

BIO 244: Unit 8

Cox's Model: Examples

Let's illustrate what we have been discussing with data from the AZT-ddI study. Consider the 425 subjects who were randomized to receive AZT ($\mathbf{rx}=\mathbf{0}$) or lower-dose ddI ($\mathbf{rx}=\mathbf{1}$), and who had not yet developed AIDS upon entry into the study ($\mathbf{as-ar-a}=\mathbf{0}$ or $\mathbf{1}$) (as-ar-a stands for asymptomatic-AIDS related complex-AIDS).

We will focus on the patients' gender, age, CD4 count, and treatment at entry, and examine their association with the main endpoint of the trial: time to AIDS or death (whichever comes first). The observed outcome (U, δ) is represented by the variables \mathbf{Tad} and \mathbf{ad} in the data file. There were 209 patients in the AZT group, of whom 75 were observed to fail, and 216 in the ddI group, of whom 49 were observed to fail.

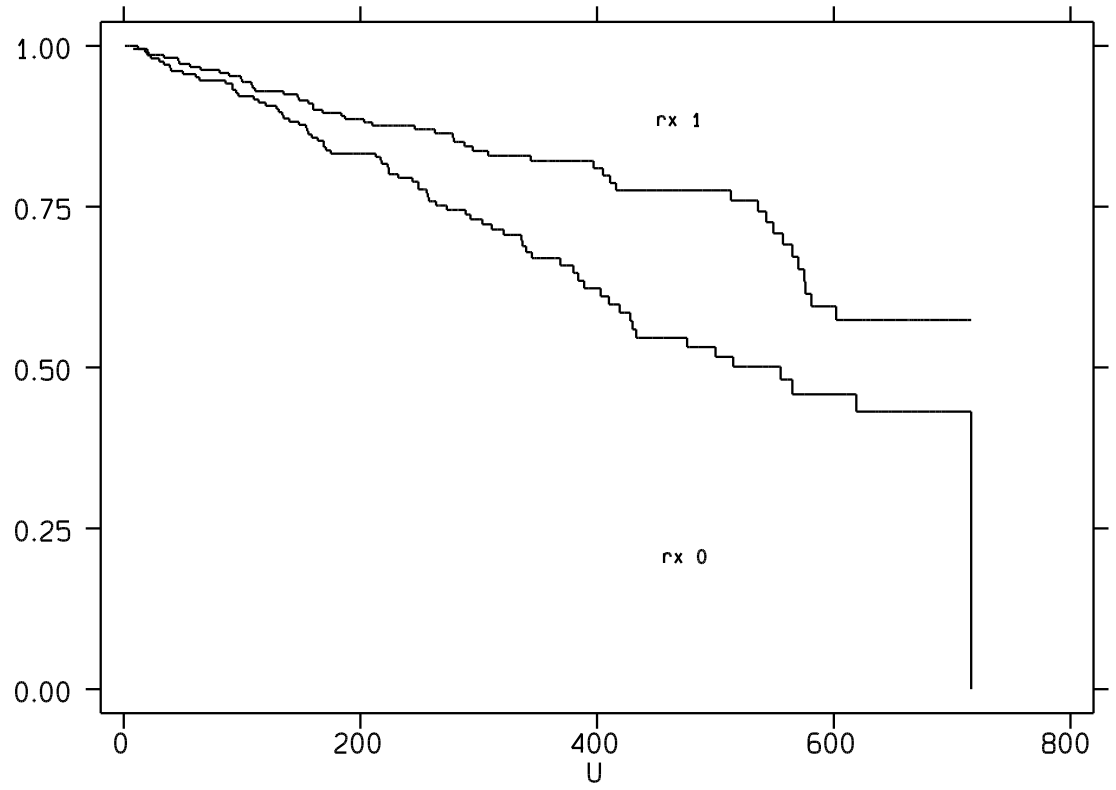
- gender $\begin{matrix} 0 & = & \text{F} \\ 1 & = & \text{M} \end{matrix}$
- CD4 cell count
(a measure of the strength of the immune system)
- age (years)
- rx = $\begin{cases} 0 & \text{AZT} \\ 1 & \text{ddI-500mg.} \end{cases}$

I. We first compute the logrank test of AZT vs. ddI, getting:

$$\begin{aligned} O - E &= 49 - 67.55 = -18.59 \\ Z &= -3.36 \quad (Z^2 = 11.26) \quad \text{P (2-sided)} = .0008. \end{aligned}$$

There were 49 failures in the ddI group compared to 67.55 'expected' under H_0 , and the groups were significantly different. The difference is evident in the following figure, which gives the Kaplan-Meier estimators of the 2 treatment groups.

Kaplan-Meier survival estimates, by rx



II. Now, let's fit a Cox PH model to the data, with rx as the only covariate.
This gives:

$$h(t|\mathbf{R}_x) = h_0(t)e^{\beta \cdot (\mathbf{R}_x)}$$

$$\rightarrow \hat{\beta} = -.608565$$

$$\hat{I}(\hat{\beta})^{-1} = \hat{var}(\hat{\beta}) = .0339$$

$$\therefore s.e.(\hat{\beta}) = .184$$

$$95\% \text{ CI} = \hat{\beta} \pm 1.96(.184) = [-.97, -.25].$$

The estimated ddI:AZT hazard ratio (HR), also called 'relative risk' (RR), is:

$$HR = \frac{h(t|\mathbf{R}_x = 1)}{h(t|\mathbf{R}_x = 0)} = e^{\beta}.$$

$$\text{Estimated HR} = e^{\hat{\beta}} = e^{-.608565} = .544.$$

Thus, the estimated hazard function for ddI is $\approx 54\%$ that for AZT.

$$\text{Wald test of } H_0: \beta = 0: Z = \frac{\hat{\beta}}{s.e.(\hat{\beta})} = -3.30$$

$$P(2\text{-sided}) = .001.$$

Note: This is very similar to the logrank test result.

III. Now let's fit a Cox PH model with 4 covariates:

$$\left. \begin{array}{l} Z_1 = \text{age} \\ Z_2 = \text{CD4} \\ Z_3 = \text{Rx} \\ Z_4 = \text{gender} \end{array} \right\} h_0(t)e^{\beta_1(\text{age}) + \beta_2(\text{CD4}) + \beta_3(\text{Rx}) + \beta_4(\text{gender})}.$$

The resulting estimated regression coefficients and Wald tests are:

$$H_1 : \beta_1 = 0 \implies P = .008, \hat{\beta}_1 = .0264$$

$$H_2 : \beta_2 = 0 \implies P < .001, \hat{\beta}_2 = -.011002$$

$$H_3 : \beta_3 = 0 \implies P = .001, \hat{\beta}_3 = -.6382$$

$$H_4 : \beta_4 = 0 \implies P = .99, \hat{\beta}_4 = .0016.$$

Thus, age, CD4 and treatment are significantly associated with failure, but gender is not.

Let's drop gender, and refit age, CD4, and rx. This gives:

	$\hat{\beta}$	<i>s.e.</i>	<i>Z</i>	<i>P</i>	comment
age	.0264	.010	2.69	.007	risk ↑ with age
CD4	-.011	.0014	-7.71	< .0001	risk ↓ with ↑ CD4
Rx	-.638	.186	-3.44	.001	risk ↓ with ddI

e.g., age: The estimated HR increase corresponding to a 10-year increase in age is:

$$= e^{10(\hat{\beta}_1)} = 1.30$$

i.e., for every additional 10 years of age, risk (or hazard) increases by 30%.

CD4: Similarly, the estimated HR for 100 CD4 cell drop is:

$$e^{\hat{\beta}_2(-100)} = 3.00$$

\implies hazard (or risk) of failure increases by a factor of 3 for every 100 CD4 cells lost.

NOTE: Estimated treatment effect is relatively unchanged after adjustment for age and CD4.

Questions that Naturally Arise

1. Do the AZT and ddI groups have PH? What about logrank or Cox analyses if they don't?
2. What if we don't adjust properly for age (e.g., we should have fitted age^2 instead of age) or age and CD4 (e.g., we should have included an interaction term)?
3. Does the ddI : AZT HR vary with CD4 or age?

STATA Commands

The STATA command for running the Cox PH model is `stcox`. One uses this after executing the `stset` command to define U and δ . For example,

`.stcox rx` fits a Cox model with only treatment, and displays the results in terms of hazard ratios (that is, estimates of $exp\beta$ instead of β). To display the regression coefficients, use:

```
.stcox rx, nohr
```

The commands used to do all the analyses in this unit are as follows (after opening AZT-ddI):

(a) first generate an indicator variable for the patient subset of interest:

```
.gen ind=1  
.replace ind=0 if(rx==2)  
.replace ind=0 if(as-ar-a==2)
```

(b) Define (U, δ) and run the logrank test:

```
.stset Tad ad  
.sts test rx if(ind==1)
```

(c) Plot Kaplan-Meier curves:

```
.sts graph if(ind==1),by(rx)
```

(d) Cox analyses:

```
.stcox rx if(ind==1),nohr to get estimates of the regression coefficients, or
```

```
.stcox rx if(ind==1) to get estimated hazard ratios.
```

Replace `rx` in the above with `age cd4 rx gender` to fit all 4 covariates.

STATA uses as a default the Breslow method for handling ties. Also, when one excludes the **nohr** option with the **stcox** command, the confidence intervals presented for the hazard ratio are obtained by finding a CI for the regression coefficient, based on the usual normal approximation, and then exponentiating these intervals. Note that this is different from using the delta method to approximate the variance of $\exp(\hat{\beta})$, and then using this to get a CI.

SAS Commands

The SAS procedure for running the Cox PH model is `proc phreg`.

The commands used to do all the analyses in this unit are as follows. To read in the data, and check whether that went ok, use, for the AZT-ddI data:

```
data AZTddI;
  infile 'D:\teaching\Unit8\AZT-ddI.dat';
  input id gender cd4 age as-ar-a rx dead Tdead ad Tad;
run;
proc print data=AZTddI (OBS=100);
run;
```

(a) first generate an indicator variable for the patient subset of interest:

```
data AZTddI;
  set AZTddI;
  ind=1;
  if (rx=2) then ind=0;
  if (as-ar-a=2) then ind=0;
run;
```

(b) create a dataset with only the subjects included in the analysis:

```
data analysis;
  set AZTddI;
  if (ind=1);
run;
```


(c) Run the logrank test and plot Kaplan Meier curves (note that ad=0 indicates censoring):

```
proc lifetest data=analysis plots=(s);
  time Tad *ad(0);
  strata rx;
run;
```

(d) Cox proportional hazards analysis: estimate of the regression coefficient, the hazard ratio and confidence intervals for the hazard ratio:

```
proc phreg data=analysis;
  model Tad*ad(0)=rx /rl;
run;
```

Replace **rx** in the above with **age cd4 rx gender** to fit all 4 covariates.

SAS uses as a default the Breslow method for handling ties. Also, when one uses /rl, the confidence intervals presented for the hazard ratio are obtained by finding a CI for the regression coefficient, based on the usual normal approximation, and then exponentiating these intervals. Note that this is different from using the delta method to approximate the variance of $exp(\hat{\beta})$, and then using this to get a CI.

Exercises

1. Verify (to yourself) that an analysis based on Cox's partial likelihood is invariant to any rank-preserving transformation of the U_i .
2. Develop a graphical GOF (Goodness Of Fit) test for proportional hazards of 2 groups (e.g., AZT vs. ddI).
3. Experiment with this subset of 425 patients from AZT-ddI to get a feel for Cox's model and its relation to the logrank test.
4. Consider a set of censored survival data from a clinical trial where we have information on treatment group ($Z_1=0$ for group 0, $Z_1=1$ for group 1), gender ($Z_2=0$ for females, $Z_2=1$ for males), and age group ($X=0$ if <50 years, $X=1$ if ≥ 50 years). Consider the model where the hazard function at time t is given by

$$h(t \mid Z_1, Z_2, X) = h_X(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 (X Z_1)\}.$$

- (a) Describe in words how this model allows age group to modify the joint association of treatment and gender on survival.
- (b) How do you specify the hypothesis that the treatment relative risk is the same in each of the 4 categories of age and gender?
- (c) How do you express the hypothesis that age group is not associated with survival?
- (d) How do you express the hypothesis of no treatment effect on survival?