanHouwelingen HC, Ibrahim JG, Scheike TH, eds. *Handbook of Survival Analysis.* Boca Raton, FL: CRC Press; 2014:157-177.

Blanche P, Dartigues J-F, Jacqmin-Gadda H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal.* 2013a; 55(5):687-704.

Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine.* 2013b; 32(30):5381-5397.

Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of t-year predicted risks 2016, University of Copenhagen, Denmark.

Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine.* 2006; 25(20):3474-3486.

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association.* 1999; 94(446):496-509.

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

Gardiner JC, Reed PL, Bonner JD, Haggerty DK, Hale DG. Incidence of hospital-acquired pressure ulcers - a population-based cohort study. *International Wound Journal.* 2016; 13(5):809-820.

Gardiner JC. Competing Risks Analysis: Overview of Regression models. SAS Global Forum 2016; Las Vegas, NV.

Gerds TA, Kattan MW, Schumacher M, Yu C. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine.* 2013; 32(13): 2173-2184.

Guo C, So Y, Jang W. Evaluating Predictive Accuracy of Survival Models with PROC PHREG. SAS Global Forum; 2017; Orlando, FL.

Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine.*1996; 15(4):361-387.

Harrell FE. *Regression Modeling Strategies: With Application to Linear Models, Logistic Regression, and Survival Analysis. Second Edition* New York, NY: Springer-Verlag; 2015.

Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics.* 2000; 56(2):337-344.

Heagerty PJ, Zheng YY. Survival model predictive accuracy and ROC curves. *Biometrics.* 2005; 61(1):92-105.

Hung H, Chiang C-T. Optimal Composite Markers for Time-Dependent Receiver Operating Characteristic Curves with Censored Survival Data. *Scandinavian Journal of Statistics.* 2010a; 37(4):664-679.

Hung H, Chiang C-T. Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics-Revue Canadienne De Statistique.* 2010b; 38(1):8-26.

Hung H, Chiang C-T. Nonparametric methodology for the time-dependent partial area under the ROC curve. *Journal of Statistical Planning and Inference.* 2011; 141(12):3829-3838.

Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Medical Research Methodology.* 2017; 17.

Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. *Blood.* 2008; 111(10):4902-4907.

Lu L, Liu C. Using the Time Dependent ROC Curve to Build Better Survival Model in SAS. Northeast SAS User Group; 2006; Philadelphia, PA.

Matange S. *Clinical Graphs Using SAS*. Cary, NC: SAS Institute Inc; 2016.

Nelson R. *An Introduction to Copulas, Second Edition.* New York, NY: Springer-Verlag; 2006.

Paniagua R, Amato D, Vonesh E, et al. Effects of increased peritoneal clearances on mortality rates in peritoneal dialysis: ADEMEX, a prospective, randomized, controlled trial. *Journal of the American Society of Nephrology.* 2002; 13(5).

Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* New York, NY: Oxford University Press; 2003.

Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating* New York: Springer-Verlag; 2009.

Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association.* 2007; 102(478):527-537.

Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine.* 2011; 30(10):1105-1117.

Vexler A, Hutson AD, Chen X. *Statistical Testing Strategies in the Health Sciences.* Boca Raton, FL: CRC Press; 2016.

Vonesh EF. *Generalized Linear and Nonlinear Models for Correlated Data: Theory and Applications Using SAS*. Cary, NC: SAS Institute Inc; 2012.

ACKNOWLEDGMENTS

We thank Changbin Guo and Ying So of the SAS Institute for clarifying the options in PROC PHREG.

CONTACT INFORMATION

We welcome your comments and questions. Please contact

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