EXAMPLE (continued)

Model 2 is best model.

$\widehat{HR}$=3.648 statistically significant

95% CI for HR: (1.5, 8.3)

Our analysis of the output for the three models has led us to conclude that model 2 is the best of the three models and that, using model 2, we get a statistically significant hazard ratio of 3.648 for the effect of the treatment, with a 95% confidence interval ranging between 1.5 and 8.3.

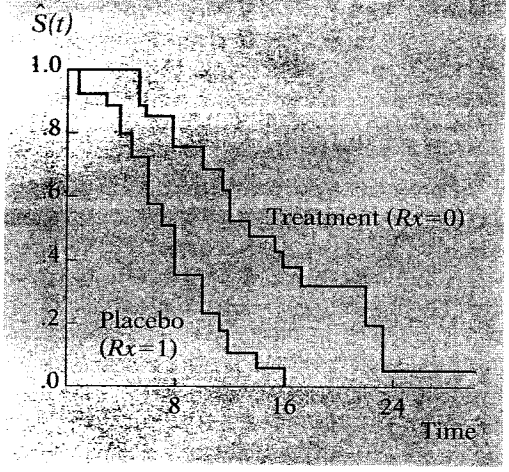
Cox model formulae not specified

Note that we were able to carry out this analysis without actually specifying the formulae for the Cox PH models being fit. Also, the strategy and methods used with the output provided have been completely analogous to the strategy and methods one uses when fitting logistic regression models (see Kleinbaum, *Logistic Regression*, Chapters 6 and 7, 1994), and very similar to carrying out a classical linear regression analysis (see Kleinbaum et al., *Applied Regression Analysis*, 2d ed., Chapter 16, 1987).

Analysis strategy and methods for Cox model analogous to those for logistic and classical linear models.

EXAMPLE (continued)

Survival Curves Adjusted for log WBC (Model 2)



In addition to the above analysis of this data, we can also obtain survival curves for each treatment group, **adjusted** for the effects of log WBC and based on the model 2 output. Such curves, sketched here at the left, give additional information to that provided by estimates and tests about the hazard ratio. In particular, these curves describe how the treatment groups compare over the time period of the study.

For these data, the survival curves show that the treatment group consistently has higher survival probabilities than the placebo group after adjusting for log WBC. Moreover, the difference between the two groups appears to widen over time.

Adjusted survival curves	KM curves
Adjusted for covariates	No covariates
Use fitted Cox model	No Cox model fitted

Note that adjusted survival curves are mathematically different from Kaplan–Meier (KM) curves. KM curves do not adjust for covariates and, therefore, are not computed using results from a fitted Cox PH model.

EXAMPLE (continued)

For these data, KM and adjusted survival plots have similar appearance.

Nevertheless, for these data, the plotted KM curves (which were described in Chapter 2) are similar in appearance to the adjusted survival curves.

P(PH) OUTPUT

Model 2:

Column name	P(PH)
Rx	0.944
log(WBC)	0.917

P(PH): gives p-value for evaluating PH assumption for each variable in model; derived from $N(0,1)$ statistic

P(PH) large (e.g., $P > 0.10$) $\Rightarrow$ *PH* satisfied

P(PH) small (e.g., $P < 0.05$) $\Rightarrow$ *PH* not satisfied

Before concluding this section, we point out one other piece of information provided in the output that we have not mentioned until now. We refer to the *P(PH)* values provided in the last column of the printout, as shown here for model 2.

The *P(PH)* information allows one to evaluate the proportional hazards (PH) assumption. The value given is a p-value derived from a standard normal statistic computed from the model output. A nonsignificant (i.e., large) p-value, say greater than 0.10, indicates that the PH assumption is satisfied, whereas a small p-value, say less than 0.05, indicates that the variable being tested does not satisfy this assumption. We discuss the PH assumption in more detail later in this presentation.

EXAMPLE (continued)

Model 2:

P(PH) nonsignificant for both variables, so PH is satisfied

The *P(PH)* output for model 2 yields nonsignificant p-values for both variables, thus indicating that the PH assumption is satisfied for both variables.

Three approaches for evaluating PH (Chapter 4)

Procedures when PH not satisfied (Chapters 5 and 6)

The *P(PH)* information provides one of three approaches for evaluating the PH assumption. All three approaches will be discussed and compared in Chapter 4. In addition, Chapters 5 and 6 describe procedures to use when the PH assumption is not satisfied.

Remainder:
- Cox model formula
- basic characteristics of Cox model
- meaning of PH assumption

In the remainder of this presentation, we describe the Cox PH formula and its basic characteristics, including what is the meaning of the PH assumption.

II. The Formula for the Cox PH Model

$$h(t,\mathbf{X}) = h_0(t)\, e^{\sum_{i=1}^{p} \beta_i X_i}$$

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

The Cox PH model is usually written in terms of the hazard model formula shown here at the left. 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 $\mathbf{X}$. That is, the bold $\mathbf{X}$ represents a collection (sometimes called a "vector") of predictor variables that is being modeled to predict an individual's hazard.

$$h_0(t) \times e^{\sum_{i=1}^{p} \beta_i X_i}$$

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 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. In contrast, the exponential expression shown here, 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 which 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.**

Time-dependent variables: Chapter 6

The use of time-dependent variables is discussed in Chapter 6. For the remainder of this presentation, we will consider time-independent X's only.

Time-independent variable:
Values for a given individual do not change over time; e.g., SEX and SMK

Assumed not to change once measured

A time-independent variable is defined to be any variable whose values for a given individual do not change over time. Examples are SEX and smoking status (SMK). Note, however, that a person's smoking status may actually change over time, but for purposes of the analysis, the SMK variable is assumed not to change once it is measured, so that only one value per individual is used.

AGE and WGT values do not change much, or effect on survival depends on one measurement.

Also note that although variables like AGE and weight (WGT) change over time, it may be appropriate to treat such variables as time-independent in the analysis if their values do not change much over time or if the effect of such variables on survival risk depends essentially on the value at only one measurement.

$$X_1 = X_2 = \cdots = X_k = 0$$

$$h(t,\mathbf{X}) = h_0(t)\, e^{\sum_{i=1}^{p} \beta_i X_i}$$

$$= h_0(t)\, e^{0}$$

$$= h_0(t)$$

Baseline hazard

The Cox model formula has the property that if all the X's are equal to zero, the formula reduces to the baseline hazard function. That is, the exponential part of the formula becomes e to the zero, which is 1. This property of the Cox model is the reason why $h_0(t)$ is called the baseline function.

No X's in model: $h(t,\mathbf{X}) = h_0(t)$.

Or, from a slightly different perspective, the Cox model reduces to the baseline hazard when no X's are in the model. Thus, $h_0(t)$ may be considered as a starting or "baseline" version of the hazard function, prior to considering any of the X's.

$h_0(t)$ is unspecified.

Cox model: **nonparametric**

Another important property of the Cox model is that the baseline hazard, $h_0(t)$, is an unspecified function. It is this property that makes the Cox model a **nonparametric** model.

EXAMPLE: Parametric Model

Weibull:

$$h(t,\mathbf{X}) = \lambda t^{\alpha-1}\, e^{\sum_{i=1}^{p} \beta_i X_i}$$

where $h_0(t) = \lambda t^{\alpha-1}$

In contrast, a **parametric** model is one whose functional form is completely specified, except for the values of the unknown parameters. For example, the Weibull hazard model is a parametric model and has the form shown here, where the unknown parameters are λ, α, and the β_i's. Note that for the Weibull model, $h_0(t)$ is given by $\lambda t^{\alpha-1}$.

Nonparametric property

⇓

Popularity of the Cox model

One of the reasons why the Cox model is so popular is that it is nonparametric. We discuss this and other reasons in the next section (III) concerning why the Cox model is so widely used.

III. Why the Cox PH Model Is Popular

Cox PH model is "robust":
Will closely approximate correct parametric model

A key reason for the popularity of the Cox model is that, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data situations. Another way of saying this is that the Cox PH model is a "robust" model, so that the results from using the Cox model will closely approximate the results for the correct parametric model.

If correct model is:

Weibull $\Rightarrow$ Cox model will approximate Weibull

Exponential $\Rightarrow$ Cox model will approximate exponential

For example, if the correct parametric model is Weibull, then use of the Cox model typically will give results comparable to those obtained using a Weibull model. Or, if the correct model is exponential, then the Cox model results will closely approximate the results from fitting an exponential model.

Prefer parametric model if sure of correct model, e.g., use goodness-of-fit test (Lee, 1982).

We would prefer to use a parametric model if we were sure of the correct model. Although there are various methods for assessing goodness of fit of a parametric model (for example, see Lee, *Statistical Methods for Survival Data Analysis*, 1982), we may not be completely certain that a given parametric model is appropriate.

When in doubt, the Cox model is a "safe" choice.

Thus, when in doubt, as is typically the case, the Cox model will give reliable enough results so that it is a "safe" choice of model, and the user does not need to worry about whether the wrong parametric model is chosen.

In addition to the general "robustness" of the Cox model, the specific form of the model is attractive for several reasons.

$$h(t,\mathbf{X}) = \underbrace{h_0(t)}_{\text{Baseline hazard}} \times \underbrace{e^{\sum_{i=1}^{p}\beta_i X_i}}_{\text{Exponential}} \Downarrow$$

As described previously, the specific form of the Cox model gives the hazard function as a product of a baseline hazard involving t and an exponential expression involving the X's without t. The exponential part of this product is appealing because it ensures that the fitted model will always give estimated hazards that are nonnegative.

$0 \le h(t,\mathbf{X}) < \infty$ always

$$h_0(t) \times \underbrace{\sum_{i=1}^{k}\beta_i X_i}_{\text{Linear}} \Downarrow \text{Might be} < 0$$

We want such nonnegative estimates because, by definition, the values of any hazard function must range between zero and plus infinity, that is, a hazard is always nonnegative. If, instead of an exponential expression, the X part of the model were, for example, linear in the X's, we might obtain negative hazard estimates, which are not allowed.

Even though $h_0(t)$ is unspecified, we can estimate the β's.

Measure of effect: hazard ratio (*HR*) involves only β's, without estimating $h_0(t)$.

Another appealing property of the Cox model is that, even though the baseline hazard part of the model is unspecified, it is still possible to estimate the β's in the exponential part of the model. As we will show later, all we need are estimates of the β's to assess the effect of explanatory variables of interest. The measure of effect, which is called a hazard ratio, is calculated without having to estimate the baseline hazard function.

Can estimate $h(t,\mathbf{X})$ and $S(t,\mathbf{X})$ for Cox model using a minimum of assumptions.

Note that the hazard function $h(t,\mathbf{X})$ and its corresponding survival curves $\mathbf{S}(t,X)$ can be estimated for the Cox model even though the baseline hazard function is not specified. Thus, with the Cox model, using a minimum of assumptions, we can obtain the primary information desired from a survival analysis, namely, a hazard ratio and a survival curve.

Cox model preferred to **logistic** model.

⇓

Uses survival times and censoring	Uses (0,1) outcome; ignores survival times and censoring

One last point about the popularity of the Cox model is that it is preferred over the logistic model when survival time information is available and there is censoring. That is, the Cox model uses more information—the survival times—than the logistic model, which considers a (0,1) outcome and ignores survival times and censoring.

IV. ML Estimation of the Cox PH Model

$$h(t,\mathbf{X}) = h_0(t)\, e^{\sum_{i=1}^{p}\beta_i X_i}$$

ML estimates: $\hat{\beta}_i$

Column name	Coeff	StErr	p-value	*HR*
Rx	1.294	0.422	0.002	3.648
log WBC	1.604	0.329	0.000	4.975
n:42	%Cen: 28.571		−2 log *L*: 144.559	

We now describe how estimates are obtained for the parameters of the Cox model. The parameters are the β's in the general Cox model formula shown here. The corresponding estimates of these parameters are called maximum likelihood (ML) estimates and are denoted as β_i "hat."

As an example of ML estimates, we consider once again the computer output for one of the models (model 2) fitted previously from remission data on 42 leukemia patients.

Model 2:

$h(t,\mathbf{X}) = h_0(t)e^{\beta_1 Rx + \beta_2 \log \text{WBC}}$

Estimated model:

$\hat{h}(t,\mathbf{X}) = \hat{h}_0(t)e^{1.294Rx + 1.604 \log \text{WBC}}$

The Cox model for this example involves two parameters, one being the coefficient of the treatment variable (denoted here as *Rx*) and the other being the coefficient of the log WBC variable. The expression for this model is shown at the left, and below this formula we show the estimated model, which contains the estimated coefficients 1.294 for *Rx* and 1.604 for log white blood cell count.

ML estimates: maximize likelihood function L

L = joint probability of observed data
$= L(\beta)$

As with logistic regression, the ML estimates of the Cox model parameters are derived by maximizing a likelihood function, usually denoted as L. The likelihood function is a mathematical expression which describes the joint probability of obtaining the data actually observed on the subjects in the study as a function of the unknown parameters (the β's) in the model being considered. L is sometimes written notationally as $L(\beta)$ where β denotes the collection of unknown parameters.

L is
- complicated mathematically
- formula built into computer program
- capable of obtaining estimates without seeing formula

We will **not** show you here the explicit mathematical expression for L for the Cox model. This expression is quite complicated mathematically; moreover, in practice, the formula for L is built into the computer program you will be using, so you never have to see it in order to obtain the ML estimates.

L is a partial likelihood:
- considers probabilities only for subjects who fail
- does not consider probabilities for subjects who are censored

The formula for the Cox model likelihood function is actually called a "partial" likelihood function rather than a (complete) likelihood function. The term "partial" likelihood is used because the likelihood formula considers probabilities only for those subjects who fail, and does not explicitly consider probabilities for those subjects who are censored. Thus the likelihood for the Cox model does not consider probabilities for all subjects, and so it is called a "partial" likelihood.

Number of failure times

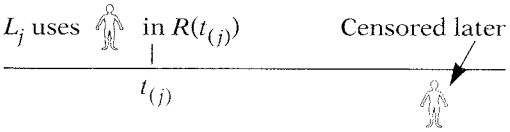
$$L = L_1 \times L_2 \times L_3 \times \cdots \times L_k = \prod_{j=1}^{k} L_j$$

where

L_j = portion of L for the jth failure time given the risk set $R(t_{(j)})$

In particular, the partial likelihood can be written as the product of several likelihoods, one for each of, say, k failure times. Thus, at the jth failure time, L_j denotes the likelihood of failing at this time, given survival up to this time. Note that the set of individuals at risk at the jth failure time is called the "risk set," $R(t_{(j)})$, and this set will change—actually get smaller in size—as the failure time increases.

Information on censored subjects used prior to censorship.



Thus, although the partial likelihood focuses on subjects who fail, survival time information prior to censorship is used for those subjects who are censored. That is, a person who is censored *after* the jth failure time is part of the risk set used to compute L_j, even though this person is censored later.

Steps for obtaining ML estimates:
- form L from model
- maximize ln L by solving

$$\frac{\partial L}{\partial \beta_i} = 0,\ i = 1,\ldots, p(\#\text{of parameters})$$

Once the likelihood function is formed for a given model, the next step for the computer is to maximize this function. This is generally done by maximizing the natural log of L, which is computationally easier.

Solution by iteration:
- guess at solution
- modify guess in successive steps
- stop when solution is obtained

The maximization process is carried out by taking partial derivatives of L with respect to each parameter in the model, and then solving a system of equations as shown here. This solution is carried out using **iteration.** That is, the solution is obtained in a stepwise manner, which starts with a guessed value for the solution, and then successively modifies the guessed value until a solution is finally obtained.

Statistical inferences for hazard ratios:
(See Section I, pages 84–92)

Test hypotheses	Confidence intervals
Wald test LR test	Large sample 95% CI

$\widehat{HR} = e^{\hat{\beta}}$ for a (0,1) exposure variable (no interaction)

Once the ML estimates are obtained, we are usually interested in carrying out statistical inferences about hazard ratios defined in terms of these estimates. We illustrated previously how to test hypotheses and form confidence intervals for the hazard ratio in Section I above. There, we described how to compute a Wald test and a likelihood ratio (LR) test. We also illustrated how to calculate a large sample 95% confidence interval for a hazard ratio. The estimated hazard ratio (HR) was computed by exponentiating the coefficient of a (0,1) exposure variable of interest. Note that the model contained no interaction terms involving exposure.

V. Computing the Hazard Ratio

$$\widehat{HR} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})}$$

where

$$\mathbf{X}^* = \left(X_1^*, X_2^*, \cdots, X_p^*\right)$$

and

$$\mathbf{X} = \left(X_1, X_2, \cdots, X_p\right)$$

denote the set of X's for two individuals

In general, a hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the X's.

We can write the hazard ratio as the estimate of $h(t,\mathbf{X}^*)$ divided by the estimate of $h(t,\mathbf{X})$, where $\mathbf{X}^*$ denotes the set of predictors for one individual, and $\mathbf{X}$ denotes the set of predictors for the other individual.

To interpret $\widehat{HR}$, want $\widehat{HR} \geq 1$, i.e., $\hat{h}(t,\mathbf{X}^*) \geq \hat{h}(t,\mathbf{X})$.

Typical coding: $\mathbf{X}^*$: unexposed group
$\mathbf{X}$: exposed group

EXAMPLE: Remission Data

$\mathbf{X}^* = (X_1^* = 1, X_2^*, \ldots, X_p^*)$, where $X_1^* = 1$ denotes **placebo** group.

$\mathbf{X} = (X_1 = 0, X_2, \ldots, X_p)$, where $X_1 = 0$ denotes **treatment** group.

Note that, as with an odds ratio, it is easier to interpret an *HR* that exceeds the null value of 1 than an *HR* that is less than 1. Thus, the X's are typically coded so that the group with the larger hazard—typically an unexposed or placebo group—corresponds to $\mathbf{X}^*$, and the group with the smaller hazard corresponds to $\mathbf{X}$. As an example, for the remission data described previously, the placebo group is coded as $X_1^* = 1$, and the treatment group is coded as $X_1 = 0$.

$$\widehat{HR} = \frac{\hat{h}(t,\mathbf{X}^*)}{\hat{h}(t,\mathbf{X})} = \frac{\cancel{\hat{h}_0(t)}\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i^*}}{\cancel{\hat{h}_0(t)}\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i}}$$

We now obtain an expression for the *HR* formula in terms of the regression coefficients by substituting the Cox model formula into the numerator and denominator of the hazard ratio expression. This substitution is shown here. Notice that the only difference in the numerator and denominator are the X^*'s versus the X's. Notice also that the baseline hazards will cancel out.

$$\widehat{HR} = \frac{\hat{h}_0(t)\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i^*}}{\hat{h}_0(t)\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i}} = e^{\sum_{i=1}^{p} \hat{\beta}_i \left(X_i^* - X_i\right)}$$

Using algebra involving exponentials, the hazard ratio formula simplifies to the exponential expression shown here. Thus, the hazard ratio is computed by exponentiating the sum of each β_i "hat" times the difference between X_i^* and X_i.

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

An alternative way to write this formula, using exponential notation, is shown here. We will now illustrate the use of this general formula through a few examples.

EXAMPLE

$\mathbf{X} = (X_1, X_2, \ldots, X_p) = (X_1)$, where X_1 denotes (0,1) exposure status ($p = 1$)

$X_1^* = 1, X_1 = 0$

$\widehat{HR} = \exp[\hat{\beta}_1 (X_1^* - X_1)]$

$= \exp[\hat{\beta}_1 (1-0)] = e^{\hat{\beta}_1}$

Suppose, for example, there is only one X variable of interest, X_1, which denotes (0,1) exposure status, so that $p = 1$. Then, the hazard ratio comparing exposed to unexposed persons is obtained by letting $X_1^* = 1$ and $X_1 = 0$ in the hazard ratio formula. The estimated hazard ratio then becomes e to the quantity β_1 "hat" times 1 minus 0, which simplifies to e to the β_1 "hat."

EXAMPLE (continued)

Model 1:

Covariate	Coeff	StErr	p-value	*HR*
Rx	1.509	0.410	0	4.523

Recall the remission data printout for model 1, which contains only the *Rx* variable, again shown here. Then the estimated hazard ratio is obtained by exponentiating the coefficient 1.509, which gives the value 4.523 shown in the *HR* column of the output.

EXAMPLE 2

Model 2:

Covariate	Coeff	StErr	p-value	RRisk
Rx	1.294	0.422	0.002	3.648
log WBC	1.604	0.329	0.000	4.975

$\mathbf{X}^*$ = (1, log WBC), $\mathbf{X}$ = (0, log WBC)

HR for effect of *Rx* adjusted for log WBC:

$$\widehat{HR} = \exp[\hat{\beta}_1(X_1^* - X_1) + \hat{\beta}_2(X_2^* - X_2)]$$

$$= \exp[1.294(1-0) + 1.604(\log \text{WBC} - \log \text{WBC})]$$

$$= \exp[1.294(1) + 1.604(0)] = 1.294$$

As a second example, consider the output for model 2, which contains two variables, the *Rx* variable and log WBC. Then to obtain the hazard ratio for the effect of the *Rx* variable adjusted for the log WBC variable, we let the vectors $\mathbf{X}^*$ and $\mathbf{X}$ be defined as $\mathbf{X}^*$ = (1, log WBC) and $\mathbf{X}$ = (0, log WBC). Here we assume that log WBC is fixed, though unspecified.

The estimated hazard ratio is then obtained by exponentiating the sum of two quantities, one involving the coefficient 1.294 of the *Rx* variable, and the other involving the coefficient 1.604 of the log WBC variable. Since the log WBC value is fixed, however, this portion of the exponential is zero, so that the resulting estimate is simply e to the 1.294.

General rule: If X_1 is a (0,1) exposure variable, then

$\widehat{\text{HR}} = e^{\hat{\beta}_1}$ (= effect of exposure adjusted for other X's)

provided no other X's are product terms involving exposure.

This second example illustrates the general rule that the hazard ratio for the effect of a (0,1) exposure variable which adjusts for other variables is obtained by exponentiating the estimated coefficient of the exposure variable. This rule has the proviso that the model does not contain any product terms involving exposure.

EXAMPLE 3

Model 3:

Covariate	Coeff	StErr	p-value	RRisk
Rx	2.355	1.681	0.161	10.537
log WBC	1.803	0.447	0.000	6.067
Rx × log WBC	−0.342	0.520	0.510	0.710

We now give a third example which illustrates how to compute a hazard ratio when the model does contain product terms. We consider the printout for model 3 of the remission data shown here.