SURVIVAL ANALYSIS

Chapter 3. The Cox Proportional Hazards Model and Its Characteristics

Instructor: Seokho Lee

2013Fall at HUFS

A Computer Example Using the Cox PH Model

EXAMPLE

Leukemia Remission Data

Group 1($n = 21$)		Group 2($n = 21$)	
t(weeks)	log WBC	t(weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

+ denotes censored observation

- We introduce the Cox PH model using computer output from the analysis of remission time data (Freireich et al., Blood, 1963), which we previously discussed in Chapters 1 and 2. The data set is listed here at the left.
- These data involve two groups of leukemia patients, with 21 patients in each group. Group 1 is the treatment group, and group 2 is the placebo group. The data set also contains the variable log WBC, which is a well-known prognostic indicator of survival for leukemia patients.
- For this example, the basic question of interest concerns comparing the survival experience of the two groups adjusting for the possible confounding and/or interaction effects of log WBC.

A Computer Example Using the Cox PH Model

EXAMPLE (continued)

T = weeks until going out of remission
X_1 = group status = E
X_2 = log WBC (confounding?)

Interaction?
$X_3 = X_1 \times X_2$ = group status × log WBC

Computer results for three Cox PH models using the R package

Other computer packages provide similar information.

- We are thus considering a problem involving two explanatory variables as predictors of survival time T, where T denotes "weeks until going out of remission." We label the explanatory variables X_1 (for group status) and X_2 (for log WBC). The variable X_1 is the primary study or exposure variable of interest. The variable X_2 is an extraneous variable that we are including as a possible confounder or effect modifier.

- Note that if we want to evaluate the possible interaction effect of log WBC on group status, we would also need to consider a third variable, that is, the product of X_1 and X_2.

- For this dataset, the computer results from fitting three different Cox proportional hazards models are presented below. The computer package used is R. This is one of several packages that have procedures for carrying out a survival analysis using the Cox model. The information printed out by different packages will not have exactly the same format, but they will provide similar information.

A Computer Example Using the Cox PH Model

```
library(survival) # loading survival package

# survival time
t<-c(6,6,6,7,10,13,16,22,23,6,9,10,11,17,19,20,25,32,32,34,35,1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)

# censorship index
delta<-c(1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)

# independent variable: group information
x1<-rep(c(0,1),c(21,21))

# covariate: log(WBC)
x2<-c(2.31,4.06,3.28,4.43,2.96,2.88,3.60,2.32,2.57,3.20,2.80,2.70,2.60,2.16,2.05,2.01,
1.78,2.20,2.53,1.47,1.45,2.80,5.00,4.91,4.48,4.01,4.36,2.42,3.49,3.97,3.52,3.05,2.32,
3.26,3.49,2.12,1.50,3.06,2.30,2.95,2.73,1.97)

# create 'survival' object using the function Surv()
remission<-Surv(t,delta)

model1<-coxph(remission~x1, method='breslow')
model2<-coxph(remission~x1+x2, method='breslow')
model3<-coxph(remission~x1*x2, method='breslow')

summary(model1)
summary(model2)
summary(model3)
```

A Computer Example Using the Cox PH Model

```
> summary(model1)
Call:
coxph(formula = remission ~ x1, method = "breslow")

  n= 42, number of events= 30

     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.5092    4.5231   0.4096 3.685 0.000229 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

   exp(coef) exp(-coef) lower .95 upper .95
x1     4.523     0.2211     2.027     10.09

Rsquare= 0.304   (max possible= 0.989 )
Likelihood ratio test= 15.21  on 1 df,   p=9.615e-05
Wald test            = 13.58  on 1 df,   p=0.0002288
Score (logrank) test = 15.93  on 1 df,   p=6.571e-05



> summary(model2)
Call:
coxph(formula = remission ~ x1 + x2, method = "breslow")

  n= 42, number of events= 30

     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.2941    3.6476   0.4221 3.066  0.00217 **
x2 1.6043    4.9746   0.3293 4.872 1.11e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
   exp(coef) exp(-coef) lower .95 upper .95
x1     3.648     0.2742     1.595     8.343
x2     4.975     0.2010     2.609     9.486

Rsquare= 0.644   (max possible= 0.989 )
Likelihood ratio test= 43.41  on 2 df,   p=3.744e-10
Wald test            = 31.78  on 2 df,   p=1.254e-07
Score (logrank) test = 42.94  on 2 df,   p=4.743e-10


> summary(model3)
Call:
coxph(formula = remission ~ x1 * x2, method = "breslow")

  n= 42, number of events= 30

        coef exp(coef) se(coef)      z Pr(>|z|)
x1     2.3549   10.5375   1.6810  1.401    0.161
x2     1.8028    6.0665   0.4467  4.036 5.45e-05 ***
x1:x2 -0.3422    0.7102   0.5197 -0.658    0.510
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

      exp(coef) exp(-coef) lower .95 upper .95
x1      10.5375     0.0949    0.3907   284.201
x2       6.0665     0.1648    2.5275    14.561
x1:x2    0.7102     1.4080    0.2564     1.967

Rsquare= 0.648   (max possible= 0.989 )
Likelihood ratio test= 43.84  on 3 df,   p=1.633e-09
Wald test            = 30.6  on 3 df,   p=1.030e-06
Score (logrank) test = 45.9  on 3 df,   p=5.95e-10
```

A Computer Example Using the Cox PH Model

EDITED OUTPUT FROM R

Model 1:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.5092    4.5231   0.4096 3.685 0.000229 ***
```

Model 2:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.2941    3.6476   0.4221 3.066  0.00217 **
x2 1.6043    4.9746   0.3293 4.872 1.11e-06 ***
```

Model 3:

```
         coef exp(coef) se(coef)      z Pr(>|z|)
x1     2.3549   10.5375   1.6810  1.401    0.161
x2     1.8028    6.0665   0.4467  4.036 5.45e-05 ***
x1:x2 -0.3422    0.7102   0.5197 -0.658    0.510
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- We now describe how to use the computer printout to evaluate the possible effect of treatment status on remission time adjusted for the potential confounding and interaction effects of the covariate log WBC. For now, we focus only on six columns of information provided in the printout, as presented at the left for all three models.
- For each model, the first column identifies the **variables** that have been included in the model. The second column gives estimates of **regression coefficients** corresponding to each variable in the model. The fourth column gives standard errors of the estimated regression coefficients. The fifth and sixth columns gives **test statistics** and **p-values** for testing the significance of each coefficient. The third column gives exponents of coefficient, which are **hazard ratios** for the effect of each variable adjusted for the other variables in the model.
- Except for the hazard ratio column, these computer results are typical of output found in standard linear regression printouts. As the printout suggests, we can analyze the results from a Cox model in a manner similar to the way we would analyze a linear regression model.

A Computer Example Using the Cox PH Model

EXAMPLE (continued)

Same dataset for each model

$n = 42$ subjects

T = time (weeks) until out of remission

Model 1: *Rx* only

Model 2: *Rx* and logWBC

Model 3: *Rx*, logWBC, and *Rx*×logWBC

- We now distinguish among the output for the three models shown here. All three models are using the same remission time data on 42 subjects. The outcome variable for each model is the same—time in weeks until a subject goes out of remission. However, the independent variables are different for each model. Model 1 contains only the treatment status variable, indicating whether a subject is in the treatment or placebo group. Model 2 contains two variables—treatment status and log WBC. And model 3 contains an interaction term defined as the product of treatment status and log WBC.

A Computer Example Using the Cox PH Model

EDITED OUTPUT

```
         coef exp(coef) se(coef)      z Pr(>|z|)
x1     2.3549   10.5375   1.6810  1.401    0.161
x2     1.8028    6.0665   0.4467  4.036 5.45e-05 ***
x1:x2 -0.3422    0.7102   0.5197 -0.658    0.510

Rsquare= 0.648   (max possible= 0.989 )
Likelihood ratio test= 43.84  on 3 df,   p=1.633e-09
Wald test            = 30.6  on 3 df,   p=1.030e-06
Score (logrank) test = 45.9  on 3 df,   p=5.95e-10
```

EXAMPLE (continued)

LR (interaction in model 3)

$= -2 \ln L_{\text{model with no explanatory variables}} - (-2 \ln L_{\text{model 3}})$

In general:

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

- We now focus on the output for model 3. The method of estimation used to obtain the coefficients for this model, as well as the other two models, is maximum likelihood (ML) estimation. Note that a p-value of 0.510 is obtained for the coefficient of the product term for the interaction of treatment with log WBC. This p-value indicates that there is no significant interaction effect, so that we can drop the product term from the model and consider the other two models instead.
- The p-value of 0.510 that we have just described is obtained by dividing the coefficient –0.3422 of the product term by its standard error of 0.5197, which gives –0.658, and then assuming that this quantity is approximately a standard normal or Z variable. This Z statistic is known as a **Wald statistic**, which is one of two test statistics typically used with ML estimates. The other test statistic, called the **likelihood ratio**, or LR statistic, makes use of the log likelihood statistic. The log likelihood statistic is obtained by multiplying the “Log likelihood” by –2 to get –2 ln L.
- Likelihood ratio test, shown in the output, is the difference between the log likelihood statistic of the reduced model which does not contain 3 explanatory terms and the log likelihood statistic of the full model containing 3 explanatory terms.

A Computer Example Using the Cox PH Model

EXAMPLE (continued)

LR (interaction in model 3)
= -2 ln $L_{model\ 2}$ - (-2 ln $L_{model\ 3}$)
= LR_{model3} - LR_{model2}
= 43.84 - 43.41 = 0.43

(LR is χ^2 with 1 d.f. under H_0: no interaction)
P-value is 0.520, **not significant**
Wald test P-value = 0.510

- To obtain the LR statistic in this example, we compute 43.84 minus 43.41 to obtain 0.43. Under the null hypothesis of no interaction effect, the test statistic has a chi-square distribution with p degrees of freedom, where p denotes the number of predictors being assessed. The p-value for this test is 0.520, which indicates no significant interaction. Although the p-values for the Wald test (0.510) and the LR test are not exactly the same, both p-values lead to the same conclusion.

LR $\neq$ Wald

When in doubt, use the LR test.

- In general, the LR and Wald statistics may not give exactly the same answer. Statisticians have shown that of the two test procedures, the LR statistic has better statistical properties, so when in doubt, you should use the LR test.

A Computer Example Using the Cox PH Model

OUTPUT

Model 2:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.2941    3.6476   0.4221 3.066  0.00217 **
x2 1.6043    4.9746   0.3293 4.872 1.11e-06 ***
```

Three statistical objectives.

1. **test for significance of effect**
2. **point estimate of effect**
3. **confidence interval for effect**

EXAMPLE (continued)

Test for treatment effect:

Wald statistic: P=0.00217 (highly significant)

LR statistic: compare

LR statistic from model 2 with

LR statistic from model without Rx (x1)

Conclusion: treatment effect is significant, after adjusting for log WBC

- We now focus on how to assess the effect of treatment status adjusting for log WBC using the model 2 output, again shown here.

- There are three statistical objectives typically considered. One is to **test for the significance** of the treatment status variable, adjusted for log WBC. Another is to obtain a **point estimate of the effect** of treatment status, adjusted for log WBC. And a third is to obtain a **confidence interval for this effect**. We can accomplish these three objectives using the output provided, without having to explicitly describe the formula for the Cox model being used.

- To test for the significance of the treatment effect, the p-value provided in the table for the Wald statistic is 0.00217, which is highly significant. Alternatively, a likelihood ratio (LR) test could be performed by comparing the log likelihood statistic (43.41) for model 2 with the log likelihood statistic for a model which does not contain the treatment variable. This latter model, which should contain only the log WBC variable, is not provided here, so we will not report on it other than to note that the LR test is also very significant. Thus, these test results show that using model 2, the treatment effect is significant, after adjusting for log WBC.

A Computer Example Using the Cox PH Model

EXAMPLE (continued)

Point estimate:

$$\widehat{HR} = 3.648 = e^{1.294}$$

Coefficient of treatment variable

- A point estimate of the effect of the treatment is provided in the exp(coef) column by the value 3.648. This value gives the estimated hazard ratio (HR) for the effect of the treatment; in particular, we see that the hazard for the placebo group is 3.648 times the hazard for the treatment group. Note that the value 3.648 is calculated as e to the coefficient of the treatment variable; that is, e to the 1.294 equals 3.648.
- To describe the confidence interval for the effect of treatment status, we consider the output for the extended table for model 2 given earlier.

OUTPUT

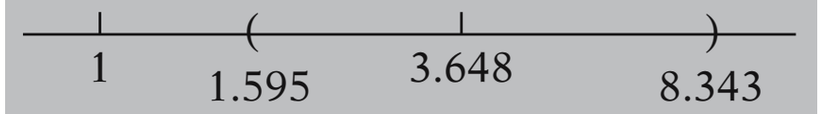
Model 2:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.2941    3.6476   0.4221 3.066  0.00217 **
x2 1.6043    4.9746   0.3293 4.872 1.11e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

   exp(coef) exp(-coef) lower .95 upper .95
x1     3.648     0.2742     1.595     8.343
x2     4.975     0.2010     2.609     9.486

Rsquare= 0.644   (max possible= 0.989 )
Likelihood ratio test= 43.41  on 2 df,   p=3.744e-10
Wald test            = 31.78  on 2 df,   p=1.254e-07
Score (logrank) test = 42.94  on 2 df,   p=4.743e-10
```

A Computer Example Using the Cox PH Model

EXAMPLE (continued)

95% confidence interval for the *HR*: (1.595, 8.343)



95% CI for β_1: $1.294 \pm (1.96)\ (0.422)$

95% CI for $HR = e^{\beta_1}$:

$$\exp[\hat{\beta}_1 \pm 1.96 s_{\hat{\beta}_1}] = e^{1.294 \pm 1.96(0.422)}$$

- From the table, we see that a 95% confidence interval for the treatment effect is given by the range of values 1.595 ~ 8.343. This is a confidence interval for the hazard ratio (HR), which surrounds the point estimate of 3.648 previously described. Notice that this confidence interval is fairly wide, indicating that the point estimate is somewhat unreliable. As expected from the low p-value of 0.002, the confidence interval for HR does not contain the null value of 1.

- The calculation of the confidence interval for HR is carried out as follows:
  1. Compute a 95% confidence interval for the regression coefficient of the Rx(x1) variable (β_1). The large sample formula is 1.294 plus or minus 1.96 times the standard error 0.422, where 1.96 is the 97.5 percentile of the standard normal or Z distribution.
  2. Exponentiate the two limits obtained for the confidence interval for the regression coefficient of Rx.

A Computer Example Using the Cox PH Model

EDITED OUTPUT

Model 1:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.5092    4.5231   0.4096 3.685 0.000229 ***
```

Model 2:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.2941    3.6476   0.4221 3.066  0.00217 **
x2 1.6043    4.9746   0.3293 4.872 1.11e-06 ***
```

EXAMPLE (continued)

HR for model 1 (4.523) is higher than *HR* for model 2 (3.648).

Confounding: crude versus adjusted $\widehat{HR}$*'s* are meaningfully different.

Confounding due to log WBC
⇒ must control for log WBC, i.e., prefer model 2 to model 1.

If no confounding, then consider precision: e.g., if 95% CI is narrower for model 2 than model 1, we prefer model 2.

- To this point, we have made use of information from outputs for models 2 and 3, but have not yet considered the model 1 output. Note that model 1 contains only the treatment status variable, whereas model 2 contains log WBC in addition to treatment status. Model 1 is sometimes called the "crude" model because it ignores the effect of potential covariates of interest, like log WBC.
- Model 1 can be used in comparison with model 2 to evaluate the potential confounding effect of the variable log WBC. In particular, notice that the value in the HR column for the treatment status variable is 4.5231 for model 1, but only 3.6476 for model 2. Thus, the crude model yields an estimated hazard ratio that is somewhat higher than the corresponding estimate obtained when we adjust for log WBC. If we decide that the crude and adjusted estimates are meaningfully different, we then say that there is confounding due to log WBC.
- Once we decide that confounding is present, we then **must** control for the confounder—in this case, log WBC—in order to obtain a valid estimate of the effect. Thus, we prefer model 2, which controls for log WBC, to model 1, which does not.
- Note that if we had decided that there is no "meaningful" confounding, then we would not need to control for log WBC to get a valid answer. Nevertheless, we might wish to control for log WBC anyhow, to obtain a more precise estimate of the hazard ratio. That is, if the confidence interval for the HR is narrower when using model 2 than when using model 1, we would prefer model 2 to model 1 for **precision** gain.

A Computer Example Using the Cox PH Model

EDITED OUTPUT: Confidence Intervals

Model 1:

	exp(coef)	exp(-coef)	lower .95	upper .95
x1	4.523	0.2211	2.027	10.09

width=8.067

Model 2:

	exp(coef)	exp(-coef)	lower .95	upper .95
x1	3.648	0.2742	1.595	8.343
x2	4.975	0.2010	2.609	9.486

width=6.748

- The confidence intervals for x1 in each model are shown here at the left. The interval for x1 in model 1 has width equal to 10.09 minus 2.027, or **8.067**; for model 2, the width is 8.343 minus 1.595, or **6.748**. Therefore, model 2 gives a more precise estimate of the hazard ratio than does model 1.

EXAMPLE (continued)

Model 2 is best model.

$\widehat{HR} = 3.648$ statistically significant

95% CI for *HR*: (1.6, 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 3.6476 for the effect of the treatment, with a 95% confidence interval ranging between 1.6 and 8.3.

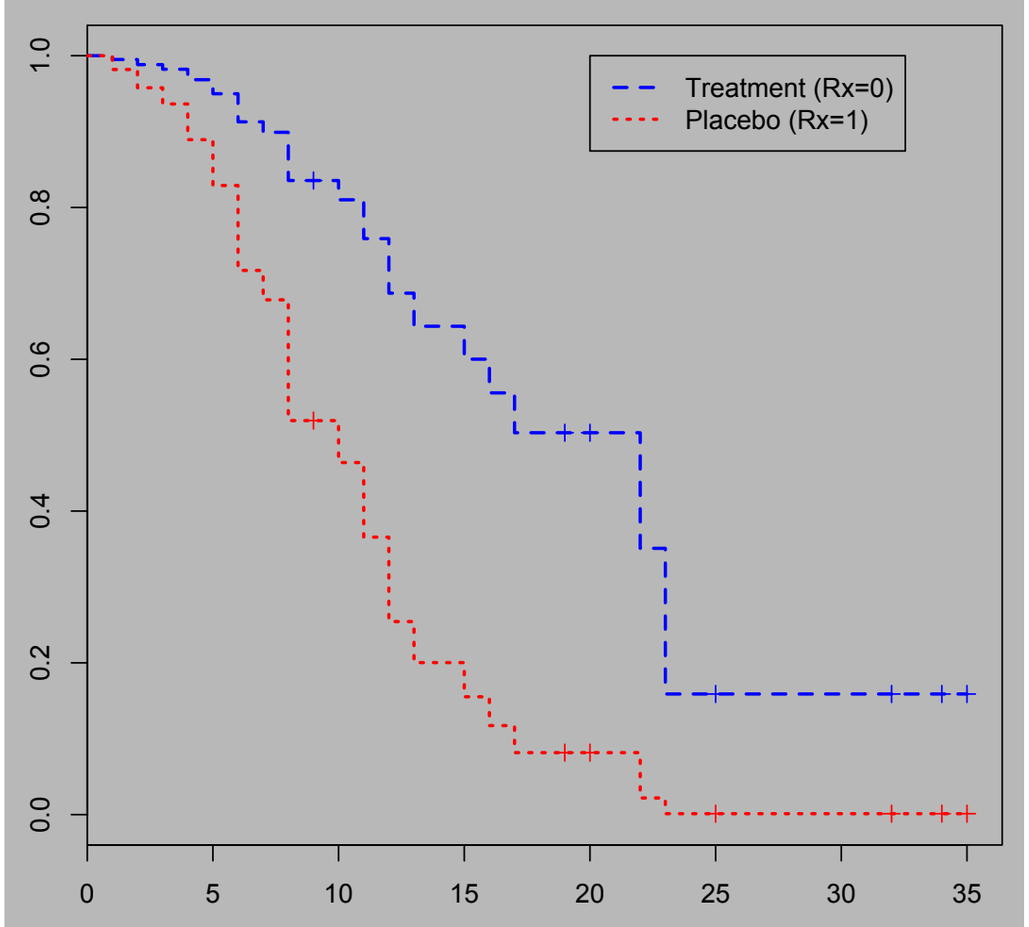
Cox model formulae not specified

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

- 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 and Klein, *Logistic Regression*, Chapters 6 and 7, 2002), and very similar to carrying out a classical linear regression analysis (see Kleinbaum et al., *Applied Regression Analysis*, 3rd ed., Chapter 16, 1997).

A Computer Example Using the Cox PH Model

EXAMPLE (continued)




```
data.x1<-data.frame(x1=c(0,1),x2=rep(mean(x2),2))
plot(survfit(model2,newdata=data.x1), lty=c(2,3),
col=c('blue','red'), lwd=2)
legend(20, 1, c('Treatment (Rx=0)','Placebo (Rx=1)'),
lty=c(2,3), col=c('blue','red'), lwd=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.

A Computer Example Using the Cox PH Model

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.

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

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 the meaning of the PH assumption and the Cox likelihood.

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

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

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

- The Cox 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 of predictor variables that is being modeled to predict an individual's hazard.
- 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.

The Formula for the Cox PH Model

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 value for a given individual does 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.

The Formula for the Cox PH Model

$$X_1 = X_2 = \cdots = X_p = 0$$

$$\begin{aligned} h(t,\mathbf{X}) &= h_0(t)\, e^{\sum_{i=1}^{p} \beta_i X_i} \\ &= h_0(t)\, e^{0} \\ &= h_0(t) \end{aligned}$$

Baseline hazard

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

$h_0(t)$ is unspecified.

Cox model: **semiparametric**

- 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.

- 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.

- 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 **semiparametric** model.

The Formula for the Cox PH Model

EXAMPLE: Parametric Model

Weibull:

$$h(t, \mathbf{X}) = \lambda p t^{p-1}$$

where $\lambda = \exp[\sum_{i=1}^{p} \beta_i X_i]$

and $h_0(t) = p t^{p-1}$

Semiparametric property

$\Downarrow$

Popularity of the Cox model

- 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, p, and the β_i 's. Note that for the Weibull model, $h_0(t)$ is given by pt^{p-1} (see Chapter 7).
- One of the reasons why the Cox model is so popular is that it is semiparametric. We discuss this and other reasons in the next section concerning why the Cox model is so widely used.

Why the Cox PH Model Is Popular?

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

If correct model is:

Weibull $\Rightarrow$	Cox model will approximate Weibull
Exponential $\Rightarrow$	Cox model will approximate exponential

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

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

- 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.
- 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.
- 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.
- 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.

Why the Cox PH Model Is Popular?

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

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

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

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 $h_0(t)$.

- 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 non-negative.
- 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.
- 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.

Why the Cox PH Model Is Popular?

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 $S(t,\mathbf{X})$ 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.

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: $\hat{\beta}_i$

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
x1	1.2941	3.6476	0.4221	3.066	0.00217	**
x2	1.6043	4.9746	0.3293	4.872	1.11e-06	***

Estimated model:

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

ML estimates: maximize likelihood function L

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

- 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 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 eukemia patients.
- The Cox model for this example involves two parameters, one being the coefficient of the treatment variable (denoted here as x1 or Rx) and the other being the coefficient of the x2 variable (log WBC). The expression for this model is shown at the left, which contains the estimated coefficients 1.294 for x1 and 1.604 for log white blood cell count.
- 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.
- The expression for the likelihood is developed at the end of the chapter. However, we give a brief overview below.

ML Estimation of the Cox PH Model

L is a partial likelihood:

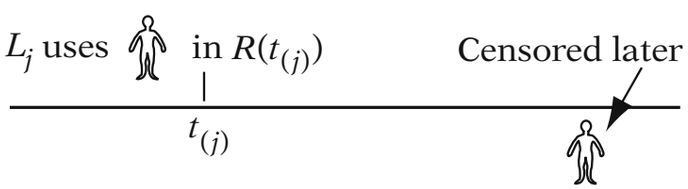
- considers probabilities only for subjects who fail
- does not consider probabilities for subjects who are censored

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 $R(t_{(j)})$

Information on censored subjects used prior to censorship.



- 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.

- 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.

- 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.

ML Estimation of the Cox PH Model

Steps for obtaining ML estimates:

- form L from model
- maximize $\ln L$ by solving

$$\frac{\partial \ln L}{\partial \beta_i} = 0$$

$$i = 1, \ldots, p(\#\text{ of parameters})$$

Solution by iteration:

- guess at solution
- modify guess in successive steps
- stop when solution is obtained

Statistical inferences for hazard ratios:

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 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.
- The maximization process is carried out by taking partial derivatives of log 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.
- 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 the earlier section. 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.

Computing the Hazard Ratio

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

where

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

and

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

denote the set of X's for two individuals

To interpret $\widehat{HR}$, want $\widehat{HR} \geq 1$, i.e.,

$\hat{h}(t,\mathbf{X}^*) \geq \hat{h}(t,\mathbf{X})$.

Typical coding: $\mathbf{X}^*$: group with larger h

$\mathbf{X}$: group with smaller h

EXAMPLE: Remission Data

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

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

- 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.

- 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 group with the larger hazard 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 $X_1 = 0$.

Computing the Hazard Ratio

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

- 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.

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

- 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 .

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

- 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.

Computing the Hazard Ratio

EXAMPLE (continued)

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

$X_1^* = 1, \; X_1 = 0$

$$\widehat{HR} = \exp[\hat{\beta}_1(X_1^* - X_1)]$$
$$= \exp[\hat{\beta}_1(1-0)] = e^{\hat{\beta}_1}$$

Model 1:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.5092    4.5231   0.4096 3.685 0.000229 ***
```

- 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 "hat" times 1 minus 0, which simplifies to e to the β_1 "hat."

- Recall the remission data printout for model 1, which contains only the x1 variable, again shown here. Then the estimated hazard ratio is obtained by exponentiating the coefficient 1.5092, which gives the value 4.5231 shown in the exp(coef) column of the output.

Computing the Hazard Ratio

EXAMPLE 2

Model 2:

```
     coef exp(coef) se(coef)     z Pr(>|z|)
x1 1.2941    3.6476   0.4221 3.066  0.00217 **
x2 1.6043    4.9746   0.3293 4.872 1.11e-06 ***
```

$\mathbf{X}^* = (1, \text{logWBC}),\ \mathbf{X} = (0, \text{logWBC})$

HR for effect of x1 adjusted for x2 (log WBC):

- As a second example, consider the output for model 2, which contains two variables, the x1 variable and x2. Then to obtain the hazard ratio for the effect of the x1 variable adjusted for the x2 variable, we let the vectors X^* and X be defined as X^*=(1, logWBC) and X=(0, logWBC). Here we assume that log WBC is the same for X^* and X though unspecified.

EXAMPLE 2 (continued)

$$\widehat{HR} = \exp[\hat{\beta}_1(X_1^* - X_1) + \hat{\beta}_1 (X_2^* - X_2)]$$
$$= \exp[1.294(1-0)$$
$$+ 1.604 (\log \text{WBC} - \log \text{WBC})]$$
$$= \exp[1.294(1) + 1.604(0)] = e^{1.294}$$

- The estimated hazard ratio is then obtained by exponentiating the sum of two quantities, one involving the coefficient 1.294 of the x1 variable, and the other involving the coefficient 1.604 of the x2 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.

Computing the Hazard Ratio

EXAMPLE 3

Model 3:

```
          coef exp(coef) se(coef)      z Pr(>|z|)
x1      2.3549   10.5375   1.6810  1.401    0.161
x2      1.8028    6.0665   0.4467  4.036 5.45e-05 ***
x1:x2 -0.3422    0.7102   0.5197 -0.658    0.510
```

EXAMPLE 3

Want *HR* for effect of *Rx* adjusted for log WBC.

Placebo subject:

$$\boldsymbol{X}^* = (X_1^* = 1,\ X_2^* = \log \text{WBC},\ X_3^* = 1 \times \log \text{WBC})$$

Treated subject:

$$\boldsymbol{X} = (X_1 = 0,\ X_2 = \log \text{WBC},\ X_3 = 0 \times \log \text{WBC})$$

$$\widehat{HR} = \exp\left[\sum_{i=1}^{3} \hat{\beta}_i (X_i^* - X_i)\right]$$

$$\begin{aligned}\widehat{HR} = \exp[&2.355(1-0) \\ &+ 1.803\ (\log \text{WBC} - \log \text{WBC}) \\ &+ (-0.342)(1 \times \log \text{WBC} \\ &- 0 \times \log \text{WBC})]\end{aligned}$$

$$= \exp[2.355 - 0.342 \log \text{WBC}]$$

- 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.

- To obtain the hazard ratio for the effect of x1 adjusted for x2 using model 3, we consider X^* and X vectors which have three components, one for each variable in the model. The $\mathbf{X}^*$ vector, which denotes a placebo subject, has components X_1^*=1, X_2^*= log WBC and X_3^*=1 times log WBC. The $\mathbf{X}$ vector, which denotes a treated subject, has components X_1=0, X_2 = log WBC and X_3 = 0 times log WBC. Note again that, as with the previous example, the value for log WBC is treated as fixed, though unspecified.

- Using the general formula for the hazard ratio, we must now compute the exponential of the sum of three quantities, corresponding to the three variables in the model. Substituting the values from the printout and the values of the vectors $\mathbf{X}^*$ and $\mathbf{X}$ into this formula, we obtain the exponential expression shown here. Using algebra, this expression simplifies to the exponential of 2.355 minus 0.342 times log WBC.

Computing the Hazard Ratio

EXAMPLE 3 (continued)

log WBC = 2:

$$\widehat{HR} = \exp[2.355 - 0.342\ (2)]$$
$$= e^{1.671} = 5.32$$

log WBC = 4:

$$\widehat{HR} = \exp[2.355 - 0.342\ (4)]$$
$$= e^{0.987} = 2.68$$

General rule for (0,1) exposure variables when there are product terms:

$$\widehat{HR} = \exp\left[\hat{\beta} + \sum \hat{\delta}_j W_j\right]$$

where

$\hat{\beta}$ = coefficient of E

$\hat{\delta}_j$ = coefficient of $E \times W_j$

($\widehat{HR}$ does not contain coefficients of non-product terms)

EXAMPLE

Model 3:

E → $\hat{\beta}$ = coefficient of Rx

W_1 → $\hat{\delta}_1$ = coefficient of $Rx \times$ log WBC

$\widehat{HR}$ (model 3) = $\exp[\hat{\beta} + \hat{\delta}_1 \log \text{WBC}]$
= $\exp[2.355 - 0.342 \log \text{WBC}]$

- In order to get a numerical value for the hazard ratio, we must specify a value for log WBC. For instance, if log WBC = 2, the estimated hazard ratio becomes 5.32, whereas if log WBC = 4, the estimated hazard ratio becomes 2.68. Thus, we get different hazard ratio values for different values of log WBC, which should make sense since log WBC is an effect modifier in model 3.
- The example we have just described using model 3 illustrates a general rule which states that the hazard ratio for the effect of a (0,1) exposure variable in a model which contains product terms involving this exposure with other X's can be written as shown here. Note that β "hat" denotes the coefficient of the exposure variable and the δ "hats" are coefficients of product terms in the model of the form $E \times W_j$. Also note that this formula does not contain coefficients of nonproduct terms other than those involving E.
- For model 3, β "hat" is the coefficient of the Rx variable (treatment variable), and there is only one δ "hat" in the sum, which is the coefficient of the product term Rx × log WBC. Thus, there is only one W, namely W_1 = log WBC. The hazard ratio formula for the effect of exposure is then given by exponentiating β "hat" plus δ_1 "hat" times log WBC. Substituting the estimates from the printout into this formula yields the expression obtained previously, namely the exponential of 2.355 minus 0.342 times log WBC.

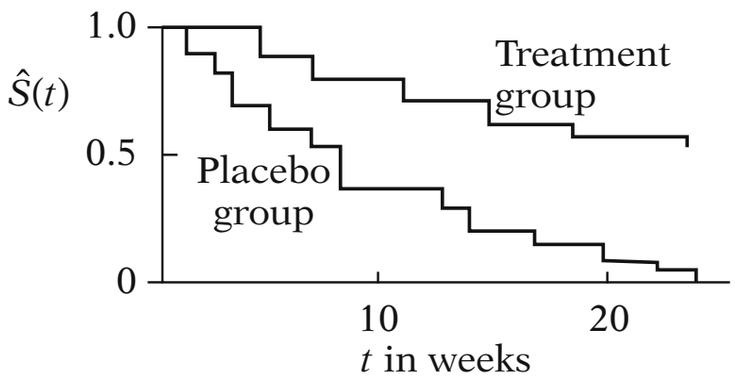
Adjusted Survival Curves Using the Cox PH Model

Two primary quantities:

1. estimated hazard ratios
2. estimated survival curves

- The two primary quantities desired from a survival analysis point of view are estimated hazard ratios and estimated survival curves. Having just described how to compute hazard ratios, we now turn to estimation of survival curves using the Cox model.

No model: use KM curves



- Recall that if no model is used to fit survival data, a survival curve can be estimated using a Kaplan–Meier method. Such KM curves are plotted as step functions as shown here for the remission data example.

Cox model: adjusted survival curves (also step functions).

- When a Cox model is used to fit survival data, survival curves can be obtained that adjust for the explanatory variables used as predictors. These are called **adjusted survival curves**, and, like KM curves, these are also plotted as step functions.

Adjusted Survival Curves Using the Cox PH Model

Cox model hazard function:

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

Cox model survival function:

$$S(t,\mathbf{X}) = \left[S_0(t)\right]^{e^{\sum_{i=1}^{p} \beta_i X_i}}$$

- The hazard function formula for the Cox PH model, shown here again, can be converted to a corresponding survival function formula as shown below. This survival function formula is the basis for determining adjusted survival curves. Note that this formula says that the survival function at time t for a subject with vector $\mathbf{X}$ as predictors is given by a baseline survival function $S(t)$ raised 0 to a power equal to the exponential of the sum of β_i times X_i.

Estimated survival function:

$$\hat{S}(t,\mathbf{X}) = \left[\hat{S}_0(t)\right]^{e^{\sum_{i=1}^{p} \hat{\beta}_i X_i}}$$

$\hat{S}_0(t)$ and $\hat{\beta}_i$ are provided by the computer program. The X_i must be specified by the investigator.

- The expression for the estimated survival function can then be written with the usual “hat” notation as shown here.

- The estimates of $\hat{S}_0(t)$ and $\hat{\beta}_i$ are provided by the computer program that fits the Cox model. The X’s, however, must first be specified by the investigator before the computer program can compute the estimated survival curve.

Adjusted Survival Curves Using the Cox PH Model

EXAMPLE: Model 2 Remission Data

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

$$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{\exp(1.294\,Rx\, + 1.604\,\log \text{WBC})}$$

Specify values for $\mathbf{X} = (Rx,\ \log \text{WBC})$:

$Rx = 1$, log WBC = 2.93:

$$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{\exp(1.294\,(1)\, + 1.604\,(2.93))}$$

$$= [\hat{S}_0(t)]^{\exp(5.99)} = \boxed{[\hat{S}_0(t)]^{400.9}}$$

$Rx = 0$, log WBC = 2.93:

$$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{\exp(1.294\,(0)\, + 1.604\,(2.93))}$$

$$= [\hat{S}_0(t)]^{\exp(4.70)} = \boxed{[\hat{S}_0(t)]^{109.9}}$$

Adjusted Survival Curves

$Rx = 1$, log WBC = 2.93:
$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{400.9}$

$Rx = 0$, log WBC = 2.93:
$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{109.9}$

- For example, if we consider model 2 for the remission data, we can obtain a specific survival curve by specifying values for the vector $\mathbf{X}$, whose component variables are Rx and log WBC.

- For instance, if Rx = 1 and log WBC = 2.93, the estimated survival curve is obtained by substituting these values in the formula as shown here, and carrying out the algebra to obtain the expression circled. Note that the value 2.93 is the overall mean log WBC for the entire dataset of 42 subjects.

- Also, if Rx=0 and log WBC = 2.93, the estimated survival curve is obtained as shown here.

- Each of the circled expressions gives **adjusted** survival curves, where the adjustment is for the values specified for the X's. Note that for each expression, a survival probability can be obtained for any value of t.

- The two formulae just obtained, again shown here, allow us to compare survival curves for different treatment groups adjusted for the covariate log WBC. Both curves describe estimated survival probabilities over time assuming the same value of log WBC, in this case, the value 2.93.

Adjusted Survival Curves Using the Cox PH Model

Typically, use $X = \overline{X}$ or X_{median}.

Computer uses $\bar{X}$.

EXAMPLE (continued)

Remission data ($n = 42$):

$\overline{\log \text{WBC}} = 2.93$

General formulae for adjusted survival curves comparing two groups:

Exposed subjects:

$$\hat{S}(t, \mathbf{X}_1) = [\hat{S}_0(t)]^{\exp[\hat{\beta}_1(1) + \sum_{i \neq 1} \hat{\beta}_i \overline{X}_i]}$$

Unexposed subjects:

$$\hat{S}(t, \mathbf{X}_0) = [\hat{S}_0(t)]^{\exp[\hat{\beta}_1(0) + \sum_{i \neq 1} \hat{\beta}_i \overline{X}_i]}$$

General formula for adjusted survival curve for all covariates in the model:

$$\hat{S}(t, \overline{\mathbf{X}}) = [\hat{S}_0(t)]^{\exp[\sum \hat{\beta}_i \overline{X}_i]}$$

- Typically, when computing adjusted survival curves, the value chosen for a covariate being adjusted is an average value like an arithmetic mean or a median. In fact, most computer programs for the Cox model automatically use the mean value over all subjects for each covariate being adjusted.

- In our example, the mean log WBC for all 42 subjects in the remission data set is 2.93. That is why we chose this value for log WBC in the formulae for the adjusted survival curve.

- More generally, if we want to compare survival curves for two levels of an exposure variable, and we want to adjust for several covariates, we can write the formula for each curve as shown here. Note that we are assuming that the exposure variable is variable X_1, whose estimated coefficient is β_1 "hat," and the value of X_1 is 1 for exposed and 0 for unexposed subjects.

- Also, if we want to obtain an adjusted survival curve which adjusts for all covariates in the model, the general formula which uses the mean value for each covariate is given as shown here. This formula will give a single adjusted survival curve rather than different curves for each exposure group.

Adjusted Survival Curves Using the Cox PH Model

EXAMPLE (continued)

Single survival curve for Cox model containing Rx and log WBC:

$\overline{Rx} = 0.50$

$\overline{\log \text{WBC}} = 2.93$

$$\hat{S}(t,\overline{\mathbf{X}}) = [\hat{S}_0(t)]^{\exp(\hat{\beta}_1\overline{Rx} + \hat{\beta}_2\overline{\log \text{WBC}})}$$

$$= [\hat{S}_0(t)]^{\exp(1.294(0.5) + 1.604(2.93))}$$

$$= [\hat{S}_0(t)]^{\exp(5.35)} = [\hat{S}_0(t)]^{210.6}$$

Compute survival probability by specifying value for t in $\hat{S}(t,\overline{\mathbf{X}}) = [\hat{S}_0(t)]^{210.6}$

Computer uses t's which are failure times.

- To illustrate this formula, suppose we again consider the remission data, and we wish to obtain a single survival curve that adjusts for both Rx and log WBC in the fitted Cox model containing these two variables. Using the mean value of each covariate, we find that the mean value for Rx is 0.5 and the mean value for log WBC is 2.93, as before.

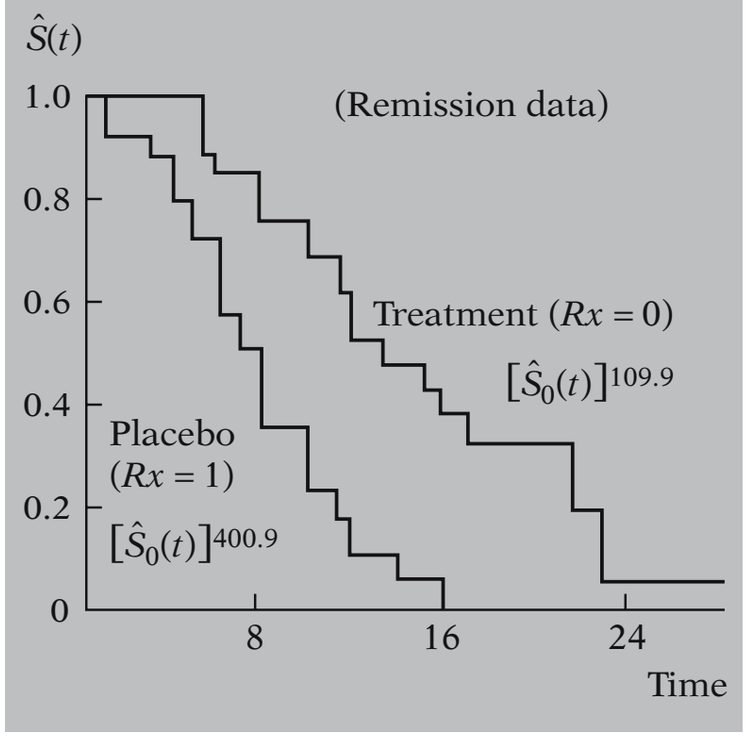
- To obtain the single survival curve that adjusts for Rx and log WBC, we then substitute the mean values in the formula for the adjusted survival curve for the model fitted. The formula and the resulting expression for the adjusted survival curve are shown here. (Note that for the remission data, where it is of interest to compare two exposure groups, the use of a single survival curve is not appropriate.)

- From this expression for the survival curve, a survival probability can be computed for any value of t that is specified. When graphing this survival curve using a computer package, the values of t that are chosen are the failure times of all persons in the study who got the event. This process is automatically carried out by the computer without having the user specify each failure time.

Adjusted Survival Curves Using the Cox PH Model

EXAMPLE

Adjusted Survival Curves for Treatment and Placebo Groups



Next section: PH assumption

- explain meaning
- when PH **not** satisfied

Later presentations:

- how to evaluate PH
- analysis when PH not met

- The graph of adjusted survival curves obtained from fitting a Cox model is usually plotted as a step function. For example, we show here the step functions for the two adjusted survival curves obtained by specifying either 1 or 0 for treatment status and letting log WBC be the mean value 2.93.

- We now turn to the concept of the proportional hazard (PH) assumption. In the next section, we explain the meaning of this assumption and we give an example of when this assumption is not satisfied.

- In later presentations, we expand on this subject, describing how to evaluate statistically whether the assumption is met and how to carry out the analysis when the assumption is not met.

The Meaning of the PH Assumption

PH: HR is constant over time, i.e.,
$\hat{h}(t,\mathbf{X}^*) = \text{constant} \times \hat{h}(t,\mathbf{X})$

$$\widehat{HR} = \frac{\hat{h}(t,\mathbf{X}^*)}{\hat{h}(t,\mathbf{X})} = \frac{\hat{h}_0(t)\exp\left[\sum \hat{\beta}_i X_i^*\right]}{\hat{h}_0(t)\exp\left[\sum \hat{\beta}_i X_i\right]} = \exp\left[\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)\right]$$

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

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

does not involve t.

- The PH assumption requires that the HR is constant over time, or equivalently, that the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time.

- To understand the PH assumption, we need to reconsider the formula for the HR that compares two different specifications $\mathbf{X}^*$ and $\mathbf{X}$ for the explanatory variables used in the Cox model. We derived this formula previously, and we show this derivation again here. Notice that the baseline hazard function $\hat{h}_0(t)$ appears in both the numerator and denominator of the hazard ratio and cancels out of the formula.

- The final expression for the hazard ratio therefore involves the estimated coefficients β_1 "hat" and the values of $\mathbf{X}^*$ and $\mathbf{X}$ for each variable. However, because the baseline hazard has canceled out, the final expression does not involve time t.

The Meaning of the PH Assumption

Let (Constant)

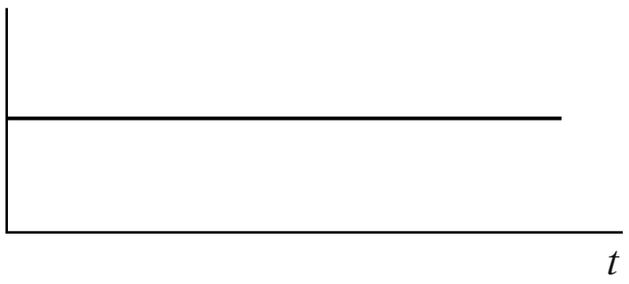
$$\hat{\theta} = \exp\left[\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)\right]$$

then

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

- Thus, once the model is fitted and the values for $\mathbf{X}^*$ and $\mathbf{X}$ are specified, the value of the exponential expression for the estimated hazard ratio is a constant, which does not depend on time. If we denote this constant by θ "hat," then we can write the hazard ratio as shown here. This is a mathematical expression which states the proportional hazards assumption.

$\widehat{HR}$ ($\mathbf{X}^*$ versus $\mathbf{X}$)



- Graphically, this expression says that the estimated hazard ratio comparing any two individuals plots as a constant over time.

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

Proportionality constant
(not dependent on time)

- Another way to write the proportional hazards assumption mathematically expresses the hazard function for individual $\mathbf{X}^*$ as θ "hat" times the hazard function for individual $\mathbf{X}$, as shown here. This expression says that the hazard function for one individual is proportional to the hazard function for another individual, where the proportionality constant is θ "hat," which does not depend on time.

The Meaning of the PH Assumption

EXAMPLE: Remission Data

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

$$\widehat{HR} = \frac{\hat{h}(t, Rx = 1, \log \text{WBC} = 2.93)}{\hat{h}(t, Rx = 0, \log \text{WBC} = 2.93)}$$

$$= \exp[1.294] = 3.65 \text{ Constant}$$

Placebo

$$\hat{h}(t, Rx = 1, \log \text{WBC} = 2.93)$$

$$= 3.65 \hat{h}(t, Rx = 0, \log \text{WBC} = 2.93)$$

Treatment

3.65 = proportionality constant

- To illustrate the proportional hazard assumption, we again consider the Cox model for the remission data involving the two variables Rx and log WBC. For this model, the estimated hazard ratio that compares placebo (Rx = 1) with treated (Rx = 0) subjects controlling for log WBC is given by e to the 1.294, which is 3.65, a constant.

- Thus, the hazard for placebo group (Rx = 1) is 3.65 times the hazard for the treatment group (Rx = 0), and the value, 3.65, is the same regardless of time. In other words, using the above model, the hazard for the placebo group is proportional to the hazard for the treatment group, and the proportionality constant is 3.65.

The Meaning of the PH Assumption

EXAMPLE: PH Not Satisfied

Cancer patients → Surgery
Cancer patients → Radiation with no surgery

$$E = \begin{cases} 0 \text{ if surgery} \\ 1 \text{ if no surgery} \end{cases}$$

$$h(t,\mathbf{X}) = h_0(t)e^{\beta E}$$

Is the above Cox PH model appropriate?

Note:

Serious surgery $\Rightarrow$ High risk for death early

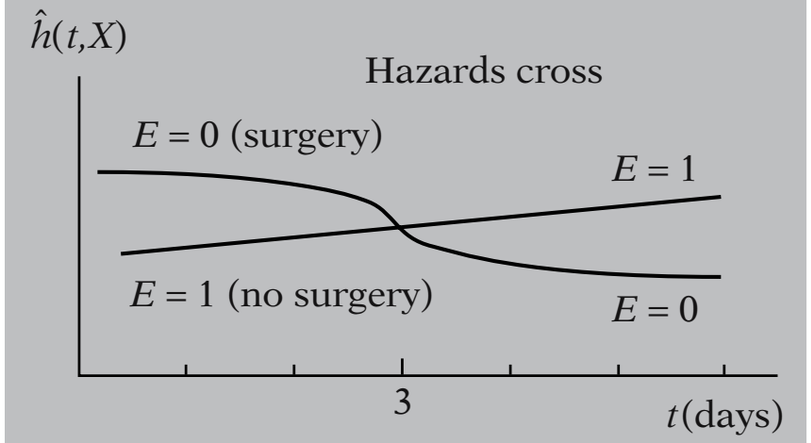
- To further illustrate the concept of proportional hazards, we now provide an example of a situation for which the proportional hazards assumption is not satisfied.

- For our example, we consider a study in which cancer patients are randomized to either surgery or radiation therapy without surgery. Thus, we have a (0,1) exposure variable denoting surgery status, with 0 if a patient receives surgery and 1 if not. Suppose further that this exposure variable is the only variable of interest, so that a Cox PH model for the analysis of this data, as shown here, will contain only the one variable E, denoting exposure.

- Now the question we consider here is whether the above Cox model containing the variable E is an appropriate model to use for this situation. To answer this question we note that when a patient undergoes serious surgery, as when removing a cancerous tumor, there is usually a high risk for complications from surgery or perhaps even death early in the recovery process, and once the patient gets past this early critical period, the benefits of surgery, if any, can then be observed.

The Meaning of the PH Assumption

EXAMPLE: PH Not Satisfied



2 days: $\dfrac{\hat{h}(t=2, E=1)}{\hat{h}(t=2, E=0)} < 1$

but

5 days: $\dfrac{\hat{h}(t=5, E=1)}{\hat{h}(t=5, E=0)} > 1$

- Thus, in a study that compares surgery to no surgery, we might expect to see hazard functions for each group that appear as shown here. Notice that these two functions cross at about three days, and that prior to three days the hazard for the surgery group is higher than the hazard for the no surgery group, whereas after three days, the hazard for the surgery group is lower than the hazard for the no surgery group.

- Looking at the above graph more closely, we can see that at 2 days, when $t = 2$, the hazard ratio of nonsurgery ($E = 1$) to surgery ($E = 0$) patients yields a value less than 1. In contrast, at t = 5 days, the hazard ratio of nonsurgery to surgery yields a value greater than 1.

The Meaning of the PH Assumption

EXAMPLE: (continued)

Given the above description, ***HR* is not constant over time.**

Cox PH model inappropriate because PH model assumes constant *HR*:

$$h(t,\mathbf{X}) = h_0(t)e^{\beta E}$$

$$\widehat{HR} = \frac{\hat{h}(t, E=1)}{\hat{h}(t, E=0)} = e^{\hat{\beta}}$$

General rule:

If the hazards cross, then a Cox PH model is not appropriate.

Analysis when Cox PH model not appropriate? See Chapters 5 and 6.

- Thus, if the above description of the hazard functions for each group is accurate, the hazard ratios are not constant over time. That is, the hazard ratio is some number less than 1 before three days and greater than 1 after three days.

- It is therefore inappropriate to use a Cox PH model for this situation, because the PH model assumes a constant hazard ratio across time, whereas our situation yields a hazard ratio that varies with time.

- In fact, if we use a Cox PH model, shown here again, the estimated hazard ratio comparing exposed to unexposed patients at any time is given by the constant value e to the β "hat," which does not vary over time.

- This example illustrates the general rule that if the hazards cross, then the PH assumption cannot be met, so that a Cox PH model is inappropriate.

- It is natural to ask at this point, if the Cox PH model is inappropriate, how should we carry out the analysis? The answer to this question is discussed in Chapters 5 and 6. However, we will give a brief reply with regard to the surgery study example just described.

The Meaning of the PH Assumption

EXAMPLE (continued)

Surgery study analysis options:

- stratify by exposure (use KM curves)
- start analysis at three days; use Cox PH model
- fit PH model for < 3 days and for > 3 days; get $\widehat{HR}$ (< 3 days) and $\widehat{HR}$ (> 3 days)
- include time-dependent variable (e.g., $\mathrm{E} \times t$); use extended Cox model

- Actually, for the surgery study there are several options available for the analysis. These include:
  - analyze by stratifying on the exposure variable; that is, do not fit any model, and, instead obtain Kaplan–Meier curves for each exposure group separately;
  - start the analysis at three days, and use a Cox PH model on three-day survivors;
  - fit Cox model for less than three days and a different Cox model for greater than three days to get two different hazard ratio estimates, one for each of these two time periods;
  - fit a modified Cox model that includes a time-dependent variable which measures the interaction of exposure with time. This model is called an **extended Cox model**.

The Meaning of the PH Assumption

Different options may lead to different conclusions.

Hazards cross $\Rightarrow$ PH not met

but

? $\Rightarrow$ PH met

See Chapter 4: Evaluating PH Assumption

- Further discussion of these options is given in subsequent chapters. We point out here, that different options may lead to different conclusions, so that the investigator may have to weigh the relative merits of each option in light of the data actually obtained before deciding on any particular option as best.
- One final comment before concluding this section: although we have shown that when the hazards cross, the PH assumption is not met, we have not shown how to decide when the PH assumption is met. This is the subject of Chapter 4 entitled, “Evaluating the PH Assumption.”

The Cox Likelihood

Likelihood

- Typically based on outcome distribution
- Outcome distribution not specified for Cox model
- Cox likelihood based on order of events rather than their distribution
  - Called partial likelihood

- Typically, the formulation of a likelihood function is based on the distribution of the outcome. However, one of the key features of the Cox model is that there is not an assumed distribution for the outcome variable (i.e., the time to event). Therefore, in contrast to a parametric model, a full likelihood based on the outcome distribution cannot be formulated for the Cox PH model. Instead, the construction of the Cox likelihood is based on the observed order of events rather than the joint distribution of events. Thus the Cox likelihood is called a “partial” likelihood.

Illustration

Scenario:

- Gary, Larry, Barry have lottery tickets
- Winning tickets chosen at times $t_1, t_2, \ldots$
- Each person ultimately chosen
- Can be chosen only once

Question:
What is the probability that the order chosen is as follows?

1. Barry
2. Gary
3. Larry

- To illustrate the idea underlying the formulation of the Cox model, consider the following scenario. Suppose Gary, Larry, and Barry are each given a lottery ticket. Winning tickets are chosen at times t_j ($j = 1, 2, \ldots$). Assume each person is ultimately chosen and once a person is chosen he cannot be chosen again (i.e., he is out of the risk set). What is the probability that the order each person is chosen is first Barry, then Gary, and finally Larry?

The Cox Likelihood

Answer:

$$\text{Probability} = \underset{\text{Barry}}{\frac{1}{3}} \times \underset{\text{Gary}}{\frac{1}{2}} \times \underset{\text{Larry}}{\frac{1}{1}} = \frac{1}{6}$$

Scenario:

Barry – 4 tickets
Gary – 1 ticket
Larry – 2 tickets

Question:
What is the probability that the order chosen is as follows?

1. Barry
2. Gary
3. Larry

Answer:

$$\text{Probability} = \frac{4}{7} \times \frac{1}{3} \times \frac{2}{2} = \frac{4}{21}$$

- The probability the Barry's ticket is chosen before Gary's and Larry's is one out of three. Once Barry's ticket is chosen it cannot be chosen again. The probability that Gary's ticket is then chosen before Larry's is one out of two. Once Barry's and Gary's tickets are chosen they cannot be chosen again which means that Larry's ticket must be chosen last. This yields a probability of 1/6 for this given order of events (see left).

- Now consider a modification of the previous scenario. Suppose Barry has 4 tickets, Gary has 1 ticket, and Larry has 2 tickets; now what is the probability that the order each person is chosen is first Barry, then Gary, and finally Larry?

- Barry, Gary, and Larry have 7 tickets in all and Barry owns 4 of them so Barry's probability of being chosen first is 4 out of 7. After Barry is chosen, Gary has 1 of the 3 remaining tickets and after Barry and Gary are chosen, Larry owns the remaining 2 tickets. This yields a probability of 4/21 for this order (see left).

The Cox Likelihood

For this scenario

Subject's number of tickets affects probability

For Cox model

Subject's pattern of covariates affects likelihood of ordered events

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

- For this scenario, the probability of a particular order is affected by the number of tickets held by each subject. For a Cox model, the likelihood of the observed order of events is affected by the pattern of covariates of each subject.

- To illustrate this connection, consider the dataset shown on the left. 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.

The Cox Likelihood

Cox PH model

$$h(t) = h_0(t)e^{\beta_1 SMOKE}$$

ID	Hazard
Barry	$h_0(t)e^{\beta_1}$
Gary	$h_0(t)e^{0}$
Harry	$h_0(t)e^{0}$
Larry	$h_0(t)e^{\beta_1}$

Individual hazards (Cox likelihood) analogous to number of tickets (lottery scenario) For example, smokers analogous to persons with extra lottery tickets

Cox Likelihood

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

Likelihood is product of 3 terms

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

$$L_1 = \left[\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1} + h_0(t)e^{0} + h_0(t)e^{0} + h_0(t)e^{\beta_1}}\right]$$

$$L_2 = \left[\frac{h_0(t)e^{0}}{h_0(t)e^{0} + h_0(t)e^{0} + h_0(t)e^{\beta_1}}\right]$$

$$L_3 = \left[\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1}}\right]$$

- Consider the Cox proportional hazards model with one predictor, SMOKE. Under this model the hazards for Barry, Gary, Harry, and Larry can be expressed as shown on the left. The individual hazards are determined by whether the subject was a smoker or nonsmoker.

- The individual level hazards play an analogous role toward the construction of the Cox likelihood as the number of tickets held by each subject plays for the calculation of the probabilities in the lottery scenario discussed earlier in this section. The subjects who smoke are analogous to persons given extra lottery tickets, thereby affecting the probability of a particular order of events.

- On the left is the Cox likelihood for these data. Notice the likelihood is a product of three terms, which correspond to the three event times. Barry got the event first at TIME = 2 years. At that time all four subjects were at risk for the event. The first product (L_1) has the sum of the four subjects' hazards in the denominator and Barry's hazard in the numerator. Gary got the event next at 3 years when Gary, Harry, and Larry were still in the risk set. Consequently, the second product (L_2) has the sum of the three hazards for the subjects still at risk in the denominator and Gary's hazard in the numerator. Harry was censored at 5 years, which occurred between the second and third event. Therefore, when Larry got the final event at 8 years, nobody else was at risk for the event. As a result, the third product (L_3) just has Larry's hazard in the denominator and the numerator.

t_1, time = 2, four at risk (L_1)
t_2, time = 3, three at risk (L_2)
t_3, time = 8, one at risk (L_3)

For each term:

Numerator—single hazard
Denominator—sum of hazards

Baseline hazard, $h_0(t)$ cancels

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

Thus, L does not depend on $h_0(t)$

Data A

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

Data B

ID	TIME	STATUS	SMOKE
Barry	1	1	1
Gary	7	1	0
Harry	8	0	0
Larry	63	1	1

Comparing datasets

- TIME variable differs
- Order of events the same
- Cox PH likelihood the same

The Cox Likelihood

- To summarize, the likelihood in our example consists of a product of three terms (L_1, L_2, and L_3) corresponding to the ordered failure times (t_1, t_2, and t_3). The denominator for the term corresponding to time t_j (j=1,2,3) is the sum of the hazards for those subjects still at risk at time t_j, and the numerator is the hazard for the subject who got the event at t_j.

- A key property of the Cox likelihood is that the baseline hazard cancels out in each term. Thus, the form of the baseline hazard need not be specified in a Cox model, as it plays no role in the estimation of the regression parameters. By factoring $h_0(t)$ in the denominator and then canceling it out of each term, the likelihood for Barry, Gary, and Larry can be rewritten as shown on the left.

- As we mentioned earlier, the Cox likelihood is determined by the order of events and censorships and not by the distribution of the outcome variable. To illustrate this point, compare datasets A and B on the left, and consider the likelihood for a Cox PH model with smoking status as the only predictor. Although the values for the variable TIME differ in the two datasets, the Cox likelihood will be the same using either dataset because the order of the outcome (TIME) remains unchanged.

The Cox Likelihood

General Approach

- K failure times
- Likelihood a product of K terms
- Construction of each term similar to Barry, Gary, and Larry

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

Obtaining maximum likelihood estimates

Solve system of equations

$$\frac{\partial \ln L}{\partial \beta_i} = 0, \quad \text{i} = 1, 2, 3, \ldots, \text{p}$$

p = # of parameters

- We have used a small dataset (four observations with three failure times) for ease of illustration. However, the approach can be generalized. Consider a dataset with K failure times and let L_j denote the contribution to the likelihood corresponding to the jth failure time. Then the Cox likelihood can be formulated as a product of each of the K terms as shown on the left. Each of the terms L_j is constructed in a similar manner as with the data for Gary, Larry, and Barry.

- Once the likelihood is formulated, the question becomes: *which values of the regression parameters would maximize* L? The process of maximizing the likelihood is typically carried out by setting the partial derivative of the natural log of L to zero and then solving the system of equations (called the score equations).

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

- This presentation is now complete.
- The next Chapter (4) describes how to evaluate the PH assumption. Chapters 5 and 6 describe methods for carrying out the analysis when the PH assumption is not met.