10 Time Dependent Covariates

Since survival data occur over time, important covariates we wish to consider may also change over time. We refer to these as time-dependent covariates. Examples of such covariates are:

- cumulative exposure to some risk factor,
- smoking status,
- heart (kidney) transplant status:

  0 prior to heart (kidney) transplant

  1 after heart (kidney) transplant

- blood pressure.

We may have a vector of such covariates, which for the ith individual in our sample we denote by $Z_i(t) = (Z_{i1}(t), ..., Z_{iq}(t))^T$, corresponding to the value of these covariates at time t. This notation allows us to use time independent covariates as well, For example, if the jth covariate is time-independent, then $Z_{ij}(t)$ is constant over time.

Modeling the hazard rate is a natural way of thinking about time-dependent covariates. If we let $Z_i^H(t)$ denote the history of the vector of the time-dependent covariates up to time t, *i.e.*, $Z_i^H(t) = \{Z_i(u), 0 \leq u \leq t\}$, then we can define the hazard rate at time t conditional on this history by

$$\lambda(t|Z_i^H(t)) = \lim_{h \to 0} \frac{P[t \leq T_i < t + h | T_i \geq t, Z_i^H(t)]}{h}.$$

This is the instantaneous rate of failure at time t, given the individual was at risk at time t with a history $Z_i^H(t)$. For such a conditional hazard rate, we may consider a proportional hazards model

$$\lambda(t|Z_i^H(t)) = \lambda_0(t)\exp(\beta^T g(Z_i^H(t))),$$

where $g(Z_i^H(t))$ is a vector of function of the history of the covariates that we feel may affect the hazard.

For example, one choice is to use

$$g(Z_i^H(t)) = Z_i(t).$$

If we assume that

$$\lambda(t|Z_i^H(t)) = \lambda_0(t)\exp(\beta^T Z_i(t)),$$

then implicitly we would be assuming that the hazard rate at time t given the entire history of the covariates up to time t is only effected by the current values of the covariates at time t. This, of course, may or may not be true. Some thought should be given when entertaining use of these models.

For example, suppose we want to consider the effect of exposure to asbestos over time on mortality. A sample of workers in a factory where asbestos is made were monitored for a period of time and data were collected on survival and asbestos exposure. For the ith individual in the sample, the data could be summarized as

$$(X_i, \Delta_i, Z_i^H(X_i)),$$

where

- $X_i = \min(T_i, C_i)$ is the observed survival time or censoring time,
- $\Delta_i = I(T_i \leq C_i)$ is the failure indicator,
- $Z_i^H(X_i)$ is the history of asbestos exposure up to time X_i. This may, for example, be daily exposure collected on that individual every six months. This data may be collected up to the point that patient dies, or until he/she is censored, or until he/she stops working at the factory.

Suppose we wish to consider the following proportional hazards model with time-dependent covariates

$$\lambda(t|Z_i^H(t)) = \lambda_0(t)\exp(\beta^T g(Z_i^H(t))).$$

What should we use for the function $g(Z_i^H(t))$?

1. We may use cumulative exposure (need extrapolation), *i.e.*,

$$g(Z_i^H(t)) = \sum_j Z_i(u_{ij})(u_{ij} - u_{i(j-1)}),$$

   where u_{ij} are days at which measurements were made prior to day t.

2. We may use average exposure up to time t,

$$g(Z_i^H(t)) = \frac{\sum_{u_{ij}<t} Z_i(u_{ij})}{\#\text{ of measurements up to } t}$$

3. We may use maximum exposure up to time t,

$$g(Z_i^H(t)) = \max\{Z_i(u_{ij}) : u_{ij} < t\}$$

We may also want to consider models such as

$$\lambda(t|Z_i^H(t)) = \lambda_0(t)\exp(\beta_1 g_1(Z_i^H(t)) + \beta_2 g_2(Z_i^H(t))),$$

where $g_1(Z_i^H(t))$ = cumulative exposure up to time t and $g_2(Z_i^H(t))$ = maximum exposure up to time t.

This model may be used if we think that both of these components of the asbestos history may have an effect on survival. It also allows us to test whether these different components of history are important on survival by testing whether the parameters β_1 or β_2 are significantly different from zero.

A cautionary note must be made when interpreting hazard rates with time-dependent covariates, the hazard function with time-dependent covariates may NOT necessarily be used to construct survival distributions.

For example, if we have a time-independent covariate Z, then the conditional survival distribution

$$S(t|Z) = P[T \geq t|Z] = \mathrm{e}^{-\int_0^t \lambda(u|Z)du}$$

is well defined and meaningful. But the following distribution

$$S(t|Z^H(t)) = P[T \geq t|Z^H(t)]$$

may not make any sense since by the very fact that $Z^H(t)$ was measured when an individual was alive at time t.

It is useful to differentiate between internal and external time-dependent covariates for this purpose.

1. An internal time-dependent covariate is one where the change of the covariate over time is related to the behavior of the individual. For example, blood pressure, disease complications, etc.

2. An external or ancillary time-dependent covariate is one whose path is generated externally. For example, levels of air pollution.

For external time-dependent covariate, we can image a process which generates the time-dependent covariate over time. Therefore, for a particular realization of the process, $Z^H(\infty)$, we can image that the following quantity exists

$$\lambda(t|Z_i^H(\infty)) = \lim_{h\to 0} \frac{P[t \leq T_i < t+h|T_i \geq t, Z_i^H(\infty)]}{h},$$

and we may be willing to assume that

$$\lambda(t|Z_i^H(\infty)) = \lambda(t|Z_i^H(t)), \quad \text{for any } t > 0.$$

Therefore, if we have an external time-dependent covariate, we can ask the question what is the survival distribution at time t given the external process which generated $Z^H(\infty)$

$$S(t|Z_i^H(\infty)) \;=\; \exp\left[-\int_0^t \lambda(u|Z_i^H(\infty))du\right]$$

$$= \exp\left[-\int_0^t \lambda(u|Z_i^H(u))du\right].$$

For internal time-dependent covariates, this conceptualization would not make sense, although the relationship of the history of the covariate process on the hazard rate does have a useful interpretation.

Once we decide on a proportional hazards model with time-dependent covariates, the estimation of the regression parameters in the model, as well as the underlying cumulative hazard function (for external time-dependent covariate), create no additional difficulties. That is, we can use the theory developed so far for time-independent covariates with only slight modification.

For example, if we consider the model

$$\lambda(t|Z^H(t)) = \lambda_0(t)\exp(\beta^T Z(t)),$$

then the partial likelihood function of β for this model is given by

$$PL(\beta) = \prod_u \left[\frac{\exp(\beta^T Z_{I(u)}(u))}{\sum_{l=1}^n \exp(\beta^T Z_l(u))Y_l(u)}\right]^{dN(u)},$$

where $I(u)$ is the indicator variable that identifies the individual label $\in \{1, 2, ..., n\}$ for the individual who dies at time u.

This formula for partial likelihood looks almost identical to the one derived for time-independent covariates.

The only difference is that at time u, the values of the time-dependent covariates at time u were used, both for the individual who dies at that time, as well as the individuals who are at risk sets at that time. Therefore, the same individual appearing in different risk sets would use the possibly different values of their covariates at those risk sets.

Estimates, standard errors, tests and all other statistical properties would then follow exactly as they did before. That is, we would compute the MPLE by maximizing the log partial likelihood

given above. The score vector and the information matrix can be obtained as the first derivative and minus second derivative of the log partial likelihood. Wald, score and likelihood ratio tests can be computed analogously.

The major difficulty with time-dependent covariates in a proportional hazards model is computing and storage. Theoretically, at each death time we need to know the exact value of the covariate at that death time for ALL individuals at risk. The management, collection and storage of such data can create some difficulties, whereas the theory is no more difficult than with time-independent covariates.

SAS has some very nice software for handling time-dependent covariates.

Example 1: Time-varying Smoking Data

Suppose we have the a small data set as follows

ID	time	status	z1	z2	z3	z4
1	2	1	1	.	.	.
2	4	1	1	1	.	.
3	5	1	0	1	0	.
4	7	0	1	0	1	.
5	8	1	1	0	0	1

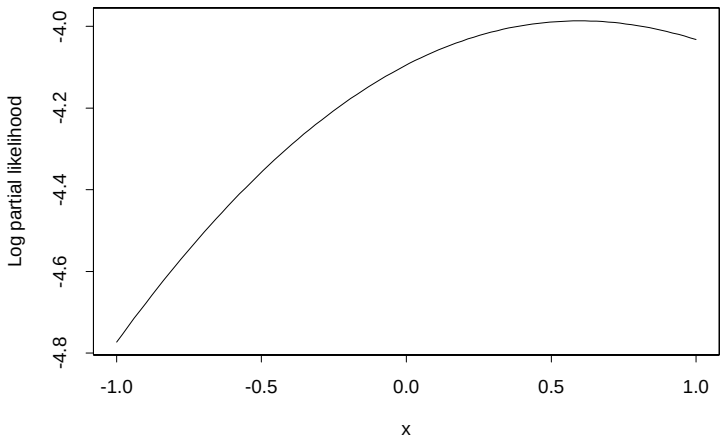
and we assume a "proportional hazards" model with time-varying smoking status:

$$\lambda(t|z_i(t)) = \lambda_0(t)e^{\beta z_i(t)},$$

where $z(t)$ is the smoking status for subject i. Then the partial likelihood function of β using the above data is

$$L(\beta; x, \delta, z(t)) = \frac{e^{\beta}}{1+4e^{\beta}} \times \frac{e^{\beta}}{2+2e^{\beta}} \times \frac{1}{2+e^{\beta}} \times \frac{e^{\beta}}{e^{\beta}}.$$

Figure 10.1: *Log partial likelihood function of β*



The log partial likelihood function of β looks like (using the following r functions:

```
  postscript(file="tvlik.ps", horizontal = F,
    height=6, width=8.5)
 # par(mfrow=c(1,2))

  x <- seq(-1, 1, length=100)
  y <- exp(x)

  y <- 2*x - log(1+4*y)-log(2+2*y)-log(2+y)
  plot(x, y, type="l", ylab="Log partial likelihood")
  box()

  dev.off()
```

The above model can be fit using the following SAS program:

```
options ls=72 ps=72;

data smoking;
  input time status z1-z4;
  cards;
  2 1 1 . . .
  4 1 1 1 . .
  5 1 0 1 0 .
  7 0 1 0 1 .
  8 1 1 0 0 1
;

proc phreg;
  model time*status(0) = smoke;
  array tt{*} t1-t4;
  array zz{*} z1-z4;
  t1 = 2;
  t2 = 4;
```

```
  t3 = 5;
  t4 = 8;
  do i=1 to 4;
   if time=tt[i] then smoke=zz[i];
  end;
run;
```

Example 2: Heart-Transplant Data

- Problem: We want to evaluate whether patients receiving heart transplants will benefit with increased survival.
- Experiment: A group of patients are recruited that are eligible for heart transplants. However, a heart has to become available and then the patients with the closest match receives this heart.
- Question: How do we evaluate the effectiveness of the heart transplant?

Here are some early attempts at answering this question in the medical literature.

1. Identify the patients that received a heart transplant and those that did not; measure their survival times from the time they entered the study and compare the survival times between these two groups using, say, a log rank test.
   Comment: Patients that died early will not have the chance to receive a heart transplant. Thus the two groups being compared are selectively biased favoring the heart transplant patients.
2. Identify patients that received a heart transplant and those that did not; measure survival times for heart transplant patients as time from transplant to death, and for the patients that did not receive a heart transplant measure survival times as time from entry to the study until death; compare two groups.
   Comment: Now the bias may go in the other direction.

Preferred Answer

Let $Z_i(t)$ denote heart transplant indicator for patient i at time t. That is,

$$Z_i(t) = \begin{cases} 1 & \text{if patient } i \text{ received a heart transplant at time } t \\ 0 & \text{otherwise} \end{cases}$$

here time t is measured as time from the entry into study to death.

Then consider the proportional hazards model with time-dependent covariate $Z_i(t)$

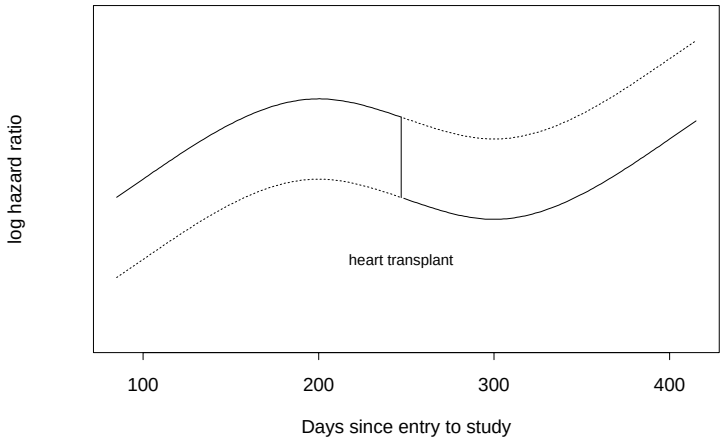
$$\lambda(t|Z_i^H(t)) = \lambda_0(t)\exp(\beta Z_i(t)).$$

This model assumes that the hazard increases by $\exp(\beta)$ after a heart transplant compared to before at time t. Therefore,

- $\beta = 0$ implies no effect on survival due to the heart transplant.
- $\beta < 0$ implies heart transplant is beneficial (hazard decreases).
- $\beta > 0$ implies heart transplant is detrimental (hazard increases).

The situation is illustrated by Figure 10.4.

Figure 10.2: *Illustration of the Effect of Heart Transplant*



If we define a variable `wait` in our data set as the time, say, in days from entry into study until receipt of heart transplant, if no heart transplant we can use `wait = .`, then we can use `Proc Phreg` in `SAS` to fit the above model. Specifically,

```
Proc Phreg data=mydata;
  model days*cens(0) = plant;
  if wait>days or wait=. then
    plant = 0;
  else
    plant = 1;
run;
```

Notice that the covariate `plant` represents the time-dependent covariate we defined in the above model and is defined after `model` statement in `Proc Phreg`. The variable `days` in the model statement is a running variable in `SAS` used to define the risk sets over time, making the variable `plant` a time-dependent covariate. Therefore, we cannot use the same `if-then-else` statement in `Data` step to define this time-dependent covariate.

Note: The model we described assumes the benefit (if $\beta < 0$) or harm (if $\beta > 0$) of heart transplant takes effect immediately after the transplant. This assumption may not be reasonable in practice. In fact, the hazard may increase right after heart transplant because of the complication due to transplant and then begin to decrease steadily.

The use of time-dependent covariates allows us to relax the proportional hazards assumption as well as giving us a framework for testing the adequacy of the proportional hazards assumption. For example, suppose we have a covariate Z (say it is time-independent) and we entertain the proportional hazards mode

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z).$$

As we know, this assumption implies that

$$\frac{\lambda(t|Z_1)}{\lambda(t|Z_0)} = \exp(\beta(Z_1 - Z_0)).$$

Suppose we wanted to test whether the hazard ratio changed over time. Consider the following model:

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z + \gamma Z g(t)),$$

where $g(t)$ is some specified function of time chosen by the data analyst. For example, we may choose $g(t) = \log(t)$.

Note: We must not include "main effect" of $g(t)$ since such a main effect will be absorbed into $\lambda_0(t)$, making it unidentifiable.

The term $\gamma Z g(t)$ is an interaction term between the covariate Z and some function $g(t)$ of time. For such a model the log hazard ratio is

$$\log\left[\frac{\lambda(t|Z_1)}{\lambda(t|Z_0)}\right] = \log\left[\frac{\lambda_0(t)\exp(\beta Z_1 + \gamma Z_1 g(t))}{\lambda_0(t)\exp(\beta Z_0 + \gamma Z_0 g(t))}\right] = (Z_1 - Z_0)(\beta + \gamma g(t)).$$

This model allows the hazard ratio to change over time giving us greater flexibility than proportional hazards assumption. In addition, testing whether or not γ is significantly from zero allows us the opportunity to evaluate the proportional hazards assumption.

The model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z + \gamma Z g(t)),$$

can be viewed as a proportional hazards model with two covariates:

1. the time-independent covariate Z,
2. the time-dependent covariate $g(t)Z$.

The term $g(t)Z$ is a simple example of an external or ancillary time-dependent covariate defined by the data analyst.

In `SAS` such a model is easy to implement. For example, suppose `days, cens` and time-independent covariate `z` are defined in a data set, then we use the following `SAS` code:

```
Proc Phreg data=mydata;
  model days*cens(0) = z zlogt;
  zlogt = z*log(days);
run;
```

In CALGB 8082, we found that nodes was the most significant prognostic factor for survival. Let us check whether the proportional hazards assumption is a reasonable representation of this relationship using $g(t) = log(t)$. Of course, we might other function $g(t)$ such as $g(t) = t, e^t$, etc.

```
title "Test PH for nodes effect using g(t)=log(t)";
proc phreg data=bcancer;
  model days*cens(0) = nodes nodelogt;
  nodelogt = nodes*log(days+1);
run;
```

```
*************************************************************************
                Test PH for nodes effect using g(t)=log(t)
                                                09:14 Sunday, April 17, 2005

                           The PHREG Procedure

                            Model Information

   Data Set                  WORK.BCANCER
   Dependent Variable        days             (censored) survival time in days
   Censoring Variable        cens             censoring indicator
   Censoring Value(s)        0
   Ties Handling             BRESLOW

              Summary of the Number of Event and Censored Values

                                                    Percent
          Total       Event    Censored    Censored

            905         490         415       45.86

                           Convergence Status

            Convergence criterion (GCONV=1E-8) satisfied.

                        Model Fit Statistics

                                  Without           With
                Criterion     Covariates     Covariates

                -2 LOG L        6264.861       6180.005
                AIC             6264.861       6184.005
                SBC             6264.861       6192.394
```

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.8564	2	<.0001
Score	116.8580	2	<.0001
Wald	114.4077	2	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
nodes	1	0.08246	0.03630	5.1613	0.0231
nodelogt	1	-0.00365	0.00521	0.4911	0.4834

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
nodes	1.086	number of positive nodes
nodelogt	0.996	

From the SAS output, there does not seem to be any problem with the proportional hazards assumption.

The model we used assumes a specific departure away from the proportional hazards assumption. That is

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z + \gamma Z\log(t)).$$

If proportional hazards failed in a fashion differently than that assumed in the above model, then the test we proposed may not detect such a deviation. Another approach is to use a more omnibus type of alternative away from the hypothesis of proportional hazards. This could be accomplished by the use of indicator functions over intervals of time.

We first partition our time axis into K intervals by choosing $(K-1)$ time points: $0 < \tau_1 < ... < \tau_{K-1}$, and then define the following indicator functions:

$$I_1(t) = I[t \in [0, \tau_1)],$$

$$I_2(t) = I[t \in [\tau_1, \tau_2)],$$

$$\vdots$$

$$I_{K-1}(t) = I[t \in [\tau_{K-2}, \tau_{K-1})],$$

$$I_K(t) = I[t \in [\tau_{K-1}, \infty)].$$

We then define the model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z + \gamma_1 Z I_1(t) + ... + \gamma_{K-1} Z I_{K-1}(t)).$$

Note: We include $K-1$ interaction terms between the covariate Z and the indicator function of time intervals. We must exclude one indicator function to avoid overparametrization.

For such a model, we have

$$\log\left[\frac{\lambda(t|Z_1)}{\lambda(t|Z_0)}\right] = (Z_1 - Z_0)\left[\beta + \gamma_1 I_1(t) + ... + \gamma_{K-1} I_{K-1}(t)\right].$$

Thus the log hazard ratio in each interval would be

$$\begin{aligned}
&(Z_1 - Z_0)\beta && \text{if } t \geq \tau_{K-1} \text{ (reference time interval)} \\
&(Z_1 - Z_0)(\beta + \gamma_1) && \text{if } 0 \leq t < \tau_1 \\
&\vdots \\
&(Z_1 - Z_0)(\beta + \gamma_{K-1}) && \text{if } \tau_{K-2} \leq t < \tau_{K-1}.
\end{aligned}$$

An omnibus test for proportional hazards assumption can be obtained by testing the hypothesis

$$H_0 : \gamma_1 = ... = \gamma_{K-1} = 0.$$

This would yield a chi-square test with $(K-1)$ degrees of freedom. As always, we can use the score test, Wald test, or likelihood ratio test to test H_0.

Note: If we find "significant" deviation from proportional hazards, then plotting $\hat{\gamma}_1, ..., \hat{\gamma}_{K-1}$ *vs.* their respective time intervals may give a suggestion on the functional form of the deviation over time.

For example, for the breast cancer data, if we consider the following proportional hazards model

$$\lambda(t|z) = \lambda_0(t)e^{\beta_1 trt + \beta_2 nn},$$

we can test PH assumption both for treatment and number of nodes using the above idea:

```
title "Test PH for treatment effect using dummy";
proc phreg data=bcancer;
  model days*cens(0) = nodes trt1 d1 d2 d3 d4/selection=forward
include=2 detail sle=0;
  if days<1000 then do;
    d1=trt1; d2=0; d3=0; d4=0;
  end;
  else if days<2000 then do;
    d1=0; d2=trt1; d3=0; d4=0;
  end;
  else if days<3000 then do;
    d1=0; d2=0; d3=trt1; d4=0;
  end;
  else if days<4000 then do;
    d1=0; d2=0; d3=0; d4=trt1;
  end;
  else do;
     d1=0; d2=0; d3=0; d4=0;
  end;
run;

title "Test PH for nodes effect using dummy";
proc phreg data=bcancer;
  model days*cens(0) = nodes trt1 d1 d2 d3 d4/selection=forward
include=2 detail sle=0;
  if days<1000 then do;
    d1=nodes; d2=0; d3=0; d4=0;
  end;
  else if days<2000 then do;
    d1=0; d2=nodes; d3=0; d4=0;
  end;
  else if days<3000 then do;
    d1=0; d2=0; d3=nodes; d4=0;
  end;
  else if days<4000 then do;
    d1=0; d2=0; d3=0; d4=nodes;
  end;
  else do;
     d1=0; d2=0; d3=0; d4=0;
  end;
run;
```

```
**********************************************************************************

                    Test PH for treatment effect using dummy
                                                  09:14 Sunday, April 17, 2005
```

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
905	490	415	45.86

The following variable(s) will be included in each model:

nodes trt1

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6264.861	6180.264
AIC	6264.861	6184.264
SBC	6264.861	6192.653

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.5972	2	<.0001
Score	113.1765	2	<.0001
Wald	111.8468	2	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
nodes	1	0.05707	0.00542	110.9645	<.0001
trt1	1	0.04168	0.09048	0.2122	0.6450

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
nodes	1.059	number of positive nodes
trt1	1.043	treatment indicator

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq	Label
d1	0.0058	0.9394	
d2	0.0222	0.8817	
d3	0.1955	0.6584	
d4	0.1396	0.7087	

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.2985	4	0.9899

Test PH for nodes effect using dummy
09:14 Sunday, April 17, 2005

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
905	490	415	45.86

The following variable(s) will be included in each model:

nodes trt1

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6264.861	6180.264
AIC	6264.861	6184.264
SBC	6264.861	6192.653

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.5972	2	<.0001
Score	113.1765	2	<.0001
Wald	111.8468	2	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
nodes	1	0.05707	0.00542	110.9645	<.0001
trt1	1	0.04168	0.09048	0.2122	0.6450

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
nodes	1.059	number of positive nodes
trt1	1.043	treatment indicator

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq	Label
d1	1.7974	0.1800	
d2	0.4795	0.4887	
d3	0.4462	0.5041	
d4	0.7319	0.3923	

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
6.8443	4	0.1443

Score Test for Proportional Hazards

Let us consider the model with indicators for time interval

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z + \gamma_1 Z I_1(t) + ... + \gamma_{K-1} Z I_{K-1}(t)).$$

We shall now derive the score test of the null hypothesis

$$H_0 : \gamma_1 = ... = \gamma_{K-1} = 0,$$

i.e., the hypothesis of proportional hazards.

We use the one of the representations for the partial likelihood of $\theta = (\beta, \gamma_1, ..., \gamma_{K-1})^T$:

$$PL(\theta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta Z_i + \sum_{j=1}^{K-1} \gamma_j Z_i I_j(X_i))}{\sum_{l=1}^{n} \exp\left[\beta Z_l + \sum_{j=1}^{K-1} \gamma_j Z_l I_j(X_i)\right] Y_l(X_i)} \right]^{\Delta_i}.$$

The log partial likelihood is equal to

$$\ell(\theta) = \sum_{i=1}^{n} \Delta_i \left[\beta Z_i + \sum_{j=1}^{K-1} \gamma_j Z_i I_j(X_i) \right] - \sum_{i=1}^{n} \Delta_i \log \left[\sum_{l=1}^{n} \exp\left(\beta Z_l + \sum_{j=1}^{K-1} \gamma_j Z_l I_j(X_i) \right) Y_l(X_i) \right].$$

To evaluate the score test of the hypothesis

$$H_0 : \gamma_1 = ... = \gamma_{K-1} = 0,$$

we need to compute the score vector with respect to the parameters $\gamma_1, ..., \gamma_{K-1}$, and evaluate these using the restricted MPLE under H_0:

$$\left.\frac{\partial \ell(\theta)}{\partial \gamma_j}\right|_{\hat{\beta}(\gamma_1 = ... = \gamma_{K-1} = 0)} = \sum_{i=1}^{n} \Delta_i \left[Z_i I_j(X_i) - \frac{\sum_{l=1}^{n} Z_l I_j(X_i) \exp(\hat{\beta} Z_l) Y_l(X_i)}{\sum_{l=1}^{n} \exp(\hat{\beta} Z_l) Y_l(X_i)} \right], \quad j = 1, ..., K-1.$$

Note: The restricted MPLE $\hat{\beta}(\gamma_1 = ... = \gamma_{K-1} = 0)$ is just the MPLE for the original proportional hazards model.

The score $\partial \ell(\theta)/\partial \gamma_j$ can be rewritten as

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \gamma_j} &= \sum_{i=1}^{n} I_j(X_i) \Delta_i \left[Z_i - \frac{\sum_{l=1}^{n} Z_l \exp(\hat{\beta} Z_l) Y_l(X_i)}{\sum_{l=1}^{n} \exp(\hat{\beta} Z_l) Y_l(X_i)} \right] \\ &= \sum \Delta_i \left[Z_i - \bar{Z}(X_i, \hat{\beta}) \right] I_j(X_i), \end{aligned}$$

where the summation is over all individuals whose value X_i is in the jth time interval, and $\bar{Z}(X_i, \hat{\beta})$ is the weighted average of the covariate for individuals at risk at time X_i.

The value $\Delta_i \left[Z_i - \bar{Z}(X_i, \hat{\beta}) \right]$ for individual i is referred to as the Schoenfeld residual or score residual (**Note**: this is not the score residue used by `SAS`); Schoenfeld (1982) *Biometrika*. The key word in `SAS` of Schoenfeld residual is `ressch` and `SAS` only calculates Schoenfeld residual for $\Delta_i = 1$.

If we denote the Schoenfeld residual by Sh_i for the ith individual, then the score can be written as

$$\frac{\partial \ell(\theta)}{\partial \gamma_j} = \sum Sh_i I(X_i \in [\tau_{j-1}, \tau_j)), \quad j = 1, ..., K-1.$$

The basis for the score test is that under the null hypothesis we expect the score vector

$$\left(\frac{\partial \ell}{\partial \gamma_1}, ..., \frac{\partial \ell}{\partial \gamma_{K-1}} \right)^T$$

to have mean zero and we reject H_0 when the score is not close to zero, *i.e.*, we reject H_0 when the quadratic form using the score vector is sufficiently large.

Therefore, Schoenfeld suggested that these residuals be plotted as a function of time; that is Sh_i *vs.* X_i.

If the proportional hazards assumption is adequate, then on average these residuals should be zero. A noticable trend away from zero may be indicative of the lack of proportional hazards.

The test we suggested earlier for $H_0 : \gamma_1 = ... = \gamma_{K-1} = 0$ can be used as a formal goodness of fit test for proportional hazards.

If we are using a model with many covariates:

$$\lambda(t|Z) = \lambda_0(t) \exp(\beta_1 Z_1 + ... + \beta_q Z_q),$$

then we can compute Schoenfeld residuals for each of the covariates; *i.e.*,

$$Sh_{ij} = \Delta_i \left[Z_{ij} - \frac{\sum_{l=1}^n Z_{lj} I_j(X_i) \exp(\hat{\beta}^T Z_l) Y_l(X_i)}{\sum_{l=1}^n \exp(\hat{\beta}^T Z_l) Y_l(X_i)} \right], \quad j = 1, ..., q.$$

For each covariate Z_j, we can then plot

$$SH_{ij} \quad vs. \quad X_i, \quad \text{for } j = 1, ..., q$$

producing q such plots.

The corresponding formal test for the jth covariate could be obtained by considering the model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta^T Z + \gamma_1 Z I_1(t) + ... + \gamma_{K-1} Z I_{K-1}(t)),$$

and testing

$$H_0 : \gamma_1 = ... = \gamma_{K-1} = 0.$$

Example: Breast cancer data revisited

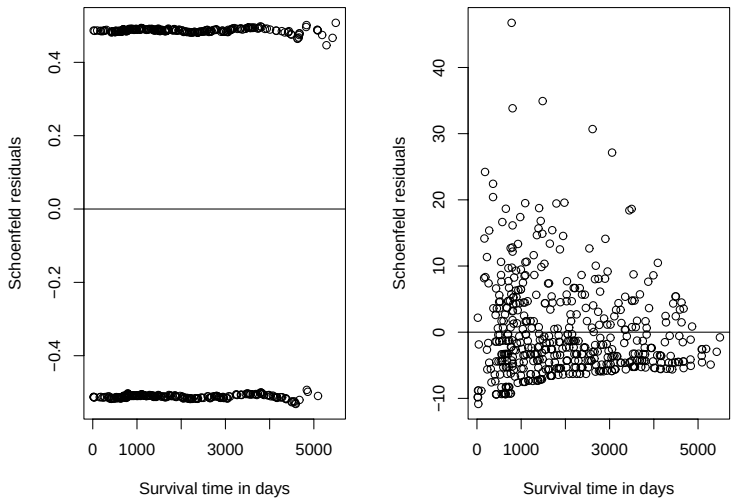
Let us consider the following proportional hazards model

$$\lambda(t|z) = \lambda_0(t) e^{\beta_1 trt + \beta_2 nn}.$$

Then we can use the following `SAS` program to output Schoenfeld residual for each covariate:

```
title "residual analysis for treatment and nodes";
proc phreg data=bcancer;
  model days*cens(0) = trt1 nodes;
  output out=residout RESSCH=trtresid nodresid;
run;

proc gplot data=residout;
  plot trtresid * days / vref=0;
  symbol1 value=circle;
run;

proc gplot data=residout;
  plot nodresid * days / vref=0;
  symbol1 value=circle;
run;
```

Figure 10.3: *Schoenfeld residual for treatment(left) and nodes(right)*



Score Test of the Functional Form of the Covariate in a Proportional Hazards Model and Martingale Residuals

Previously we discussed an omnibus alternative model that allowed deviation from proportional hazards. This model tests for proportional hazards as well as giving us the motivation for considering Schoenfeld residuals.

We also consider ways of checking adequacy of the covariate relationship in a proportional hazards model. This include using higher order polynomials or dummy variables after discretizing the continuous covariate. Using the idea of discretization, we can formally develop a hierarchical model for testing the adequacy of functional relationship to the covariate.

Suppose we entertain the model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z),$$

and we want to consider whether the relationship $\exp(\beta Z)$ is suitable. We then partition the covariate values of Z into K intervals by defining values $\xi_1, ..., \xi_{K-1}$ along the range of possible

values of the covariate Z, and define the indicator functions

$$\begin{aligned} I_1(Z) &= I[Z < \xi_1] \\ I_2(Z) &= I[\xi_1 \le Z < \xi_2] \\ &\vdots \\ I_{K-1}(Z) &= I[\xi_{K-2} Z \le \xi_{K-1}] \\ I_K(Z) &= I[Z \ge \xi_{K-1}]. \end{aligned}$$

We then consider the model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z + \sum_{j=1}^{K-1} \gamma_j I_j(Z))).$$

Remark: We use $(K-1)$ indicator variables to avoid overparametrization.

This model is an omnibus alternative away from the null model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z).$$

A formal goodness of fit test could be derived to test the adequacy of the above null model by testing the hypothesis

$$H_0 : \gamma_1 = ... = \gamma_{K-1} = 0$$

using the Wald, score or likelihood ratio tests. These tests are asymptotically chi-square with $(K-1)$ $d.f.$ if the null hypothesis were true.

The following program tests linear nodes effect for breast cancer data:

```
title "Test linear nodes effect using dummy";
proc phreg data=bcancer;
  model days*cens(0) = nodes trt1 dn1 dn2 dn3 dn4/selection=forward
include=2 detail sle=0;
run;
```

```
**********************************************************************

                    Test linear nodes effect using dummy
                                          09:35 Sunday, April 17, 2005
```

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
896	490	406	45.31

The following variable(s) will be included in each model:

nodes trt1

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6264.861	6180.264
AIC	6264.861	6184.264
SBC	6264.861	6192.653

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.5972	2	<.0001
Score	113.1765	2	<.0001
Wald	111.8468	2	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
nodes	1	0.05707	0.00542	110.9645	<.0001
trt1	1	0.04168	0.09048	0.2122	0.6450

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
nodes	1.059	number of positive nodes
trt1	1.043	treatment indicator

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq	Label
dn1	5.2307	0.0222	
dn2	0.9936	0.3189	
dn3	0.2617	0.6089	
dn4	4.7095	0.0300	

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
9.6325	4	0.0471

The above output indicates that it may not be appropriate to assume linear effect for the number of positive nodes.

Score Test for Covariate Relationship

Similar to the computation made before when we discussed Schoenfeld residual, the log partial likelihood can be derived as

$$\ell(\beta,\gamma_1,...,\gamma_{K-1})=\sum_{i=1}^{n}\Delta_i\left[\beta Z_i+\sum_{j=1}^{K-1}\gamma_j I_j(Z_i)\right]-\sum_{i=1}^{n}\Delta_i\log\left[\sum_{l=1}^{n}\exp\left(\beta Z_l+\sum_{j=1}^{K-1}\gamma_j I_j(Z_l)\right)Y_l(X_i)\right].$$

The jth element of the score vector is given by

$$\left.\frac{\partial\ell}{\partial\gamma_j}\right|_{\hat{\beta}(\gamma_1=...=\gamma_{K-1}=0)}=\sum_{i=1}^{n}\Delta_i\left[I_j(Z_i)-\frac{\sum_{l=1}^{n}I_j(Z_l)\exp(\hat{\beta}Z_l)Y_l(X_i)}{\sum_{l=1}^{n}\exp(\hat{\beta}Z_l)Y_l(X_i)}\right],\quad j=1,...=K-1.$$

This can be written as

$$\sum_{i=1}^{n}\Delta_i I_j(Z_i)-\sum_{i=1}^{n}\left[\frac{\sum_{l=1}^{n}\Delta_i I_j(Z_l)\exp(\hat{\beta}Z_l)Y_l(X_i)}{\sum_{r=1}^{n}\exp(\hat{\beta}Z_r)Y_r(X_i)}.\right]$$

If we interchange the sum's in the second term, the second term becomes

$$\sum_{l=1}^{n}I_j(Z_l)\exp(\hat{\beta}Z_l)\sum_{i=1}^{n}\left[\frac{\Delta_i I(X_i\le X_l)}{\sum_{r=1}^{n}\exp(\hat{\beta}Z_r)Y_r(X_i)}\right].$$

Note: Here we used the fact $Y_l(X_i) = I(X_i \leq X_l)$, *i.e.*, whether or subject l is still at risk at time X_i.

By reversing the index l with the index i we get

$$\sum_{i=1}^{n} I_j(Z_i)\exp(\hat{\beta}Z_i)\sum_{l=1}^{n}\left[\frac{\Delta_l I(X_l \leq X_i)}{\sum_{r=1}^{n}\exp(\hat{\beta}Z_r)Y_r(X_l)}\right].$$

Obviously, the quantity

$$\sum_{l=1}^{n}\left[\frac{\Delta_l I(X_l \leq X_i)}{\sum_{r=1}^{n}\exp(\hat{\beta}Z_r)Y_r(X_l)}\right]$$

is nothing but the Breslow estimator for the cumulative baseline hazard function evaluated at time X_i; *i.e.*, $\hat{\Lambda}_0(X_i)$ (since $\Delta_l I(X_l \leq X_i)$ is $dN_i(X_i)$, the indicator indicating whether or subject i was dead at time X_i).

Therefore, the jth element of the score vector is equal to

$$\sum_{i=1}^{n} I_j(Z_i)\left[\Delta_i - \hat{\Lambda}_0(X_i)\exp(\hat{\beta}Z_i)\right].$$

The quantity

$$\left[\Delta_i - \hat{\Lambda}_0(X_i)\exp(\hat{\beta}Z_i)\right]$$

for the ith individual is referred to as the martingale residual MR_i. Therefore, the score vector has as its jth element

$$\frac{\partial \ell}{\partial \gamma_j} = \sum_{i=1}^{n} I_j(Z_i)MR_i, \quad j = 1, ..., K-1.$$

Under

$$H_0 : \lambda(t|Z) = \lambda_0(t)\exp(\beta Z),$$

the score vector has mean zero. So a sufficiently large quadratic form of the score vector is an indication that H_0 may not be true. Therefore we reject H_0 when this quadratic form of the score vector is large.

It is recommended that the martingale residuals be plotted against the covariate Z; *i.e.*, we plot MR_i *vs.* Z_i. When there are multiple covariates in the model, we recommend plotting MR_i *vs.* $Z_i^T\hat{\beta}$ (call `xbeta` in `SAS`).

Note: The sum of the martingale residuals in jth intervals of the covariate range makes up the jth element of the score vecotr: $\partial\ell/\partial\gamma_j$.

If our model is correct, these resisuals should have mean zero and be uncorrelated with the covariate (since the score vector has mean zero conditional on the convariate). So a noticable trend away from zero may be indicative of covariate model misspecification.

Example: Breast cancer data revisited

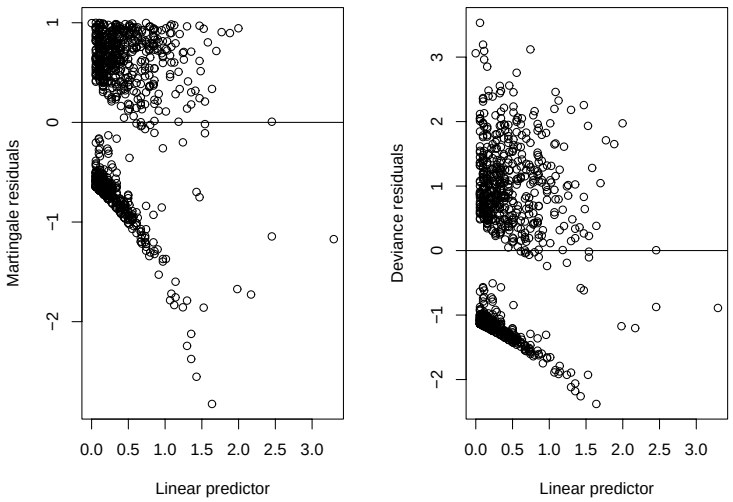
Let us consider the model on page 198 and plot the martigale residual using the the following program:

```
proc phreg data=bcancer;
  model days*cens(0) = trt1 nodes;
  output out=residout resmart=mart resdev=dev xbeta=xb;
run;

proc gplot data=residout;
  plot mart * xb / vref=0;
  symbol1 value=circle;
run;

proc gplot data=residout;
  plot dev * xb / vref=0;
  symbol1 value=circle;
run;
```

Let us return to `CAL 8082` and consider the relationship to nodes once again. The following table summarizes the results:

Model	$-2\text{Log}L$	$d.f.$
nodes	5189.80	3
nodes + dummy	5174.23	8
nodes + nodes^2	5183.30	4
nodes + nodes^2 + dummy	5173.90	9

Figure 10.4: *Martingale (left) and deviance (right) residual for the model*



Since

$$\chi^2_{0.05;5} = 11.07, \quad \chi^2_{0.01;5} = 15.09,$$

these results suggest that putting in a quadratic term when modeling nodes gives an adequate fit.

What to do if you find substantial deviation from proportional hazards

The proportional hazards model is the most popular model for censored survival data. The parameters of the model have a nice interpretation, the theoretical properties have been studied expensively, software is readily available, and the likelihood surface is easy to work with.

There will be situations however when the proportional hazards assumption is not an adequate fit to the data. What can we do in such cases?

By hierarchical model building, we can identify covariates where the proportional hazards assumption is not appropriate and by including interaction terms between functions of times and covariates get a more suitable model. However, this model building results in a loss of parsimony with results that may be difficult to interpret and difficult to explain to your collaborators.

Another alternative is to use a stratified proportional hazards model. When we are considering many covariates in a model, we may find that most of the covariates follow a proportional hazards relationship and only a few of the covariates do not. If this is the case, we may stratify our study population into categories obtained by different combinations of the covariates and then use a stratified proportional hazards model.

If we denote the number of strata by K and let l index the strata, where $l = 1, ..., K$, then the stratified proportional hazards model is given by

$$\lambda_l(t|Z) = \lambda_{0l}(t)\exp(\beta^T Z),$$

where $Z = (Z_1, ..., Z_q)^T$ is an q dimensional vector of covariates that satisfy proportional hazards. In this model, there are K unspecified baseline hazard function for each stratum; *i.e.*,

$$\lambda_{0l}(t), \quad l = 1, ..., K; t \geq 0,$$

and within each stratum, covariates Z satisfy proportional hazards assumption and the "effect" of the covariates Z are the same across K strata.

The interpretation of $\beta = (\beta_1, ..., \beta_q)^T$ is exactly the same as in an unstratified proportional hazards model. Namely, if we consider the hazard ratio resulting from an increase of one unit in the covariate Z_j, keeping all other covariates fixed (including those used to construct the strata), we get

$$\frac{\lambda(t|Z_j = z_j + 1)}{\lambda(t|Z_j = z_j)} = \exp(\beta_j),$$

independent of time t. However, the hazard ratio between strata, fixing the value of other covariates, is

$$\frac{\lambda_{0l}(t)}{\lambda_{0l'}(t)}, \quad \text{comparing strata } l \text{ to } l'.$$

Since these functions are unrestricted, any relationship of this hazard ratio over time is possible.

To obtain estimates for β, we only need a slight modification to the partial likelihood.

For stratum l, denote the data within that stratum by

$$(X_{li}, \Delta_{li}, Z_{li}), \quad i = 1, ..., n_l, l = 1, ..., K.$$

The total sample size is $n = \sum_{l=1}^{K} n_l$.

The modified partial likelihood of β is given by

$$PL(\beta) = \prod_{l=1}^{K} PL_l(\beta),$$

where $PL_l(\beta)$ is the partial likelihood of β contributed by the data from the lth stratum:

$$PL_l(\beta) = \prod_u \left[\frac{\exp(\beta^T Z_{l[i(u)]})}{\sum_{i=1}^{n_l} \exp(\beta^T Z_{li}) Y_{li}(u)} \right]^{dN_l(u)},$$

where $dN_l(u)$ is the number of deaths observed in time interval $[u, u + \Delta u)$ in the lth stratum, $Y_{li}(u) = I(X_{li} \geq u)$ is the indicator indicating whether or not subject i in stratum l is at risk at time u.

All inferential methods derived previously for the unstratified partial likelihood can be used with the stratified partial likelihood above, such as MPLE, score test, Wald likelihood ratio test, etc.

The Breslow estimator for the cumulative baseline hazard can also be used for the cumulative baseline hazard function for the lth strata; *i.e.*,

$$\hat{\Lambda}_{0l}(t) = \sum_{u \leq t} \left[\frac{dN_l(u)}{\sum_{i=1}^{n_l} \exp(\hat{\beta}^T Z_{li}) Y_{li}(u)} \right], \quad l = 1, ..., K.$$

For example, in the breast cancer data, if we suspect the proportional hazards assumption for `er`, then we can stratify on this covariate. The following is the `SAS` program and output:

```
options ps=200 ls=80;

data bcancer;
  infile "cal8082.dat";
  input days cens trt meno tsize nodes er;
  trt1 = trt - 1;
  label days="(censored) survival time in days"
    cens="censoring indicator"
```

```
    trt="treatment"
    meno="menopausal status"
    tsize="size of largest tumor in cm"
    nodes="number of positive nodes"
    er="estrogen receptor status"
    trt1="treatment indicator";
run;

title "Model 1: Univariate analysis of treatment";
proc phreg;
  model days*cens(0) = trt1;
run;

title "Model 2: Univariate analysis of treatment stratified on ER";
proc phreg;
  model days*cens(0) = trt1;
  strata er;
run;

title "Model 3: Log-rank test of treatment effect stratified on ER";
proc lifetest notable;
  time days*cens(0);
  strata er;
  test trt1;
run;
```

```
*******************************************************************************
```

Model 1: Univariate analysis of treatment 1
12:05 Saturday, April 16, 2005

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
905	497	408	45.08

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6362.858	6362.421
AIC	6362.858	6364.421
SBC	6362.858	6368.629

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.4375	1	0.5083
Score	0.4375	1	0.5083
Wald	0.4374	1	0.5084

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
trt1	1	0.05935	0.08973	0.4374	0.5084

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
trt1	1.061	treatment indicator

Model 2: Univariate analysis of treatment stratified on ER 2
12:05 Saturday, April 16, 2005

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Stratum	er	Total	Event	Censored	Percent Censored
1	0	278	170	108	38.85
2	1	513	258	255	49.71
Total		791	428	363	45.89

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	4781.604	4780.894
AIC	4781.604	4782.894
SBC	4781.604	4786.953

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.7100	1	0.3994
Score	0.7099	1	0.3995
Wald	0.7089	1	0.3998

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
trt1	1	0.08150	0.09680	0.7089	0.3998

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
trt1	1.085	treatment indicator

Model 3: Log-rank test of treatment effect stratified on ER 3
12:05 Saturday, April 16, 2005

The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

Stratum	er	Total	Failed	Censored	Percent Censored
1	0	278	170	108	38.85
2	1	513	258	255	49.71
Total		791	428	363	45.89

NOTE: There were 114 observations with missing values, negative time values or frequency values less than 1.

Model 3: Log-rank test of treatment effect stratified on ER 4
12:05 Saturday, April 16, 2005

The LIFETEST Procedure

Testing Homogeneity of Survival Curves for days over Strata

Rank Statistics

er	Log-Rank	Wilcoxon
0	44.090	33623
1	-44.090	-33623

Covariance Matrix for the Log-Rank Statistics

er	0	1
0	88.6179	-88.6179
1	-88.6179	88.6179

Covariance Matrix for the Wilcoxon Statistics

er	0	1
0	30007365	-3.001E7
1	-3.001E7	30007365

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	21.9360	1	<.0001
Wilcoxon	37.6743	1	<.0001
-2Log(LR)	19.6431	1	<.0001

Rank Tests for the Association of days with Covariates Pooled over Strata

Univariate Chi-Squares for the Log-Rank Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
trt1	-8.7104	10.3381	0.7099	0.3995	treatment indicator

Covariance Matrix for the Log-Rank Statistics

Variable	trt1
trt1	106.877

Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment	Label
trt1	1	0.7099	0.3995	0.7099	0.3995	treatment indicator

Note: We note from the above output that the score test of treatment effect stratified on `er` is the same as the stratified log-rank test of treatment stratified on `er`.