

# Proportional Hazards Regression with Censored Data: Basics

Dave Harrington

May 14 - 18, 2018

Introduction

The Cox proportional hazards (PH) model

The partial likelihood

Additional details on the Cox model

Time-varying (time-dependent) covariates

Derivations

# Introduction

# REGRESSION MODELS FOR SURVIVAL DATA

These notes focus on proportional hazards regression models.

*Proportional Hazards (PH) models*

$$\lambda(t; \mathbf{Z}) = \lambda_0(t)\Psi(\mathbf{Z})$$

- The second term is usually written  $\Psi(Z) = e^{\beta Z}$ .
- Suppose  $Z = 1$  for treated subjects and  $Z = 0$  for untreated subjects. The hazard is increased by a factor of  $e^{\beta}$  for treated subjects versus untreated subjects.
- This is a *semi-parametric* model.

## OTHER REGRESSION MODELS FOR SURVIVAL DATA

### *Accelerated Failure Time (AFT) models*

$$\log(T) = \mu + \beta \mathbf{Z} + \sigma W,$$

- $W$  is an *error term*.
- Typically (but not always), there is a *parametric* model for  $W$ :
  - exponential, Weibull, gamma, or log-normal

### *Additive Hazards Regression models*

$$\lambda(t; Z(t)) = \beta_0(t) + \sum_{k=1}^p \beta_k(t) Z_k(t)$$

Klein & Moeschberger have excellent treatments of both of these models

# COVARIATES

In general,  $\mathbf{Z}$  is a  $p$ -dimensional vector of covariates.  $\mathbf{Z}$  may include

- continuous factors (e.g., age, blood pressure)
- discrete factors (e.g., sex, marital status)
- possible interactions (e.g., age by sex interaction)

Two covariates,  $A$  and  $B$ , interact if

- hazard of death depends on the combination of  $A$  and  $B$
- best to only include interactions if the corresponding main effects are also included

## COVARIATES...

A categorical covariate  $A$  with  $a$  levels is usually modeled with

- $a$  binary variables<sup>1</sup> ( $U_1, U_2, \dots, U_a$ ) such that  $U_j = 1$  if  $A = j$ .
- Then,

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_2 U_2 + \beta_3 U_3 + \dots + \beta_a U_a)$$

- In the above model, the subgroup with  $A = 1$  or  $U_1 = 1$  is the reference group and is not included in the model.

Alternatively, in R, categorical covariates can (and should) be stored or converted to *factor* variables.

---

<sup>1</sup>These are sometimes called dummy variables or 0-1 variables.

## The Cox proportional hazards (PH) model



# THE COX MODEL

The Cox (1972) proportional hazards model is given by

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp(\beta \mathbf{Z}).$$

This is the most common model used for survival data.<sup>2</sup>

- Natural interpretation
- Allows for flexible choice of covariates
- Fairly easy to fit
- Robust software exists in nearly all standard packages

## References

- Collett, Chapter 3
- Cox and Oakes, Chapter 7
- Klein and Moeschberger, Chapters 8 & 9

---

<sup>2</sup>Some books use  $h(t; \mathbf{X})$  as their standard notation instead of  $\lambda(t; \mathbf{Z})$ .

## WHY THE NAME PROPORTIONAL HAZARDS?

Suppose in two groups,  $Z = 1$  for treated and  $Z = 0$  for control. If

- $\lambda_1(t)$  is the hazard rate for the treated group and
- $\lambda_0(t)$  is the hazard for control, then

$$\lambda_1(t) = \lambda(t; Z = 1) = \lambda_0(t) \exp(\beta Z) = \lambda_0(t) \exp(\beta)$$

The ratio of the two hazards is a constant,  $\phi$ , which does not depend on time;  $\phi$  is referred to as the *hazard ratio*.

$$\phi = \frac{\lambda_1(t)}{\lambda_0(t)} = e^\beta$$

The hazards of the two groups remain proportional over time.

# THE BASELINE HAZARD FUNCTION

In general,  $\lambda_0(t)$  is called the *baseline hazard function*.

- It is the underlying hazard for subjects with all covariates  $Z_1, \dots, Z_p$  equal to 0 (i.e., the “reference group”).

General form of the PH model:

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p)$$

When all of the  $Z_j$ 's equal 0:

$$\begin{aligned}\lambda(t; \mathbf{Z} = 0) &= \lambda_0(t) \exp(\beta_1(0) + \beta_2(0) + \dots + \beta_p(0)) \\ &= \lambda_0(t)\end{aligned}$$

## ROLE OF THE EXPLANATORY VARIABLES

In the general case, the  $i$ -th individual has a set of covariates in a column vector  $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, \dots, Z_{pi})$ . The hazard rate is modeled as some multiple of the baseline hazard:

$$\lambda_i(t; \mathbf{Z}_i) = \lambda_0(t) \exp(\beta_1 Z_{1i} + \dots + \beta_p Z_{pi})$$

## WHY THE NAME PROPORTIONAL HAZARDS? (AGAIN!)

The log of the hazard ratio for the  $i$ -th individual compared to the reference group can be written as

$$\log \left( \frac{\lambda_i(t)}{\lambda_0(t)} \right) = \beta_1 Z_{1i} + \beta_2 Z_{2i} + \cdots + \beta_p Z_{pi}.$$

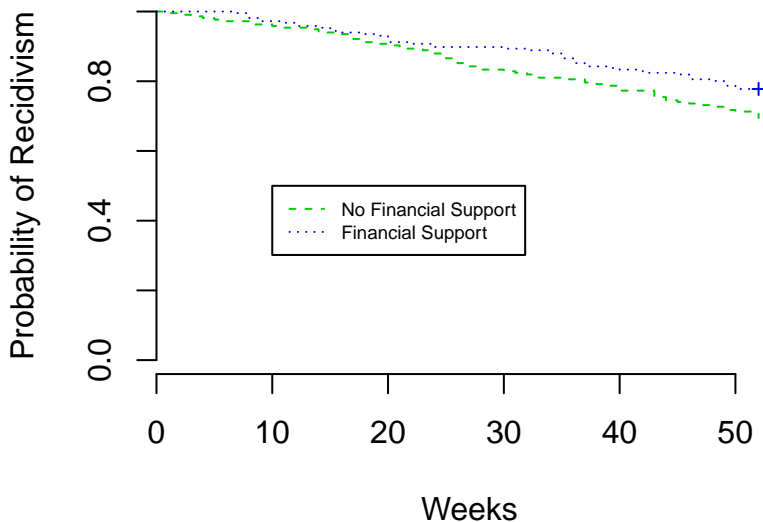
The Cox PH model is a linear model for the log of the hazard ratio.

- Surprisingly, the parameter  $\beta$  can be estimated without having to make assumptions about  $\lambda_0(t)$ .
- This is the reason for the terminology *semi-parametric*.

Estimating equations coming after an example.

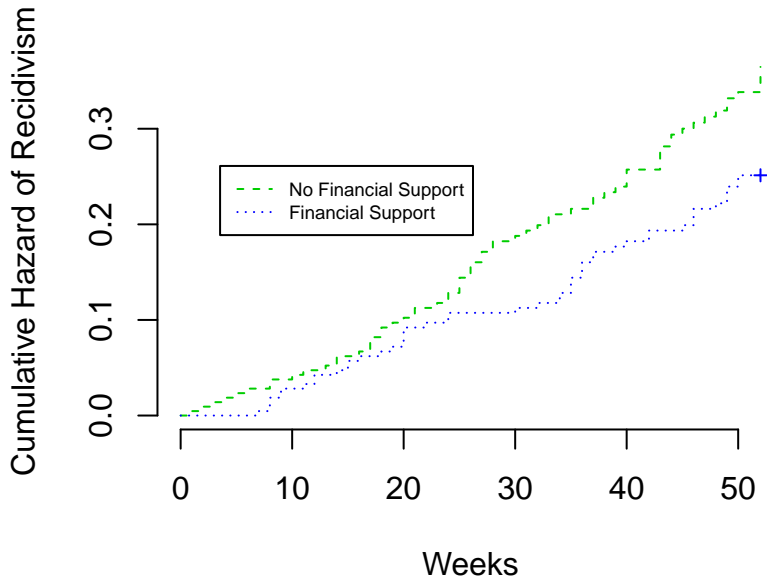
## EXAMPLE: ROSSI RECIDIVISM DATA, SURVIVAL CURVES

**KM of Recidivism Probability**



# ROSSI DATA, CUMULATIVE HAZARDS

## Cumulative Risk of Recidivism Probability



## EXAMPLE: ROSSI RECIDIVISM DATA ...

Let

- $T$  be time to recidivism (the variable `week`)
- $Z$  denote randomization to financial incentive (the variable `fin`)

Estimate and interpret the model

$$\lambda(t; Z) = \lambda_0 \exp(\beta Z)$$



## EXAMPLE: ROSSI RECIDIVISM DATA ...

```
library(survival)
library(eventtimedata)
data(rossi)
coxph(Surv(week, arrest) ~ fin, data = rossi)
```

```
## Call:
## coxph(formula = Surv(week, arrest) ~ fin, data = rossi)
##
##              coef exp(coef) se(coef)      z      p
## finyes -0.369      0.691    0.190 -1.95 0.052
##
## Likelihood ratio test=3.84  on 1 df, p=0.0501
## n= 432, number of events= 114
```

## INTERPRETATION

The term coef is the  $\beta$  coefficient for fin.

- There are 114 events (re-arrests) in 432 cases.
- The value -0.369 implies that the hazard for re-arrest is lower in the financial incentive group by a factor of  $\exp(-0.369) = 0.691$ .
- The standard error for the coefficient is 0.190.
- The Wald statistic  $p$ -value for a test of  $H_0 : \beta = 0$  is 0.052.
  - The Wald test statistic used to test  $H_0 : \beta = 0$  is given by

$$\frac{\hat{\beta}}{\text{s.e.}(\hat{\beta})}$$

- The likelihood ratio (LR) test of the same hypothesis has a similar result. Definition of the LR test coming later.

There is *no* information about  $\lambda_0(t)$  in the output. That follows from the method of estimation.

# LIKELIHOOD ESTIMATION FOR THE PH MODEL

There have been several derivations of the likelihood used for the PH model.

Cox (1972) derived a *partial likelihood*.

Suppose the data for an individual  $i$  are  $(X_i, \delta_i, \mathbf{Z}_i)$ , where

- $X_i$  is a censored failure time random variable
- $\delta_i$  is the failure/censoring indicator (1=fail, 0=censor)
- $\mathbf{Z}_i$  represents a set of covariates

The covariates may be continuous, discrete, or time-varying.

Suppose there are  $K$  distinct failure times, and let  $\tau_1, \dots, \tau_K$  represent the  $K$  ordered, distinct death times.

For now, assume there are no tied death times.

# THE PARTIAL LIKELIHOOD

Let  $\mathcal{R}(t) = \{i : x_i \geq t\}$  denote the set of individuals who are “at risk” for failure at time  $t$ .

Notation for risk sets:

- $\mathcal{R}(\tau_j)$  is the risk set at the  $j$ -th failure time.
- $\mathcal{R}(X_i)$  is the risk set at the failure time of individual  $i$  when  $\delta_i = 1$
- $r_j$  is the number of individuals in  $\mathcal{R}(\tau_j)$ , while  $\mathcal{R}(\tau_j)$  identifies the subjects at risk.

Intuitively, the partial likelihood is a product over the set of observed failure times of the conditional probabilities of seeing the observed deaths, given the set of individuals at risk at those times.

## THE PARTIAL LIKELIHOOD . . .

At each failure time  $\tau_j$ , the contribution to the likelihood is:

$$\begin{aligned} L_j(\beta) &= P(\text{individual } j \text{ fails} | 1 \text{ failure from } \mathcal{R}(\tau_j)) \\ &= \frac{P(\text{individual } j \text{ fails} | \text{at risk at } \tau_j)}{\sum_{\ell \in \mathcal{R}(\tau_j)} P(\text{individual } \ell \text{ fails} | \text{at risk at } \tau_j)} \\ &= \frac{\lambda(\tau_j; \mathbf{Z}_j)}{\sum_{\ell \in \mathcal{R}(\tau_j)} \lambda(\tau_j; \mathbf{Z}_\ell)} \end{aligned}$$

## THE PARTIAL LIKELIHOOD . . .

Under the PH assumption,  $\lambda(t; \mathbf{Z}) = \lambda_0(t)e^{\beta\mathbf{z}}$ , so

$$\begin{aligned} L^{partial}(\beta) &= \prod_{j=1}^K \frac{\lambda_0(\tau_j)e^{\beta\mathbf{z}_j}}{\sum_{\ell \in \mathcal{R}(\tau_j)} \lambda_0(\tau_j)e^{\beta\mathbf{z}_\ell}} \\ &= \prod_{j=1}^K \frac{e^{\beta\mathbf{z}_j}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta\mathbf{z}_\ell}} \end{aligned}$$

This is equivalent to the partial likelihood as defined by Cox

$$L^{partial}(\beta) = \prod_{j=1}^n \left[ \frac{e^{\beta\mathbf{z}_j}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta\mathbf{z}_\ell}} \right]^{\delta_j}$$

Contributions to the partial likelihood occur only at failure times.

## A SIMPLE EXAMPLE

individual	$X_i$	$\delta_i$	$Z_i$
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

Now compile the pieces that go into the partial likelihood contributions at each failure time ...

## A SIMPLE EXAMPLE ...

$j$	ordered failure		$i_j$	Likelihood contribution $\left[ e^{\beta Z_i} / \sum_{j \in \mathcal{R}(X_i)} e^{\beta Z_j} \right]^{\delta_i}$
	time $X_i$	$\mathcal{R}(X_i)$		
1	6	$\{1,2,3,4\}$	3	$e^{7\beta} / [e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
2	8	$\{1,2,4\}$	2	$(e^{5\beta} / [e^{4\beta} + e^{5\beta} + e^{3\beta}])^0 = 1$
3	9	$\{1,4\}$	1	$e^{4\beta} / [e^{4\beta} + e^{3\beta}]$
4	10	$\{4\}$	4	$e^{3\beta} / e^{3\beta} = 1$

The partial likelihood is the product of these four terms.

The estimate for  $\beta$  is the value  $\hat{\beta}$  that maximizes the partial likelihood.



## A MORE REALISTIC EXAMPLE: ROSSI DATA

```
library(survival)
library(eventtimedata)
data(rossi)
coxph(Surv(week, arrest) ~ fin + age + race + mar,
      data = rossi)
```

Output next slide.

We will discuss interpretation in class.

## A MORE REALISTIC EXAMPLE: ROSSI DATA

```
## Call:
## coxph(formula = Surv(week, arrest) ~ fin + age + race + mar,
##       data = rossi)
##
##               coef exp(coef) se(coef)      z      p
## finyes          -0.351    0.704    0.190 -1.84 0.0654
## age             -0.065    0.937    0.021 -3.10 0.0019
## raceother       -0.244    0.784    0.306 -0.80 0.4258
## marnot married  0.508    1.663    0.373  1.36 0.1735
##
## Likelihood ratio test=21.2  on 4 df, p=0.000289
## n= 432, number of events= 114
```

# CATEGORICAL VARIABLES CODED AS NUMERIC

Education (educ) is coded on the dataset as an integer variable with values 2 - 6.<sup>3</sup>

The next two slides show two models.

- Model with educ fitted as a numeric variable
- Model with educ converted to a factor variable

---

<sup>3</sup>See the documentation for definitions.

## CATEGORICAL VARIABLES CODED AS NUMERIC ...

```
library(survival)
library(eventtimedata)
data(rossi)
coxph(Surv(week, arrest) ~ fin + educ, data = rossi)
```

```
## Call:
```

```
## coxph(formula = Surv(week, arrest) ~ fin + educ, data = rossi)
```

```
##
```

	coef	exp(coef)	se(coef)	z	p
## finyes	-0.351	0.704	0.190	-1.85	0.065
## educ	-0.256	0.774	0.120	-2.14	0.033

```
##
```

```
## Likelihood ratio test=8.7 on 2 df, p=0.0129
```

```
## n= 432, number of events= 114
```

## CATEGORICAL VARIABLES CODED AS NUMERIC: BE CAREFUL ...

```
library(survival)
library(eventtimedata)
data(rossi)
coxph(Surv(week, arrest) ~ fin + as.factor(educ), data = rossi)
```

```
## Call:
```

```
## coxph(formula = Surv(week, arrest) ~ fin + as.factor(educ), data = r
```

```
##
```

##		coef	exp(coef)	se(coef)	z	p
## finyes		-0.402	0.669	0.190	-2.11	0.035
## as.factor(educ)3		0.841	2.318	0.514	1.64	0.102
## as.factor(educ)4		0.459	1.583	0.537	0.86	0.393
## as.factor(educ)5		-0.187	0.830	0.671	-0.28	0.781
## as.factor(educ)6		-0.534	0.586	1.119	-0.48	0.633

```
##
```

```
## Likelihood ratio test=16.3 on 5 df, p=0.00605
```

```
## n= 432, number of events= 114
```

## CATEGORICAL VARIABLES CODED AS NUMERIC ...

The second model is preferable.

- Why?

To assess whether `educ` significantly adds to the second model:

- Conduct a likelihood ratio test using the change in  $\log(\text{partial likelihood})$  when `educ` is added to a model with `fin`
- Some theory is needed first.

## The partial likelihood

# REVIEW OF LIKELIHOOD-BASED INFERENCE

Suppose

- $L(\beta)$  is the likelihood for a one dimensional parameter  $\beta$  in a model, and
- $\ell(\beta) = \log(L(\beta))$

Tests of  $H_0 : \beta = \beta_0$  can be based on three asymptotically equivalent statistics.



## REVIEW OF LIKELIHOOD-BASED INFERENCE...

Wald statistic:

$$\frac{\hat{\beta} - \beta_0}{\text{s.e.}(\hat{\beta})} \sim N(0, 1).$$

Log-likelihood ratio statistic:

$$-2 \left[ \log(L(\hat{\beta})) - \log(L(\beta_0)) \right] \sim \chi^2(1).$$

Score test statistic:

$$U(\beta_0) = \frac{\partial}{\partial \beta} \Big|_{\beta=\beta_0} \ell(\beta) \sim N(0, 1)$$

All have multivariate versions for  $p$ -dimensional parameter vectors  $\beta$ , not shown here.

## REVIEW OF LIKELIHOOD-BASED INFERENCE. . .

The statistical literature has results on which test is preferred, but that is not covered here.

The tests are usually similar enough that choice of a statistic is based on convenience.

Example coming later.

## PARTIAL LIKELIHOOD INFERENCE

Inference can be conducted by treating the partial likelihood as though it satisfies all the regular likelihood properties.

The *log(partial likelihood)* is:

$$\begin{aligned}\ell(\beta) &= \log \left[ \prod_{j=1}^n \frac{e^{\beta \mathbf{z}_j}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{z}_\ell}} \right]^{\delta_j} \\ &= \log \left[ \prod_{j=1}^K \frac{e^{\beta \mathbf{z}_j}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{z}_\ell}} \right] \\ &= \sum_{j=1}^K \left[ \beta \mathbf{z}_j - \log \left[ \sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{z}_\ell} \right] \right] \\ &= \sum_{j=1}^K l_j(\beta)\end{aligned}$$

where  $l_j$  is the log-partial likelihood contribution at the  $j$ -th ordered death time.

## PARTIAL LIKELIHOOD INFERENCE

Suppose there is only one covariate ( $\beta$  is one-dimensional):

The *partial likelihood score equations* are:

$$U(\beta) = \frac{\partial}{\partial \beta} \ell(\beta) = \sum_{j=1}^n \delta_j \left[ Z_j - \frac{\sum_{\ell \in \mathcal{R}(\tau_j)} Z_{\ell} e^{\beta Z_{\ell}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_{\ell}}} \right]$$

$U(\beta)$  is a sum of *observed* minus *expected* values:

$$U(\beta) = \frac{\partial}{\partial \beta} \ell(\beta) = \sum_{j=1}^n \delta_j (Z_j - \bar{Z}_j)$$

where  $\bar{Z}_j$  is the *weighted average* of  $Z$  over the individuals in the risk set at time  $\tau_j$ .

Note that  $\beta$  is involved through the term  $\bar{Z}_j$ .

The maximum partial likelihood estimators can be found by solving  $U(\beta) = 0$ .

## PARTIAL LIKELIHOOD INFERENCE . . .

Analogous to standard likelihood theory, it can be shown that

$$\frac{(\hat{\beta} - \beta)}{\text{s.e.}(\hat{\beta})} \sim N(0, 1)$$

The variance of  $\hat{\beta}$  can be obtained by inverting the second derivative of the partial likelihood,

$$\text{var}(\hat{\beta}) \sim \left[ -\frac{\partial^2}{\partial \beta^2} \ell(\beta) \right]^{-1}$$

From the above expression for  $U(\beta)$ :

$$\frac{\partial^2}{\partial \beta^2} \ell(\beta) = \sum_{j=1}^n \delta_j \left[ -\frac{\sum_{\ell \in \mathcal{R}(\tau_j)} (Z_j - \bar{Z}_j)^2 e^{\beta Z_\ell}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}} \right]$$

The actual variance of  $\hat{\beta}$  is a function of the unknown  $\beta$ .

The *observed* information is calculated by substituting the partial likelihood estimate of  $\beta$  into the above formula for the variance.

## EXAMPLE: ROSSI RECIDIVISM DATA

Suppose the primary interest is in evaluating the effect of financial incentive on changing the rate of recidivism, after adjusting for

- age, race, mar, educ

```
library(survival)
library(eventtimedata)
data(rossi)
coxph(Surv(week, arrest) ~ fin + age + race + mar
      + as.factor(educ), data = rossi)
```

## EXAMPLE: ROSSI RECIDIVISM DATA ...

```
## Call:
## coxph(formula = Surv(week, arrest) ~ fin + age + race + mar +
##       as.factor(educ), data = rossi)
##
##               coef exp(coef) se(coef)      z      p
## finyes          -0.3853    0.6802   0.1921 -2.01 0.0448
## age             -0.0555    0.9460   0.0211 -2.63 0.0087
## raceother       -0.3405    0.7114   0.3103 -1.10 0.2725
## marnot married   0.4715    1.6024   0.3751  1.26 0.2087
## as.factor(educ)3 0.6344    1.8858   0.5200  1.22 0.2225
## as.factor(educ)4 0.2362    1.2665   0.5420  0.44 0.6629
## as.factor(educ)5 -0.2572    0.7732   0.6724 -0.38 0.7021
## as.factor(educ)6 -0.5018    0.6054   1.1219 -0.45 0.6546
##
## Likelihood ratio test=30  on 8 df, p=0.000212
## n= 432, number of events= 114
```

# INTERPRETATION

- Financial incentive appears to significantly reduce the rate of recidivism by approximately 32% (adjusted hazard ratio 0.68).
- The likelihood ratio ( $LR$ ) test (30 on 8 df) is highly significant; this full set of covariates is significantly associated with the rate of recidivism.
  - Formally, the  $LR$  test is a test of  $H_0 : \beta = \mathbf{0}$ .
  - The statistic is  $-2 \left[ \ell(\hat{\beta}) - \ell(\mathbf{0}) \right]$ .
  - It has 8 df, since the difference in the number of parameters between the full model and the null model is 8.
  - The  $LR$  test is analogous to an  $F$ -test in linear regression



## INTERPRETATION...

- Age is significantly associated with recidivism: younger participants have a rate that is about 5% lower.
  - Why is the  $p$ -value so small when the effect size is also very small?
- Confidence intervals for the adjusted hazard ratio for age can be found by calculating a confidence interval for  $\hat{\beta}$  then exponentiating.<sup>4</sup>

```
beta.hat = -0.0555; se.beta.hat = 0.0211  
m = 1.96*se.beta.hat  
exp(beta.hat - m); exp(beta.hat + m)
```

```
## [1] 0.9076867
```

```
## [1] 0.9859556
```

---

<sup>4</sup>Possible to do this entirely within R without any 'hand' substitutions; method shown later.

## INTERPRETATION . . .

- Race and marital status do not seem to be associated with a change in rate of recidivism.
- The coefficients for these variables estimate the change in rate compared to the baseline categories of `race = black` and `mar = married`.

## INTERPRETATION. . .

Use a  $LR$  test to assess whether `educ` should be in the model.

- The models with and without `educ` are *nested* models.
  - Suppose that  $\beta_1, \dots, \beta_q$  are the parameter values for the 4 variables that do not include `educ` ( $q = 4$ ), and that
  - $\beta_{q+1}, \dots, \beta_p$  are the 4 parameter values for `educ`.
- In the full model, the  $LR$  statistic is

$$-2 \left[ \ell(\hat{\beta}) - \ell(\mathbf{0}) \right]$$

- In the model that does not include `educ`, the  $LR$  statistic is

$$-2 \left[ \ell(\hat{\beta}_1, \dots, \hat{\beta}_q) - \ell(\mathbf{0}) \right]$$

- The difference between the  $LR$  statistics for the full and reduced models is

$$-2 \left[ \ell(\hat{\beta}_1, \dots, \hat{\beta}_p) - \ell(\hat{\beta}_1, \dots, \hat{\beta}_q) \right]$$

## INTERPRETATION. . .

- The likelihood ratio test for educ can be calculated as the difference in the likelihood test statistics between the two models.
  - The difference has  $p - q$  degrees of freedom. In this case,  $p - q = 4$ .
- This general principle applies to  $LR$  tests in nested models.

## INTERPRETATION...

Fit the smaller model without educ to obtain the  $LR$  statistic.

```
## Call:
## coxph(formula = Surv(week, arrest) ~ fin + age + race + mar,
##       data = rossi)
##
##               coef exp(coef) se(coef)      z      p
## finyes          -0.351    0.704    0.190 -1.84 0.0654
## age             -0.065    0.937    0.021 -3.10 0.0019
## raceother       -0.244    0.784    0.306 -0.80 0.4258
## marnot married   0.508    1.663    0.373  1.36 0.1735
##
## Likelihood ratio test=21.2  on 4 df, p=0.000289
## n= 432, number of events= 114
```

- The difference in  $LR$  statistics is  $30 - 21.2 = 8.8$ ; with 4 degrees of freedom,  $p = 0.066$ .
- Does education seem associated with a change in rate of recidivism?

# DOING THE CALCULATIONS DIRECTLY IN R

```
library(survival)
library(eventtimedata)
data(rossi)
full.model <- coxph(Surv(week, arrest) ~ fin + age + race + mar
                    + as.factor(educ), data = rossi)
smaller.model <- coxph(Surv(week, arrest) ~ fin + age + race + mar,
                      data = rossi)
anova(full.model, smaller.model)

## Analysis of Deviance Table
## Cox model: response is Surv(week, arrest)
## Model 1: ~ fin + age + race + mar + as.factor(educ)
## Model 2: ~ fin + age + race + mar
##      loglik  Chisq Df P(>|Chi|)
## 1 -660.38
## 2 -664.78 8.7961 4 0.0664 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## DOING THE CALCULATIONS DIRECTLY IN R ...

```
library(survival)
library(eventtimedata)
data(rossi)
full.model <- coxph(Surv(week, arrest) ~ fin + age + race + mar
  + as.factor(educ), data = rossi)
summary(full.model)$conf.int
```

##		exp(coef)	exp(-coef)	lower .95	upper .95
##	finyes	0.6802145	1.4701245	0.46682820	0.9911393
##	age	0.9460051	1.0570768	0.90760574	0.9860291
##	raceother	0.7113791	1.4057202	0.38719854	1.3069788
##	marnot married	1.6024434	0.6240470	0.76829913	3.3422202
##	as.factor(educ)3	1.8858294	0.5302707	0.68054873	5.2257132
##	as.factor(educ)4	1.2664805	0.7895897	0.43778833	3.6638090
##	as.factor(educ)5	0.7732286	1.2932786	0.20698171	2.8885765
##	as.factor(educ)6	0.6054137	1.6517631	0.06716234	5.4573103

## Additional details on the Cox model



# THE MODEL FOR TWO GROUPS

When there are no tied failure times in a model with two groups,

- The score test for  $\beta = 0$  is the log rank test. (Derivation not shown.)

Thus, in a model with two groups

- The log-rank test, the PH model estimate, and the KM or cumulative hazard provide complementary information.

## TIED FAILURE TIMES

The PH model assumes a continuous hazard (i.e., no tied failure times). There are 4 methods to adjust for tied failure times:

1. Cox's (1972) modification: "discrete" method
2. Exact method (Kalbfleisch and Prentice)
3. Peto-Breslow method
4. Efron's (1977) method

All have advantages/disadvantages.

All are available in R, but Efron is the default. It does well in most settings.

Breslow is the default in SAS and Stata.

See Kleinbaum and Moeschberger for a detailed treatment.

# STRATIFIED COX MODELS

Randomized clinical trials often stratify the randomization to force balance within strata.

The PBT01 trial stratified the randomization on cycle of complete response, which was accounted for in the log-rank statistic.

A Cox model also allows stratification.

- Suppose  $X$  is a categorical variable with  $S$  levels, and  $\mathbf{Z}$  is a covariate.
- In a stratified Cox model, the hazard for the  $s^{th}$  level of  $X$  is

$$\lambda(t; X, \mathbf{Z}) = \lambda_{0s}(t) \exp(\beta \mathbf{Z})$$

- There is a different baseline hazard for each level of  $X$  but the 'effect' of  $\mathbf{Z}$  is the same across strata.

## EXAMPLE: PBT01

```
library(survival)
library(eventtimedata)
data("pbt01")

coxph(Surv(survival,died) ~ treatment + strata(cycle.of.resp),
      data = pbt01)
```

```
## Call:
## coxph(formula = Surv(survival, died) ~ treatment + strata(cycle.of.resp),
##       data = pbt01)
##
##               coef exp(coef) se(coef)      z      p
## treatmentcontrol -0.231      0.794    0.192 -1.2 0.23
##
## Likelihood ratio test=1.45 on 1 df, p=0.228
## n= 184, number of events= 114
```

## STRATIFIED COX MODELS . . .

The algebra for the stratified  $PL$  is more difficult than a direct explanation.

1. A separate  $PL$  is calculated for each strata.
2. The separate  $PL$ s are multiplied together, then
3. Maximized with respect to the common  $\beta$ .
4. Estimated variances use the information matrix for the combined  $PL$ .

# ESTIMATING SURVIVAL FUNCTIONS WITH PH

In the Cox PH model  $\lambda_i(t, \mathbf{Z}) = \lambda_0(t) \exp(\beta \mathbf{Z})$ .

What does this imply about the survival function,  $S_z(t)$ , for the  $i$ -th individual with covariates  $\mathbf{Z}_i$ ?

Using the relationship between the hazard and the survival function, for the baseline (reference) group:

$$S_0(t) = e^{-\int_0^t \lambda_0(u) du} = e^{-\Lambda_0(t)}$$

# ESTIMATING COVARIATE-SPECIFIC SURVIVAL FUNCTIONS WITH PH ...

For the  $i$ -th patient with covariates  $\mathbf{Z}_i$ :

$$\begin{aligned} S_i(t) &= e^{-\int_0^t \lambda_i(u) du} = e^{-\Lambda_i(t)} \\ &= e^{-\int_0^t \lambda_0(u) \exp(\beta \mathbf{Z}_i) du} \\ &= e^{-\exp(\beta \mathbf{Z}_i) \int_0^t \lambda_0(u) du} \\ &= \left[ e^{-\int_0^t \lambda_0(u) du} \right]^{\exp(\beta \mathbf{Z}_i)} \quad \text{since } [e^b]^a = e^{ab} \\ &= [S_0(t)]^{\exp(\beta \mathbf{Z}_i)} \\ &= [S_0(t)]^{HR} \end{aligned}$$

## ESTIMATING COVARIATE-SPECIFIC SURVIVAL FUNCTIONS WITH PH...

So after estimating  $\hat{\beta}$  and using the *MPLE*,

$$\hat{S}_i(t) = [\hat{S}_0(t)]^{\widehat{HR}}$$



# USING R TO PLOT ESTIMATED SURVIVAL FUNCTIONS FROM A COX MODEL

Will use the PBT01 data

Easier first to examine an model not stratified by cycle of initial response.

Notice the effect on the curves of the PH model

```

library(survival)
library(eventtimedata)
data("pbt01")

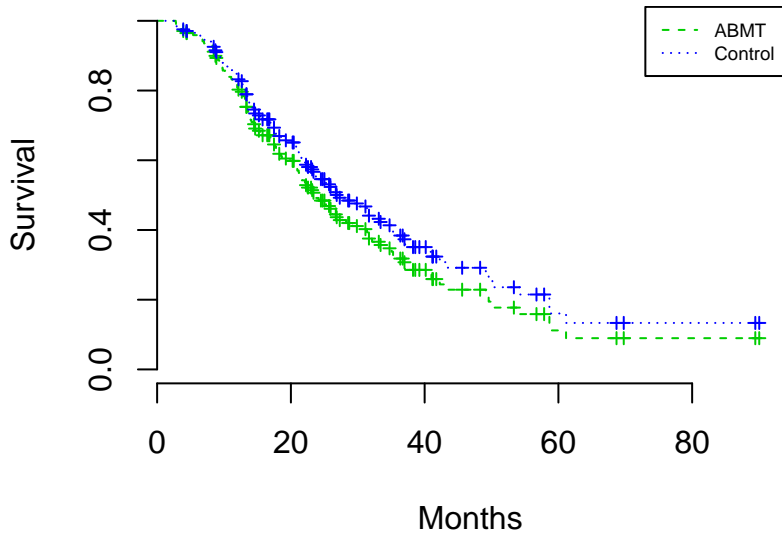
pbt01.ph = coxph(Surv(survival,died) ~ treatment,
                 data = pbt01)
newdata = data.frame(treatment = c("abmt", "control"))

plot(survfit(pbt01.ph, newdata), xlab = "Months", ylab="Survival",
     lty = 2:3, col = 3:4, mark.time = TRUE,
     axes = FALSE,
     main = "Probability of Surviving, Assuming PH",
     cex = 0.6,
     cex.main = 0.8)
axis(1)
axis(2)

legend("topright", lty = 2:3, col = 3:4,
      c("ABMT", "Control"),
      cex=0.6)

```

## Probability of Surviving, Assuming PH



## EXAMPLE: PBT01

Plot the estimated survival functions, by treatment, for each of the two strata defined by cycle of response.

- Assuming a proportional hazards effect for ABMT

## EXAMPLE: PBT01

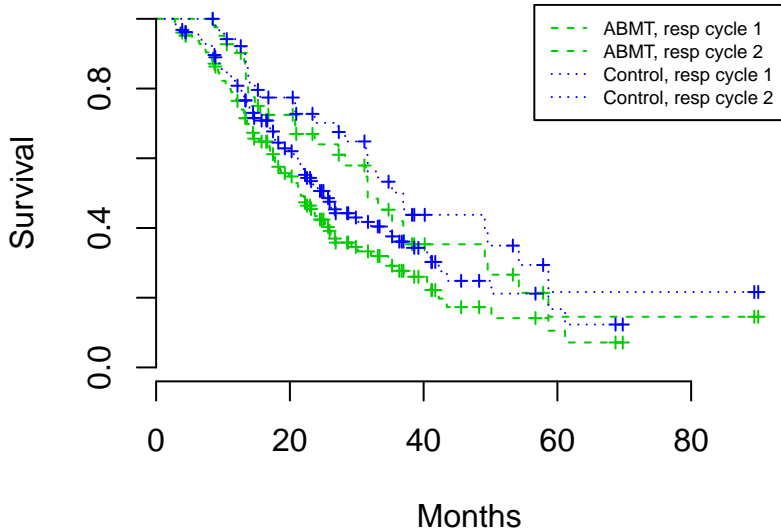
```
library(survival)
library(eventtimedata)
data("pbt01")

pbt01.ph = coxph(Surv(survival,died) ~ treatment + strata(cycle.of.resp),
                 data = pbt01)
newdata = data.frame(treatment = c("abmt", "control"))

plot(survfit(pbt01.ph, newdata), xlab = "Months", ylab="Survival",
     lty = c(2,2,3,3), col = c(3,3,4,4), mark.time = TRUE,
     axes = FALSE,
     main = "Probability of Surviving, ABMT",
     cex = 0.6,
     cex.main = 0.8)
axis(1)
axis(2)

legend("topright", lty = c(2,2,3,3), col = c(3,3,4,4),
      c("ABMT, resp cycle 1", "ABMT, resp cycle 2",
        "Control, resp cycle 1", "Control, resp cycle 2"),
      cex=0.6)
```

## Probability of Surviving, ABMT



Time-varying (time-dependent) covariates

## BACKGROUND AND INTRODUCTION

Conceptually (i.e., mathematically) there is little problem in extending the *PH* model to covariates that depend on time.

The model becomes

$$\lambda(t; Z) = \lambda_0(t) \exp[\beta Z(t)]$$

However, there can be problems in interpretation, especially when the time-dependent covariate is essentially a surrogate for the event of interest.



# ISSUES IN USING TIME-DEPENDENT COVARIATES

The coding of the time-dependent covariate must be done with care.

- Software has not always been available. Robust routines are now available in R and SAS.
- The data structure required in R also works in SAS, but SAS is more sometimes flexible.

This is best illustrated with an example: an intervention study in drug addiction.<sup>5</sup>

---

<sup>5</sup>Hosmer and Lemeshow, *Applied Survival Analysis*

## EXAMPLE: UMARU IMPACT STUDY (UIS)

Data is from the University of Massachusetts AIDS Research Unit (UMARU) IMPACT Study.<sup>6</sup>

- 5-year collaborative project comprised of two concurrent randomized trials of residential treatment for drug abuse
  - *Program A*: Randomized 444 subjects to a 3- or 6-month program of health education and relapse prevention. Clients taught to recognize “high-risk” situations that are triggers to relapse, and taught skills to cope with these situations without using drugs.
  - *Program B*: Randomized 184 participants to a 6- or 12-month program with highly structured lifestyle in a communal living setting.
- 628 participants total

---

<sup>6</sup>Dataset `uis` in the package `eventtimedata`

## VARIABLES IN UIS

id	Subject ID (1-628)
age	Age in years
beck	Beck depression score
hercoc	Heroin or cocaine use prior to entry
ivhx	IV Drug use at admission
ndrugtx	Number of previous drug treatments
race	Race (white, other)
treat	Treatment assignment (short vs long)
site	Treatment program
los	Length of stay in Treatment (days)
time	Time to return to drug use (days)
censor	Indicator of drug use relapse (1 = yes, 0 = censored)

Details about coding in package documentation.

# AN INITIAL MODEL FOR UIS

Initial model estimates treatment effect, adjusting for

- Beck depression score, age, and race
- interaction between age and site
- interaction between race and site

```
library(survival)
library(eventtimedata)
coxph(formula = Surv(time, censor) ~ treat + beck +
age * race + strata(site), data = uis)
```

Model interpretation discussed with the next slide

## AN INITIAL MODEL FOR UIS...

```
## Call:
## coxph(formula = Surv(time, censor) ~ treat + beck + age * race +
##       strata(site), data = uis)
##
##               coef exp(coef) se(coef)      z      p
## treatlong      -0.24427    0.78327  0.09286 -2.63 0.0085
## beck           0.01011    1.01017  0.00485  2.09 0.0370
## age            -0.01252    0.98756  0.00839 -1.49 0.1358
## raceother      -0.15116    0.85971  0.61994 -0.24 0.8074
## age:raceother  -0.00492    0.99509  0.01896 -0.26 0.7953
##
## Likelihood ratio test=23.2  on 5 df, p=0.000308
## n= 585, number of events= 471
##      (43 observations deleted due to missingness)
```

# WHAT IS THE EFFECT OF COMING OFF TREATMENT?

Hosmer, Lemeshow, and May speculated that the the important effect may simply be stopping treatment, regardless of whether treatment is long or short.

- Natural to look at the association between outcome and a time-dependent covariate that records when a patient stops treatment

The variable `treat.off` will take values

- 0 while a subject is on treatment
- 1 starting at the time treatment stops

# TIME-DEPENDENT COVARIATES IN R

The work for time-dependent covariates in R is done in the data structure.

If a covariate changes value over time, each case in the dataset should have a row for each interval in which the covariate is constant. Each row contains

- All fixed covariates and the outcome
- Two variables that indicate the start (usually labeled *tstart*) and stop time (labeled *tstop*) for that interval
- The value of the time-dependent covariate in that interval

## UIS DATA WITH TIME-DEPENDENT VARIABLE FOR TREATMENT

##	id	age	beck	los	time	censor	tstart	tstop	relapse	treat.off
## 1	1	39	9	123	188	1	0	123	0	0
## 2	1	39	9	123	188	1	123	188	1	1
## 3	2	33	34	25	26	1	0	25	0	0
## 4	2	33	34	25	26	1	25	26	1	1
## 5	3	33	10	7	207	1	0	7	0	0
## 6	3	33	10	7	207	1	7	207	1	1
## 7	4	32	20	66	144	1	0	66	0	0
## 8	4	32	20	66	144	1	66	144	1	1

Original file has one row per case.

Code to create this file coming later.



## TWO MODELS, INVOLVING ONLY TREATMENT

```
## Call:
## coxph(formula = Surv(time, censor) ~ treat, data = uis)
##
##               coef exp(coef) se(coef)      z      p
## treatlong -0.232      0.793   0.089 -2.6 0.0092
##
## Likelihood ratio test=6.78  on 1 df, p=0.00921
## n= 628, number of events= 508

## Call:
## coxph(formula = Surv(tstart, tstop, relapse) ~ treat * treat.off,
##       data = uis2)
##
##               coef exp(coef) se(coef)      z      p
## treatlong      -0.523      0.593   0.226 -2.32  0.020
## treat.off       2.280      9.780   0.187 12.21 <2e-16
## treatlong:treat.off 0.621      1.861   0.246  2.52  0.012
##
## Likelihood ratio test=389  on 3 df, p=0
## n= 1174, number of events= 508
```

# CODE TO CREATE TIME-DEPENDENT COVARIATE AND FIT MODELS

```
library(survival)
library(eventtimedata)
data("uis")

#model 1: original dataset
coxph(Surv(time, censor) ~ treat, data = uis)

#create file for time-dependent covariate
uis.temp = tmerge(uis, uis, id = id, relapse = event(time, censor))
uis2 = tmerge(uis.temp, uis.temp, id=id, treat.off = tdc(los))

#model 2: time-dependent covariate
coxph(Surv(tstart, tstop, relapse) ~ treat*treat.off, data = uis2)
```

## A LARGER MODEL

```
## Call:
## coxph(formula = Surv(tstart, tstop, relapse) ~ treat * treat.off +
##      beck + age + race + strata(site), data = uis2)
##
##              coef exp(coef) se(coef)      z      p
## treatlong      -0.53187    0.58751  0.24247 -2.19 0.02827
## treat.off       2.24928    9.48089  0.19729 11.40 < 2e-16
## beck           0.00927    1.00931  0.00474  1.95 0.05076
## age            -0.00291    0.99710  0.00747 -0.39 0.69729
## raceother      -0.41081    0.66311  0.11133 -3.69 0.00022
## treatlong:treat.off 0.68670    1.98715  0.26391  2.60 0.00927
##
## Likelihood ratio test=370  on 6 df, p=0
## n= 1099, number of events= 471
##      (75 observations deleted due to missingness)
```

## CODING TIME-DEPENDENT COVARIATES

The function `tmerge` works in many situations. See the [R vignette](#) by Therneau, Crowson and Atkinson for additional examples.

[Fox](#) shows R code for the Rossi dataset, converting from a wide-format to the long-format required by R for the monthly employment status variable.

# PARTIAL LIKELIHOOD WITH TIME-VARYING COVARIATES

Suppose there are  $K$  distinct failure times

- Let  $(\tau_1, \dots, \tau_K)$  represent the  $K$  ordered, distinct death times. For now, assume there are no tied death times.
- Let  $\mathcal{R}(t) = \{i : x_i \geq t\}$  denote the set of individuals who are at risk for failure at time  $t$ .
- Let  $i_j$  denote the label or identity of the individual who fails at time  $\tau_j$ , including the value of his/her time-varying covariate during their time in the study

$$\{Z_{i_j}(t), t \in [0, \tau_j]\}$$

- Let  $H_j$  denote the history of the entire data set, up to the  $j$ -th death or failure time, including the *time* of the failure, but not the identity of the case who fails, also including the values of all covariates for everyone up to and including time  $\tau_j$ .

## PARTIAL LIKELIHOOD WITH TIME-VARYING COVARIATES. . .

The partial likelihood in this setting can be written as

$$L(\beta) = \prod_{j=1}^d \frac{\exp(\beta \mathbf{Z}_{jj})}{\sum_{\ell \in \mathcal{R}(\tau_j)} \exp(\beta \mathbf{Z}_{\ell j})},$$

where  $\mathbf{Z}_{\ell j}$  is a way to denote the value of the covariate vector for the  $\ell$ -th person at the  $j$ -th death time, i.e.

$$\mathbf{Z}_{\ell j} = \mathbf{Z}_{\ell}(\tau_j)$$

Inference proceeds similarly as in the standard case. The main difference is that the values of  $\mathbf{Z}$  changes at each risk set.

## Derivations

## ADDITIONAL DERIVATION OF THE PARTIAL LIKELIHOOD

In general, the likelihood contributions for censored data fall into two categories:

- *Individual is censored at  $X_i$ :*

$$L_i(\beta) = S(X_i) = \exp\left[-\int_0^{X_i} \lambda_i(u) du\right]$$

- *Individual fails at  $X_i$ :*

$$L_i(\beta) = S(X_i)\lambda_i(X_i) = \lambda_i(X_i) \exp\left[-\int_0^{X_i} \lambda_i(u) du\right]$$

Thus, everyone contributes  $S(X_i)$  to the likelihood, and only those who fail contribute  $\lambda_i(X_i)$ .



## ADDITIONAL DERIVATION OF THE PARTIAL LIKELIHOOD . . .

The total likelihood will be :

$$L(\beta) = \prod_{i=1}^n \lambda_i(X_i)^{\delta_i} \exp\left[-\int_0^{X_i} \lambda_i(u) du\right]$$

The above likelihood holds for all censored survival data, with general hazard function  $\lambda(t)$ .

It does not depend on the PH assumption.

## ADDITIONAL DERIVATION OF THE PARTIAL LIKELIHOOD . . .

Now multiply and divide by the term  $\left[ \sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i) \right]^{\delta_i}$ :

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\lambda_i(X_i)}{\sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i)} \right]^{\delta_i} \left[ \sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i) \right]^{\delta_i} \exp \left[ - \int_0^{X_i} \lambda_i(u) du \right]$$

Cox (1972) argued that the first term in this product contains almost all the information about  $\beta$ , while the last two terms contain information about the baseline hazard  $\lambda_0(t)$ .

## ADDITIONAL DERIVATION OF THE PARTIAL LIKELIHOOD . . .

Under the Cox PH assumption, the first term can be written as

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n \left[ \frac{\lambda_i(X_i)}{\sum_{j \in \mathcal{R}(X_i)} \lambda_i(X_i)} \right]^{\delta_i} \\ &= \prod_{i=1}^n \left[ \frac{\lambda_0(X_i) \exp(\beta \mathbf{Z}_i)}{\sum_{j \in \mathcal{R}(X_i)} \lambda_0(X_i) \exp(\beta \mathbf{Z}_j)} \right]^{\delta_i} \\ &= \prod_{i=1}^n \left[ \frac{\exp(\beta \mathbf{Z}_i)}{\sum_{j \in \mathcal{R}(X_i)} \exp(\beta \mathbf{Z}_j)} \right]^{\delta_i} \end{aligned}$$

This is the partial likelihood defined by Cox; it does not depend on the underlying hazard  $\lambda_0(\cdot)$ .

Cox recommended treating this as an ordinary likelihood for making inferences about  $\beta$  in the presence of the nuisance parameter  $\lambda_0(\cdot)$ .