5

The Stratified Cox Procedure

Introduction

We begin with an example of the use of the stratified Cox procedure for a single predictor that does not satisfy the PH assumption. We then describe the general approach for fitting a stratified Cox model, including the form of the (partial) likelihood function used to estimate model parameters.

We also describe the assumption of no interaction that is typically incorporated into most computer programs that carry out the stratified Cox procedure. We show how the no-interaction assumption can be tested, and what can be done if interaction is found.

We conclude with a second example of the stratified Cox procedure in which more than one variable is stratified.

Abbreviated Outline

The outline below gives the user a preview of the material to be covered by the presentation. A detailed outline for review purposes follows the presentation.

Objectives

Upon completing the chapter, the learner should be able to:

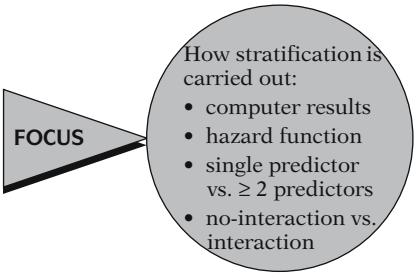
1. Recognize a computer printout for a stratified Cox procedure.
2. State the hazard form of a stratified Cox model for a given survival analysis scenario and/or a given set of computer results for such a model.
3. Evaluate the effect of a predictor of interest based on computer results from a stratified Cox procedure.
4. For a given survival analysis scenario and/or a given set of computer results involving a stratified Cox model,

   - state the no-interaction assumption for the given model;
   - describe and/or carry out a test of the no-interaction assumption;
   - describe and/or carry out an analysis when the no-interaction assumption is not satisfied.

I. Preview

Stratified Cox model:

- modification of Cox PH model
- Stratification of predictor not satisfying PH
- includes predictors satisfying PH



The "stratified Cox model" is a modification of the Cox proportional hazards (PH) model that allows for control by "stratification" of a predictor that does not satisfy the PH assumption. Predictors that are assumed to satisfy the PH assumption are included in the model, whereas the predictor being stratified is not included.

In this presentation, we focus on how stratification is carried out by describing the analysis of computer results and the form of the hazard function for a stratified Cox model. We first consider stratifying on a single predictor and then later consider stratifying on two or more predictors. Further, we distinguish between the use of a "no-interaction" version of the stratified Cox model and an alternative approach that allows interaction.

II. An Example

EXAMPLE

Clinical trial: 42 leukemia patients
Response-days in remission

	Coef.	Std. Err.	*P*(*PH*)
log WBC	1.594	0.330	0.828
Rx	1.391	0.457	0.935
Sex	0.263	0.449	0.031

- log WBC and *Rx* satisfy PH
- Sex does not satisfy PH

(Same conclusions using graphical approaches)

Stratified Cox (SC):

- control for sex (stratified);
- simultaneously include log WBC and *Rx* in the model

Consider the computer results shown here for a Cox PH model containing the three variables, log WBC, treatment group (*Rx*), and SEX. These results derive from a clinical trial of 42 leukemia patients, where the response of interest is days in remission.

From the printout, the *P*(*PH*) values for log WBC and treatment group are nonsignificant. However, the *P*(*PH*) value for SEX is significant below the .05 level. These results indicate that log WBC and treatment group satisfy the PH assumption, whereas the SEX variable does not. The same conclusions regarding the PH assumption about these variables would also be made using the graphical procedures described earlier.

Because we have a situation where one of the predictors does not satisfy the PH assumption, we carry out a stratified Cox (SC) procedure for the analysis. Using SC, we can control for the SEX variable—which does not satisfy the PH assumption—by stratification while simultaneously including in the model the log WBC and treatment variables—which do satisfy the PH assumption.

EXAMPLE (continued)

STATA OUTPUT USING SC:

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
log WBC	1.390	0.338	0.000	4.016	2.072	7.783
Rx	0.931	0.472	0.048	2.537	1.006	6.396

No. of subjects = 42 Log likelihood = –57.560 Stratified by sex

Appendix A illustrates SC procedures using Stata, SAS, and SPSS.

- Log WBC and *Rx* are included in SC model.
- SC model is stratified by SEX.

Effect of *Rx* adjusted for log WBC and SEX:

- Hazard ratio: 2.537 = $e^{0.931}$
- Interpretation: Placebo group (*Rx* = 1) has 2.5 times the hazard as the treatment group (Rx = 0)

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
log WBC	1.390	0.338	0.000	4.016	2.072	7.783
Rx	0.931	0.472	0.048	2.537	1.006	6.396

No. of subjects = 42 Log likelihood = –57.560 Stratified by sex

95% CI for *Rx* (1.006, 6.396) indicates considerable variability.

CI formula: exp(0.931 ± 1.96 × 0.472)

Wald test: P = 0.048 (two-tailed), significant at the 0.05 level.

The computer results from a SC procedure are shown here. These results come from the Stata package. (See the Computer Appendix for running a SC procedure in Stata, SAS, or SPSS).

The computer results show that the log WBC and *Rx* variables are included in the model listing, whereas the SEX variable is not included; rather, the model stratifies on the SEX variable, as indicated at the bottom of the output. Note that the SEX variable is being adjusted by stratification, whereas log WBC is being adjusted by its inclusion in the model along with *Rx*.

In the above output, we have also circled some key information that can be used to assess the effect of the *Rx* variable adjusted for both log WBC and SEX. In particular, we can see that the hazard ratio for the effect of *Rx* adjusted for log WBC and SEX is given by the value 2.537. This value can be obtained by exponentiating the coefficient 0.931 of the *Rx* variable. The hazard ratio value can be interpreted to mean that the placebo group (for which $Rx = 1$) has 2.5 times the hazard for going out of remission as the treatment group (for which $Rx = 0$).

Also, we can see from the output that a 95% confidence interval for the effect of the *Rx* variable is given by the limits 1.006 to 6.396. This is a fairly wide range, thus indicating considerable variability in the 2.537 hazard ratio point estimate. Note that these confidence limits can be obtained by exponentiating the quantity 0.931 plus or minus 1.96 times the standard error 0.472.

From the above output, a test for the significance of the *Rx* variable adjusted for log WBC and SEX is given by the Wald statistic P value of 0.048. This is a two-tailed P-value, and the test is just significant at the 0.05 level.

EXAMPLE (continued)

LR test: Output for reduced model

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
log WBC	1.456	0.320	0.000	4.289	2.291	8.03

No. of subjects = 42 Log likelihood = (−59.648) Stratified by sex

LR = (−2 × −59.648) − (−2 × −57.560)
= 119.296 − 115.120 = 4.179 (P < 0.05)

LR and Wald give same conclusion.

Hazard function for stratified Cox model:

$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 Rx + \beta_2 \log \text{WBC}]$
$g = 1,2;$

g denotes stratum #.

SC model for males and females:
Females ($g = 1$):

$h_1(t,\mathbf{X}) = h_{01}(t)\exp[\beta_1 Rx + \beta_2 \log \text{WBC}]$

Males ($g = 2$):

$h_2(t,\mathbf{X}) = h_{02}(t)\exp[\beta_1 Rx + \beta_2 \log \text{WBC}]$

Rx and log WBC in the model
Sex *not* in the model (stratified)

$\widehat{HR}$ for effect of *Rx* adjusted for log WBC and sex:

$e^{\hat{\beta}_1}$

where β_1 is the coefficient of *Rx*.

An alternative test involves a likelihood ratio (*LR*) statistic that compares the above model (full model) with a reduced model that does not contain the *Rx* variable. The output for the reduced model is shown here. The log-likelihood statistic for the reduced model is −2 times −59.648, which is to be compared with the log-likelihood statistic of −2 times −57.560 for the full model.

The *LR* statistic is therefore 119.296 minus 115.120, which equals 4.179. Under H_0, this statistic has a chi-square distribution with one degree of freedom and is significant at the 0.05 level. Thus, the *LR* and Wald tests lead to the same conclusion.

So far, we have illustrated the results from a stratified Cox procedure without actually describing the model form being used. For the remission data example, we now present the hazard function form for the stratified Cox model, as shown here. This hazard function formula contains a subscript g that indicates the gth stratum.

Thus, in our remission data example, where we have stratified on SEX, g takes on one of two values, so that we have a different baseline hazard function for males and females.

Notice that the hazard function formula contains the variables *Rx* and log WBC, but does not contain the variable SEX. SEX is not included in the model because it doesn't satisfy the PH assumption. So, instead, the SEX variable is controlled by stratification.

Because the variables *Rx* and log WBC are included in the model, we can estimate the effect of each variable adjusted for the other variable and the SEX variable using standard exponential hazard ratio expressions. For example, the estimated hazard ratio for the effect of *Rx*, adjusted for log WBC and SEX, is given by e to the β_1 "hat," where β_1 is the coefficient of the *Rx* variable.

EXAMPLE (continued)

Cannot estimate *HR* for SEX variable (SEX doesn't satisfy PH).

Nevertheless, because the SEX variable is not included in the model, it is not possible to obtain a hazard ratio value for the effect of SEX adjusted for the other two variables. This is the price to be paid for stratification on the SEX variable. Note that a single value for the hazard ratio for SEX is not appropriate if SEX doesn't satisfy the PH assumption, because the hazard ratio must then vary with time.

Different baseline hazard functions: $h_{01}(t)$ for females and $h_{02}(t)$ for males.

Same coefficients β_1 and β_2 for both female and male models.

Notice also that the hazard functions for males and females differ only insofar as they have different baseline hazard functions, namely, $h_{01}(t)$ for females and $h_{02}(t)$ for males. However, the coefficients β_1 and β_2 are the same for both female and male models.

Different baselines $\begin{cases} h_{01}(t) \Rightarrow \text{Survival curve for females} \\ h_{02}(t) \Rightarrow \text{Survival curve for males} \end{cases}$

Because there are different baseline hazard functions, the fitted stratified Cox model will yield different estimated survival curves for females and males. These curves will be described shortly.

Females and males: same β_1 and $\beta_2 \Rightarrow$ same $\widehat{HR}$'s, e.g., $e^{\hat{\beta}_1}$

No interaction assumption (see Section IV)

Note, however, that because the coefficients of *Rx* and log WBC are the same for females and males, estimates of hazard ratios, such as e to the β_1 "hat," are the same for both females and males. This feature of the stratified Cox model is called the "no-interaction" assumption. It is possible to evaluate whether this assumption is tenable and to modify the analysis if not tenable. We will discuss this assumption further in Section IV.

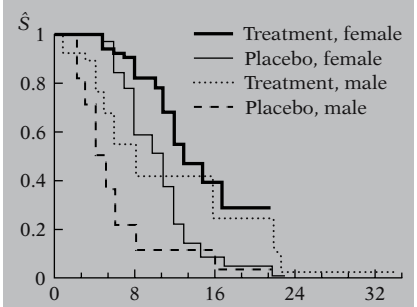
Estimates of β_1 and β_2:

Maximize partial likelihood (L), where $L = L_1 \times L_2$
L_1 is the likelihood for females derived from $h_1(t)$,
and L_2 is the likelihood for males derived from $h_2(t)$.

To obtain estimates of β_1 and β_2, a (partial) likelihood function (L) is formed from the model and the data; this function is then maximized using computer iteration. The likelihood function (L) for the stratified Cox (SC) model is different from the nonstratified Cox model. For the SC model, L is obtained by multiplying together likelihood functions for each stratum. Thus, L is equal to the product of L_1 and L_2, where L_1 and L_2 denote the female and male likelihood functions, respectively, which are derived from their respective hazard functions $h_1(t)$ and $h_2(t)$.

EXAMPLE (continued)

Adjusted Survival Curves for *Rx* from Stratified Cox Model (adjusted for log WBC)



As mentioned above, adjusted survival curves can be obtained for each stratum as shown here. Here we have shown *four* survival curves because we want to compare the survival for two treatment groups over each of two strata.

If we compare treatment and placebo group separately by sex, we can see that the treatment group has consistently better survival prognosis than the placebo group for females and males separately. This supports our findings about the hazard ratio for the treatment effect derived earlier from the computer results for the stratified Cox model.

III. The General Stratified Cox (SC) Model

Example: one binary predictor

↓

General: several predictors, several strata

$Z_1, Z_2, \ldots, Z_k$, do not satisfy PH

$X_1, X_2, \ldots, X_p$, satisfy PH

In the previous example, we illustrated the SC model for one binary predictor not satisfying the PH assumption. We now describe the general form of the SC model that allows for stratification of several predictors over several strata.

We assume that we have k variables not satisfying the PH assumption and p variables satisfying the PH assumption. The variables not satisfying the PH assumption we denote as $Z_1, Z_2, \ldots, Z_k$; the variables satisfying the PH assumption we denote as $X_1, X_2, \ldots, X_p$.

Define a single new variable Z^*:

1. categorize each Z_i
2. form combinations of categories (strata)
3. the strata are the categories of Z^*

To perform the stratified Cox procedure, we define a single new variable, which we call Z^*, from the Z's to be used for stratification. We do this by forming categories of each Z_i, including those Z_i that are interval variables. We then form combinations of categories, and these combinations are our strata. These strata are the categories of the new variable Z^*.

EXAMPLE

		Age		
		Young	Middle	Old
Treatment status	Placebo	1	2	3
	Treatment	4	5	6

Z^* = new variable with six categories
Stratify on Z^*

For example, suppose k is 2, and the two Z's are age (an interval variable) and treatment status (a binary variable). Then we categorize age into, say, three age groups—young, middle, and old. We then form six age group–by–treatment-status combinations, as shown here. These six combinations represent the different categories of a single new variable that we stratify on in our stratified Cox model. We call this new variable Z^*.

Z^* has k^* categories where k^* = total # of combinations (strata), e.g., $k^* = 6$ in above example.

In general, the stratification variable Z^* will have k^* categories, where k^* is the total number of combinations (or strata) formed after categorizing each of the Z's. In the above example, k^* is equal to 6.

The general SC model:

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p]$$

$g = 1, 2, \ldots, k^*$, strata defined from Z^*

We now present the general hazard function form for the stratified Cox model, as shown here. This formula contains a subscript g which indicates the gth stratum. The strata are defined as the different categories of the stratification variable Z^*, and the number of strata equals k^*.

Z^* not included in the model

$X_1, X_2, \ldots, X_p$ included in the model

Note that the variable Z^* is not explicitly included in the model but that the X's, which are assumed to satisfy the PH assumption, are included in the model.

Different baseline hazard functions: $h_{0g}(t), g = 1, 2, \ldots, k^*$
Same coefficients: $\beta_1, \beta_2, \ldots, \beta_p$

Note also that the baseline hazard function $h_{0g}(t)$ is allowed to be different for each stratum. However, the coefficients $\beta_1, \beta_2, \ldots, \beta_p$ are the same for each stratum.

Different baselines $\left\{\begin{array}{l}\hat{h}_{01}(t) \Rightarrow \hat{S}_1(t) \\ \hat{h}_{02}(t) \Rightarrow \hat{S}_2(t) \\ \quad\vdots \\ \hat{h}_{0k}(t) \Rightarrow \hat{S}_k(t)\end{array}\right\}$ Different survival curves

As previously described by example, the fitted SC model will yield different estimated survival curves for each stratum because the baseline hazard functions are different for each stratum.

$\widehat{HR}$ same for each stratum

(no-interaction assumption, Section IV)

However, because the coefficients of the X's are the same for each stratum, estimates of hazard ratios are the same for each stratum. This latter feature of the SC model is what we previously have called the "no-interaction" assumption to be discussed further in Section IV.

(Partial) likelihood function:

$L = L_1 \times L_2, \times \cdots \times L_{k^*}$

Strata:	1	2	…	k^*
Likelihood:	L_1	L_2	…	L_{k^*}
Hazard:	$h_1(t,\mathbf{X})$	$h_2(t,\mathbf{X})$	…	$h_{k^*}(t,\mathbf{X})$

To obtain estimates of the regression coefficients $\beta_1, \beta_2, \ldots, \beta_p$, we maximize a (partial) likelihood function L that is obtained by multiplying together likelihood functions for each stratum, as shown here. Thus, L is equal to the product of L_1 times L_2, and so on, up until L_{k^*}, where the subscripted L's denote the likelihood functions for different strata, with each of these L's being derived from its corresponding hazard function.

IV. The No-Interaction Assumption and How to Test It

We previously pointed out that the SC model contains regression coefficients, denoted as β's, that do not vary over the strata. We have called this property of the model the "no-interaction assumption." In this section, we explain what this assumption means. We also describe how to evaluate the assumption and what to do if the assumption is violated.

Stratified Cox model

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p]$$

β coefficients do not vary over strata (no-interaction assumption)

- how to evaluate
- what to do if violated

EXAMPLE

No-interaction SC model:

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
log WBC	1.390	0.338	0.000	4.016	2.072	7.783
Rx	0.931	0.472	0.048	2.537	1.006	6.396

No. of subjects = 42 Log likelihood = −57.560 Stratified by sex

Interaction by fitting separate models:

Cox regression (Females)
Analysis time _t: survt

Column name	Coeff	StErr.	p-value	*HR*	0.95	CI	*P*(*PH*)
4 log WBC	1.639	0.519	0.002	5.150	1.862	14.242	0.228
5 *Rx*	1.859	0.729	0.011	6.418	1.537	26.790	0.603

No. of subjects = 20 Log likelihood = −22.100

Cox regression (Males)
Analysis time _t: survt

Column name	Coeff	StErr.	p-value	*HR*	0.95	CI	*P*(*PH*)
4 log WBC	1.170	0.499	0.019	3.222	1.213	8.562	0.674
5 *Rx*	0.267	0.566	0.637	1.306	0.431	3.959	0.539

No. of subjects = 22 Log likelihood = −33.736

Which model is more appropriate statistically?

We return to the SC output previously illustrated. Notice that only one set of coefficients, namely, 1.390 for log WBC and 0.931 for *Rx*, are provided, even though there are two strata, one for females and one for males. These results assume no interaction of the sex variable with either log WBC or *Rx*.

If we allow for interaction, then we would expect to obtain different coefficients for each of the (SEX) strata. This would happen if we fit separate hazard models to the female and male data, with each model containing the log WBC and Rx variables. The computer results from fitting separate models are shown here.

Notice that the coefficient of log WBC is 1.639 for females but is 1.170 for males. Also, the coefficient for *Rx* is 1.859 for females but 0.267 for males. These results show different coefficients for females than for males, particularly for the *Rx* variable.

But are corresponding coefficients statistically different? That is, which model is more appropriate statistically, the no-interaction model or the interaction model? To answer this question, we must first look at the hazard function model for the interaction situation.

EXAMPLE (continued)

Interaction model:

$(\blacklozenge)\ h_g(t,\mathbf{X})$
$= h_{0g}(t)\exp[\beta_{1g}\log \text{WBC} + \beta_{2g}Rx]$
where $g = 1$ (females), $g = 2$ (males)

No-interaction model:

$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 \log \text{WBC} + \beta_2 Rx]$
where $g = 1$ (females), $g = 2$ (males)

Alternative interaction model:

$(\star)\ h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1^* \log \text{WBC} + \beta_2^* Rx + \beta_3^* (\text{SEX} \times \log \text{WBC}) + \beta_4^* \times (\text{SEX} \times Rx)]$

where $\text{SEX} = \begin{cases} 1 \text{ if female} \\ 0 \text{ if male} \end{cases}$

$h_{0g}(t)$ are different for $g = 1,2$
β^* coefficients do not involve g

Equivalence of models $(\blacklozenge)$ and $(\star)$:
$g = 1$ (females), so that sex = 1:

$h_1(t,\mathbf{X}) = h_{01}(t)\exp[\beta_1^* \log \text{WBC} + \beta_2^* Rx + \beta_3^* (1 \times \log \text{WBC}) + \beta_4^* (1 \times Rx)]$
$= h_{01}(t)\exp[(\beta_1^* + \beta_3^*) \log \text{WBC} + (\beta_2^* + \beta_4^*)\ \text{Rx}]$

$g = 2$ (males), so that sex = 0:

$h_2(t,\mathbf{X}) = h_{02}(t)\exp[\beta_1^* \log \text{WBC} + \beta_2^* Rx + \beta_3^* (0 \times \log \text{WBC}) + \beta_4^* (0 \times Rx)]$
$= h_{02}(t)\exp[\beta_1^* \log \text{WBC} + \beta_2^* Rx]$

Interaction models in same format:

Females ($g = 1$): $h_1(t,\mathbf{X})$
$(\blacklozenge) = h_{01}(t)\exp[\beta_{11}\log \text{WBC} + \beta_{21}Rx]$
$(\star) = h_{01}(t)\exp[(\beta_1^* + \beta_3^*) \log \text{WBC} + (\beta_2^* + \beta_4^*)Rx]$

Males ($g = 2$): $h_2(t,\mathbf{X})$
$(\blacklozenge) = h_{02}(t)\exp[\beta_{12}\log \text{WBC} + \beta_{22}Rx]$
$(\star) = h_{02}(t)\exp[\beta_1^* \log \text{WBC} + \beta_2^* Rx]$

One way to state the hazard model formula when there is interaction is shown here $(\blacklozenge)$. Notice that each variable in this model has a different coefficient for females than for males, as indicated by the subscript g in the coefficients β_{1g} and β_{2g}.

In contrast, in the no-interaction model, the coefficient (β_1) of log WBC is the same for females and for males; also, the coefficient (β_2) of Rx is the same for females and for males.

An alternative way to write the interaction model is shown here $(\star)$. This alternative form contains two product terms—SEX $\times$ log WBC and SEX $\times$ Rx—as well as the main effects of log WBC and Rx. We have coded the SEX so that 1 denotes female and 0 denotes male.

In this alternative model, note that although the baseline hazards $h_{0g}(t)$ are different for each sex, the β^* coefficients do not involve the subscript g and therefore are the same for each sex.

Nevertheless, this alternative formula $(\star)$ is equivalent to the interaction formula $(\blacklozenge)$ above. We show this by specifying the form that the model takes for $g = 1$ (females) and $g = 2$ (males).

Notice that the coefficients of log WBC are different in each formula, namely, $(\beta_1^* + \beta_3^*)$ for females versus β_1^* for males.

Similarly, the coefficients of Rx are different, namely, $(\beta_2^* + \beta_4^*)$ for females versus β_2^* for males.

The preceding formulae indicate that two seemingly different formulae for the interaction model—$(\blacklozenge)$ versus $(\star)$, shown earlier—can be written in the same format. We show these formulae here separately for females and males.

EXAMPLE (continued)

$$\begin{array}{lll} & (\blacklozenge) & (\star) \\ \text{Females } (g=1): & \beta_{11} = \beta_1^* + \beta_3^* \\ & \beta_{21} = \beta_2^* + \beta_4^* \end{array}$$

$$\begin{array}{lll} & (\blacklozenge) & (\star) \\ \text{Males } (g=2): & \beta_{12} = \beta_1^* \\ & \beta_{22} = \beta_2^* \end{array}$$

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
log WBC	1.170	0.499	0.019	3.222	1.213	8.562
Rx	0.267	0.566	0.637	1.306	0.431	3.959
Sex × log WBC	0.469	0.720	0.515	1.598	0.390	6.549
Sex × *Rx*	1.592	0.923	0.084	4.915	0.805	30.003

No. of subjects = 42 Log likelihood = –55.835 Stratified by sex

Females:

$$\text{log WBC} \begin{cases} \beta_{11} = \boxed{1.639} \\ \hat{\beta}_1^* + \hat{\beta}_3^* = 1.170 + 0.469 = \boxed{1.639} \end{cases}$$

$$Rx \begin{cases} \beta_{21} = \boxed{1.859} \\ \hat{\beta}_2^* + \hat{\beta}_4^* = 0.267 + 1.592 = \boxed{1.859} \end{cases}$$

Males:

$$\text{log WBC } \hat{\beta}_{12} = \boxed{1.170} = \hat{\beta}_1^*$$
$$Rx \qquad \hat{\beta}_{22} = \boxed{0.267} = \hat{\beta}_2^*$$

Interaction model:

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1^* \log \text{WBC} + \beta_2^* Rx + \beta_3^* (\text{SEX} \times \log \text{WBC}) + \beta_4^* (\text{SEX} \times Rx)]$$

Notice that for females, the coefficient β_{11} in model ($\blacklozenge$) must be equivalent to $(\beta_1^* + \beta_3^*)$ in model ($\star$) because both models have the same format, and both β_{11} and $(\beta_1^* + \beta_3^*)$ are coefficients of the same variable, log WBC. Similarly, β_{21} in model ($\blacklozenge$) is equivalent to $(\beta_2^* + \beta_4^*)$ in model ($\star$) because both are coefficients of the same variable, *Rx*.

For males, it follows in an analogous way, that the coefficient β_{12} is equivalent to β_1^*, and, similarly, β_{22} equals β_2^*.

Here we provide computer results obtained from fitting the alternative interaction model ($\star$). The estimated regression coefficients $\hat{\beta}_1^*$, $\hat{\beta}_2^*$, $\hat{\beta}_3^*$, and $\hat{\beta}_4^*$, respectively, are circled.

We have shown above that the sums $\hat{\beta}_1^* + \hat{\beta}_3^*$ and $\hat{\beta}_2^* + \hat{\beta}_4^*$ are equal to the coefficients $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$, respectively, in the original interaction model for females.

Also, we have shown that $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are equal to the coefficients $\hat{\beta}_{12}$ and $\hat{\beta}_{22}$, respectively, in the original interaction model for the males. The numerical equivalences are shown here. Note again that the coefficients of log WBC and *Rx* for females are different from males, as is to be expected if sex interacts with each variable.

We have thus seen that the interaction model can be written in a format that contains product terms involving the variable being stratified—SEX—being multiplied by each of the predictors not being stratified. We show this model involving product terms again here. We will use this model to describe a test of the no-interaction assumption.

EXAMPLE (continued)

Testing the no-interaction assumption:
$LR = -2 \ln L_R - (-2 \ln L_F)$
R = reduced (no-interaction) model
F = full (interaction) model

$LR \sim \chi^2_{2df}$ under H_0: no interaction
(2 df because two product terms tested in interaction model)

No interaction (reduced model):

Output: –2 log *L*: 115.120

$-2 \ln L_R$

Interaction (full model):

Output: –2 log *L*: 111.670

$-2 \ln L_F$

$LR = 115.120 - 111.670 = 3.45$
($P > 0.05$ not significant).
Thus, the no-interaction model is acceptable.

The test is a likelihood ratio (*LR*) test which compares log-likelihood statistics for the interaction model and the no-interaction model. That is, the *LR* test statistic is of the form $-2 \ln L_R$ minus $-2 \ln L_F$, where R denotes the reduced model, which in this case is the no-interaction model, and F denotes the full model, which is the interaction model.

This *LR* test statistic has approximately a chi-square distribution with 2 degrees of freedom under the null hypothesis that the no-interaction model is correct. The degrees of freedom here is 2 because there are two product terms being tested in the interaction model.

The log-likelihood statistic for the reduced model comes from the computer output for the no-interaction model and is equal to −2 times −57.560, or 115.120.

The log-likelihood statistic for the full model comes from the computer results for the interaction model and is equal to −2 times −55.835, or 111.670.

The *LR* statistic is therefore 115.120 minus 111.670, which equals 3.45. This value is not significant at the 0.05 level for 2 degrees of freedom. Thus, it appears that despite the numerical difference between corresponding coefficients in the female and male models, there is no statistically significant difference. We can therefore conclude for these data that the no-interaction model is acceptable (at least at the 0.05 level).

Remission data example:

- described no-interaction assumption
- evaluated assumption using *LR* test
- provided interaction model if needed

Now, we generalize this process.

Using the remission data example, we have described the no-interaction assumption, have shown how to evaluate this assumption using a likelihood ratio test, and have provided the form of an interaction model that should be used in case the no-interaction assumption does not hold. We now describe this process more generally for any stratified Cox analysis.

No-interaction SC model:

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p]$$
$$g = 1, 2, \ldots, k^*, \text{ strata defined from } Z^*$$

Recall that the general form of the no-interaction model for the stratified Cox procedure is given as shown here. This model allows for several variables being stratified through the use of a newly defined variable called Z^*, whose strata consist of combinations of categories of the variables being stratified.

SC model allowing interaction:

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_{1g} X_1 + \beta_{2g} X_2 + \cdots + \beta_{pg} X_p]$$
$$g = 1, 2, \ldots, k^*, \text{ strata defined from } Z^*$$

If, in contrast, we allow for interaction of the Z^* variable with the X's in the model, we can write the model as shown here. Notice that in this interaction model, each regression coefficient has the subscript g, which denotes the gth stratum and indicates that the regression coefficients are different for different strata of Z^*.

Alternative SC interaction model:

- uses product terms involving Z^*
- define $k^* - 1$ dummy variables $Z_1^*, Z_2^*, \ldots, Z_{k^*-1}^*$, from Z^*
- products of the form $Z_i^* \times X_j$, where $i = 1, \ldots, k^* - 1$ and $j = 1, \ldots, p$.

An alternative way to write the interaction model uses product terms involving the variable Z^* with each of the predictors. However, to write this model correctly, we need to use $k^* - 1$ dummy variables to distinguish the k^* categories of Z^*; also, each of these dummy variables, which we denote as $Z_1^*, Z_2^*, \ldots, Z_{k^*-1}^*$, needs to be involved in a product term with each of the X's.

$$\begin{aligned} h_g(t,\mathbf{X}) = h_{0g}(t)\exp[&\beta_1 X_1 + \cdots + \beta_p X_p \\ &+ \beta_{11}(Z_1^* \times X_1) + \cdots + \beta_{p1}(Z_1^* \times X_p) \\ &+ \beta_{12}(Z_2^* \times X_1) + \cdots + \beta_{p2}(Z_2^* \times X_p) \\ &+ \cdots + \beta_{1,k^*-1}(Z_{k^*-1}^* \times X_1) + \cdots \\ &+ \beta_{p,k^*-1}(Z_{k^*-1}^* \times X_p)] \end{aligned}$$
$$g = 1, 2, \ldots, k^*, \text{ strata defined from } Z^*$$

The hazard model formula alternative model is shown here. Notice that the first line of the formula contains the X's by themselves, the next line contains products of each X_j with Z_1^*, the third line contains the products with Z_2^*, and the last line contains products with Z_{k^*-1}. Note also that the subscript g occurs only with the baseline hazard function $h_{0g}(t)$, and is not explicitly used in the β coefficients.

EXAMPLE (Remission Data)

$Z^* = \text{sex}$, $k^* = 2$,
$Z_1^* = \text{sex}(0,1)$,
$X_1 = \log \text{WBC}$, $X_2 = Rx$ ($p = 2$)

$$\begin{aligned} h_g(t,\mathbf{X}) &= h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 \\ &\quad + \beta_{11}(Z_1^* \times X_1) \\ &\quad + \beta_{21}(Z_1^* \times X_2)] \\ &= h_{0g}(t)\exp[\beta_1^* \log \text{WBC} \\ &\quad + \beta_2^* \text{Rx} + \beta_3^*(\text{sex} \times \log \text{WBC}) \\ &\quad + \beta_4^*(\text{sex} \times Rx)] \end{aligned}$$

$g = 1, 2$
$\beta_1 = \beta_1^*$, $\beta_2 = \beta_2^*$, $\beta_{11} = \beta_3^*$, and $\beta_{21} = \beta_4^*$

In our previous example involving the remission data, the stratification variable (Z^*) was the variable SEX, and k^* was equal to 2; thus, we have only one dummy variable Z_1^*, which uses a (0,1) coding to indicate sex, and we have only (p equal to) two predictors—X_1 equal to log WBC and X_2 equal to Rx. The interaction model is then written in either of the forms shown here.

The latter version of the interaction model is what we previously presented for the remission data example. Because the two versions presented here are equivalent, it follows that $\beta_1^* = \beta_1$, $\beta_2 = \beta_2^*$, $\beta_{11} = \beta_3^*$, and $\beta_{21} = \beta_4^*$.

We have thus seen that the interaction model can be written in a format that contains product terms involving dummy variables (i.e., Z_i^*) for the variable being stratified being multiplied by each of the predictors (i.e., X_i) not being stratified. We will use this model to describe a test of the no-interaction assumption.

Testing the no-interaction assumption:
$LR = -2\ln L_R - (-2\ln L_F)$
$R =$ reduced (no-interaction) model
$F =$ full (interaction) model
contains product terms

$$H_0: \begin{cases} \beta_{11} = \cdots = \beta_{p1} = 0 \\ \beta_{12} = \cdots = \beta_{p2} = 0 \\ \vdots \\ \beta_{1,k^*-1} = \cdots = \beta_{p,k^*-1} = 0 \end{cases}$$

The test is a likelihood ratio (LR) test which compares log likelihood statistics for the interaction model and the no-interaction model. That is, the LR test statistic is of the form $-2\ln L_R$ minus $-2\ln L_F$, where R denotes the reduced model, which in this case is the no-interaction model, and F denotes the full model, which is the interaction model.

The no-interaction model differs from the interaction model in that the latter contains additional product terms. Thus, one way to state the null hypothesis of no interaction is that the coefficients of each of these product terms are all zero.

$LR \overset{\sim}{\sim} \chi^2_{p(k^*-1)\text{ df}}$
under H_0: no interaction

The LR test statistic has approximately a chi-square distribution with $p(k^* - 1)$ degrees of freedom under the null hypothesis. The degrees of freedom here is $p(k^* - 1)$ because this value gives the number of product terms that are being tested in the interaction model.

$p(k^* - 1)$ gives number of product terms being tested in interaction model

EXAMPLE (Remission Data)

$Z^* = \text{sex}$, $k^* = 2$,
$Z_1^* = \text{sex}(0,1)$,
$X_1 = \log \text{WBC}$, $X_2 = \text{Rx}$ ($p = 2$)
$p(k^* - 1) = 2$, so
$LR \sim \chi^2_{2\text{df}}$ under H_0: no interaction

Returning to the remission data example, for which $p = 2$ and $k^* = 2$, the value of $p(k^* - 1)$ is equal to two times $(2 - 1)$, which equals two. Thus, to test whether the SEX variable interacts with the log WBC and *Rx* predictors, the degrees of freedom for the *LR* statistic is two, as previously described.

V. A Second Example Involving Several Stratification Variables

EXAMPLE

vets.dat: survival time in days, $n = 137$

Veteran's Administration Lung Cancer Trial
Column 1: Treatment (standard = 1, test = 2)
Column 2: Cell type 1 (large = 1, other = 0)
Column 3: Cell type 2 (adeno = 1, other = 0)
Column 4: Cell type 3 (small = 1, other = 0)
Column 5: Cell type 4 (squamous = 1, other = 0)
Column 6: Survival time (days)
Column 7: Performance status (0 = worst, ..., 100 = best)
Column 8: Disease duration (months)
Column 9: Age
Column 10: Prior therapy (none = 0, some = 10)
Column 11: Status (0 = censored, 1 = died)

Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]	*P*(*PH*)
Treatment	0.290	0.207	0.162	1.336	0.890	2.006	0.628
Large cell	0.400	0.283	0.157	1.491	0.857	2.594	0.033
Adeno cell	1.188	0.301	0.000	3.281	1.820	5.915	0.081
Small cell	0.856	0.275	0.002	2.355	1.374	4.037	0.078
Perf. Stat	–0.033	0.006	0.000	0.968	0.958	0.978	0.000
Dis. Durat.	0.000	0.009	0.992	1.000	0.982	1.018	0.919
Age	–0.009	0.009	0.358	0.991	0.974	1.010	0.198
Pr. Therapy	0.007	0.023	0.755	1.007	0.962	1.054	0.145

No. of subjects = 137 Log likelihood = –475.180

Variables not satisfying PH:
- cell type (3 dummy variables)
- performance status
- prior therapy (possibly)

SC model: stratifies on cell type and performance status

The dataset "vets.dat" considers survival times in days for 137 patients from the Veteran's Administration Lung Cancer Trial cited by Kalbfleisch and Prentice in their text (*The Statistical Analysis of Survival Time Data*, Wiley, pp. 223–224, 1980). The exposure variable of interest is treatment status. Other variables of interest as control variables are cell type (four types, defined in terms of dummy variables), performance status, disease duration, age, and prior therapy status. Failure status is defined by the status variable. A complete list of the variables is shown here.

Here we provide computer output obtained from fitting a Cox PH model to these data. Using the *P*(*PH*) information in the last column, we can see that at least four of the variables listed have *P*(*PH*) values below the 0.100 level. These four variables are labeled in the output as large cell (0.033), adeno cell (0.081), small cell (0.078), and Perf. Stat (0.000). Notice that the three variables, large cell, adeno cell, and small cell, are dummy variables that distinguish the four categories of cell type.

Thus, it appears from the *P*(*PH*) results that the variables cell type (defined using dummy variables) and performance status do not satisfy the PH assumption.

Based on the conclusions just made about the PH assumption, we now describe a stratified Cox analysis that stratifies on the variables, cell type and performance status.

EXAMPLE (continued)

Z^* given by combinations of categories:

- cell type (four categories)
- performance status (interval) change to
- PSbin (two categories)

Z^* has $k^* = 4 \times 2 = 8$ categories

Four other variables considered as X's:

- treatment status
- disease duration
- age
- prior therapy

Here, we use treatment status and age as X's

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Treatment	0.125	0.208	0.548	1.134	0.753	1.706
Age	−0.001	0.010	0.897	0.999	0.979	1.019

No. of subjects = 137 Log likelihood = −262.020 Stratified by Z^*

No-interaction model

$\widehat{HR} = 1.134$ ($P = 0.548$)

Treatment effect (adjusted for age and Z^*) is nonsignificant

No-interaction model:

$h_g(t,\mathbf{X})$
$= h_{0g}(t)\exp[\beta_1 \text{ Treatment} + \beta_2 \text{ Age}]$
$g = 1, 2, \ldots, 8$ (= # of strata defined from Z^*)

Interaction model:

$h_g(t,\mathbf{X})$
$= h_{0g}(t)\exp[\beta_{1g} \text{ Treatment} + \beta_{2g} \text{ Age}]$
$g = 1, 2, \ldots, 8$

Because we are stratifying on two variables, we need to form a single new categorical variable Z^* whose categories represent combinations of categories of the two variables. The cell type variable has four categories by definition. The performance status variable, however, is an interval variable ranging between 0 for worst to 100 for best, so it needs to be categorized. We categorize this variable into two groups using a cutpoint of 60, and we denote this binary variable as PSbin. Thus, the number of categories for our Z^* variable is 4×2, or 8; that is, $k^* = 8$.

In addition to the two stratification variables, cell type and performance status, there are four other variables to be considered as predictors in the stratified Cox model. These are treatment status, disease duration, age, and prior therapy.

For illustrative purposes here, we use only treatment status and age as predictors. The other two variables, disease duration and prior therapy, are considered in exercises following this presentation.

Here we show computer output from fitting a stratified Cox model that stratifies on cell type and performance status using the eight-category stratification variable Z^*. This model also includes treatment and age as predictors. These results consider a no-interaction model, because only one regression coefficient is provided for the treatment and age predictors. Notice that the estimated hazard ratio is 1.134 for the effect of the treatment variable adjusted for age and Z^*, the latter being adjusted by stratification. The p-value for this adjusted treatment effect is 0.548, which is highly nonsignificant.

The no-interaction model we have just described has the hazard function formula shown here.

To evaluate whether the no-interaction model is appropriate, we need to define an interaction model that allows different regression coefficients for different strata. One way to write this interaction model is shown here.

EXAMPLE (continued)

Alternative interaction model:

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 \text{ Treatment} + \beta_2 \text{ Age} + \beta_{11}(Z_1^* \times \text{Treatment}) + \cdots + \beta_{17}(Z_7^* \times \text{Treatment}) + \beta_{21}(Z_1^* \times \text{Age}) + \cdots + \beta_{27}(Z_7^* \times \text{Age})]$$

$g = 1, 2, \ldots, 8$

Another version of interaction model:
Replace $Z_1^*, \ldots, Z_7^*$ by
Z_1^* = large cell (binary)
Z_2^* = adeno cell (binary)
Z_3^* = small cell (binary)
Z_4^* = PSbin (binary)

$$Z_5^* = Z_1^* \times Z_4^*$$
$$Z_6^* = Z_2^* \times Z_4^*$$
$$Z_7^* = Z_3^* \times Z_4^*$$

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 \text{ Treatment} + \beta_2 \text{ Age} + \beta_{11}(\text{tr } Z_1^*) + \beta_{12}(\text{tr } Z_2^*) + \beta_{13}(\text{tr } Z_3^*) + \beta_{14}(\text{tr } Z_4^*) + \beta_{15}(\text{tr } Z_1^* Z_4^*) + \beta_{16}(\text{tr } Z_2^* Z_4^*) + \beta_{17}(\text{tr } Z_3^* Z_4^*) + \beta_{21}(\text{AGE } Z_1^*) + \beta_{22}(\text{AGE } Z_2^*) + \beta_{23}(\text{AGE } Z_3^*) + \beta_{24}(\text{AGE } Z_4^*) + \beta_{25}(\text{AGE } Z_1^* Z_4^*) + \beta_{26}(\text{AGE } Z_2^* Z_4^*) + \beta_{27}(\text{AGE } Z_3^* Z_4^*)]$$

An alternative version of this interaction model that involves product terms is shown here. This version uses seven dummy variables denoted as Z_1^*, Z_2^* up through Z_7^* to distinguish the eight categories of the stratification variable Z^*. The model contains the main effects of treatment and age plus interaction terms involving products of each of the seven dummy variables with each of the two predictors.

Yet another version of the interaction model is to replace the seven dummy variables Z_1^* to Z_7^* by the seven variables listed here. These variables are three of the binary variables making up the cell type variable, the binary variable for performance status, plus three product terms involving each of the cell type dummy variables multiplied by the PSbin dummy variable (Z_4^*).

The latter interaction model is shown here. In this model, the variable tr Z_1^* denotes the product of treatment status with the large cell dummy Z_1^*, the variable tr Z_2^* denotes the product of treatment status with the adeno cell variable Z_2^*, and so on. Also, the variable tr $Z_1^* Z_4^*$ denotes the triple product of treatment status times the large cell variable Z_1^* times the PSbin variable Z_4^*, and so on, for the other triple product terms involving treatment. Similarly, for the terms involving age, the variable Age Z_1^* denotes the product of age with Z_1^*, and the variable Age $Z_1^* Z_4^*$ denotes the triple product of age times Z_1^* times Z_4^*.

Note that we are just considering the interaction between the stratified variables and the predictors. We could also (but do not) consider the interaction between the two predictors, treatment, and age.

EXAMPLE (continued)

Stratified Cox Regression Analysis on Variable: Z^*
Response: Surv. Time

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Treatment	0.286	0.664	0.667	1.331	0.362	4.893
Age	0.000	0.030	0.978	0.999	0.942	1.060
tr Z_1^*	2.351	1.772	0.184	10.495	0.326	337.989
tr Z_2^*	–1.158	0.957	0.226	0.314	0.048	2.047
tr Z_3^*	0.582	0.855	0.496	1.790	0.335	9.562
tr Z_4^*	–1.033	0.868	0.234	0.356	0.065	1.950
tr $Z_1^*Z_4^*$	–0.794	1.980	0.688	0.452	0.009	21.882
tr $Z_2^*Z_4^*$	2.785	1.316	0.034	16.204	1.229	213.589
tr $Z_3^*Z_4^*$	0.462	1.130	0.683	1.587	0.173	14.534
Age Z_1^*	0.078	0.064	0.223	1.081	0.954	1.225
Age Z_2^*	–0.047	0.045	0.295	0.954	0.873	1.042
Age Z_3^*	–0.059	0.042	0.162	0.943	0.868	1.024
Age Z_4^*	0.051	0.048	0.287	1.053	0.958	1.157
Age $Z_1^*Z_4^*$	–0.167	0.082	0.042	0.847	0.721	0.994
Age $Z_2^*Z_4^*$	–0.045	0.068	0.511	0.956	0.838	1.092
Age $Z_3^*Z_4^*$	0.041	0.061	0.499	1.042	0.924	1.175

No. of subjects = 137 Log likelihood = –249.972 Stratified by Z^*

Here we provide the computer results from fitting the interaction model just described. Notice that the first two variables listed are the main effects of treatment status and age. The next seven variables are product terms involving the interaction of treatment status with the seven categories of Z^*. The final seven variables are product terms involving the interaction of age with the seven categories of Z^*. As defined on the previous page, the seven variables used to define Z^* consist of three dummy variables Z_1^*, Z_2^* and Z_3^* for cell type, a binary variable Z_4^* for performance status and products of Z_4^* with each of Z_1^*, Z_2^*, and Z_3^*. Note that once the variables Z_1^*, Z_2^*, Z_3^*, and Z_4^* are specified, the values of the three product terms are automatically determined.

Eight possible combinations of Z_1^* to Z_4^*:

$g = 1$: $Z_1^* = Z_2^* = Z_3^* = Z_4^* = 0$
$g = 2$: $Z_1^* = 1, Z_2^* = Z_3^* = Z_4^* = 0$
$g = 3$: $Z_2^* = 1, Z_1^* = Z_3^* = Z_4^* = 0$
$g = 4$: $Z_3^* = 1, Z_1^* = Z_2^* = Z_4^* = 0$
$g = 5$: $Z_1^* = Z_2^* = Z_3^* = 0, \ Z_4^* = 1$
$g = 6$: $Z_1^* = 1, Z_2^* = Z_3^* = 0, \ Z_4^* = 1$
$g = 7$: $Z_2^* = 1, Z_1^* = Z_3^* = 0, \ Z_4^* = 1$
$g = 8$: $Z_3^* = 1, Z_1^* = Z_2^* = 0, \ Z_4^* = 1$

We can use these results to show that the interaction model being fit yields different regression coefficients for each of the eight categories defined by the subscript g for the stratification variable Z^*. These eight categories represent the possible combinations of the four variables Z_1^* to Z_4^*, as shown here.

$g = 1$: $Z_1^* = Z_2^* = Z_3^* = Z_4^* = 0$
(Squamous cell type and PSbin = 0)

All product terms are zero:
$h_1(t,\mathbf{X})$
$= h_{01}(t)\exp[\beta_1 \text{Treatment} + \beta_2 \text{Age}]$,
where $\hat{\beta}_1 = 0.286$,
$\hat{\beta}_2 = 0.000$, so that
$\hat{h}_1(t,\mathbf{X}) = \hat{h}_{01}(t)\exp[(0.286)\text{Treatment}]$

Consider the hazard function when the variables Z_1^* through Z_4^* are all equal to zero. This stratum is defined by the combination of squamous cell type and a binary performance status value of 0. In this case, all product terms are equal to zero and the hazard model contains only the main effect terms treatment and age. The estimated hazard function for this stratum uses the coefficients 0.286 for treatment and 0.000 for age, yielding the expression shown here. Note that age drops out of the expression because its coefficient is zero to three decimal places.

$g = 2$: $Z_1^* = 1, Z_2^* = Z_3^* = Z_4^* = 0$
(Large cell type and PSbin = 0)

Nonzero product terms	Coefficients
Age Z_1^* = Age	β_{21}
tr Z_1^* = Treatment	β_{11}

Now consider the hazard function when the variable Z_1^* equals 1 and Z_2^* through Z_4^* are equal to zero. This stratum is defined by the combination of large cell type and a PSbin value of 0. In this case, the only nonzero product terms are Age Z_1^* and tr Z_1^*, whose coefficients are β_{21} and β_{11}, respectively.

EXAMPLE (continued)

$h_2(t,\mathbf{X}) = h_{02}(t)\exp[(\beta_1 + \beta_{11})\text{Treatment} + (\beta_2 + \beta_{21})\text{ Age}]$

$\hat{\beta}_1 = 0.286$, $\hat{\beta}_2 = 0.000$

$\hat{\beta}_{11} = 2.351$, $\hat{\beta}_{21} = 0.078$

The hazard function for this second stratum is shown here. Notice that the coefficients of the treatment and age variables are $(\beta_1 + \beta_{11})$ and $(\beta_2 + \beta_{21})$, respectively. The estimated values of each of these coefficients are given here.

Hazard functions for interaction model:

$g = 1$: $(Z_1^* = Z_2^* = Z_3^* = Z_4^* = 0)$:
$\hat{h}_1(t,\mathbf{X}) = \hat{h}_{01}(t)\exp[(0.286)\text{Treatment}]$

$g = 2$: $(Z_1^* = 1, Z_2^* = Z_3^* = Z_4^* = 0)$:
$\hat{h}_2(t,\mathbf{X}) = \hat{h}_{02}(t)\exp[(2.637)\text{Treatment} + (0.078)\text{Age}]$

$g = 3$: $(Z_2^* = 1, Z_1^* = Z_3^* = Z_4^* = 0)$:
$\hat{h}_3(t,\mathbf{X}) = \hat{h}_{03}(t)\exp[(-0.872)\text{Treatment} + (-0.047)\text{Age}]$

$g = 4$: $(Z_3^* = 1, Z_1^* = Z_2^* = Z_4^* = 0)$:
$\hat{h}_4(t,\mathbf{X}) = \hat{h}_{04}(t)\exp[(0.868)\text{Treatment} + (-0.059)\text{Age}]$

$g = 5$: $(Z_1^* = Z_2^* = Z_3^* = 0, Z_4^* = 1)$:
$\hat{h}_5(t,\mathbf{X}) = \hat{h}_{05}(t)\exp[(-0.747)\text{Treatment} + (0.051)\text{Age}]$

$g = 6$: $(Z_1^* = 1, Z_2^* = Z_3^* = 0, Z_4^* = 1)$:
$\hat{h}_6(t,\mathbf{X}) = \hat{h}_{06}(t)\exp[(0.810)\text{Treatment} + (-0.038)\text{Age}]$

$g = 7$: $(Z_2^* = 1, Z_1^* = Z_3^* = 0, Z_4^* = 1)$:
$\hat{h}_7(t,\mathbf{X}) = \hat{h}_{07}(t)\exp[(0.880)\text{Treatment} + (-0.041)\text{Age}]$

$g = 8$: $(Z_3^* = 1, Z_1^* = Z_2^* = 0, Z_4^* = 1)$:
$\hat{h}_8(t,\mathbf{X}) = \hat{h}_{08}(t)\exp[(0.297)\text{Treatment} + (0.033)\text{Age}]$

The corresponding *estimated* hazard function for the second stratum (i.e., $g = 2$) is shown here. For comparison, we repeat the estimated hazard function for the first stratum.

The estimated hazard functions for the remaining strata are provided here. We leave it up to the reader to verify these formulae. Notice that the coefficients of treatment are all different in the eight strata, and the coefficients of age also are all different in the eight strata.

LR test to compare no-interaction model with interaction model:

We have presented computer results for both the no-interaction and the interaction models. To evaluate whether the no-interaction assumption is satisfied, we need to carry out a likelihood ratio test to compare these two models.

H_0: no-interaction model acceptable, i.e.,
Treatment: $\beta_{11} = \beta_{12} = \cdots = \beta_{17} = 0$
and Age: $\beta_{21} = \beta_{22} = \cdots = \beta_{27} = 0$

The null hypothesis being tested is that the no-interaction model is acceptable. Equivalently, this null hypothesis can be stated by setting the coefficients of all product terms in the interaction model to zero. That is, the seven coefficients of product terms involving treatment and the seven coefficients of the product terms involving age are set equal to zero as shown here.

14 coefficients $\Rightarrow$ df = 14

Because the null hypothesis involves 14 coefficients, the degrees of freedom of the LR chi-square statistic is 14. The test statistic takes the usual form involving the difference between log-likelihood statistics for the reduced and full models, where the reduced model is the no-interaction model and the full model is the interaction model.

$LR = -2 \ln L_R - (2 \ln L_F)$

R = reduced (no-interaction) model

F = full (interaction) model

EXAMPLE (continued)

$LR \sim \chi^2_{14df}$ under H_0: no interaction

$LR = (-2 \times -262.020) - (-2 \times -249.972)$
$= 524.040 - 499.944 = 24.096$
$P = 0.045$ (significant at 0.05)
Conclusion:
Reject H_0: interaction model is preferred.

Might use further testing to simplify interaction model, e.g., test for seven products involving treatment or test for seven products involving age.

Thus, under the null hypothesis, the *LR* statistic is approximately chi-square with 14 degrees of freedom.

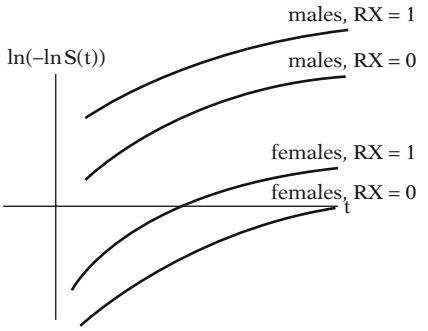
The computer results for the no-interaction and interaction models give log-likelihood values of 524.040 and 499.944, respectively. The difference is 24.096. A chi-square value of 24.096 with 14 degrees of freedom yields a p-value of 0.045, so that the test gives a significant result at the 0.05 level. This indicates that the no-interaction model is not acceptable and the interaction model is preferred.

Note, however, that it may be possible from further statistical testing to simplify the interaction model to have fewer than 14 product terms. For example, one might test for only the seven product terms involving treatment or only the seven product terms involving age.

VI. A Graphical View of the Stratified Cox Approach

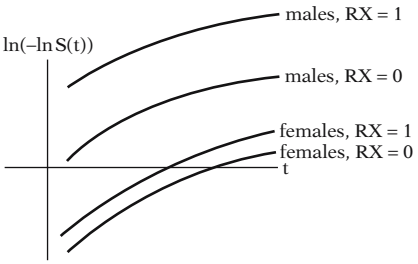
In this section we examine four log–log survival plots illustrating the assumptions underlying a stratified Cox model with or without interaction. Each of the four models considers two dichotomous predictors: treatment (coded RX = 1 for placebo and RX = 0 for new treatment) and SEX (coded 0 for females and 1 for males). The four models are as follows (see left).

a. $h(t) = h_0(t)\exp(\beta_1 RX + \beta_2 SEX)$
$\ln(-\ln S(t)) = \ln(-\ln S_0(t)) + \beta_1 RX + \beta_2 SEX$



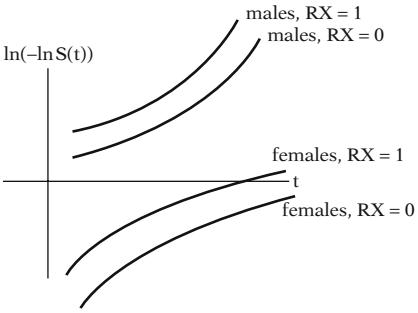
a. $h_0(t)\exp(\beta_1 RX + \beta_2 SEX)$. This model assumes the PH assumption for both RX and SEX and also assumes no interaction between RX and SEX. Notice all four log–log curves are parallel (PH assumption) and the effect of treatment is the same for females and males (no interaction). The effect of treatment (controlling for SEX) can be interpreted as the distance between the log–log curves from RX = 1 to RX = 0, for males and for females, separately.

b. $h(t) = h_0(t)\exp(\beta_1 RX + \beta_2 SEX + \beta_3 RX \times SEX)$
$\ln(-\ln S(t)) = \ln(-\ln S_0(t)) + \beta_1 RX + \beta_2 SEX + \beta_3 RX \times SEX$



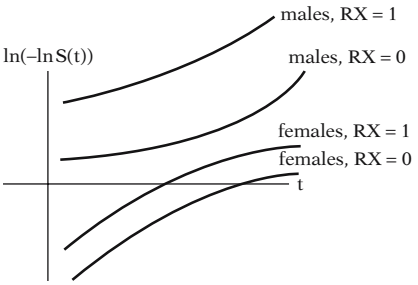
b. $h(t) = h_0(t)\exp(\beta_1 RX + \beta_2 SEX + \beta_3 RX \times SEX)$. This model assumes the PH assumption for both RX and SEX and allows for interaction between these two variables. All four log–log curves are parallel (PH assumption) but the effect of treatment is larger for males than females as the distance from RX = 1 to RX = 0 is greater for males.

c. $h(t) = h_{0g}(t)\exp(\beta_1 RX)$ (g = 1 for males, g = 0 for females)
$\ln(-\ln S(t)) = \ln(-\ln S_{0g}(t)) + \beta_1 RX$



c. $h(t) = h_{0g}(t)\exp(\beta_1 RX)$, where g = 1 for males, g = 0 for females. This is a stratified Cox model in which the PH assumption is not assumed for SEX. Notice the curves for males and females are not parallel. However, the curves for RX are parallel within each stratum of SEX indicating that the PH assumption is satisfied for RX. The distance between the log–log curves from RX = 1 to RX = 0 is the same for males and females indicating no interaction between RX and SEX.

d. $h(t) = h_{0g}(t)\exp(\beta_1 RX + \beta_2 RX \times SEX)$ (g = 1 for males, g = 0 for females)
$\ln(-\ln S(t)) = \ln(-\ln S_{0g}(t)) + \beta_1 RX + \beta_2 RX \times SEX$



d. $h(t) = h_{0g}(t)\exp(\beta_1 RX + \beta_2 RX \times SEX)$, where g = 1 for males, g = 0 for females. This is a stratified Cox model allowing for interaction of RX and SEX. The curves for males and females are not parallel although the PH assumption is satisfied for RX within each stratum of SEX. The distance between the log–log curves from RX = 1 to RX = 0 is greater for males than females indicating interaction between RX and SEX.

VII. Summary

We now summarize the most important features of the stratified Cox (SC) model described in this presentation.

Stratified Cox (SC) model:

- stratification of predictors not satisfying PH assumption
- includes predictors satisfying PH
- does not include stratified variables

The SC model is a modification of the Cox PH model to allow for control by "stratification" of predictors not satisfying the PH assumption. Variables that are assumed to satisfy the assumption are included in the model as predictors; the stratified variables are not included in the model.

Computer Results

The computer results for a SC model provides essentially the same type of output as provided for a Cox PH model without stratification. An example of SC output using the remission data is shown here. The variables included as predictors in the model are listed in the first column followed by their estimated coefficients, standard errors, p-values, hazard ratio values, and 95% confidence limits. Such information cannot be provided for the variables being stratified, because these latter variables are not explicitly included in the model.

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf. Interval]	
log WBC	1.390	0.338	0.000	4.016	2.072	7.783
RX	0.931	0.472	0.048	2.537	1.006	6.396
No. of subjects = 42		Log likelihood = −57.560			Stratified by sex	

Hazard function for stratified Cox model:

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p]$$

$g = 1, 2, \ldots, k^*$, strata defined from Z^*

Z^* has k^* categories

$X_1, X_2, \ldots, X_p$ satisfy PH

The general hazard function form for the stratified Cox model is shown here. This formula contains a subscript g that indicates the gth stratum, where the strata are different categories of the stratification variable Z^* and the number of strata equals k^*. Notice that the baseline hazard functions are different in each stratum.

Stratification variable Z^*:

- identify $Z_1, Z_2, \ldots, Z_k$ not satisfying PH
- categorize each Z
- form combinations of categories (strata)
- each combination is a stratum of Z^*

The variable Z^* is defined by first identifying the Z_i variables not satisfying the PH assumption. We then categorize each Z and form combinations of categories of each of the Z's. Each combination represents a different stratum making up the variable Z^*.

No-interaction model:
Same coefficients $\beta_1, \beta_2, \ldots, \beta_p$ for each g, i.e., Z^* does not interact with the X's.

The above model is designated as a "no-interaction" model because the β's in the model are the same for each subscript g. The no-interaction assumption means that the variables being stratified are assumed *not* to interact with the X's in the model.

Different baselines
$$\begin{cases} h_{01}(t) \Rightarrow \hat{S}_1(t) \\ h_{02}(t) \Rightarrow \hat{S}_2(t) \\ \vdots \\ h_{0k}(t) \Rightarrow \hat{S}_{k^*}(t) \end{cases}$$
Different survival curves

For the no-interaction model, the fitted SC model will yield different estimated survival curves for each stratum because the baseline hazard functions are different for each stratum.

$\widehat{HR}$ same for each stratum

However, because the coefficients of the X's are the same for each stratum, estimates of hazard ratios are the same for each stratum.

(Partial) likelihood function:

$$L = L_1 \times L_2 \times \cdots \times L_{k^*}$$

Regression coefficients in the SC model are estimated by maximizing a partial likelihood function that is obtained by multiplying likelihood functions for each stratum.

Stratified Cox model allowing interaction:

$$h_g(t, \mathbf{X}) = h_{0g}(t) \exp[\beta_{1g} X_1 + \beta_{2g} X_2 + \cdots + \beta_{pg} X_p]$$
$g = 1, 2, \ldots, k^*$, strata defined from Z^*.

In order to evaluate the no-interaction assumption, we must define an interaction model for comparison. One version of the interaction model is shown here. This version shows regression coefficients with different subscripts in different strata; that is, each β coefficient has a subscript g.

Alternative stratified Cox interaction model:

- uses product terms involving Z^*
- define $k^* - 1$ dummy variables from Z^*
- products of the form $Z_i^* \times X_j$

An alternative way to write the interaction model uses product terms involving the Z^* variable with each predictor. This model uses k^*-1 dummy variables to distinguish the k^* categories of Z^*. Each of these dummy variables is included as a product term with each of the X's.

Testing the no-interaction assumption:

$LR = -2 \ln L_R - (2 \ln L_F)$
$R =$ reduced (no-interaction) model
$F =$ full (interaction) model contains product terms
$LR \overset{\sim}{\sim} \chi^2_{p(k^*-1)\text{df}}$ under H_0: no interaction

To evaluate the no-interaction assumption, we can perform a likelihood ratio test that compares the (reduced) no-interaction model to the (full) interaction model. The null hypothesis is that the no-interaction assumption is satisfied. The test statistic is given by the difference between the log-likelihood statistics for the no-interaction and interaction models. This statistic is approximately chi-square under the null hypothesis. The degrees of freedom is $p(k^*-1)$ where p denotes the number of X's and k^* is the number of categories making up Z^*.

PRESENTATION COMPLETE!

Chapters

1. Introduction to Survival Analysis
2. Kaplan–Meier Survival Curves and the Log–Rank Test
3. The Cox Proportional Hazards Model and Its Characteristics
4. Evaluating the Proportional Hazards Assumption

√5. The Stratified Cox Procedure

Next:

6. Extension of the Cox Proportional Hazards Model for Time-Dependent Variables

This presentation is now complete. We suggest that the reader review this presentation using the detailed outline that follows. Then answer the practice exercises and the test that follow.

The next Chapter (6) is entitled "Extension of the Cox PH Model for Time-Dependent Variables." There we show how an "extended" Cox model can be used as an alternative to the stratified Cox model when one or more predictors do not satisfy the PH assumption. We also discuss more generally what is a time-dependent variable, and show how such a variable can be evaluated using an extended Cox model.

Detailed Outline

I. Preview (page 176)

A. Focus on how stratified Cox (SC) procedure is carried out:
- analysis of computer results from SC procedure;
- hazard function for SC model;
- stratifying on a single predictor versus two or more predictors;
- no-interaction versus interaction models.

II. An Example (pages 176–180)

A. Cox PH results for remission data yield $P(PH) = 0.031$ for SEX.

B. SC model used: control for SEX (stratified); include log WBC and *Rx* in model.

C. Analysis of *Rx* effect from stratified Cox results:
$\widehat{HR} = 2.537$; 95% CI: (1.006,6.396); LR and Wald tests: $P < 0.05$.

D. Hazard model: $h_g(\mathrm{t}, \mathbf{X}) = h_{0g}(t) \exp[\beta_1 \log \mathrm{WBC} + \beta_2 Rx], g = 1,2$
- different baseline hazard functions and survival curves for females and males;
- same coefficients β_1 and β_2 for both females and males (no-interaction assumption);
- obtain estimates by maximizing partial likelihood $L = L_1 \times \mathrm{L}_2$.

E. Graph of four adjusted survival curves for *Rx* (adjusted for log WBC).

III. The General Stratified Cox (SC) Model (pages 180–181)

A. $$\boxed{h_g(t,\mathbf{X}) = h_{0g}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p], \quad g = 1, 2, \cdots, k^*},$$
where the strata are defined from the stratification variable Z^*.

B. Z^* defined from $Z_1, Z_2, \ldots, Z_k$ variables that do not satisfy PH:
- categorize each Z_i
- form combinations of categories
- each combination is a stratum of Z^*

C. Different baseline hazard functions and survival curves for each stratum.

D. Assumes no interaction: same coefficients $\beta_1, \beta_2, \ldots, \beta_p$ for each g; i.e., Z^* does not interact with the X's; i.e., estimated HR is same for each stratum.

E. Obtain estimates by maximizing partial likelihood $L = L_1 \times L_2 \times \cdots \times L_{k^*}$, where L_i is likelihood for ith stratum.

IV. The No-Interaction Assumption and How to Test It (pages 182–188)

A. Assumes same coefficients $\beta_1, \beta_2, \ldots, \beta_p$ for each g.

B. Interaction model:

$$\boxed{h_g(t, \mathbf{X}) = h_{0g}(t)\mathbf{exp}[\beta_{1g} X_1 + \beta_{2g} X_2 + \cdots + \beta_{pg} X_p]},$$

$g = 1, 2, \ldots, k^*$ strata defined from Z^*.

C. Alternative stratified Cox interaction model:

- uses product terms involving Z^*
- define k^*-1 dummy variables $Z_1^*, Z_2^*, \ldots, Z_{k^*-1}{}^*$ from Z^*
- products of the form $Z_i^* \times X_j$, where $i = 1, \ldots, k^* - 1; j = 1, \ldots, p$
- hazard function: $g = 1, 2, \ldots, k^*$ strata defined from Z^*

$$\boxed{\begin{aligned} h_g(t,\mathbf{X}) = h_{0g}(t)\mathbf{exp}[&\beta_1 X_1 + \cdots + \beta_p X_p + \beta_{11}(Z_1^* \times \mathrm{X}_1) \\ &+ \cdots + \beta_{p1}(Z_1^* \times X_p) + \beta_{12}(Z_2^* \times X_1) + \cdots + \beta_{p2}(Z_2^* \times X_p) \\ &+ \cdots + \beta_{1,k^*-1}(Z_{k^*-1}^* \times X_1) + \cdots + \beta_{p,k^*-1}(Z_{k^*-1}^* \times X_p)] \end{aligned}}$$

D. Testing the no-interaction assumption: use LR statistic given by $LR = -2 \ln L_R - (-2 \ln L_F)$ where R = reduced (no interaction) model and F = full (interaction) model
$LR \dot{\sim} \chi^2_{p(k^*-1)\mathrm{df}}$ under H_0: no interaction,
i.e., $\beta_{11} = \beta_{21} = \ldots = \beta_{p,k^*-1} = 0$

V. A Second Example Involving Several Stratification Variables (pages 188–193)

A. Dataset "vets.dat" from Veteran's Administration Lung Cancer Trial; $n = 137$; survival time in days.

B. Variables are: treatment status, cell type (four types), performance status, disease duration, age, and prior therapy status.

C. Cox PH results indicate [using $P(PH)$] that cell type and performance status do not satisfy PH assumption.

D. Example stratifies on cell type and performance status using four categories of cell type and two categories of performance status, so that Z^* has $k^* = 8$ strata.

E. X's considered in model are treatment status and age.

F. Computer results for no-interaction model: estimated HR for effect of treatment adjusted for age and Z^* is 1.134 ($P = 0.548$); not significant.

G. Hazard function for no-interaction model: $h_g(t,\mathbf{X}) = h_{0g}(t)\,\mathbf{exp}[\beta_1\,\text{Treatment} + \beta_2\,\text{Age}]$, $g = 1, 2, \ldots, 8$

H. Hazard function for interaction model: $h_g(t,\mathbf{X}) = h_{0g}(t)\,\mathbf{exp}[\beta_{1g}\,\text{Treatment} + \beta_{2g}\,\text{Age}]$, $g = 1, 2, \ldots, 8$

I. Alternative version of interaction model: $h_g(t,\mathbf{X}) = h_{0g}(t)\,\mathbf{exp}[\beta_1\,\text{Treatment} + \beta_2\,\text{Age} + \beta_{11}(Z_1^* \times \text{Treatment}) + \cdots + \beta_{17}(Z_7^* \times \text{Treatment}) + \beta_{21}(Z_1^* \times \text{Age}) + \cdots + \beta_{27}(Z_7^* \times \text{Age})]$, $g = 1, 2, \ldots, 8$
where $Z_1^* =$ large cell (binary), $Z_2^* =$ adeno cell (binary), $Z_3^* =$ small cell (binary), $Z_4^* =$ PSbin (binary), $Z_5^* = Z_1^* \times Z_4^*$, $Z_6^* = Z_2^* \times Z_4^*$, $Z_7^* = Z_3^* \times Z_4^*$

J. Demonstration that alternative interaction version (in item I) is equivalent to original interaction formulation (in item H) using computer results for the alternative version.

K. Test of no-interaction assumption:
- null hypothesis: $\beta_{11} = \beta_{12} = \ldots = \beta_{17} = 0$ and $\beta_{21} = \beta_{22} = \ldots = \beta_{27} = 0$
- $LR \sim \chi^2_{14\,\text{df}}$ under H_0: no interaction
- $LR = 524.040 - 499.944 = 24.096$ ($P = 0.045$)
  Conclusion: Reject null hypothesis; interaction model is preferred.

VI. A Graphical View of the Stratified Cox Approach (pages 193–194)
Comparison of log–log survival curves

1. Describe interaction of Rx and Sex.
2. Describe violation of PH assumption for Sex.

VII. Summary (pages 195–196)

Practice Exercises

The following questions derive from the dataset **vets.dat** concerning the Veteran's Administration Lung Cancer Trial that we previously considered in the presentation on the stratified Cox model. Recall that survival times are in days and that the study size contains 137 patients. The exposure variable of interest is treatment status (standard = 1, test = 2). Other variables of interest as control variables are cell type (four types, defined in terms of dummy variables), performance status, disease duration, age, and prior therapy status. Failure status is defined by the status variable (0 = censored, 1 = died).

1. Consider the following two edited printouts obtained from fitting a Cox PH model to these data.

Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]	*P*(*PH*)
Treatment	0.290	0.207	0.162	1.336	0.890	2.006	0.628
Large cell	0.400	0.283	0.157	1.491	0.857	2.594	0.033
Adeno cell	1.188	0.301	0.000	3.281	1.820	5.915	0.081
Small cell	0.856	0.275	0.002	2.355	1.374	4.037	0.078
Perf.Stat	−0.033	0.006	0.000	0.968	0.958	0.978	0.000
Dis.Durat.	0.000	0.009	0.992	1.000	0.982	1.018	0.919
Age	−0.009	0.009	0.358	0.991	0.974	1.010	0.198
Pr.Therapy	0.007	0.023	0.755	1.007	0.962	1.054	0.145

No. of subjects = 137 Log likelihood = −475.180

Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]	*P*(*PH*)
Treatment	0.298	0.197	0.130	1.347	0.916	1.981	0.739
Small cell	0.392	0.210	0.062	1.481	0.981	2.235	0.382
Perf.Stat	−0.033	0.005	0.000	0.968	0.958	0.978	0.000
Dis.Durat.	−0.001	0.009	0.887	0.999	0.981	1.017	0.926
Age	−0.006	0.009	0.511	0.994	0.976	1.012	0.211
Pr.Therapy	−0.003	0.023	0.884	0.997	0.954	1.042	0.146

No. of subjects = 137 Log likelihood = −487.770

How do the printouts differ in terms of what the *P*(*PH*) information says about which variables do not satisfy the PH assumption?

2. Based on the above information, if you were going to stratify on the cell type variable, how would you define the strata? Explain.

3. Consider a stratified analysis that stratifies on the variables Z_1 = "small cell" and Z_2 = "performance status." The small cell variable is one of the dummy variables for cell type defined above. The performance status variable is dichotomized into high (60 or above) and low (below 60) and is denoted as PSbin. The stratification variable which combines categories from Z_1 and Z_2 is denoted as SZ^* and consists of four categories. The predictors included (but not stratified) in the analysis are treatment status, disease duration, age, and prior therapy. The computer results are as follows:

Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Treatment	0.090	0.197	0.647	1.095	0.744	1.611
Dis.Durat.	0.000	0.010	0.964	1.000	0.982	1.019
Age	0.002	0.010	0.873	1.002	0.983	1.021
Pr.Therapy	−0.010	0.023	0.656	0.990	0.947	1.035

No. of subjects = 137 Log likelihood = −344.848 Stratified by SZ*

Based on these results, describe the point and interval estimates for the hazard ratio for the treatment effect adjusted for the other variables, including SZ^*. Is this hazard ratio meaningfully and/or statistically significant? Explain.

4. State the form of the hazard function for the model being fit in question 3. Why does this model assume no interaction between the stratified variables and the predictors in the model?
5. State two alternative ways to write the hazard function for an "interaction model" that allows for the interaction of the stratified variables with the treatment status variable, but assumes no other type of interaction.
6. State two alternative versions of the hazard function for an interaction model that allows for the interaction of the stratified variables (small cell and performance status) with each of the predictors treatment status, disease duration, age, and prior therapy.
7. For the interaction model described in question 6, what is the formula for the hazard ratio for the effect of treatment adjusted for the other variables? Does this formula give a different hazard ratio for different strata? Explain.

8. State two alternative versions of the null hypothesis for testing whether the no-interaction assumption is satisfied for the stratified Cox model. Note that one of these versions should involve a set of regression coefficients being set equal to zero.
9. State the form of the likelihood ratio statistic for evaluating the no-interaction assumption. How is this statistic distributed under the null hypothesis, and with what degrees of freedom?
10. Provided below are computer results for fitting the interaction model described in question 6. In this printout the variable Z_1^* denotes the small cell variable and the variable Z_2^* denotes the PSbin variable. The variable DDZ_1^* denotes the product of Z_1^* with disease duration, and other product terms are defined similarly.

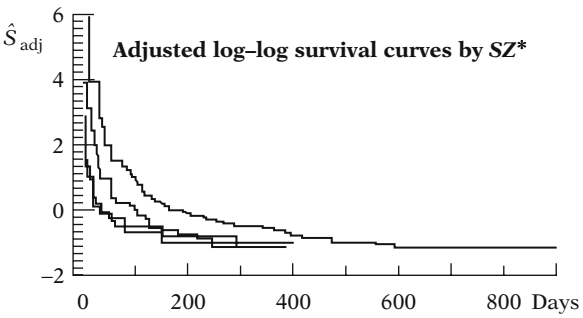
Stratified Cox regression
Analysis time _t: survt

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Treatment	0.381	0.428	0.374	1.464	0.632	3.389
Dis.Durat.	0.015	0.021	0.469	1.015	0.975	1.057
Age	0.000	0.017	0.994	1.000	0.968	1.033
Pr.Therapy	0.023	0.041	0.571	1.023	0.944	1.109
DDZ_1^*	−0.029	0.024	0.234	0.971	0.926	1.019
$AgeZ_1^*$	−0.055	0.037	0.135	0.946	0.880	1.018
PTZ_1^*	0.043	0.075	0.564	1.044	0.901	1.211
DDZ_2^*	0.025	0.032	0.425	1.026	0.964	1.092
$AgeZ_2^*$	0.001	0.024	0.956	1.001	0.956	1.049
PTZ_2^*	−0.078	0.054	0.152	0.925	0.831	1.029
$DDZ_1Z_2^*$	−0.071	0.059	0.225	0.931	0.830	1.045
$AgeZ_1Z_2^*$	0.084	0.049	0.084	1.088	0.989	1.196
$PTZ_1Z_2^*$	−0.005	0.117	0.963	0.995	0.791	1.250
trZ_1^*	0.560	0.732	0.444	1.751	0.417	7.351
trZ_2^*	−0.591	0.523	0.258	0.554	0.199	1.543
$trZ_1Z_2^*$	−0.324	0.942	0.731	0.723	0.114	4.583

No. of subjects = 137 Log likelihood = −335.591 Stratified by SZ^*

Use the above computer results to state the form of the **estimated** hazard model for each of the four strata of the stratification variable SZ^*. Also, for each strata, compute the hazard ratio for the treatment effect adjusted for disease duration, age, and prior therapy.

11. Carry out the likelihood ratio test to evaluate the no-interaction model described in question 4. In carrying out this test, make sure to state the null hypothesis in terms of regression coefficients being set equal to zero in the interaction model fitted in question 10. Also, determine the p-value for this test and state your conclusions about significance as well as which model you prefer, the no-interaction model or the interaction model.
12. The adjusted log–log survival curves for each of the four strata defined by the stratification variable SZ^* (adjusted for treatment status, disease duration, age, and prior therapy) are presented below.



Using this graph, what can you conclude about whether the PH assumption is satisfied for the variables, small cell type and PSbin?

13. Comment on what you think can be learned by graphing adjusted survival curves that compare the two treatment groups for each of the four strata of SZ^*.

Test

The following questions consider a dataset from a study by Caplehorn et al. ("Methadone Dosage and Retention of Patients in Maintenance Treatment," *Med. J. Aust.*, 1991). These data comprise the times in days spent by heroin addicts from entry to departure from one of two methadone clinics. Two other covariates, namely, prison record and maximum methadone dose, are believed to affect the survival times. The dataset name is **addicts.dat.** A listing of the variables is given below:

Column 1: Subject ID
Column 2: Clinic (1 or 2)

Column 3: Survival status (0 = censored, 1 = departed from clinic)
Column 4: Survival time in days
Column 5: Prison record (0 = none, 1 = any)
Column 6: Maximum methadone dose (mg/day)

1. The following edited printout was obtained from fitting a Cox PH model to these data:

Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]	*P*(*PH*)
clinic	−1.009	0.215	0.000	0.365	0.239	0.556	0.001
prison	0.327	0.167	0.051	1.386	0.999	1.924	0.332
dose	−0.035	0.006	0.000	0.965	0.953	0.977	0.341

No. of subjects = 238 Log likelihood = −673.403

Based on the *P*(*PH*) information in the above printout, it appears that clinic does not satisfy the PH assumption; this conclusion is also supported by comparing log–log curves for the two clinics and noticing strong nonparallelism. What might we learn from fitting a stratified Cox (SC) model stratifying on the clinic variable? What is a drawback to using a SC procedure that stratifies on the clinic variable?

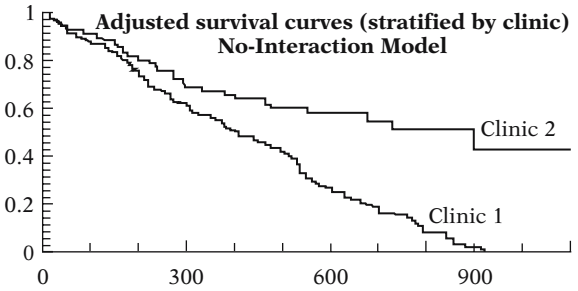
2. The following printout was obtained from fitting a SC PH model to these data, where the variable being stratified is clinic:

Stratified Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Prison	0.389	0.169	0.021	1.475	1.059	2.054
Dose	−0.035	0.006	0.000	0.965	0.953	0.978

No. of subjects = 238 Log likelihood = −597.714 Stratified by clinic

Using the above fitted model, we can obtain the adjusted curves below that compare the adjusted survival probabilities for each clinic (i.e., stratified by clinic) adjusted for the variables, prison and maximum methadone dose.



Based on these adjusted survival curves, what conclusions can you draw about whether the survival experience is different between the two clinics? Explain.

3. State the hazard function model being estimated in the above computer results. Why is this model a no-interaction model?
4. Using the above computer results, provide point and interval estimates for the effect of prison adjusted for clinic and dose. Is this adjusted prison effect significant? Explain.
5. The following computer results consider a SC model that allows for interaction of the stratified variable clinic with each of the predictors, prison and dose. Product terms in the model are denoted as clinpr = clinic × prison and clindos = clinic × dose.

Stratified Cox regression
Analysis time _t: survt

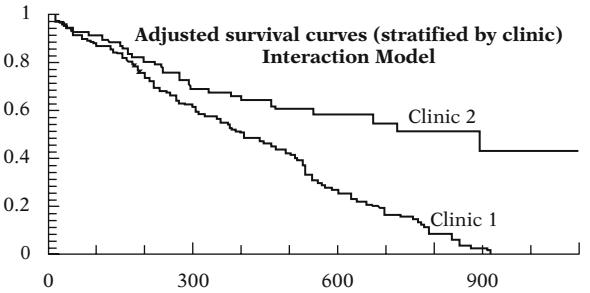
	Coef.	Std. Err.	P > |z|	Haz. Ratio	[95% Conf.	Interval]
prison	1.087	0.539	0.044	2.966	1.032	8.523
dose	−0.035	0.020	0.079	0.966	0.929	1.004
clinpr	−0.585	0.428	0.172	0.557	0.241	1.290
clindos	−0.001	0.015	0.942	0.999	0.971	1.028

No. of subjects = 238 Log likelihood = −596.779 Stratified by clinic

State two alternative versions of the interaction model being estimated by the above printout, where one of these versions should involve the product terms used in the above printout.

6. Using the computer results above, determine the estimated hazard models for each clinic. (Note that the clinics are coded as 1 or 2.)

7. Below are the adjusted survival curves for each clinic based on the interaction model results above. These curves are adjusted for the prison and dose variables.



   Compare the survival curves by clinic obtained for the interaction model with the corresponding curves previously shown for the no-interaction model. Do both curves indicate the similar conclusions about the clinic effect? Explain.

8. Carry out a likelihood ratio test to determine whether the no-interaction model is appropriate. In doing so, make use of the computer information described above, state the null hypothesis, state the form of the likelihood statistic and its distribution under the null hypothesis, and compute the value of the likelihood statistic and evaluate its significance. What are your conclusions?

Answers to Practice Exercises

1. The first printout indicates that the variables large cell, adeno cell, small cell, and performance status do not satisfy the PH assumption at the 0.10 level. The second printout considers a different model that does not contain the large cell and adeno cell variables. This latter printout indicates that small cell satisfies the PH assumption, in contrast to the first printout. The performance status variable, however, does not satisfy the PH assumption as in the first printout.

2. The cell type variable is defined to have four categories, as represented by the three dummy variables in the first printout. The "small cell" variable dichotomizes the cell type variable into the categories small cell type versus the rest. From the second printout, the small cell variable does not appear by itself to violate the PH assumption. This result conflicts with the results of the first printout, for which the cell type variable considered in four categories does not

satisfy the PH assumption at the 0.10 level of significance. We therefore think it is more appropriate to use a SC procedure only if four strata are to be used. A drawback to using four strata, however, is that the number of survival curves to be plotted is larger than for two strata; consequently, a large number of curves is more difficult to interpret graphically than when there are only two curves. Thus, for convenience of interpretation, we may choose to dichotomize the cell type variable instead of considering four strata. We may also consider dichotomies other than those defined by the small cell variable. For instance, we might consider dichotomizing on either the adeno or large cell variables instead of the small cell variable. Alternatively, we may combine categories so as to compare, say, large and adeno cell types with small and squamous types. However, a decision to combine categories should not be just a statistical decision, but should also be based on biologic considerations.

3. $\widehat{HR}_{\text{adj}} = 1.095$, 95% CI: (0.744,1.611), two-tailed P-value is 0.647, not significant. The estimated hazard ratio for treatment is neither meaningfully or statistically significant. The point estimate is essentially 1, which says that there is no meaningful effect of treatment adjusted for the predictors in the model and for the stratified predictor SZ^*.

4. $h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1 \text{Treatment} + \beta_2 DD + \beta_3 \text{Age} + \beta_4 PT]$, $g = 1, \ldots, 4$, where the strata are defined from the stratification variable SZ^*, $DD =$ disease duration, and $PT =$ prior therapy. This model assumes no interaction because the coefficient of each predictor in the model is not subscripted by g, i.e., the regression coefficients are the same for each stratum.

5. Version 1: $h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_{1g}\ \text{Treatment} + \beta_2 DD + \beta_3\ \text{Age} + \beta_4 PT]$, $g = 1, \ldots, 4$.

   Version 2: $h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1\ \text{Treatment} + \beta_2 DD + \beta_3\ \text{Age} + \beta_4 PT + \beta_5(Z_1^* \times \text{Treatment}) + \beta_6(Z_2^* \times\ \text{Treatment}) + \beta_7(Z_1^* \times Z_2^* \times \text{Treatment})]$, where $Z_1^* =$ small cell type (0, 1), $Z_2^* =$ PSbin (0, 1), and $g = 1, \ldots, 4$.

6. Version 1: $h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_{1g}\ \text{Treatment} + \beta_{2g}\,DD + \beta_{3g}\ \text{Age} + \beta_{4g}\,PT]$, $g = 1, \ldots, 4$.

   Version 2: $h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_1\ \text{Treatment} + \beta_2\,DD + \beta_3\ \text{Age} + B_4\,PT + \beta_5(Z_1^* \times \text{Treatment}) + \beta_6(Z_1^* \times DD) + \beta_7(Z_1^* \times \text{Age}) + \beta_8(Z_1^* \times PT) + \beta_9(Z_2^* \times \text{Treatment}) + \beta_{10}(Z_2^* \times DD) + \beta_{11}(Z_2^* \times \text{Age}) + \beta_{12}(Z_2^* \times PT) + \beta_{13}(Z_1^* \times Z_2^* \times \text{Treatment}) + \beta_{14}(Z_1^* \times Z_2^* \times DD) + \beta_{15}(Z_1^* \times Z_2^* \times \text{Age}) + \beta_{16}(Z_1^* \times Z_2^* \times PT)]$, $g = 1, \ldots, 4$.

7. $HR_g = \exp^{(\beta_{1g})}$, using version 1 model form. Yes, this formula gives different hazard ratios for different strata because the value of the hazard ratio changes with the subscript g.

8. H_0: No interaction assumption is satisfied.

   H_0: $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14}$, $\beta_{21} = \beta_{22} = \beta_{23} = \beta_{24}$, $\beta_{31} = \beta_{32} = \beta_{33} = \beta_{34}$, $\beta_{41} = \beta_{42} = \beta_{43} = \beta_{44}$ from version 1.

   H_0: $\beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = \beta_{16} = 0$ from version 2.

9. $LR = -2\ln L_R - (-2\ln L_F)$, where R denotes the reduced (no-interaction) model and F denotes the full (interaction) model. Under the null hypothesis, LR is approximately a chi-square with 12 degrees of freedom.

10. Estimated hazard models for each stratum:

    $g = 1; Z_1^* = Z_2^* = 0$:
    $\hat{h}_1(t,\mathbf{X}) = \hat{h}_{01}(t)\exp[(0.381)\text{Treatment} + (0.015)DD + (0.000)\text{Age} + (0.023)PT]$

    $g = 2; Z_1^* = 1, Z_2^* = 0$:
    $\hat{h}_2(t,\mathbf{X}) = \hat{h}_{02}(t)\exp[(0.941)\text{Treatment} + (-0.014)DD + (-0.055)\text{Age} + (0.066)PT]$

    $g = 3; Z_1^* = 0, Z_2^* = 1$:
    $\hat{h}_3(t,\mathbf{X}) = \hat{h}_{03}(t)\exp[(-0.210)\text{Treatment} + (0.040)DD + (0.001)\text{Age} + (-0.055)PT]$

    $g = 4; Z_1^* = 1, Z_2^* = 1$:
    $\hat{h}_4(t,\mathbf{X}) = \hat{h}_{04}(t)\exp[(0.026)\text{Treatment} + (-0.060)DD + (0.030)\text{Age} + (-0.017)PT]$

Estimated hazard ratios for treatment effect adjusted for DD, Age, and PT:

$$g = 1: \widehat{HR}_1 = \exp(0.381) = 1.464$$
$$g = 2: \widehat{HR}_2 = \exp(0.941) = 2.563$$
$$g = 3: \widehat{HR}_3 = \exp(-0.210) = 0.811$$
$$g = 4: \widehat{HR}_4 = \exp(0.026) = 1.026$$

11. H_0: $\beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = \beta_{16} = 0$

    $LR = 689.696 - 671.182 = 18.514$, which is approximately chi-square with 12 df.

    $P = 0.101$, which is not significant below the .05 level.
    Conclusion: Accept the null hypothesis and conclude that the no-interaction model is preferable to the interaction model.

12. The three curves at the bottom of the graph appear to be quite non-parallel. Thus, the PH assumption is not satisfied for one or both of the variables, small cell type and PSbin. Note, however, that because both these variables have been stratified together, it is not clear from the graph whether only one of these variables fails to satisfy the PH assumption.

13. If we graph adjusted survival curves that compare the two treatment groups for each of the four strata, we will be able to see graphically how the treatment effect, if any, varies over time within each strata. The difficulty with this approach, however, is that eight adjusted survival curves will be produced, so that if all eight curves are put on the same graph, it may be difficult to see what is going on.

6

Extension of the Cox Proportional Hazards Model for Time-Dependent Variables

Introduction

We begin by defining a time-dependent variable and providing some examples of such a variable. We also state the general formula for a Cox model that is extended to allow time dependent variables, followed by a discussion of the characteristics of this model, including a description of the hazard ratio.

In the remainder of the presentation, we give examples of models with time-dependent variables, including models that allow for checking the PH assumption for time-independent variables. In particular, we describe a method that uses "heaviside functions" to evaluate the PH assumption for time-independent variables. We also describe two computer applications of the extended Cox model, one concerning a study on the treatment of heroin addiction and the other concerning the Stanford heart transplant study.

Abbreviated Outline

The outline below gives the user a preview of the material to be covered by the presentation. A detailed outline for review purposes follows the presentation.

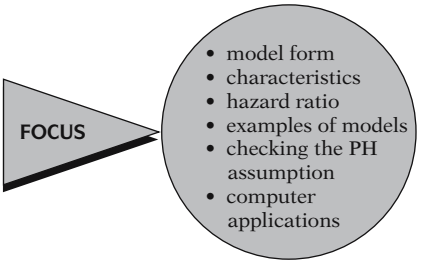
Objectives

Upon completing the chapter, the learner should be able to:

1. State or recognize the general form of the Cox model extended for time-dependent variables.
2. State the specific form of an extended Cox model appropriate for the analysis, given a survival analysis scenario involving one or more time-dependent variables.
3. State the formula for a designated hazard ratio of interest, given a scenario describing a survival analysis using an extended Cox model.
4. State the formula for an extended Cox model that provides a method for checking the PH assumption for one more of the time-independent variables in the model, given a scenario describing a survival analysis involving time-independent variables.
5. State the formula for an extended Cox model that uses one or more heaviside functions to check the PH assumption for one more of the time-independent variables in the model, given a scenario describing a survival analysis involving time-independent variables.
6. State the formula for the hazard ratio during different time interval categories specified by the heaviside function(s), for a model involving heaviside function(s).
7. Carry out an appropriate analysis of the data to evaluate the effect of one or more of the explanatory variables in the model(s) being used, given computer results for a survival analysis involving time-dependent variables. Such an analysis will involve:
   - computing and interpreting any hazard ratio(s) of interest;
   - carrying out and interpreting appropriate test(s) of hypotheses for effects of interest;
   - obtaining confidence intervals for hazard ratios of interest;
   - evaluating interaction and confounding involving one or more covariates.

Presentation

I. Preview



This presentation describes how the Cox proportional hazards (PH) model can be extended to allow time-dependent variables as predictors. Here, we focus on the model form, characteristics of this model, the formula for and interpretation of the hazard ratio, and examples of the extended Cox model. We also show how the extended Cox model can be used to check the PH assumption for time-independent variables, and we provide computer applications to illustrate different types of time-dependent variables. Finally, we describe the extended cox likelihood and how it contrasts with the Cox PH likelihood function.

II. Review of the Cox PH Model

$$h(t,\mathbf{X}) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i\right]$$

$\mathbf{X} = (X_1, X_2, \ldots, X_p)$
Explanatory/predictor variables

The general form of the Cox PH model is shown here. This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by the bold **X**. That is, the bold **X** represents a collection (sometimes called a "vector") of predictor variables that is being modeled to predict an individual's hazard.

$$h_0(t) \times \exp\left[\sum_{i=1}^{p}\beta_i X_i\right]$$

Baseline hazard	Exponential
Involves t but not X's	Involves X's but not t (X's are time-independent)

The Cox model formula says that the hazard at time t is the product of two quantities. The first of these, $h_0(t)$, is called the **baseline hazard** function. The second quantity is the exponential expression $\boldsymbol{e}$ to the linear sum of $\beta_i X_i$, where the sum is over the p explanatory X variables.

An important feature of this formula, which concerns the proportional hazards (PH) assumption, is that the baseline hazard is a function of t but does not involve the X's, whereas the exponential expression involves the X's but does not involve t. The X's here are called **time-independent** X's.

X's involving t: time dependent

Requires extended Cox model (no PH)

It is possible, nevertheless, to consider X's that do involve t. Such X's are called **time-dependent** variables. If time-dependent variables are considered, the Cox model form may still be used, but such a model no longer satisfies the PH assumption and is called the **extended Cox model**. We will discuss time-dependent variables and the corresponding extended Cox model beginning in the next section.

Hazard ratio formula:

$$\widehat{HR} = \exp\left[\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)\right]$$

where $\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_p^*)$ and $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ denote the two sets of X's.

From the Cox PH model, we can obtain a general formula, shown here, for estimating a hazard ratio that compares two specifications of the X's, defined as $\mathbf{X}^*$ and $\mathbf{X}$.

PH assumption:

$$\frac{\hat{h}(t,\mathbf{X}^*)}{\hat{h}(t,\mathbf{X})} = \hat{\theta} \text{ (a constant over } t)$$

i.e., $\hat{h}(t,\mathbf{X}^*) = \hat{\theta}\hat{h}(t,\mathbf{X})$

The (PH) assumption underlying the Cox PH model is that the hazard ratio comparing any two specifications of $\mathbf{X}$ predictors is constant over time. Equivalently, this means that the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time.

Hazards cross $\Rightarrow$ PH not met

Hazards don't cross $\nRightarrow$ PH met

An example of when the PH assumption is not met is given by any study situation in which the hazards for two or more groups cross when graphed against time. However, even if the hazard functions do not cross, it is possible that the PH assumption is not met.

Three approaches:

- graphical
- time-dependent variables
- goodness-of-fit test

As described in more detail in Chapter 4, there are three general approaches for assessing the PH assumption. These are

- a graphical approach;
- the use of time-dependent variables in an extended Cox model; and
- the use of a goodness-of-fit test.

Time-dependent covariates:

Extend Cox model: add product term(s) involving some function of time

When time-dependent variables are used to assess the PH assumption for a time-independent variable, the Cox model is extended to contain **product** (i.e., interaction) **terms** involving the time-independent variable being assessed and some function of time.

EXAMPLE

$h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 \text{sex} + \beta_2(\text{sex} \times t)]$

$H_0: \beta_2 = 0 \Rightarrow$ PH assumption satisfied

For example, if the PH assumption is being assessed for gender, a Cox model might be extended to include the variable sex $\times$ t in addition to sex. If the coefficient of the product term turns out to be non-significant, we can conclude that the PH assumption is satisfied for sex provided that the variable sex $\times$ t is an appropriate choice of time-dependent variable.

Options when PH assumption not satisfied:

- Use a stratified Cox (SC) model.
- Use time-dependent variables.

There are two options to consider if the PH assumption is not satisfied for one or more of the predictors in the model. In Chapter 5, we described the option of using a stratified Cox (SC) model, which stratifies on the predictor(s) not satisfying the PH assumption, while keeping in the model those predictors that satisfy the PH assumption. In this chapter, we describe the other option, which involves using time-dependent variables.

Time-dependent variables may be:

- inherently time-dependent
- defined to analyze a time-independent predictor not satisfying the PH assumption.

Note that a given study may consider predictors that are inherently defined as time-dependent, as we will illustrate in the next section. Thus, in addition to considering time-dependent variables as an option for analyzing a time-independent variable not satisfying the PH assumption, we also discuss predictors which are inherently defined as time-dependent.

III. Definition and Examples of Time-Dependent Variables

A time-dependent variable is defined as any variable whose value for a given subject may differ over time (t). In contrast, a time-independent variable is a variable whose value for a given subject remains constant over time.

As a simple example, the variable RACE is a time-independent variable, whereas the variable RACE $\times$ time is a time-dependent variable.

Definition:

Time-dependent	Time-independent
Value of variable differs over time	Value of variable is constant over time
Example:	
Race $\times$ t	Race

EXAMPLES OF DEFINED VARIABLES

Defined variable: RACE × t

Time-independent
Race = 1 ⇒ Race × t = t
Race = 0 ⇒ Race × t = 0 (at any t)

$E \times (\log t - 3)$
Function of t
[E denotes a (0,1) exposure variable].

$E \times g(t)$ where $g(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0 \end{cases}$

Heavyside function

$t \geq t_0$: $E \times g(t) = E$ [1]

$t < t_0$: $E \times g(t) = 0$ [0]

Heavyside functions used when PH assumptions not met.

The variable RACE × time is an example of what is called a "defined" time-dependent variable. Most defined variables are of the form of the product of a time-independent variable (e.g., RACE) multiplied by time or some function of time. Note that after RACE is determined for a given subject, all the values of the RACE × time variable are completely defined over a specified time interval of study.

A second example of a defined variable is given by $E \times (\log t - 3)$, where E denotes, say, a (0,1) exposure status variable determined at one's entry into the study. Notice that here we have used a function of time—that is, $\log t - 3$—rather than time alone.

Yet another example of a defined variable, which also involves a function of time, is given by $E \times g(t)$, where $g(t)$ is defined to take on the value 1 if t is greater than or equal to some specified value of t, called t_0, and takes on the value 0 if t is less than t_0.

The function $g(t)$ is called a "heaviside" function. Note that whenever t is greater than or equal to t_0, $g(t)$ equals 1, so $E \times g(t) = E$; however, whenever t is less than t_0, $g(t) = 0$, so the value of E × $g(t)$ is always 0. We will later return to illustrate how heaviside functions may be used as one method for the analysis when a time-independent variable like E does not satisfy the proportional hazards assumption.

Internal variable:

EXAMPLES OF INTERNAL VARIABLES

$E(t)$, $EMP(t)$, $SMK(t)$, $OBS(t)$,

Values change because of "internal" characteristics or behavior of the individual.

Another type of time-dependent variable is called an "internal" variable. Examples of such a variable include exposure level E at time t, employment status (EMP) at time t, smoking status (SMK) at time t, and obesity level (OBS) at time t.

All these examples consider variables whose values may change over time for any subject under study; moreover, for internal variables, the reason for a change in value depends on "internal" characteristics or behavior specific to the individual.

"Ancillary" variable:
Value changes because of "external" characteristics.

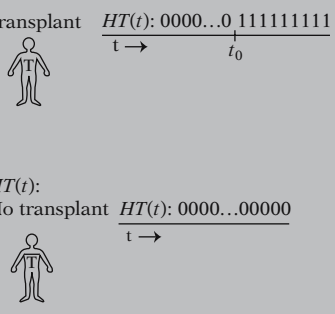
EXAMPLES OF ANCILLARY VARIABLES

Air pollution index at time t; $EMP(t)$

ANOTHER EXAMPLE

Heart transplant status at time t:

$$HT(t) = \begin{cases} 1 & \text{if received transplant at some time } t_0 \le t \\ 0 & \text{if did not receive transplant by time } t \end{cases}$$



Heart transplant status = $HT(t)$

Internal: Status determined from individual traits

Ancillary: Status determined from external availability of a donor

In contrast, a variable is called an "ancillary" variable if its value changes primarily because of "external" characteristics of the environment that may affect several individuals simultaneously. An example of an ancillary variable is air pollution index at time t for a particular geographical area. Another example is employment status (EMP) at time t, if the primary reason for whether someone is employed or not depends more on general economic circumstances than on individual characteristics.

As another example, which may be part internal and part ancillary, we consider heart transplant status (HT) at time t for a person identified to have a serious heart condition, making him or her eligible for a transplant. The value of this variable HT at time t is 1 if the person has already received a transplant at some time, say t_0, prior to time t. The value of HT is 0 at time t if the person has not yet received a transplant by time t.

Note that once a person receives a transplant, at time t_0, the value of HT remains at 1 for all subsequent times. Thus, for a person receiving a transplant, the value of HT is 0 up to the time of transplant, and then remains at 1 thereafter. In contrast, a person who never receives a transplant has HT equal to 0 for all times during the period he or she is in the study.

The variable "heart transplant status," $HT(t)$, can be considered essentially an internal variable, because individual traits of an eligible transplant recipient are important determinants of the decision to carry out transplant surgery. Nevertheless, the availability of a donor heart prior to tissue and other matching with an eligible recipient can be considered an "ancillary" characteristic external to the recipient.

Computer commands differ for defined vs. internal vs. ancillary.

But, the form of extended Cox model and procedures for analysis are the same regardless of variable type.

The primary reason for distinguishing among defined, internal, or ancillary variables is that the computer commands required to define the variables for use in an extended Cox model are somewhat different for the different variable types, depending on the computer program used. Nevertheless, the form of the extended Cox model is the same regardless of variable type, and the procedures for obtaining estimates of regression coefficients and other parameters, as well as for carrying out statistical inferences, are also the same.

IV. The Extended Cox Model for Time-Dependent Variables

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i X_i + \sum_{j=1}^{p_2}\delta_j X_j(t)\right]$$

$$\mathbf{X}(t) = \underbrace{(X_1, X_2, \ldots X_{p_1}}_{\text{Time-independent}},$$

$$\underbrace{X_1(t), X_2(t), \ldots X_{p_2}(t))}_{\text{Time-dependent}}$$

Given a survival analysis situation involving both time-independent and time-dependent predictor variables, we can write the extended Cox model that incorporates both types as shown here at the left. As with the Cox PH model, the extended model contains a baseline hazards function $h_0(t)$ which is multiplied by an exponential function. However, in the extended model, the exponential part contains both time-independent predictors, as denoted by the X_i variables, and time-dependent predictors, as denoted by the $X_j(t)$ variables. The entire collection of predictors at time t is denoted by the bold $\mathbf{X}(t)$.

EXAMPLE

$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta(E \times t)]$,
$p_1 = 1, p_2 = 1$,
$\mathbf{X}(t) = (X_1 = E, X_1(t) = E \times t)$

As a simple example of an extended Cox model, we show here a model with one time-independent variable and one time-dependent variable. The time-independent variable is exposure status E, say a (0,1) variable, and the time-dependent variable is the product term $E \times t$.

Estimating regression coefficients:
ML procedure:
Maximize (partial) L.
Risk sets more complicated than for PH model.

As with the simpler Cox PH model, the regression coefficients in the extended Cox model are estimated using a maximum likelihood (ML) procedure. ML estimates are obtained by maximizing a (partial) likelihood function L. However, the computations for the extended Cox model are more complicated than for the Cox PH model, because the risk sets used to form the likelihood function are more complicated with time-dependent variables. The extended Cox likelihood is described later in this chapter.

Computer programs for the extended Cox model:

Stata (Stcox)
SAS (PHREG)
SPSS (COXREG)
} Computer Appendix

Computer packages that include programs for fitting the extended Cox model include Stata, SAS, and SPSS. See the Computer Appendix at the end of this text for a comparison of the Stata, SAS, and SPSS procedures applied to the same dataset.

Statistical inferences:
Wald and/or *LR* tests
Large sample confidence intervals

Methods for making statistical inferences are essentially the same as for the PH model. That is, one can use Wald and/or likelihood ratio (*LR*) tests and large sample confidence interval methods.

Assumption of the model:
The hazard at time t depends on the value of $X_j(t)$ at that same time.

An important assumption of the extended Cox model is that the effect of a time-dependent variable $X_j(t)$ on the survival probability at time t depends on the value of this variable at that *same* time t, and not on the value at an earlier or later time.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i X_i + \sum_{j=1}^{p_2}\delta_j X_j(t)\right]$$

One coefficient for $X_j(t)$

Note that even though the values of the variable $X_j(t)$ may change over time, the hazard model provides only one coefficient for each time-dependent variable in the model. Thus, at time t, there is only one value of the variable $X_j(t)$ that has an effect on the hazard, that value being measured at time t.

Can modify for lag-time effect

It is possible, nevertheless, to modify the definition of the time-dependent variable to allow for a "lag-time" effect.

Lag-time effect:

EXAMPLE

$EMP(t)$ = employment status at week t

Model without lag-time:
$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta EMP(t)]$
Same week

Model with 1-week lag-time:
$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta^* EMP(t-1)]$
One-week earlier

To illustrate the idea of a lag-time effect, suppose, for example, that employment status, measured weekly and denoted as $EMP(t)$, is the time-dependent variable being considered. Then, an extended Cox model that does *not* consider lag-time assumes that the effect of employment status on the probability of survival at week t depends on the observed value of this variable at the same week t, and not, for example, at an earlier week.

However, to allow for, say, a time-lag of one week, the employment status variable may be modified so that the hazard model at time t is predicted by the employment status at week $t-1$. Thus, the variable $EMP(t)$ is replaced in the model by the variable $EMP(t-1)$.

General lag-time extended model:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i X_i + \sum_{j=1}^{p_2}\delta_j X_j(t-L_j)\right]$$

$X_j(t-L_j)$ replaces $X_j(t)$

More generally, the extended Cox model may be alternatively written to allow for a lag-time modification of any time-dependent variable of interest. If we let L_j denote the lag-time specified for time-dependent variable j, then the general "lag-time extended model" can be written as shown here. Note that the variable $X_j(t)$ in the earlier version of the extended model is now replaced by the variable $X_j(t-L_j)$.

V. The Hazard Ratio Formula for the Extended Cox Model

We now describe the formula for the hazard ratio that derives from the extended Cox model. The most important feature of this formula is that the proportional hazards assumption is no longer satisfied when using the extended Cox model.

PH assumption is not satisfied for the extended Cox model.

$$\widehat{HR}(t) = \frac{\hat{h}(t,\mathbf{X}^*(t))}{\hat{h}(t,\mathbf{X}(t))} = \exp\left[\sum_{i=1}^{p_1}\hat{\beta}_i\left[X_i^* - X_i\right] + \sum_{j=1}^{p_2}\delta_j\left[X_j^*(t) - X_j(t)\right]\right]$$

The general hazard ratio formula for the extended Cox model is shown here. This formula describes the ratio of hazards at a particular time t, and requires the specification of two sets of predictors at time t. These two sets are denoted as bold $\mathbf{X}^*(t)$ and bold $\mathbf{X}(t)$.

Two sets of predictors:

$$\mathbf{X}^*(t) = (X_1^*, X_2^*, \ldots, X_{p_1}^*, X_1^*(t), X_2^*(t), \ldots, X_{p_2}^*(t))$$

$$\mathbf{X}(t) = (X_1, X_2, \ldots, X_{p_1}, X_1(t), X_2(t), \ldots, X_{p_2}(t))$$

The two sets of predictors, $\mathbf{X}^*(t)$ and $\mathbf{X}(t)$, identify two specifications at time t for the combined set of predictors containing both time-independent and time-dependent variables. The individual components for each set of predictors are shown here.

EXAMPLE

$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta(E \times t)]$

$$E = \begin{cases} 1 & \text{if exposed} \\ 0 & \text{if unexposed} \end{cases}$$

$\mathbf{X}^*(t) = (E = 1, E \times t = t)$
$\mathbf{X}(t) = (E = 0, E \times t = 0)$

$$\widehat{HR}(t) = \frac{\hat{h}(t, E = 1)}{\hat{h}(t, E = 0)} = \exp[\hat{\beta}(1 - 0) + \hat{\delta}((1 \times t) - (0 \times t))] = \exp[\hat{\beta} + \hat{\delta}t]$$

$\hat{\delta} > 0 \Rightarrow \widehat{HR}(t) \uparrow$ as $t \uparrow$
PH assumption *not* satisfied

As a simple example, suppose the model contains only one time-independent predictor, namely, exposure status E, a (0,1) variable, and one time-dependent predictor, namely, $E \times t$. Then, to compare exposed persons, for whom $E = 1$, with unexposed persons, for whom $E = 0$, at time t, the bold $\mathbf{X}^*(t)$ set of predictors has as its two components $E = 1$ and $E \times t = t$; the bold $\mathbf{X}(t)$ set has as its two components $E = 0$ and $E \times t = 0$.

If we now calculate the estimated hazard ratio that compares exposed to unexposed persons at time t, we obtain the formula shown here; that is, *HR* "hat" equals the exponential of β "hat" plus δ "hat" times t. This formula says that the hazard ratio is a function of time; in particular, if δ "hat" is positive, then the hazard ratio increases with increasing time. Thus, the hazard ratio in this example is certainly not constant, so that the PH assumption is not satisfied for this model.

$$\widehat{HR}(t) = \exp\left[\sum_{i=1}^{p_1} \hat{\beta}_i\left[X_i^* - X_i\right] + \sum_{j=1}^{p_2} \boxed{\hat{\delta}_j}\left(\left[X_j^*(t) - X_j(t)\right]\right)\right]$$

A function of time

More generally, because the general hazard ratio formula involves differences in the values of the time-dependent variables at time t, this hazard ratio is a function of time. Thus, in general, the extended Cox model does not satisfy the PH assumption if any δ_j is not equal to zero.

In general, PH assumption not satisfied for extended Cox model.

$\hat{\delta}_j$ is not time-dependent.

$\hat{\delta}_j$ represents "overall" effect of $X_j(t)$.

Note that, in the hazard ratio formula, the coefficient δ_j "hat" of the difference in values of the jth time-dependent variable is itself not time-dependent. Thus, this coefficient represents the "overall" effect of the corresponding time-dependent variable, considering all times at which this variable has been measured in the study.

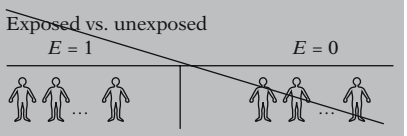
EXAMPLE

$E(t)$ = chemical exposure status at time t (weekly)

$$= \begin{cases} 0 & \text{if unexposed at time } t \\ 1 & \text{if exposed at time } t \end{cases}$$

$E(t)$	0 1 0 1 1
t	1 2 3 4 5…

$E(t)$	1 1 0 1 0
t	1 2 3 4 5…



$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta E(t)]$$

One coefficient

δ represents the overall effect of $E(t)$.

$$\widehat{HR}(t) = \frac{\hat{h}(t,E(t)=1)}{\hat{h}(t,E(t)=0)} = \exp\left[\hat{\delta}[1-0]\right] = e^{\hat{\delta}}, \text{ a fixed number}$$

But, PH is *not* satisfied:
$\widehat{HR}(t)$ is time-dependent because $E(t)$ is time-dependent.

As another example to illustrate the formula for the hazard ratio, consider an extended Cox model containing only one variable, say a weekly measure of chemical exposure status at time t. Suppose this variable, denoted as $E(t)$, can take one of two values, 0 or 1, depending on whether a person is unexposed or exposed, respectively, at a given weekly measurement.

As defined, the variable $E(t)$ can take on different patterns of values for different subjects. For example, for a five-week period, subject A's values may be 01011, whereas subject B's values may be 11010.

Note that in this example, we do not consider two separate groups of subjects, with one group always exposed and the other group always unexposed throughout the study. This latter situation would require a (0,1) time-independent variable for exposure, whereas our example involves a time-dependent exposure variable.

The extended Cox model that includes only the variable $E(t)$ is shown here. In this model, the values of the exposure variable may change over time for different subjects, but there is only one coefficient, δ, corresponding to the one variable in the model. Thus, δ represents the overall effect on survival time of the time-dependent variable $E(t)$.

Notice, also, that the hazard ratio formula, which compares an exposed person to an unexposed person at time t, yields the expression e to the δ "hat."

Although this result is a fixed number, the PH assumption is not satisfied. The fixed number gives the hazard ratio at a given time, assuming that the exposure status at that time is 1 in the numerator and is 0 denominator. Thus, the hazard ratio is time-dependent, because exposure status is time-dependent, even though the formula yields a single fixed number.

VI. Assessing Time-Independent Variables That Do Not Satisfy the PH Assumption

Use an extended Cox model to

- check PH assumption;
- assess effect of variable not satisfying PH assumption.

We now discuss how to use an extended Cox model to check the PH assumption for time-independent variables and to assess the effect of a variable that does not satisfy the PH assumption.

Three methods for checking PH assumption:

1. graphical
2. extended Cox model
3. GOF test

As described previously (see Chapter 4), there are three methods commonly used to assess the PH assumption: (1) graphical, using, say, log–log survival curves; (2) using an extended Cox model; and (3) using a goodness-of-fit (GOF) test. We have previously (in Chapter 4) discussed items 1 and 3, but only briefly described item 2, which we focus on here.

Cox PH model for p time-independent X's:

$$h(t,\mathbf{X}) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i\right]$$

If the dataset for our study contains several, say p, time-independent variables, we might wish to fit a Cox PH model containing each of these variables, as shown here.

Extended Cox model:
Add product terms of the form:

$X_i \times g_i(t)$

However, to assess whether such a PH model is appropriate, we can extend this model by defining several product terms involving each time-independent variable with some function of time. That is, if the ith time-independent variable is denoted as X_i, then we can define the ith product term as $X_i \times g_i(t)$ where $g_i(t)$ is some function of time for the ith variable.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \sum_{i=1}^{p}\delta_i X_i g_i(t)\right]$$

The extended Cox model that simultaneously considers all time-independent variables of interest is shown here.

EXAMPLE

$g_i(t) = 0$ for all i implies no time-dependent variable involving X_i, i.e.,

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i\right]$$

In using this extended model, the crucial decision is the form that the functions $g_i(t)$ should take. The simplest form for $g_i(t)$ is that all $g_i(t)$ are identically 0 at any time; this is another way of stating the original PH model, containing no time-dependent terms.

EXAMPLE 2

$$g_i(t) = t \Rightarrow X_i g_i(t) = X_i \times t$$

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \sum_{i=1}^{p}\delta_i (X_i \times t)\right]$$

Another choice for the $g_i(t)$ is to let $g_i(t) = t$. This implies that for each X_i in the model as a main effect, there is a corresponding time-dependent variable in the model of the form $X_i \times t$. The extended Cox model in this case takes the form shown here.

EXAMPLE 3: one variable at a time

$$X_L \text{ only} \Rightarrow \begin{cases} g_L(t) = t, \\ g_i(t) = 0 \text{ for other } i \end{cases}$$

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \delta_L (X_L \times t)\right]$$

Suppose, however, we wish to focus on a particular time-independent variable, say, variable X_L. Then $g_i(t) = t$ for $i = L$, but equals 0 for all other i. The corresponding extended Cox model would then contain only one product term $X_L \times t$, as shown here.

EXAMPLE 4

$$g_i(t) = \ln t \Rightarrow X_i g_i(t) = X_i \times \ln t$$

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \sum_{i=1}^{p}\delta_i (X_i \times \ln t)\right]$$

Another choice for the $g_i(t)$ is the log of t, rather than simply t, so that the corresponding time-dependent variables will be of the form $X_i \times \ln t$.

EXAMPLE 5: Heaviside Function

$$g_i(t) = \begin{cases} 0 & \text{if } t \geq t_0 \\ 1 & \text{if } t < t_0 \end{cases}$$

And yet another choice would be to let $g_i(t)$ be a "heaviside function" of the form $g_i(t) = 1$ when t is at or above some specified time, say t_0, and $g_i(t) = 0$ when t is below t_0. We will discuss this choice in more detail shortly.

Extended Cox model:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \sum_{i=1}^{p}\delta_i X_i g_i(t)\right]$$

Given a particular choice of the $g_i(t)$, the corresponding extended Cox model, shown here again in general form, may then be used to check the PH assumption for the time-independent variables in the model. Also, we can use this extended Cox model to obtain a hazard ratio formula that considers the effects of variables not satisfying the PH assumption.

- Check PH assumption.
- Obtain hazard ratio when PH assumption not satisfied.

To check the PH assumption using a statistical test, we consider the null hypothesis that all the δ terms, which are coefficients of the $X_i g_i(t)$ product terms in the model, are zero.

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_p = 0$$

Under H_0, the model reduces to PH model:

$$h(t,\mathbf{X}) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i\right]$$

Under this null hypothesis, the model reduces to the PH model.

$$\begin{aligned} LR &= -2\ln L_{\text{PH model}} \\ &\quad -(-2\ln L_{\text{ext. Cox model}}) \\ &\dot{\sim}\chi^2_p \text{ under } H_0 \end{aligned}$$

This test can be carried out using a likelihood ratio (LR) test which computes the difference between the log likelihood statistic, $-2\ln L$, for the PH model and the log likelihood statistic for the extended Cox model. The test statistic thus obtained has approximately a chi-square distribution with p degrees of freedom under the null hypothesis, where p denotes the number of parameters being set equal to zero under H_0.

EXAMPLE

$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta(E \times t)]$
H_0: $\delta = 0$ (i.e., PH assumption is satisfied)

Reduced model:
$h(t,\mathbf{X}) = h_0(t)\exp[\beta E]$

$LR = -2\ln L_R - (-2\ln L_F)$
$\dot{\sim}\chi^2$ with 1 df under H_0

F = full (extended), R = reduced (PH)

As an example of this test, suppose we again consider an extended Cox model that contains the product term $E \times t$ in addition to the main effect of E, where E denotes a (0,1) time-independent exposure variable.

For this model, a test for whether or not the PH assumption is satisfied is equivalent to testing the null hypothesis that $\delta = 0$. Under this hypothesis, the reduced model is given by the PH model containing the main effect E only. The likelihood ratio statistic, shown here as the difference between log-likelihood statistics for the full (i.e., extended model) and the reduced (i.e., PH) model, will have an approximate chi-square distribution with one degree of freedom in large samples.

SAS: **PHREG** fits both PH and extended Cox models.
Stata: **Stcox** fits both PH and extended Cox models.

Note that to carry out the computations for this test, two different types of models, a PH model and an extended Cox model, need to be fit.

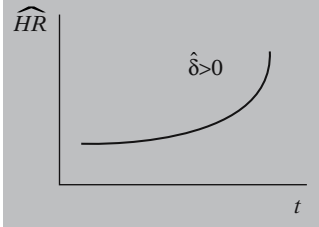
If PH test significant: Extended Cox model is preferred; HR is time-dependent.

If the result of the test for the PH assumption is significant, then the extended Cox model is preferred to the PH model. Thus, the hazard ratio expression obtained for the effect of an exposure variable of interest is time-dependent. That is, the effect of the exposure on the outcome cannot be summarized by a single HR value, but can only be expressed as a function of time.

EXAMPLE

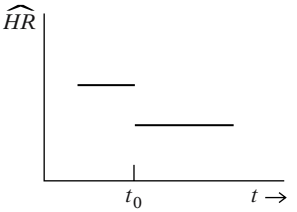
$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta(E \times t)]$

$\widehat{HR} = \exp[\hat{\beta} + \hat{\delta}t]$



We again consider the previous example, with the extended Cox model shown here. For this model, the estimated hazard ratio for the effect of exposure is given by the expression e to the quantity β "hat" plus δ "hat" times t. Thus, depending on whether δ "hat" is positive or negative, the estimated hazard ratio will increase or decrease exponentially as t increases. The graph shown here gives a sketch of how the hazard ratio varies with time if δ "hat" is positive.

Heaviside function:



We now provide a description of the use of a "heaviside" function. When such a function is used, the hazard ratio formula yields constant hazard ratios for different time intervals, as illustrated in the accompanying graph.

$$g(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0 \end{cases}$$

Recall that a heaviside function is of the form $g(t)$, which takes on the value 1 if t is greater than or equal to some specified value of t, called t_0, and takes on the value 0 if t is less than t_0. An extended Cox model which contains a single heaviside function is shown here.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta E g(t)]$$

$t \geq t_0$: $g(t) = 1 \Rightarrow E \times g(t) = E$

$$h(t,\mathbf{X}) = h_0(t)\exp[(\beta+\delta)E]$$

$$\widehat{HR} = \exp[\hat{\beta} + \hat{\delta}]$$

Note that if $t \geq t_0$, $g(t) = 1$, so the value of $E \times g(t) = E$; the corresponding hazard function is of the form $h_0(t) \times e$ to the quantity $(\beta + \delta)$ times E, and the estimated hazard ratio for the effect of E has the form e to the sum of β "hat" plus δ "hat."

$t < t_0$: $g(t) = 0 \Rightarrow E \times g(t) = 0$

$$h(t,\mathbf{X}) = h_0(t)\exp[\beta E]$$

$$\widehat{HR} = \exp[\hat{\beta}]$$

If $t < t_0$, $g(t) = 0$, the corresponding hazard ratio is simplified to e to the β "hat."

A single heaviside function in the model

$$h(t,\mathbf{X}) = h_0(t)\exp[\beta E + \delta(E \times g(t))]$$

yields two hazard ratios:

$t \geq t_0$: $\widehat{HR} = \exp(\hat{\beta} + \hat{\delta})$

$t < t_0$: $\widehat{HR} = \exp(\hat{\beta})$

Thus, we have shown that the use of a single heaviside function results in an extended Cox model which gives two hazard ratio values, each value being constant over a fixed time interval.

Alternative model with two heaviside functions:

$$h(t,\mathbf{X}) = h_0(t)\exp[\delta_1(E \times g_1(t)) + \delta_2(E \times g_2(t))]$$

$$g_1(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0 \end{cases}$$

$$g_2(t) = \begin{cases} 1 & \text{if } t < t_0 \\ 0 & \text{if } t \geq t_0 \end{cases}$$

Note: Main effect for E not in model.

There is actually an equivalent way to write this model that uses two heaviside functions in the same model. This alternative model is shown here. The two heaviside functions are called $g_1(t)$ and $g_2(t)$. Each of these functions are in the model as part of a product term with the exposure variable E. Note that this model does not contain a main effect term for exposure.

Two *HR*'s from the alternative model:

$t \geq t_0: g_1(t) = 1, g_2(t) = 0$

$$h(t,\mathbf{X}) = h_0(t)\exp[\delta_1(E \times 1) + \delta_2(E \times 0)] = h_0(t)\exp[\delta_1 E]$$

so that $\widehat{HR} = \exp(\hat{\delta}_1)$

For this alternative model, as for the earlier model with only one heaviside function, two different hazard ratios are obtained for different time intervals. To obtain the first hazard ratio, we consider the form that the model takes when $t \geq t_0$. In this case, the value of $g_1(t)$ is 1 and the value of $g_2(t)$ is 0, so the exponential part of the model simplifies to $\delta_1 \times E$; the corresponding formula for the estimated hazard ratio is then e to the δ_1 "hat."

$t < t_0: g_1(t) = 0, g_2(t) = 1$

$$h(t,\mathbf{X}) = h_0(t)\exp[\delta_1(E \times 0) + \delta_2(E \times 1)] = h_0(t)\exp[\delta_2 E]$$

so that $\widehat{HR} = \exp(\hat{\delta}_2)$

When $t < t_0$, the value of $g_1(t)$ is 0 and the value of $g_2(t)$ is 1. Then, the exponential part of the model becomes $\delta_2 \times E$, and the corresponding hazard ratio formula is e to the δ_2 "hat."

Alternative model:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta_1(E \times g_1(t)) + \delta_2(E \times g_2(t))]$$

Original model:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta(E \times g(t))]$$

$t \geq t_0: \widehat{HR} = \exp(\hat{\delta}_1) = \exp(\hat{\beta} + \hat{\delta})$

$t < t_0: \widehat{HR} = \exp(\hat{\delta}_2) = \exp(\hat{\beta})$

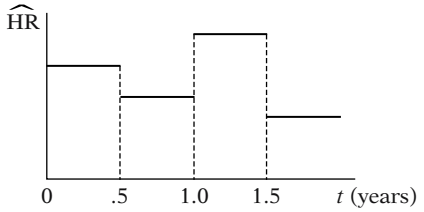
Thus, using the alternative model, again shown here, we obtain two distinct hazard ratio values. Mathematically, these are the same values as obtained from the original model containing only one heaviside function. In other words, δ_1 "hat" in the alternative model equals β "hat" plus δ "hat" in the original model (containing one heaviside function), and δ_2 "hat" in the alternative model equals β "hat" in the original model.

Heaviside functions:

- two $\widehat{HR}$'s constant within two time intervals
- *extension*: several $\widehat{\text{HR}}$'s constant within several time intervals

We have thus seen that heaviside functions can be used to provide estimated hazard ratios that remain constant within each of two separate time intervals of follow-up. We can also extend the use of heaviside functions to provide several distinct hazard ratios that remain constant within several time intervals.

Four time intervals:

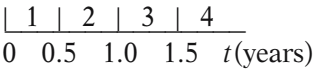


Suppose, for instance, that we wish to separate the data into *four* separate time intervals, and for each interval we wish to obtain a different hazard ratio estimate as illustrated in the graph shown here.

Extended Cox model contains either

- $E, E \times g_1(t), E \times g_2(t), E \times g_3(t)$
  or
- $E \times g_1(t), E \times g_2(t), E \times g_3(t), E \times g_4(t)$

We can obtain four different hazard ratios using an extended Cox model containing *a main effect of exposure and three heaviside functions* in the model as products with exposure. Or, we can use a model containing *no main effect* exposure term, but with product terms involving exposure with *four heaviside functions*.



To illustrate the latter model, suppose, as shown on the graph, that the first time interval goes from time 0 to 0.5 of a year; the second time interval goes from 0.5 to 1 year; the third time interval goes from 1 year to a year and a half; and the fourth time interval goes from a year and a half onward.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta_1 E g_1(t) + \delta_2 E g_2(t) + \delta_3 E g_3(t) + \delta_4 E g_4(t)]$$

where

$$g_1(t) = \begin{cases} 1 & \text{if } 0 \le t < 0.5 \text{ year} \\ 0 & \text{if otherwise} \end{cases}$$

$$g_2(t) = \begin{cases} 1 & \text{if } 0.5 \text{ year} \le t < 1.0 \text{ year} \\ 0 & \text{if otherwise} \end{cases}$$

$$g_3(t) = \begin{cases} 1 & \text{if } 1.0 \text{ year} \le t < 1.5 \text{ years} \\ 0 & \text{if otherwise} \end{cases}$$

$$g_4(t) = \begin{cases} 1 & \text{if } t \ge 1.5 \text{ years} \\ 0 & \text{if otherwise} \end{cases}$$

Then, an appropriate extended Cox model containing the four heaviside functions $g_1(t), g_2(t), g_3(t)$, and $g_4(t)$ is shown here. This model assumes that there are four different hazard ratios identified by three cutpoints at half a year, one year, and one and a half years. The formulae for the four hazard ratios are given by separately exponentiating each of the four estimated coefficients, as shown below:

$$4\,\widehat{\text{HR}}\text{'s} \begin{cases} 0 \le t < 0.5: \widehat{HR} = \exp(\hat{\delta}_1) \\ 0.5 \le t < 1.0: \widehat{HR} = \exp(\hat{\delta}_2) \\ 1.0 \le t < 1.5: \widehat{HR} = \exp(\hat{\delta}_3) \\ t \ge 1.5: \widehat{HR} = \exp(\hat{\delta}_4) \end{cases}$$

VII. An Application of the Extended Cox Model to An Epidemiologic Study on the Treatment of Heroin Addiction

EXAMPLE

1991 Australian study (Caplehorn et al.) of heroin addicts

- two methadone treatment clinics
- T = days remaining in treatment (= days until drop out of clinic)
- clinics differ in treatment policies

Dataset name: ADDICTS
Column 1: Subject ID
Column 2: Clinic (1 or 2) ← E
Column 3: Survival status (0 = censored, 1 = departed clinic)
Column 4: Survival time in days
Column 5: Prison Record (0 = none, 1 = any) ← covariates
Column 6: Maximum Methadone Dose (mg/day)

$$h(t,\mathbf{X}) = h_0(t)\exp[\beta_1(\text{clinic}) + \beta_2(\text{prison}) + \beta_3(\text{dose})]$$

	Coef.	Std. Err.	p>|z|	Haz. Ratio	P(PH)
Clinic	–1.009	0.215	0.000	0.365	0.001
Prison	0.327	0.167	0.051	1.386	0.332
Dose	–0.035	0.006	0.000	0.965	0.347

$P(PH)$ for the variables prison and dose are nonsignificant ⇒ remain in model

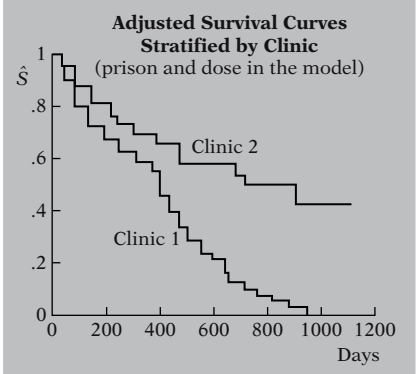
A 1991 Australian study by Caplehorn et al., compared retention in two methadone treatment clinics for heroin addicts. A patient's survival time (T) was determined as the time in days until the patient dropped out of the clinic or was censored at the end of the study clinic. The two clinics differed according to their overall treatment policies.

A listing of some of the variables in the dataset for this study is shown here. The dataset name is called "ADDICTS," and survival analysis programs in the Stata package are used in the analysis. Note that the survival time variable is listed in column 4 and the survival status variable, which indicates whether a patient departed from the clinic or was censored, is listed in column 3. The primary exposure variable of interest is the clinic variable, which is coded as 1 or 2. Two other variables of interest are prison record status, listed in column 5 and coded as 0 if none and 1 if any, and maximum methadone dose, in milligrams per day, which is listed in column 6. These latter two variables are considered as covariates.

One of the first models considered in the analysis of the addicts dataset was a Cox PH model containing the three variables, clinic, prison record, and dose. An edited printout of the results for this model is shown here. What stands out from this printout is that the $P(PH)$ value for the clinic variable is zero to three significant places, which indicates that the clinic variable does not satisfy the proportional hazard assumption.

Since the $P(PH)$ values for the other two variables in the model are highly nonsignificant, this suggests that these two variables, namely, prison and dose, can remain in the model.

EXAMPLE (continued)



Results:

- Curve for clinic 2 consistently lies above curve for clinic 1.
- Curves diverge, with clinic 2 being vastly superior after one year.

Stratifying by **clinic:** cannot obtain hazard ratio for **clinic**

Hazard ratio for **clinic** requires **clinic** in the model.

Extended Cox model:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{clinic}) + \beta_2(\text{prison}) + \beta_3(\text{dose}) + \delta(\text{clinic})g(t)]$$

where

$$g(t) = \begin{cases} 1 & \text{if } t \geq 365 \text{ days} \\ 0 & \text{if } t < 365 \text{ days} \end{cases}$$

and

$$\text{clinic} = \begin{cases} 1 & \text{if clinic 1} \\ 0 & \text{if clinic 2} \end{cases}$$

Note: Previously clinic = 2 for clinic 2

$t \geq 365$ days: $HR = \exp(\hat{\beta}_1 + \hat{\delta})$

$t < 365$ days: $HR = \exp(\hat{\beta}_1)$

Further evidence of the PH assumption not being satisfied for the clinic variable can be seen from a graph of adjusted survival curves stratified by clinic, where the prison and dose variables have been kept in the model. Notice that the two curves are much closer together at earlier times, roughly less than one year (i.e., 365 days), but the two curves diverge greatly after one year. This indicates that the hazard ratio for the clinic variable will be much closer to one at early times but quite different from one later on.

The above graph, nevertheless, provides important results regarding the comparison of the two clinics. The curve for clinic 2 consistently lies above the curve for clinic 1, indicating that clinic 2 does better than clinic 1 in retaining its patients in methadone treatment. Further, because the two curves diverge after about a year, it appears that clinic 2 is vastly superior to clinic 1 after one year but only slightly better than clinic 1 prior to one year.

Unfortunately, because the clinic variable has been stratified in the analysis, we cannot use this analysis to obtain a hazard ratio expression for the effect of clinic, adjusted for the effects of prison and dose. We can only obtain such an expression for the hazard ratio if the clinic variable is in the model.

Nevertheless, we can obtain a hazard ratio using an alternative analysis with an extended Cox model that contains a heaviside function, $g(t)$, together with the clinic variable, as shown here. Based on the graphical results shown earlier, a logical choice for the cutpoint of the heaviside function is one year (i.e., 365 days). The corresponding model then provides two hazard ratios: one that is constant above 365 days and the other that is constant below 365 days.

Note that in the extended Cox model here, we have coded the clinic variable as 1 if clinic 1 and 0 if clinic 2, whereas previously we had coded clinic 2 as 2. The reason for this change in coding, as illustrated by computer output below, is to obtain hazard ratio estimates that are greater than unity.

EXAMPLE (continued)

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_2(\text{prison}) + \beta_3(\text{dose}) + \delta_1(\text{clinic})g_1(t) + \delta_2(\text{clinic})g_2(t)]$$

where

$$g_1(t) = \begin{cases} 1 & \text{if } t < 365 \text{ days} \\ 0 & \text{if } t \geq 365 \text{ days} \end{cases}$$

and

$$g_2(t) = \begin{cases} 1 & \text{if } t \geq 365 \text{ days} \\ 0 & \text{if } t < 365 \text{ days} \end{cases}$$

$t < 365$ days: $\widehat{HR} = \exp(\hat{\delta}_1)$

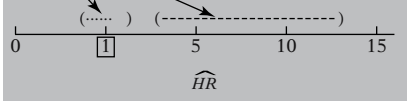
$t \geq 365$ days: $\widehat{HR} = \exp(\hat{\delta}_2)$

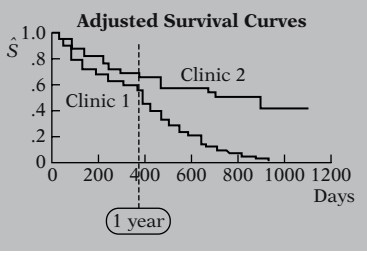
	Coef.	Std. Err.	p>|z|	Haz. Ratio	[95% Conf.	Interval]
Prison	0.378	0.168	0.025	1.459	1.049	2.029
Dose	–0.036	0.006	0.000	0.965	0.953	0.977
Clinic × g_1	0.460	0.255	0.072	1.583	0.960	2.611
Clinic × g_2	1.828	0.386	0.000	6.223	2.921	13.259

$t < 365$ days: $\widehat{HR} = e^{0.460} = 1.583$

$t \geq 365$ days: $\widehat{HR} = e^{1.828} = 6.223$

95% confidence intervals for clinic effect:
$t < 365$ days: (0.960, 2.611)
$t \geq 365$ days: (2.921, 13.259)





An equivalent way to write the model is to use two heaviside functions, $g_1(t)$ and $g_2(t)$, as shown here. This latter model contains product terms involving clinic with each heaviside function, and there is no main effect of clinic.

Corresponding to the above model, the effect of clinic is described by two hazard ratios, one for time less than 365 days and the other for greater than 365 days. These hazard ratios are obtained by separately exponentiating the coefficients of each product term, yielding e to the δ_1 "hat" and e to the δ_2 "hat," respectively.

A printout of results using the above model with two heaviside functions is provided here. The results show a borderline nonsignificant hazard ratio ($P = 0.072$) of 1.6 for the effect of clinic when time is less than 365 days in contrast to a highly significant (P = 0.000 to three decimal places) hazard ratio of 6.2 when time exceeds 365 days.

Note that the estimated hazard ratio of 1.583 from the printout is computed by exponentiating the estimated coefficient 0.460 of the product term "clinic $\times$ g_1" and that the estimated hazard ratio of 6.223 is computed by exponentiating the estimated coefficient 1.828 of the product term "clinic $\times$ g_2".

Note also that the 95% confidence interval for the clinic effect prior to 365 days—that is, for the product term "clinic $\times$ $g_1(t)$"—is given by the limits 0.960 and 2.611, whereas the corresponding confidence interval after 365 days—that is, for the product term "clinic $\times$ g_2"—is given by the limits 2.921 and 13.259. The latter interval is quite wide, showing a lack of precision when t exceeds 365 days; however, when t precedes 365 days, the interval includes the null hazard ratio of 1, suggesting a chance effect for this time period.

The results we have just shown support the observations obtained from the graph of adjusted survival curves. That is, these results suggest a large difference in clinic survival times after one year in contrast to a small difference in clinic survival times prior to one year, with clinic 2 always doing better than clinic 1 at any time.

EXAMPLE (continued)

One other analysis:
Use an extended Cox model that provides for diverging survival curves

There is, nevertheless, at least one other approach to the analysis using time-dependent variables that we now describe. This approach considers our earlier graphical observation that the survival curves for each clinic continue to diverge from one another even after one year. In other words, it is reasonable to consider an extended Cox model that allows for such a divergence, rather than a model that assumes the hazard ratios are constant before and after one year.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{clinic}) + \beta_2(\text{prison}) + \beta_3(\text{dose}) + \delta(\text{clinic} \times t)]$$

One way to define an extended Cox model that provides for diverging survival curves is shown here. This model includes, in addition to the clinic variable by itself, a time-dependent variable defined as the product of the clinic variable with time (i.e. clinic $\times$ t). By including this product term, we are able to estimate the effect of clinic on survival time, and thus the hazard ratio, for any specified time t.

$$\widehat{\text{HR}} = \exp(\hat{\beta}_1 + \hat{\delta}t)$$

$\widehat{\text{HR}}$ changes over time.

To demonstrate how the hazard ratio changes over time for this model, we consider what the model and corresponding estimated hazard ratio expression are for different specified values of t.

$t = 91$ days

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{clinic}) + \beta_2(\text{prison}) + \beta_3(\text{dose}) + \delta(\text{clinic})(91)]$$

So

$$\widehat{\text{HR}} = \exp(\hat{\beta}_1 + 91\hat{\delta})$$

For example, if we are interested in the effect of clinic on survival on day 91, so that $t = 91$, the exponential part of the model simplifies to terms for the prison and dose variables plus β_1 times the clinic variable plus δ times the clinic variable times 91: the corresponding estimated hazard ratio for the clinic effect is then e to the power β_1 "hat" plus δ "hat" times t = 91.

$t = 274$:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{clinic}) + \beta_2(\text{prison}) + \beta_3(\text{dose}) + \delta(\text{clinic})(274)]$$

$$\widehat{\text{HR}} = \exp(\hat{\beta}_1 + 274\hat{\delta})$$

At 274 days, the exponential part of the model contains the prison, dose, and clinic main effect terms as before, plus δ times the clinic variable times 274: the corresponding hazard ratio for the clinic effect is then e to β_1 "hat" plus 274 δ "hat".

$t = 458.5$:

$$\widehat{\text{HR}} = \exp(\hat{\beta}_1 + 458.5\hat{\delta})$$

$t = 639$:

$$\widehat{\text{HR}} = \exp(\hat{\beta}_1 + 639\hat{\delta})$$

$t = 821.5$:

$$\widehat{\text{HR}} = \exp(\hat{\beta}_1 + 821.5\hat{\delta})$$

The formulae for the estimated hazard ratio for other specified days are shown here. Notice that the estimated hazard ratio appears to be increase over the length of the follow-up period. Thus, if δ "hat" is a positive number, then the estimated hazard ratios will increase over time.

$$\hat{\delta} > 0 \Rightarrow \widehat{HR} \uparrow \text{as time} \uparrow$$

EXAMPLE (continued)

Computer results for extended Cox model involving $T(t)$:

	Coef.	Std. Err.	P>|z|	Haz. Ratio	[95% Conf.	Interval]
prison	0.390	0.169	0.021	1.476	1.060	2.056
dose	–0.035	0.006	0.000	0.965	0.953	0.978
clinic	–0.0183	0.347	0.958	0.982	0.497	1.939
clinic × t	0.003	0.001	0.001	1.003	1.001	1.005

$\widehat{\text{cov}}(\hat{\beta}_1, \hat{\delta}) = -.000259$ Log likelihood = –667.642

$\hat{\beta}_1 = -0.0183 \qquad \hat{\delta} = 0.003$

$\widehat{\text{HR}}$ depends on $\hat{\beta}_1$ and $\hat{\delta}$.

$t = 91.5$: $\widehat{\text{HR}} = \exp(\hat{\beta}_1 + \hat{\delta}t) = 1.292$

$t = 274$: $\widehat{\text{HR}} = \exp(\hat{\beta}_1 + \hat{\delta}t) = 2.233$

$t = 458.5$: $\widehat{\text{HR}} = \exp(\hat{\beta}_1 + \hat{\delta}t) = 3.862$

$t = 639$: $\widehat{\text{HR}} = \exp(\hat{\beta}_1 + \hat{\delta}t) = 6.677$

$t = 821.5$: $\widehat{\text{HR}} = \exp(\hat{\beta}_1 + \hat{\delta}t) = 11.544$

$$\exp\left[\hat{\beta}_1 + \hat{\delta}t \pm 1.96\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\delta}t)}\right]$$

$$\text{Var}(\hat{\beta}_1 + \hat{\delta}t) = s^2_{\hat{\beta}_1} + t^2 s^2_{\hat{\delta}} + 2t\,\widehat{\text{cov}}(\hat{\beta}_1, \hat{\delta})$$

$$\uparrow \qquad \uparrow \qquad \uparrow$$

$$(0.347)^2 \quad (0.001)^2 \quad (-.000259)$$

Time (days)	$\widehat{\text{HR}}$	95% CI
91.5	1.292	(0.741, 2.250)
274	2.233	(1.470, 3.391)
458.5	3.862	(2.298, 6.491)
639	6.677	(3.102, 14.372)
821.5	11.544	(3.976, 33.513)

We now show edited results obtained from fitting the extended Cox model we have just been describing, which contains the product of clinic with time. The covariance estimate shown at the bottom of the table will be used below to compute confidence intervals.

From these results, the estimated coefficient of the clinic variable is β_1 "hat" equals -0.0183, and the estimated coefficient δ "hat" obtained for the product term equals 0.003. For the model being fit, the hazard ratio depends on the values of both β_1 "hat" and δ "hat."

On the left, the effect of the variable clinic is described by five increasing hazard ratio estimates corresponding to each of five different values of t. These values, which range between 1.292 at 91.5 days to 11.544 at 821.5 days, indicate how the effect of clinic diverges over time for the fitted model.

We can also obtain 95% confidence intervals for each of these hazard ratios using the large sample formula shown here. The variance expression in the formula is computed using the variances and covariances which can be obtained from the computer results given above. In particular, the variances are $(0.347)^2$ and $(0.001)^2$ for β_1 "hat" and δ "hat," respectively; the covariance value is -0.000259.

A table showing the estimated hazard ratios and their corresponding 95% confidence intervals for the clinic effect is given here. Note that all confidence intervals are quite wide.

VIII. An Application of the Extended Cox Model to the Analysis of the Stanford Heart Transplant Data

EXAMPLE

Patients identified as eligible for heart transplant:
T = time until death or censorship
65 patients receive transplants
38 patients do not receive transplants
n = 103 patients

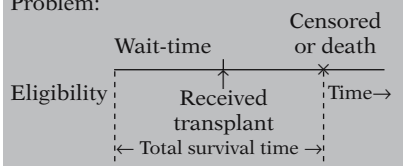
Goal: Do patients receiving transplants survive longer than patients not receiving transplants?

We now consider another application of the extended Cox model which involves the use of an internally defined time-dependent variable. In a 1977 report (Crowley and Hu, *J. Amer. Statist. Assoc*.) on the Stanford Heart Transplant Study, patients identified as being eligible for a heart transplant were followed until death or censorship. Sixty-five of these patients received transplants at some point during follow-up, whereas thirty-eight patients did not receive a transplant. There were, thus, a total of $n = 103$ patients. The goal of the study was to assess whether patients receiving transplants survived longer than patients not receiving transplants.

One approach:
Compare two separate groups: 65 transplants vs. 38 nontransplants

One approach to the analysis of this data was to separate the dataset into two separate groups, namely, the 65 heart transplant patients and the 38 patients not receiving transplants, and then to compare survival times for these groups.

Problem:



A problem with this approach, however, is that those patients who received transplants had to wait from the time they were identified as eligible for a transplant until a suitable transplant donor was found. During this "wait-time" period, they were at risk for dying, yet they did not have the transplant. Thus, the wait-time accrued by transplant patients contributes information about the survival of nontransplant patients. Yet, this wait-time information would be ignored if the *total* survival time for each patient were used in the analysis.

Note: Wait-time contributes to survival time for nontransplants.

Covariates:
Tissue mismatch score } prognostic only
Age at transplant } for transplants

Age at eligibility: not considered prognostic for nontransplants

Another problem with this approach is that two covariates of interest, namely, *tissue mismatch score* and *age at transplant*, were considered as prognostic indicators of survival only for patients who received transplants. Note that *age at eligibility* was not considered an important prognostic factor for the nontransplant group.

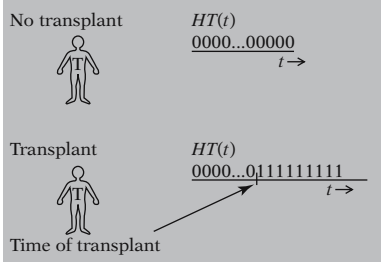
EXAMPLE (continued)

Problems:
- wait-time of transplant recipients
- prognostic factors for transplants only

Alternative approach:
Uses an extended Cox model

Exposure variable:
Heart transplant status at time t, defined as

$$HT(t) = \begin{cases} 0 & \text{if did not receive transplant by time } t, \text{ i.e., if } t < \text{wait-time} \\ 1 & \text{if received transplant prior to time } t, \text{ i.e., if } t \geq \text{wait-time} \end{cases}$$



Wait-time for transplants contributes to survival for nontransplants.

In addition to $HT(t)$, two time-dependent covariates included in model.

Because of the problems just described, which concern the wait-time of transplants and the effects of prognostic factors attributable to transplants only, an alternative approach to the analysis is recommended. This alternative involves the use of time-dependent variables in an extended Cox model.

The exposure variable of interest in this extended Cox model is heart transplant status at time t, denoted by $HT(t)$. This variable is defined to take on the value 0 at time t if the patient has not received a transplant at this time, that is, if t is less than the wait-time for receiving a transplant. The value of this variable is 1 at time t if the patient has received a transplant prior to or at time t, that is, if t is equal to or greater than the wait-time.

Thus, for a patient who did not receive a transplant during the study, the value of $HT(t)$ is 0 at all times. For a patient receiving a transplant, the value of $HT(t)$ is 0 at the start of eligibility and continues to be 0 until the time at which the patient receives the transplant; then, the value of $HT(t)$ changes to 1 and remains 1 throughout the remainder of follow-up.

Note that the variable $HT(t)$ has the property that the wait-time for transplant patients contributes to the survival experience of nontransplant patients. In other words, this variable treats a transplant patient as a nontransplant patient prior to receiving the transplant.

In addition to the exposure variable $HT(t)$, two other time-dependent variables are included in our extended Cox model for the transplant data. These variables are covariates to be adjusted for in the assessment of the effect of the $HT(t)$ variable.

EXAMPLE (continued)

Covariates:

$$TMS(t) = \begin{cases} 0 & \text{if } t < \text{wait-time} \\ TMS & \text{if } t \geq \text{wait-time} \end{cases}$$

$$AGE(t) = \begin{cases} 0 & \text{if } t < \text{wait-time} \\ AGE & \text{if } t \geq \text{wait-time} \end{cases}$$

These covariates are denoted as $TMS(t)$ and $AGE(t)$ and they are defined as follows: $TMS(t)$ equals 0 if t is less than the wait-time for a transplant but changes to the "tissue mismatch score" (TMS) at the time of the transplant if t is equal to or greater than the wait-time. Similarly, $AGE(t)$ equals 0 if t is less than the wait-time but changes to AGE at time of transplant if t is equal to or greater than the wait-time.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta_1 HT(t) + \delta_2 TMS(t) + \delta_3 AGE(t)]$$

The extended Cox model for the transplant data is shown here. The model contains the three time-dependent variables $HT(t)$, $TMS(t)$ and $AGE(t)$ as described above.

Focus:
Assessing the effect of $HT(t)$ adjusted for $TMS(t)$ and $AGE(t)$.

Note: $HT(t)$ does not satisfy PH assumption.

For this model, since $HT(t)$ is the exposure variable of interest, the focus of the analysis concerns assessing the effect of this variable adjusted for the two covariates. Note, however, that because the $HT(t)$ variable is time-dependent by definition, this variable does not satisfy the PH assumption, so that any hazard ratio estimate obtained for this variable is technically time-dependent.

Variable	Coef.	Std. Err.	P>|z|	Haz. Ratio
$HT(t)$	−3.1718	1.1861	(0.008)	(0.0417)
$TMS(t)$	0.4442	0.2802	0.112	1.5593
$AGE(t)$	0.0552	0.0226	0.014	1.0567

A summary of computer results for the fit of the above extended Cox model is shown here. These results indicate that the exposure variable $HT(t)$ is significant below the one percent significance level (i.e., the two-sided p-value is 0.008). Thus, transplant status appears to be significantly associated with survival.

$$\widehat{\text{HR}} = e^{-3.1718} = 0.0417 = \frac{1}{23.98}$$

$$\widehat{\text{HR}} = \frac{\hat{h}(\text{transplants})}{\hat{h}(\text{nontransplants})} \approx \frac{1}{24}?$$

Not appropriate!

To evaluate the strength of the association, note that e to the coefficient of $HT(t)$ equals 0.0417. Since 1 over 0.0417 is 23.98, it appears that there is a 24-fold increase in the hazard of nontransplant patients to transplant patients. The preceding interpretation of the value 0.0417 as a hazard ratio estimate is not appropriate, however, as we shall now discuss further.

EXAMPLE (continued)

23.98 is inappropriate as a $\widehat{HR}$:

- does not compare two *separate* groups
- exposure variable is *not* time-independent
- wait-time on transplants contributes to survival on nontransplants

Alternative interpretation:
At time t,
$\hat{h}$("not yet received transplant")
$\approx 24\ \hat{h}$("already received transplant")

More appropriate:

Hazard ratio formula should account for *TMS* and *AGE*.

Transplant?	$HT(t)$	$TMS(t)$	$AGE(t)$
Yes	1	*TMS*	*AGE*
No	0	0	0

i denotes ith transplant patient

$\mathbf{X}^*(t) = (HT(t) = 1, TMS(t) = TMS_i, AGE(t) = AGE_i)$
$\mathbf{X}(t) = (HT(t) = 0, TMS(t) = 0, AGE(t) = 0)$

$$\widehat{HR}(t) = \exp[\hat{\delta}_1(1-0) + \hat{\delta}_2(TMS_i - 0) + \hat{\delta}_3(AGE_i - 0)]$$
$$= \exp[\hat{\delta}_1 + \hat{\delta}_2 TMS_i + \hat{\delta}_3 AGE_i]$$
$$= \exp[-3.1718 + 0.4442\ TMS_i + 0.0552\ AGE_i]$$

First, note that the value of 23.98 inappropriately suggests that the hazard ratio is comparing two separate groups of patients. However, the exposure variable in this analysis is *not* a time-independent variable that distinguishes between two separate groups. In contrast, the exposure variable is time-dependent, and uses the wait-time information on transplants as contributing to the survival experience of non-transplants.

Since the exposure variable is time-dependent, an alternative interpretation of the hazard ratio estimate is that, at any given time t, the hazard for a person *who has not yet received a transplant* (but may receive one later) is approximately 24 times the hazard for a person *who already has received a transplant by that time*.

Actually, we suggest that a more appropriate hazard ratio expression is required to account for a transplant's *TMS* and *AGE* score. Such an expression would compare, at time t, the values of each of the three time-dependent variables in the model. For a person who received a transplant, these values are 1 for $HT(t)$ and *TMS* and *AGE* for the two covariates. For a person who has not received a transplant, the values of all three variables are 0.

Using this approach to compute the hazard ratio, the $\mathbf{X}^*(t)$ vector, which specifies the predictors for a patient i who received a transplant at time t, has the values 1, TMS_i and AGE_i for patient i; the $\mathbf{X}(t)$ vector, which specifies the predictors at time t for a patient who has not received a transplant at time t, has values of 0 for all three predictors.

The hazard ratio formula then reduces to e to the sum of δ_1 "hat" plus δ_2 "hat" times TMS_i plus δ_3 "hat" times AGE_i, where the δ "hat's" are the estimated coefficients of the three time-dependent variables. Substituting the numerical values for these coefficients in the formula gives the exponential expression circled here.

EXAMPLE (continued)

$\widehat{HR}(t)$ is time-dependent, i.e., its value at time t depends on TMS_i and AGE_i at time t

TMS range: (0–3.05)
AGE range: (12–64)

The resulting formula for the hazard ratio is time-dependent in that its value depends on the *TMS* and *AGE* values of the *i*th patient at the time of transplant. That is, different patients can have different values for *TMS* and *AGE* at time of transplant. Note that in the dataset, *TMS* ranged between 0 and 3.05 and *AGE* ranged between 12 and 64.

We end our discussion of the Stanford Heart Transplant Study at this point. For further insight into the analysis of this dataset, we refer the reader to the 1977 paper by Crowley and Hu (*J. Amer. Statist. Assoc.*).

IX. The Extended Cox Likelihood

At the end of the presentation from Chapter 3 (Section VIII), we illustrated the Cox likelihood using the dataset shown on the left. In this section we extend that discussion to illustrate the Cox likelihood with a time-dependent variable.

ID	TIME	STATUS	SMOKE
Barry	2	1	1
Gary	3	1	0
Harry	5	0	0
Larry	8	1	1

SURVT = Survival time (in years)
STATUS = 1 for event, 0 for censorship
SMOKE = 1 for a smoker, 0 for a nonsmoker

To review: The data indicate that Barry got the event at TIME = 2 years. Gary got the event at 3 years, Harry was censored at 5 years, and Larry got the event at 8 years. Furthermore, Barry and Larry were smokers whereas Gary and Harry were nonsmokers.

Cox PH model: $h(t) = h_0(t)e^{\beta_1 SMOKE}$

Cox PH Likelihood

$$L = \left[\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1} + h_0(t)e^0 + h_0(t)e^0 + h_0(t)e^{\beta_1}}\right] \times \left[\frac{h_0(t)e^0}{h_0(t)e^0 + h_0(t)e^0 + h_0(t)e^{\beta_1}}\right] \times \left[\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1}}\right]$$

In Chapter 3 we constructed the Cox likelihood with one predictor SMOKE in the model. The model and the likelihood are shown on the left. The likelihood is a product of three terms, one term for each event time t_j (TIME = 2, 3, and 8). The denominator of each term is the sum of the hazards from the subjects still in the risk set at time t_j, including the censored subject Harry. The numerator of each term is the hazard of the subject who got the event at t_j. The reader may wish to reread Section VIII of Chapter 3.

Cox extended model

$$h(t) = h_0(t)e^{\beta_1 SMOKE + \beta_2 SMOKE \times TIME}$$

Time-dependent covariate
(its value changes over time)

Now consider an extended Cox model, which contains the predictor SMOKE, and a time-dependent variable SMOKE × TIME. For this model it is not only the baseline hazard that may change over time but also the value of the predictor variables. This can be illustrated by examining Larry's hazard at each event time.

Larry got the event at TIME = 8

Larry's hazard at each event time

TIME	Larry's Hazard
2	$h_0(t)e^{\beta_1+2\beta_2}$
3	$h_0(t)e^{\beta_1+3\beta_2}$
8	$h_0(t)e^{\beta_1+8\beta_2}$

Larry, a smoker, got the event at TIME = 8. However at TIME = 2, 3, and 8, the covariate SMOKE × TIME changes values, thus affecting Larry's hazard at each event time (see left). Understanding how the expression for an individual's hazard changes over time is the key addition toward understanding how the Cox extended likelihood differs from the Cox PH likelihood.

Cox extended model

$$L = \left[\frac{h_0(t)e^{\beta_1+2\beta_2}}{h_0(t)e^{\beta_1+2\beta_2} + h_0(t)e^{0} + h_0(t)e^{0} + h_0(t)e^{\beta_1+2\beta_2}}\right] \times \left[\frac{h_0(t)e^{0}}{h_0(t)e^{0} + h_0(t)e^{0} + h_0(t)e^{\beta_1+3\beta_2}}\right] \times \left[\frac{h_0(t)e^{\beta_1+8\beta_2}}{h_0(t)e^{\beta_1+8\beta_2}}\right]$$

The likelihood for the extended Cox model is constructed in a similar manner to that of the likelihood for the Cox PH model. The difference is that the expression for the subject's hazard is allowed to vary over time. The extended Cox likelihood for these data is shown on the left.

Likelihood is product of 3 terms:

$$L = L_1 \times L_2 \times L_3$$

Barry ($t = 2$) Gary ($t = 3$) Larry ($t = 8$)

Just as with the Cox PH likelihood shown previously, the extended Cox likelihood is also a product of three terms, corresponding to the three event times ($L = L_1 \times L_2 \times L_3$). Barry got the event first at $t = 2$, then Gary at $t = 3$, and finally Larry at $t = 8$. Harry, who was censored at $t = 5$, was still at risk when Barry and Gary got the event. Therefore, Harry's hazard is still in the denominator of L_1 and L_2.

SMOKE × TIME = 0 for nonsmokers

SMOKE × TIME changes over time for smokers

Larry's hazard changes over L_1, L_2, L_3.

The inclusion of the time-varying covariate SMOKE × TIME does not change the expression for the hazard for the nonsmokers (Gary and Harry) because SMOKE is coded 0 for nonsmokers. However, for smokers (Barry and Larry), the expression for the hazard changes with time. Notice how Larry's hazard changes in the denominator of L_1, L_2 and L_3 (see dashed arrows above).

$h_0(t)$ cancels in L

$$L = \left[\frac{e^{\beta_1+2\beta_2}}{e^{\beta_1+2\beta_2}+e^0+e^0+e^{\beta_1+2\beta_2}}\right] \times \left[\frac{e^0}{e^0+e^0+e^{\beta_1+3\beta_2}}\right] \times \left[\frac{e^{\beta_1+8\beta_2}}{e^{\beta_1+8\beta_2}}\right]$$

The baseline hazard cancels in the extended Cox likelihood as it does with the Cox PH likelihood. Thus, the form of the baseline hazard need not be specified, as it plays no role in the estimation of the regression parameters.

Incorrent coding of SMOKE × TIME

ID	TIME	STATUS	SMOKE	SMOKE × TIME
Barry	2	1	1	2
Gary	3	1	0	0
Harry	5	0	0	0
Larry	8	1	1	8

Coded as time-independent, not time-dependent

A word of caution for those planning to run a model with a time-varying covariate: it is incorrect to create a product term with TIME in the data step by multiplying each individual's value for SMOKE with his survival time. In other words, SMOKE × TIME should not be coded like the typical interaction term. In fact, if SMOKE × TIME were coded as it is on the left, then SMOKE × TIME would be a time-independent variable. Larry's value for SMOKE × TIME is incorrectly coded at a constant value of 8 even though Larry's value for SMOKE × TIME changes in the likelihood over L_1, L_2, and L_3.

Incorrectly coded SMOKE × TIME

- Time independent
- Probably highly significant
- Survival time should predict survival time
- But not meaningful

If the incorrectly coded time-independent SMOKE × TIME were included in a Cox model it would not be surprising if the coefficient estimate were highly significant even if the PH assumption were not violated. It would be expected that a product term with each individual's survival time would predict the outcome (his survival time), but it would not be meaningful. Nevertheless, this is a common mistake.

Correctly coding SMOKE × TIME

- Time dependent
- Computer packages allow definition in the analytic procedure
- See Computer Appendix for details

To obtain a correctly defined SMOKE × TIME time-dependent variable, computer packages typically allow the variable to be defined within the analytic procedure. See Computer Appendix to see how time-dependent variables are defined in Stata, SAS, and SPSS.

Coding SMOKE × TIME as time-dependent

Multiple Observations per Subject

ID	TIME	STATUS	SMOKE	SMOKE × TIME
Barry	2	1	1	2
Gary	2	0	0	0
Gary	3	1	0	0
Harry	2	0	0	0
Harry	3	0	0	0
Harry	5	0	0	0
Larry	2	0	1	2
Larry	3	0	1	3
Larry	5	0	1	5
Larry	8	1	1	8

Coded as time-dependent

When a time-dependent variable is defined within the Cox analytic procedure, the variable is defined internally such that the user may not see the time-dependent variable in the dataset. However, the dataset on the left will provide a clearer idea of the correct definition of SMOKE × TIME. The dataset contains multiple observations per subject. Barry was at risk at t = 2 and got the event at that time. Gary was at risk at t = 2 and t = 3. Gary didn't get the event at t = 2 but did get the event at t = 3. Harry was at risk at t = 2, t = 3, t = 5 and didn't get the event. Larry was at risk at t = 2, t = 3, t = 5, t = 8 and got the event at t = 8. Notice how the SMOKE × TIME variable changes values for Larry over time.

Multiple observations per subject: revisited in Chapter 8 (recurrent events)

Survival analysis datasets containing multiple observations per subject are further discussed in Chapter 8 on recurrent events. With recurrent event data, subjects may remain at risk for subsequent events after getting an event.

X. Summary

Review Cox PH model.

Define time-dependent variable: defined, internal, ancillary.

A summary of this presentation on time-dependent variables is now provided. We began by reviewing the main features of the Cox PH model. We then defined a time-dependent variable and illustrated three types of these variables—defined, internal, and ancillary.

Extended Cox model:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i X_i + \sum_{j=1}^{p_2}\delta_j X_j(t)\right]$$

Next, we gave the form of the "extended Cox model," shown here again, which allows for time-dependent as well as time-independent variables.

$$\widehat{HR}(t) = \exp\left[\sum_{i=1}^{p_1}\hat{\beta}_i\left[X_i^* - X_i\right] + \sum_{j=1}^{p_2}\hat{\delta}_j\left[\boxed{X_j^*(t) - X_j(t)}\right]\right]$$

Function of time

We then described various characteristics of this extended Cox model, including the formula for the hazard ratio. The latter formula is time-dependent so that the PH assumption is not satisfied.

Model for assessing PH assumption:

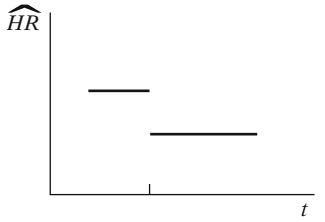
$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \sum_{i=1}^{p}\delta_i X_i g_i(t)\right]$$

We also showed how to use time-dependent variables to assess the PH assumption for time-independent variables. A general formula for an extended Cox model that simultaneously considers all time-independent variables of interest is shown here.

Examples of $g_i(t)$:
t, $\log t$, heaviside function

The functions $g_i(t)$ denote functions of time for the ith variable that are to be determined by the investigator. Examples of such functions are $g_i(t) = t$, $\log t$, or a heaviside function.

Heaviside functions:



The use of heaviside functions were described and illustrated. Such functions allow for the hazard ratio to be constant within different time intervals.

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta E g(t)]$$

where

$$g(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0 \end{cases}$$

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1 E g_1(t) + \beta_2 E g_2(t)]$$

where

$$g_1(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0 \end{cases}$$

$$g_2(t) = \begin{cases} 1 & \text{if } t < t_0 \\ 0 & \text{if } t \geq t_0 \end{cases}$$

For two time intervals, the model can take either one of two equivalent forms as shown here. The first model contains a main effect of exposure and only one heaviside function. The second model contains two heaviside functions without a main effect of exposure. Both models yield two distinct and equivalent values for the hazard ratio.

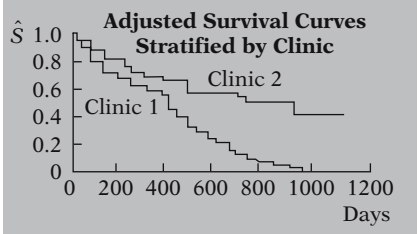
EXAMPLE 1

1991 Australian study of heroin addicts

- two methadone maintenance clinics
- *addicts* dataset file
- clnic variable did not satisfy PH assumption

We illustrated the use of time-dependent variables through two examples. The first example considered the comparison of two methadone maintenance clinics for heroin addicts. The dataset file was called *addicts*. In this example, the clinic variable, which was a dichotomous exposure variable, did not satisfy the PH assumption.

EXAMPLE (continued)



$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_2(\text{prison}) + \beta_3(\text{dose}) + \delta_1(\text{clinic})g_1(t) + \delta_2(\text{clinic})g_2(t)]$$

Heaviside functions

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_2(\text{prison}) + \beta_3(\text{dose}) + \beta_1(\text{clinic}) + \delta(\text{clinic} \times t)]$$

where
$T(t)$ 1, 3, 5, 7, 9 in half-year intervals

Adjusted survival curves stratified by clinic showed clinic 2 to have consistently higher survival probabilities than clinic 1, with a more pronounced difference in clinics after one year of follow-up. However, this stratification did not allow us to obtain a hazard ratio estimate for clinic. Such an estimate was possible using an extended Cox model containing interaction terms involving clinic with time.

Two extended Cox models were considered. The first used heaviside functions to obtain two distinct hazard ratios, one for the first year of follow-up and the other for greater than one year of follow-up. The model is shown here.

The second extended Cox model used a time-dependent variable that allowed for the two survival curves to diverge over time. This model is shown here.

Both models yielded hazard ratio estimates that agreed reasonably well with the graph of adjusted survival curves stratified by clinic.

EXAMPLE 2: Stanford Heart Transplant Study

Goals: Do patients receiving transplants survive longer than patients not receiving transplants?

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta_1 HT(t) + \delta_2 TMS(t) + \delta_3 AGE(t)]$$

Exposure variable

The second example considered results obtained in the Stanford Heart Transplant Study. The goal of the study was to assess whether patients receiving transplants survived longer than patients not receiving transplants.

The analysis of these data involved an extended Cox model containing three time-dependent variables. One of these, the exposure variable, and called $HT(t)$, was an indicator of transplant status at time t. The other two variables, $TMS(t)$ and $AGE(t)$, gave tissue mismatch scores and age for transplant patients when time t occurred after receiving a transplant. The value of each of these variables was 0 at times prior to receiving a transplant.

EXAMPLE (continued)

Results: $HT(t)$ highly significant, i.e., transplants have better prognosis than nontransplants.
Hazard ratio estimate problematic:

$$\widehat{HR} = e^{\hat{\delta}_1} = \frac{1}{23.98}$$

More appropriate formula:

$$\widehat{HR} = \exp[\, -3.1718 + 0.4442\, TMS_i + 0.0552\, AGE_i]$$

The results from fitting the above extended Cox model yielded a highly significant effect of the exposure variable, thus indicating that survival prognosis was better for transplants than for nontransplants.

From these data, we first presented an inappropriate formula for the estimated hazard ratio. This formula used the exponential of the coefficient of the exposure variable, which gave an estimate of 1 over 23.98. A more appropriate formula considered the values of the covariates $TMS(t)$ and $AGE(t)$ at time t. Using the latter, the hazard ratio estimate varied with the tissue mismatch scores and age of each transplant patient.

Chapters

1. Introduction to Survival Analysis
2. Kaplan–Meier Curves and the Log–Rank Test
3. The Cox Proportional Hazards Model
4. Evaluating the Proportional Hazards Assumption
5. The Stratified Cox Procedure
6. ✓ Extension of the Cox Proportional Hazards Model for Time-Dependent Variables

Next:

7. Parametric models

This presentation is now complete. We suggest that the reader review the detailed outline that follows and then answer the practice exercises and test that follow the outline.

A key property of Cox models is that the distribution of the outcome, survival time, is unspecified. In the next chapter, parametric models are presented in which the underlying distribution of the outcome is specified. The exponential, Weibull, and log-logistic models are examples of parametric models.

Detailed Outline

I. Preview (page 214)

II. Review of the Cox PH Model (pages 214–216)

A. The formula for the Cox PH model:

$$h(t,\mathbf{X}) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i\right]$$

B. Formula for hazard ratio comparing two individuals:

$\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_p^*)$ and $\mathbf{X} = (X_1, X_2, \ldots, X_p)$:

$$\frac{h(t,\mathbf{X}^*)}{h(t,\mathbf{X})} = \exp\left[\sum_{i=1}^{p}\beta_i(X_i^* - X_i)\right]$$

C. The meaning of the PH assumption:

- Hazard ratio formula shows that the hazard ratio is independent of time:

$$\frac{h(t,\mathbf{X}^*)}{h(t,\mathbf{X})} = \theta$$

- Hazard ratio for two X's are proportional:

$$h(t,\mathbf{X}^*) = \theta h(t,\mathbf{X})$$

D. Three methods for checking the PH assumption:

i. *Graphical:* Compare ln–ln survival curves or observed versus predicted curves

ii. *Time-dependent covariates:* Use product (i.e., interaction) terms of the form $X \times g(t)$.

iii. *Goodness-of-fit test:* Use a large sample Z statistic.

E. Options when the PH assumption is not met:

i. Use a stratified Cox procedure.

ii. Use an extended Cox model containing a time-dependent variable of the form $X \times g(t)$.

III. Definition and Examples of Time-Dependent Variables (pages 216–219)

A. Definition: any variable whose values differ over time

B. Examples of defined, internal, and ancillary time-dependent variables

IV. The Extended Cox Model for Time-Dependent Varibles (pages 219–221)

A. $$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i X_i + \sum_{j=1}^{p_2}\delta_j X_j(t)\right]$$

where $\mathbf{X}(t) = (X_1, X_2, \ldots, X_{p_1}, X_1(t), X_2(t), \ldots, X_{p_2}(t))$ denotes the entire collection of predictors at time t, X_i denotes the ith time-independent variable, and $X_j(t)$ denotes the jth time-dependent variable.

B. ML procedure used to estimate regression coefficients.

C. List of computer programs for the extended Cox model.

D. Model assumes that the hazard at time t depends on the value of $X_j(t)$ at the *same* time.

E. Can modify model for lag-time effect.

V. The Hazard Ratio Formula for the Extended Cox Model (pages 221–223)

A. $$HR(t) = \exp\left[\sum_{i=1}^{p_1}\hat{\beta}_i\left[X_i^* - X_i\right] + \sum_{j=1}^{p_2}\hat{\delta}_j\left[X_j^*(t) - X_j(t)\right]\right]$$

B. Because $HR(t)$ is a function of time, the PH assumption is not satisfied.

C. The estimated coefficient of $X_j(t)$ is time-independent, and represents an "overall" effect of $X_j(t)$.

VI. Assessing Time-Independent Variables That Do Not Satisfy the PH Assumption (pages 224–229)

A. General formula for assessing PH assumption:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i + \sum_{i=1}^{p}\delta_i X_i g_i(t)\right]$$

B. $g_i(t)$ is a function of time corresponding to X_i

C. Test H_0: $\delta_1 = \delta_2 = \ldots = \delta_p = 0$

D. Heaviside function:

$$g(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0 \end{cases}$$

E. The model with a single heaviside function:

$$\boxed{h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta E + \delta E g(t)]}$$

F. The model with two heaviside functions:

$$\boxed{h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta_1 E g_1(t) + \delta_2 E g_2(t)]}$$

where

$$g_1(t) = \begin{cases} 1 \text{ if } t \geq t_0 \\ 0 \text{ if } t < t_0 \end{cases} \quad \text{and} \quad g_2(t) = \begin{cases} 1 \text{ if } t < t_0 \\ 0 \text{ if } t \geq t_0 \end{cases}$$

G. The hazard ratios:

$t \geq t_0$: $\widehat{HR} = \exp(\hat{\beta} + \hat{\delta}) = \exp(\hat{\delta}_1)$

$t < t_0$: $\widehat{HR} = \exp(\hat{\beta}) = \exp(\hat{\delta}_2)$

H. Several heaviside functions: examples given with four time-intervals:

- Extended Cox model contains either $\{E, E \times g_1(t), E \times g_2(t), E \times g_3(t)\}$ or $\{E \times g_1(t), E \times g_2(t), E \times g_3(t), E \times g_4(t)\}$
- The model using four product terms and no main effect of *E:*

$$\boxed{\begin{aligned} h(t,\mathbf{X}(t)) = h_0(t)\exp[&\delta_1 E g_1(t) + \delta_2 E g_2(t) \\ &+ \delta_3 E g_3(t) + \delta_4 E g_4(t)] \end{aligned}}$$

where

$$g_i(t) = \begin{cases} 1 & \text{if } t \text{ is within interval } i \\ 0 & \text{if otherwise} \end{cases}$$

VII. An Application of the Extended Cox Model to an Epidemiologic Study on the Treatment of Heroin Addiction (pages 230–234)

A. 1991 Australian study of heroin addicts

- two methadone maintenance clinics
- *addicts* dataset file
- clinic variable did not satisfy PH assumption

B. Clinic 2 has consistently higher retention probabilities than clinic 1, with a more pronounced difference in clinics after one year of treatment.

C. Two extended Cox models were considered:

- Use heaviside functions to obtain two distinct hazard ratios, one for less than one year and the other for greater than one year.
- Use a time-dependent variable that allows for the two survival curves to diverge over time.

VIII. An Application of the Extended Cox Model to the Analysis of the Stanford Heart Transplant Data (pages 235–239)

A. The goal of the study was to assess whether patients receiving transplants survived longer than patients not receiving transplants.

B. We described an extended Cox model containing three time-dependent variables:

$$h(t,\mathbf{X}(t)) = h_0(t)\exp[\delta_1 HT(t) + \delta_2 TMS(t) + \delta_3 AGE(t)]$$

C. The exposure variable, called $HT(t)$, was an indicator of transplant status at time t. The other two variables, $TMS(t)$ and AGE(t), gave tissue mismatch scores and age for transplant patients when time t occurred after receiving a transplant.

D. The results yielded a highly significant effect of the exposure variable.

E. The use of a hazard ratio estimate for this data was problematical.

- An inappropriate formula is the exponential of the coefficient of $HT(t)$, which yields 1/23.98.
- An alternative formula considers the values of the covariates $TMS(t)$ and $AGE(t)$ at time t.

IX. Extended Cox Likelihood (pages 239–242)

A. Review of PH likelihood (Chapter 3).

B. Barry, Gary, Larry, example of Cox likelihood.

X. Summary (pages 242–245)

Practice Exercises

The following dataset called "anderson.dat" consists of remission survival times on 42 leukemia patients, half of whom receive a new therapy and the other half of whom get a standard therapy (Freireich et al., *Blood*, 1963). The exposure variable of interest is treatment status ($Rx = 0$ if new treatment, $Rx = 1$ if standard treatment). Two other variables for control are log white blood cell count (i.e., log WBC) and sex. Failure status is defined by the relapse variable (0 if censored, 1 if failure). The dataset is listed as follows:

Subj	Surv	Relapse	Sex	log WBC	*Rx*
1	35	0	1	1.45	0
2	34	0	1	1.47	0
3	32	0	1	2.2	0
4	32	0	1	2.53	0
5	25	0	1	1.78	0
6	23	1	1	2.57	0

(*Continued on next page*)

Subj	Surv	Relapse	Sex	log WBC	*Rx*
7	22	1	1	2.32	0
8	20	0	1	2.01	0
9	19	0	0	2.05	0
10	17	0	0	2.16	0
11	16	1	1	3.6	0
12	13	1	0	2.88	0
13	11	0	0	2.6	0
14	10	0	0	2.7	0
15	10	1	0	2.96	0
16	9	0	0	2.8	0
17	7	1	0	4.43	0
18	6	0	0	3.2	0
19	6	1	0	2.31	0
20	6	1	1	4.06	0
21	6	1	0	3.28	0
22	23	1	1	1.97	1
23	22	1	0	2.73	1
24	17	1	0	2.95	1
25	15	1	0	2.3	1
26	12	1	0	1.5	1
27	12	1	0	3.06	1
28	11	1	0	3.49	1
29	11	1	0	2.12	1
30	8	1	0	3.52	1
31	8	1	0	3.05	1
32	8	1	0	2.32	1
33	8	1	1	3.26	1
34	5	1	1	3.49	1
35	5	1	0	3.97	1
36	4	1	1	4.36	1
37	4	1	1	2.42	1
38	3	1	1	4.01	1
39	2	1	1	4.91	1
40	2	1	1	4.48	1
41	1	1	1	2.8	1
42	1	1	1	5	1

The following edited printout gives computer results for fitting a Cox PH model containing the three predictives *Rx*, log WBC, and Sex.

Cox regression Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf. Interval]		*P*(*PH*)
Sex	0.263	0.449	0.558	1.301	0.539	3.139	0.042
log WBC	1.594	0.330	0.000	4.922	2.578	9.397	0.714
Rx	1.391	0.457	0.002	4.018	1.642	9.834	0.500

No. of subjecs = 42 Log likelihood = −72.109

1. Which of the variables in the model fitted above are time-independent and which are time-dependent?

2. Based on this printout, is the PH assumption satisfied for the model being fit? Explain briefly.

3. Suppose you want to use an extended Cox model to assess the PH assumption for all three variables in the above model. State the general form of an extended Cox model that will allow for this assessment.

4. Suppose you wish to assess the PH assumption for the Sex variable using a heaviside function approach designed to yield a constant hazard ratio for less than 15 weeks of follow-up and a constant hazard ratio for 15 weeks or more of follow-up. State two equivalent alternative extended Cox models that will carry out this approach, one model containing one heaviside function and the other model containing two heaviside functions.

5. The following is an edited printout of the results obtained by fitting an extended Cox model containing two heaviside functions:

   Time-Dependent Cox Regression Analysis

Analysis time_t: survt	Coef.	Std. Err.	$p > \|z\|$	Haz. Ratio	[95% Conf. Interval]	
log WBC	1.567	0.333	0.000	4.794	2.498	9.202
Rx	1.341	0.466	0.004	3.822	1.533	9.526
0–15 wks	0.358	0.483	0.459	1.430	0.555	3.682
15+ wks	−0.182	0.992	0.855	0.834	0.119	5.831

   No. of subjects = 42 Log likelihood = −71.980

   Using the above computer results, carry out a test of hypothesis, estimate the hazard ratio, and obtain 95% confidence interval for the treatment effect adjusted for log WBC and the time-dependent Sex variables. What conclusions do you draw about the treatment effect?

6. We now consider an alternative approach to controlling for Sex using an extended Cox model. We define an interaction term between sex and time that allows for diverging survival curves over time.

   For the situation just described, write down the extended Cox model, which contains *Rx*, log WBC, and Sex as main effects plus the product term sex × time.

7. Using the model described in question 6, express the hazard ratio for the effect of Sex adjusted for *Rx* and log WBC at 8 and 16 weeks.

8. The following is an edited printout of computer results obtained by fitting the model described in question 6.

Time-Dependent Cox Regression Analysis

Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Sex	1.820	1.012	0.072	6.174	0.849	44.896
log WBC	1.464	0.336	0.000	4.322	2.236	8.351
Rx	1.093	0.479	0.022	2.984	1.167	7.626
Sex × Time	−0.345	0.199	0.083	0.708	0.479	1.046

No. of subjects = 42 Log likelihood = −70.416

Based on the above results, describe the hazard ratio estimate for the treatment effect adjusted for the other variables in the model, and summarize the results of the significance test and interval estimate for this hazard ratio. How do these results compare with the results previously obtained when a heaviside function approach was used? What does this comparison suggest about the drawbacks of using an extended Cox model to adjust for variables not satisfying the PH assumption?

9. The following gives an edited printout of computer results using a stratified Cox procedure that stratifies on the Sex variable but keeps *Rx* and log WBC in the model.

Stratified Cox regression

Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
log WBC	1.390	0.338	0.000	4.016	2.072	7.783
Rx	0.931	0.472	0.048	2.537	1.006	6.396

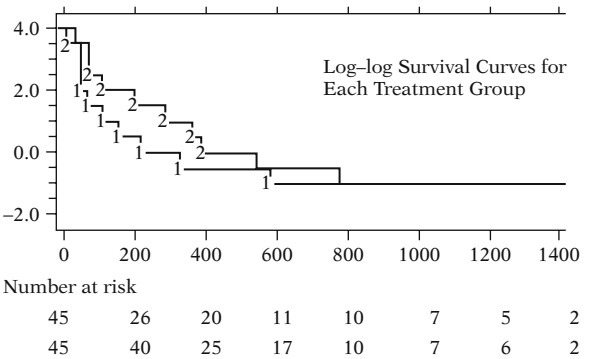
No. of subjects = 42 Log likelihood = −57.560 Stratified by sex

Compare the results of the above printout with previously provided results regarding the hazard ratio for the effect of *Rx*. Is there any way to determine which set of results is more appropriate? Explain.

Test

The following questions consider the analysis of data from a clinical trial concerning gastric carcinoma, in which 90 patients were randomized to either chemotherapy (coded as 2) alone or to a combination of chemotherapy and radiation (coded as 1). See Stablein et al., "Analysis of Survival Data with Nonproportional Hazard Functions," *Controlled Clinical Trials*, vol. 2, pp. 149–159 (1981). A listing of the dataset (called chemo) is given at the end of the presentation.

1. A plot of the log–log Kaplan–Meier curves for each treatment group is shown below. Based on this plot, what would you conclude about the PH assumption regarding the treatment group variable? Explain.



2. The following is an edited printout of computer results obtained when fitting the PH model containing only the treatment group variable. Based on these results, what would you conclude about the PH assumption regarding the treatment group variable? Explain.

Cox regression Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf. Interval]		*P*(*PH*)
Tx	−0.267	0.233	0.253	0.766	0.485	1.21	0

No. of subjects = 90 Log likelihood = −282.744

3. The following printout shows the results from using a heaviside function approach with an extended Cox model to fit these data. The model used product terms of the treatment variable (*Tx*) with each of three heaviside functions. The first product term (called Time1) involves a heaviside function for the period from 0 to 250 days, the second product term (i.e., Time2) involves the period from 250 to 500 days, and the third product term (i.e., Time3) involves the open-ended period from 500 days and beyond.

Time-Dependent Cox Regression Analysis

Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Time1	−1.511	0.461	0.001	0.221	0.089	0.545
Time2	0.488	0.450	0.278	1.629	0.675	3.934
Time3	0.365	0.444	0.411	1.441	0.604	3.440

No. of subjects = 90 Log likelihood = −275.745

Write down the hazard function formula for the extended Cox model being used, making sure to explicitly define the heaviside functions involved.

4. Based on the printout, describe the hazard ratios in each of the three time intervals, evaluate each hazard ratio for significance, and draw conclusions about the extent of the treatment effect in each of the three time intervals considered.
5. Inspection of the printout provided in question 3 indicates that the treatment effect in the second and third intervals appears quite similar. Consequently, another analysis was considered that uses only two intervals, from 0 to 250 days versus 250 days and beyond. Write down the hazard function formula for the extended Cox model that considers this situation (i.e., containing two heaviside functions). Also, write down an equivalent alternative hazard function formula which contains the main effect of treatment group plus one heaviside function variable.
6. For the situation described in question 5, the computer results are provided below. Based on these results, describe the hazard ratios for the treatment effect below and above 250 days, summarize the inference results for each hazard ratio, and draw conclusions about the treatment effect within each time interval.

Time-Dependent Cox Regression Analysis

Analysis time_t: survt Column name	Coeff	StErr	p-value	*HR*	0.95	CI
Time1	−1.511	0.461	0.001	0.221	0.089	0.545
Time2	0.427	0.315	0.176	1.532	0.826	2.842

No. of subjects = 90 Log likelihood = −275.764

Answers to Practice Exercises

1. All three variables in the model are time-independent variables.

2. The computer results indicate that the Sex variables do not satisfy the PH assumption because the $P(PH)$ value is 0.042, which is significant at the 0.05 level.

3. $$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{sex}) + \beta_2(\log \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex})g_1(t) + \delta_2(\log \text{WBC})g_2(t) + \delta_3(Rx)g_3(t)]$$
   where the $g_i(t)$ are functions of time.

4. Model 1 (one heaviside function)
   $$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{sex}) + \beta_2(\log \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex})g_1(t)]$$
   where
   $$g_1(t) = \begin{cases} 1 & \text{if } 0 \le t < 15 \text{ weeks} \\ 0 & \text{if } t \ge 15 \text{ weeks} \end{cases}$$
   Model 2 (two heaviside functions):
   $$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_2(\log \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex})g_1(t) + \delta_2(\text{sex})g_2(t)]$$
   where
   $$g_1(t) = \begin{cases} 1 & \text{if } 0 \le t < 15 \text{ weeks} \\ 0 & \text{if } t \ge 15 \text{ weeks} \end{cases}$$
   and
   $$g_2(t) = \begin{cases} 0 & \text{if } t \ge 15 \text{ weeks} \\ 1 & \text{if } 0 \le t < 15 \text{ weeks} \end{cases}$$

5. The estimated hazard ratio for the effect of *Rx* is 3.822; this estimate is adjusted for log WBC and for the Sex variable considered as two time-dependent variables involving heaviside functions. The Wald test for significance of *Rx* has a p-value of 0.004, which is highly significant. The 95% confidence interval for the treatment effect ranges between 1.533 and 9.526, which is quite wide, indicating considerable unreliability of the 3.822 point estimate. Nevertheless, the results estimate a statistically significant treatment effect of around 3.8.

6. $$h(t,\mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{sex}) + \beta_2(\log \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex} \times \text{t})]$$

7. The hazard ratio for the effect of Sex in each time interval, controlling for *Rx* and log WBC is given as follows:

   $t = 8$ weeks $\quad \widehat{HR} = \exp[\hat{\beta}_1 + 8\hat{\delta}_1]$
   $t = 16$ weeks $\quad \widehat{HR} = \exp[\hat{\beta}_1 + 16\hat{\delta}_1]$

8. Using the model containing Sex, log WBC, *Rx*, and Sex $\times$ Time, the estimated hazard ratio for the treatment effect is given by 2.984, with a p-value of 0.022 and a 95% confidence interval ranging between 1.167 and 7.626. The point estimate of 2.984 is quite different from the point estimate of 3.822 for the heaviside function model, although the confidence intervals for both models are wide enough to include both estimates. The discrepancy between point estimates demonstrates that when a time-dependent variable approach is to be used to account for a variable not satisfying the PH assumption, different results may be obtained from different choices of time-dependent variables.

9. The stratified Cox analysis yields a hazard ratio of 2.537 with a p-value of 0.048 and a 95% CI ranging between 1.006 and 6.396. The point estimate is much closer to the 2.984 for the model containing the Sex $\times$ Time product term than to the 3.822 for the model containing two heaviside functions. One way to choose between models would be to compare goodness-of-fit test statistics for each model; another way is to compare graphs of the adjusted survival curves for each model and determine by eye which set of survival curves fits the data better.