Part [1.0] – Measures of Classification Accuracy for the Prediction of Survival Times

- Patrick J. Heagerty PhD
- Department of Biostatistics
- University of Washington

Review: Cox Regression Model

- Introduction
  - Cox (1972)
- Model
  - hazard model
  - log hazard, survival models
  - PH assumption
  - Interpretation of coefficients
  - Specific examples

Cox Regression Model

- Estimation
  - Coefficients
  - Partial likelihood
  - Approximation for ties
  - Survival curve(s)
  - Hazard curve(s)
- Stratification
  - Using covariate
  - Using "true" stratification

Cox (1972)

- D.R. Cox (1972) "Regression Models and Life-Tables" (with discussion) *JRSS-B*, 74: 187-220.
- "The present paper is largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables and more generally to the incorporation of regression-like arguments into life-table analysis." (p. 187)
- $\boxed{\text{Model proposed:}}$ $\lambda(t \mid X) = \lambda_0(t) \cdot \exp(X\beta)$
- "In the present paper we shall, however, concentrate on exploring the consequence of allowing $\lambda_0(t)$ to be arbitrary, main interest being in the regression parameters." (p. 190)
- "A Conditional Likelihood" – later called Partial Likelihood.
- Score Test = LogRank Test

- Discussion:
  - "Mr. Richard Peto (Oxford University): I have greatly enjoyed Professor Cox's paper. It seems to me to formulate and to solve the problem of regression of prognosis on other factors perfectly, and it is very pretty."
- Impact:
  - Science Citation Index: 29,140 citations (29 June 2015)
  - David R. Cox is knighted in 1985 in recognition of his scientific contributions.

Sir David R. Cox

Cox Regression Model

- $\boxed{\text{Response Variable:}}$
  - ▷ Observed: (Y_i, δ_i)
  - ▷ Of Interest: T_i, or $\lambda(t)$
- T_i survival, with distribution given by:
  - ▷ Survival function: $S(t)$
  - ▷ Hazard function: $\lambda(t)$
- $\boxed{\text{Observed Covariates:}}$ $X_1, X_2, \ldots, X_k$
  - ▷ For subject j we observe: $(Y_j, \delta_j), X_{1j}, X_{2j}, \ldots, X_{kj}$
- **IDEA**: same as with other regression models – Model relates the covariates $X_1, \ldots, X_k$ to the distribution (either $S(t)$ or $\lambda(t)$) of the response variable of interest, T.

Cox Regression Model

- $\boxed{\text{Model:}}$

$$\lambda(t \mid X_1, X_2, \ldots, X_k) \;=\; \lambda_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$$

- $\boxed{\text{Model:}}$ alternatively expressed as

$$\log \lambda(t \mid X_1, \ldots, X_k) \;=\; \log \lambda_0(t) + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

$$S(t \mid X_1, \ldots, X_k) \;=\; [S_0(t)]^{[\exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)]}$$

- Note definitions:
  - ▷ $\lambda_0(t) = \lambda(t \mid X_1 = 0, X_2 = 0, \ldots, X_k = 0)$
  - ▷ $S_0(t) = S(t \mid X_1 = 0, X_2 = 0, \ldots, X_k = 0)$

Interpreting Cox Regression Coefficients

- $\boxed{\text{Proportional Hazards:}}$

$$\begin{aligned} \text{RR} &= \frac{\lambda(t \mid X_1, X_2, \ldots, X_k)}{\lambda(t \mid X_1 = 0, X_2 = 0, \ldots, X_k = 0)} \\ &= \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) \end{aligned}$$

- `RR` above is: "Relative risk, or hazard, of death comparing subjects with covariate values $(X_1, X_2, \ldots, X_k)$ to subjects with covariate values $(0, 0, \ldots, 0)$."

Interpreting Cox Regression Coefficients

- $\boxed{\text{In General:}}$
  - ▷ β_m is the log RR (or log hazard ratio, log HR) comparing subjects with $X_m = (x+1)$ to subjects with $X_m = x$, given that all other covariates are constant (ie. the same for the groups compared).

$$\frac{\lambda(t \mid X_1, \ldots, \overbrace{X_m = (x+1)}^{\texttt{here}} \ldots, X_k)}{\lambda(t \mid X_1, \ldots, \underbrace{X_m = (\ x\ ),}_{\texttt{here}} \ldots, X_k)} =$$

$$\frac{\lambda_0(t)\exp(\beta_1 X_1 + \ldots \beta_m(x+1) + \ldots + \beta_k X_k)}{\lambda_0(t)\exp(\beta_1 X_1 + \ldots \beta_m(x) + \ldots + \beta_k X_k)} = \exp(\beta_m)$$

Interpreting Cox Regression Coefficients

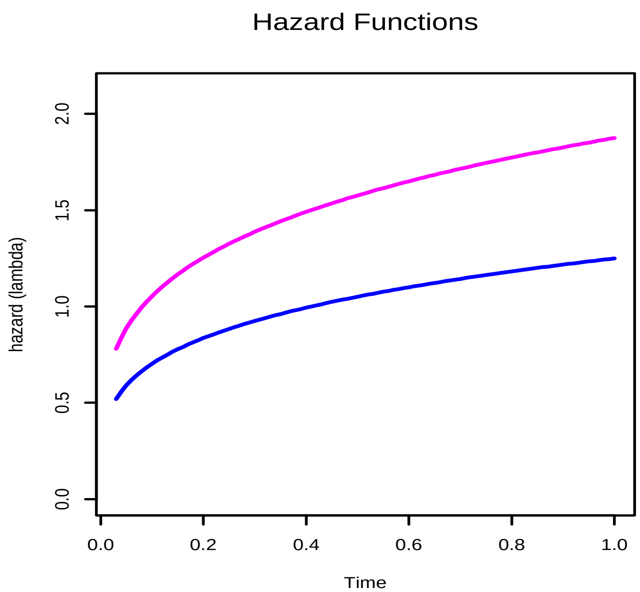
- $\boxed{\text{The RR Comparing 2 Covariate Values (vectors):}}$

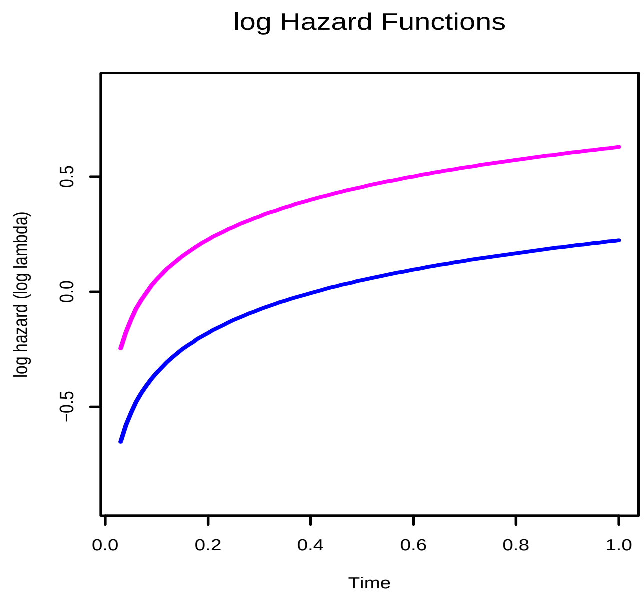
  ▷ RR comparing $(X_1, X_2, \ldots, X_k)$ to $(X'_1, X'_2, \ldots, X'_k)$.

$$\begin{aligned}
\texttt{RR}(X \texttt{ vs. } X') &= \frac{\lambda(t \mid X_1, X_2, \ldots, X_k)}{\lambda(t \mid X'_1, X'_2 \ldots, X'_k)} \\
&= \exp\Big[\ \beta_1 \cdot (X_1 - X'_1) + \\
&\qquad\quad \beta_2 \cdot (X_2 - X'_2)\ + \\
&\qquad\quad \ldots + \\
&\qquad\quad \beta_k \cdot (X_k - X'_k)\ \Big]
\end{aligned}$$

Cox Model Examples

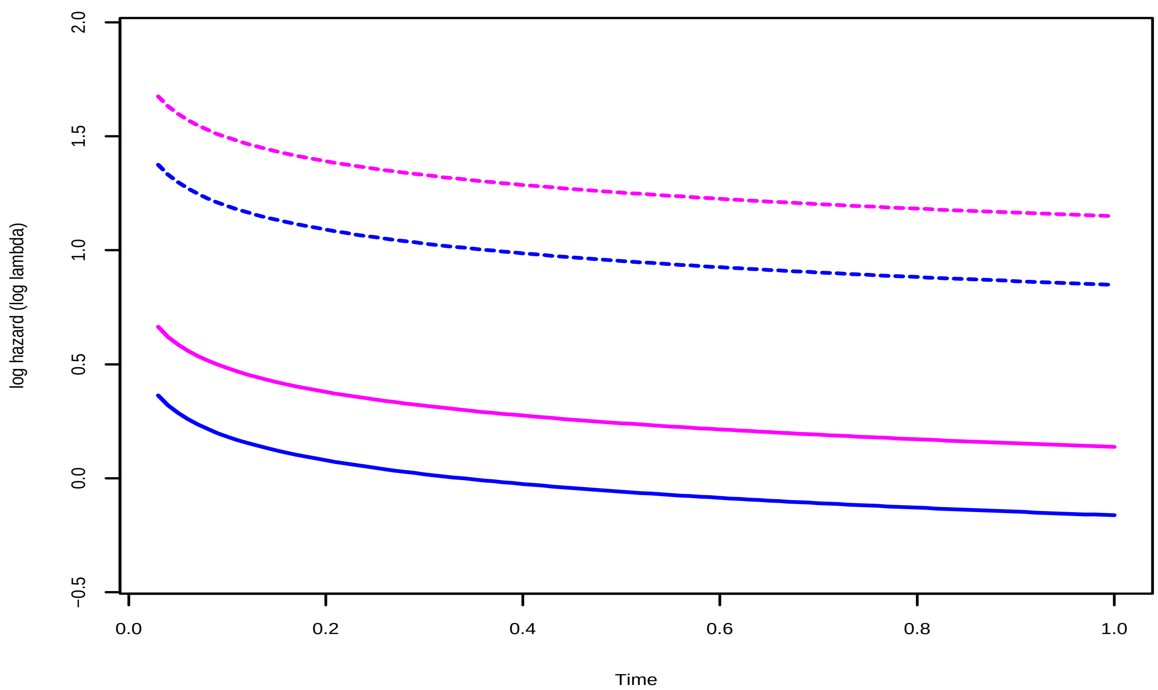
- $\boxed{\textcolor{red}{1}:}$ One dichotomous covariate
  - $X_E = 1$ if exposed; $X_E = 0$ if not exposed.
  - $\lambda(t \mid X_E) = \lambda_0(t) \exp(\beta X_E)$





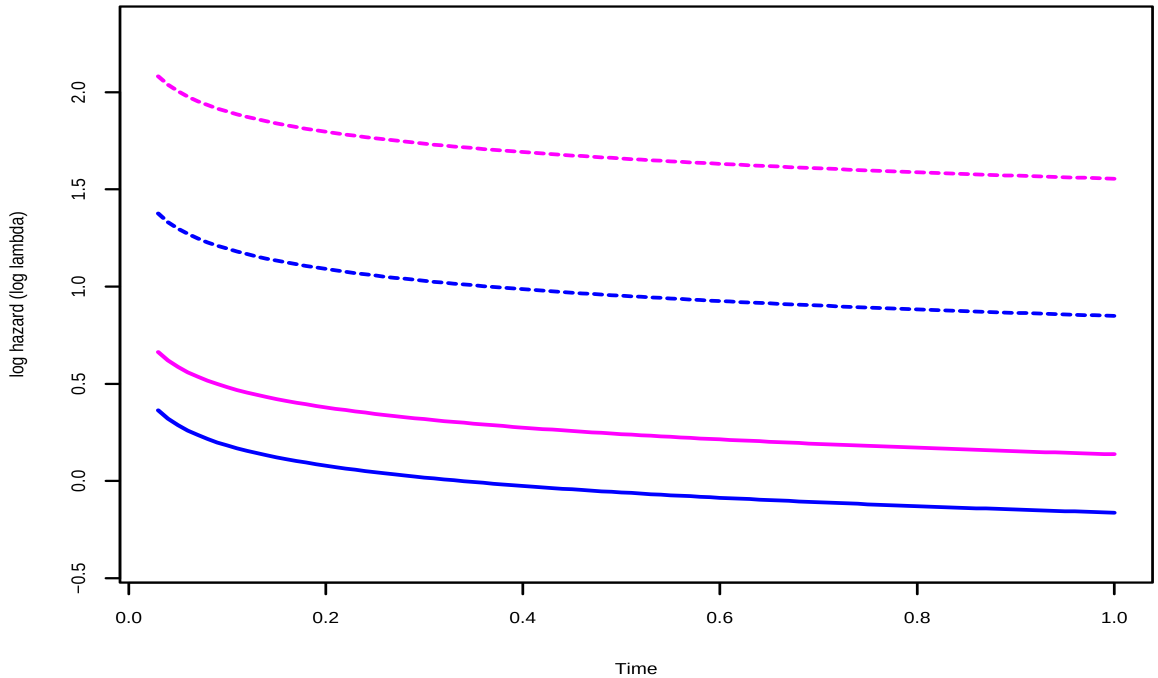
Cox Model Examples

- $\boxed{\mathbf{2}:}$ Dichotomous covariate; Dichotomous confounder
  - $X_C = 1$ if level 2; $X_C = 0$ if level 1.
  - $\lambda(t \mid X_E, X_C) = \lambda_0(t) \exp(\beta_1 X_E + \beta_2 X_C)$



Cox Model Examples

- $\boxed{\textbf{3}:}$ Dichotomous covariate; confounder; (interaction)
  - ▷ With interaction
  - ▷ $\lambda(t \mid X_E, X_C) = \lambda_0(t)\exp(\beta_1 X_E + \beta_2 X_C + \beta_3 X_E X_C)$



Cox Models: Comments

- In each example the hazard functions are "parallel" – that is, the change in hazard over time was the same for each covariate value.
- For regression models there are different possible tests for a hypothesis about coefficients: likelihood ratio; score; Wald. (more later!)
- The score test for example (1) with $H_0 : \beta = 0$ is the LogRank Test.
- The score test using "dummy variables" to code (4) groups with $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ is the same as the K-sample Heterogeneity test (generalization of LogRank).

Fitting the Cox Model

- Obtain estimates of $\beta_1, \beta_2, \ldots, \beta_k$ by maximizing the "partial likelihood" function:

$$PL(\beta_1, \beta_2, \ldots, \beta_k).$$

  - $\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_k$ are MPLE's
  - CI's for β_j using:

$$\widehat{\beta}_j \pm Z_{1-\alpha/2}\mathsf{SE}(\widehat{\beta}_j).$$

  - CI's for hazard ratio (HR) using:

$$\exp[\widehat{\beta}_j - Z_{1-\alpha/2}\mathsf{SE}(\widehat{\beta}_j)],\ \exp[\widehat{\beta}_j + Z_{1-\alpha/2}\mathsf{SE}(\widehat{\beta}_j)]$$

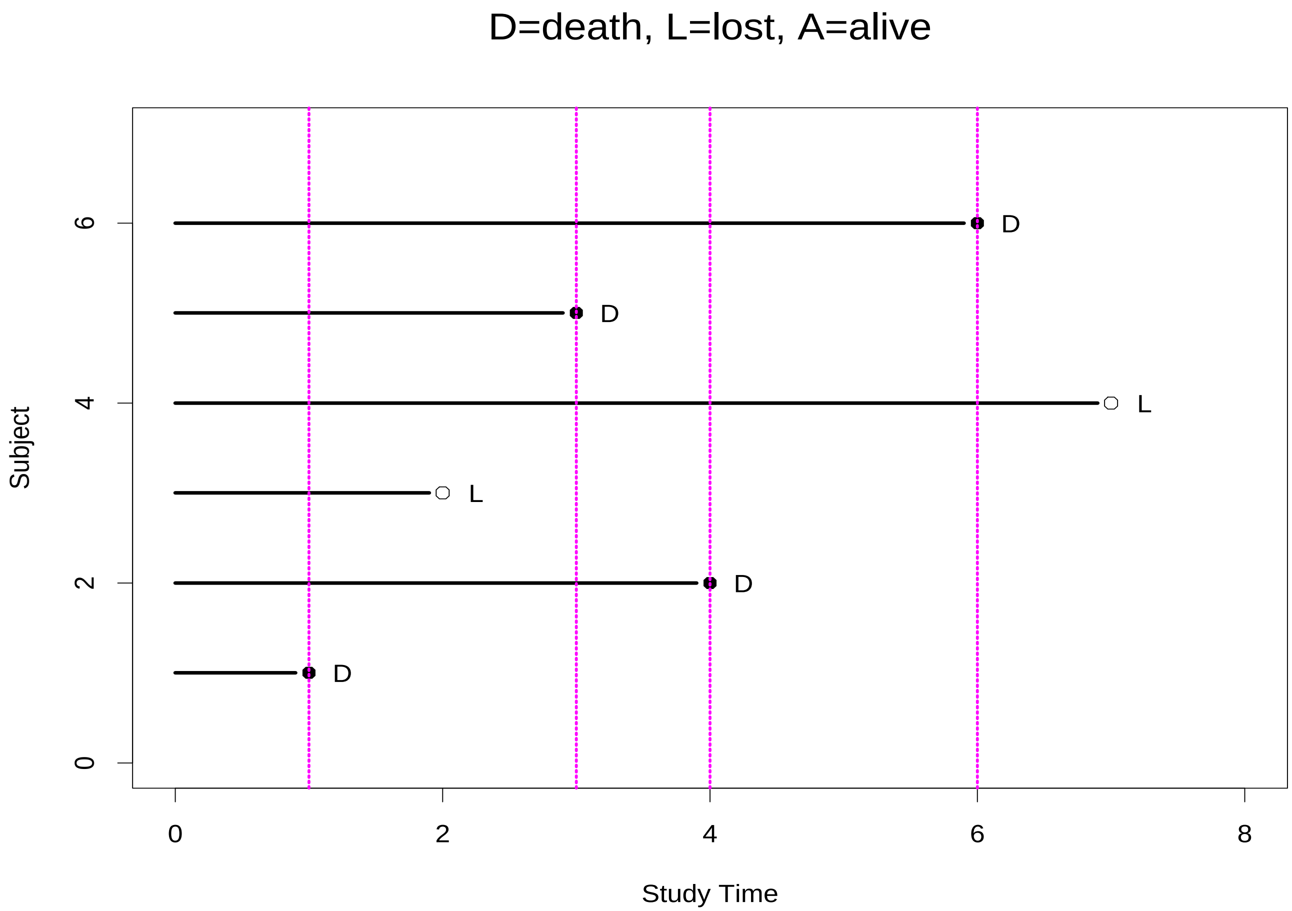
  - Wald test, score test, and likelihood ratio test similar to logistic regression. Now using the partial likelihood.

Partial Likelihood

- $\boxed{\text{Model:}}$ $\lambda(t \mid X_1, \ldots, X_k) = \lambda_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k)$
- $\boxed{\text{Order Data:}}$
  - ▷ $t_{(i)}$ is the ith ordered failure time.
  - ▷ Assume no **ties**, and let $X_{(i)} = (X_{1(i)}, X_{2(i)}, \ldots, X_{k(i)})$ be the covariates for the subject who dies at time $t_{(i)}$.
  - ▷ Let $\mathcal{R}_i$ denote the "risk set" at time $t_{(i)}$, which denotes all subjects with $Y_j \geq t_{(i)}$.
- Partial Likelihood: (no ties)

$$PL(\beta_1, \ldots, \beta_k) = \prod_{i=1}^{J} \frac{\exp(\beta_1 X_{1(i)} + \beta_2 X_{2(i)} + \ldots + \beta_k X_{k(i)})}{\sum_{j \in \mathcal{R}_i} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj})}$$

Risk Set Illustration



Risk Set Illustration

- Failure times: $t_{(1)} = 1, t_{(2)} = 3, t_{(3)} = 4, t_{(4)} = 6$.
- Risk sets:
  - $\mathcal{R}_1 = \{ \qquad \}$
  - $\mathcal{R}_2 = \{ \qquad \}$
  - $\mathcal{R}_3 = \{ \qquad \}$
  - $\mathcal{R}_4 = \{ \qquad \}$

Partial Likelihood – Justification

- Cox (1972) – "No information can be contributed about β by time intervals in which no failures occur because the component $\lambda_0(t)$ might conceivably be identically zero in such intervals."

- Cox (1972) – "We therefore argue conditionally on the set $\{t_{(i)}\}$ of instants at which failure occur."

- Cox (1972) – "For the particular failure at time $t_{(i)}$ conditional on the risk set, $\mathcal{R}_i$, the probability that the failure is on the individual as observed is:

$$\frac{\exp(\beta_1 X_{1(i)} + \beta_2 X_{2(i)} + \ldots + \beta_k X_{k(i)})}{\sum_{j \in \mathcal{R}_i} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj})}.$$

- $\boxed{\text{Note:}}$ This likelihood contribution has the exact same form as a (matched) logistic regression conditional likelihood.

Partial Likelihood – Justification

- **Q**: What is the probability of the observed data at time $t_{(i)}$ given that one person was observed to die among the risk set?

$$\text{Note} \quad : \quad P[T \in (t, t+\Delta t] \mid T \geq t] \approx \lambda(t) \cdot \Delta t$$

$$\text{Person who died} \quad : \quad \lambda_0(t) \exp(\beta_1 X_{1(i)} + \ldots + \beta_k X_{k(i)}) \Delta t = \ P_{(i)}$$

$$\text{Generic } j \text{ in } \mathcal{R}_i \quad : \quad \lambda_0(t) \exp(\beta_1 X_{1j} + \ldots + \beta_k X_{kj}) \Delta t = \ P_j$$

Partial Likelihood – Justification

- $\boxed{\text{Probability One Death, Was } (i) :}$

$$P_{(i)} \times (1 - P_1) \times (1 - P_2) \ldots \times \texttt{skip}(\texttt{i}) \times (1 - P_k)$$

- $\boxed{\text{Probability of One Death:}}$

$$\begin{aligned}
\text{P(One Death)} &= \text{P(} 1 \text{ died, others lived)} + \\
&\quad \text{P(} 2 \text{ died, others lived)} + \\
&\quad \ldots + \\
&\quad \text{P(} k \text{ died, others lived)} \\
\text{P(} j \text{ died, others lived)} &= P_j \times \prod_{k \neq j} (1 - P_k)
\end{aligned}$$

- $\boxed{\text{Note:}}$ $(1 - P_j) \approx 1$ for small Δt.

Partial Likelihood – Justification

- Now calculate the desired quantity:

$$
\begin{aligned}
\text{P(Observed Data} \mid \text{1 death)} &= \frac{\text{P(Only (i) Dies)}}{\text{P(One Death)}} \\
&= \frac{P_{(i)} \prod_{k \neq (i)} (1 - P_k)}{\sum_{j \in \mathcal{R}_i} P_j \prod_{k \neq j} (1 - P_k)} \\
&\approx \frac{P_{(i)}}{\sum_{j \in \mathcal{R}_i} P_j}
\end{aligned}
$$

$$
\begin{aligned}
\frac{P_{(i)}}{\sum_{j \in \mathcal{R}_i} P_j} &= \frac{\lambda_0(t) \exp(\beta_1 X_{1(i)} + \beta_2 X_{2(i)} + \ldots + \beta_k X_{k(i)}) \cdot \Delta t}{\sum_{j \in \mathcal{R}_i} \lambda_0(t) \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj}) \cdot \Delta t} \\
&= \frac{\exp(\beta_1 X_{1(i)} + \beta_2 X_{2(i)} + \ldots + \beta_k X_{k(i)})}{\sum_{j \in \mathcal{R}_i} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj})}
\end{aligned}
$$

Partial Likelihood – Comments

- Notice that our model is equivalent to

$$\log \lambda(t \mid X_1 \dots X_k) = \alpha(t) + \beta_1 X_1 + \dots \beta_k X_k$$

  where $\alpha(t) = \log \lambda_0(t)$, but the PL does not depend on $\alpha(t)$.

- Using the partial likelihood (PL) to estimate parameters provides estimates of the regression coefficients, β_j, only.
- The model is called "semi-parametric" since we only need to parameterize the effect of covariates, and do not say anything about the baseline hazard.
- **Q**: Why not just use standard maximum likelihood?
- **A**: To do so would require choosing a model for the baseline hazard, but we actually don't need to do that!

Partial Likelihood and Ties

- If there is more than one death at time $t_{(i)}$ then the denominator for the partial likelihood contribution will involve a large number of terms. For example if there are 20 people at risk at time $t_{(i)}$ and 3 die then there are "20 choose 3" $=$ 1140 terms.

- $\boxed{\text{Approximation}}$ (Breslow, Peto) `default in STATA`
  - ▷ The numerator can be calculated and represented using:
    - ∗ Sum X_1 for deaths: $s_{1i} = \sum_{j:Y_j=t_{(i)},\delta_j=1} X_{1j}$
    - ∗ Sum X_2 for deaths: $s_{2i} = \sum_{j:Y_j=t_{(i)},\delta_j=1} X_{2j}$ etc.
  - ▷ The approximation with D_i deaths at time $t_{(i)}$ is:

$$PL_A = \prod_{i=1}^{J} \frac{\exp(\beta_1 s_{1i} + \beta_2 s_{2i} + \ldots + \beta_k s_{ki})}{\left[\sum_{j\in\mathcal{R}_i} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj})\right]^{D_i}}$$

Comments on Ties

- If **continuous** times, T_i, then ties should not be an issue.
  - ▷ Time recorded in (days,minutes).
  - ▷ Modest sample size.
- If **discrete** times, $T_i \in [t_k, t_{k+1})$, recorded then consider methods appropriate for discrete-time data (e.g. variants on logistic regression)
  - ▷ See Singer & Willett (2003) chpts 10–12; H& L pp. 268-9.

Comments on Ties

- However, there is plenty of room between continuous and discrete.
  - ▷ $\boxed{\text{Example:}}$ **USRDS Data** = 200,000 subjects.
    - ∗ 25% annual mortality = 50,000 deaths/year.
    - ∗ 50,000 deaths/365 days = 137 deaths/day.
- Kalbfleisch & Prentice (2002), section 4.2.3 summarize options and relative pros/cons.
  - ▷ "Breslow method" – simple to implement/justify; some bias if discrete.
  - ▷ "Efron method" – also simple comp; performs well.
  - ▷ "exact method" – justified; comp challenge.
  - ▷ Should be minor issue in general, and if not then perhaps a discrete-time approach should be considered.

(Partial) Likelihood Ratio Tests

- $\boxed{\text{Full Model:}}$

$$\lambda(t|X) = \lambda_0(t)\exp(\beta_1 X_1 + \ldots + \beta_p X_p + \underbrace{\beta_{p+1} X_{p+1} + \ldots \beta_k X_k}_{\texttt{extra}})$$

- $\boxed{\text{Reduced Model:}}$

$$\lambda(t|X) = \lambda_0(t)\exp(\beta_1 X_1 + \ldots + \beta_p X_p)$$

- In order to test:
  - ▷ H_0 : Reduced model $\Leftrightarrow$ $H_0 : \beta_{p+1} = \ldots = \beta_k = 0$
  - ▷ H_1 : Full model $\Leftrightarrow$ H_1 : extra coeff $\neq 0$ somewhere
- Use the partial likelihood ratio statistic

$$X^2_{PLR} = [2\log P\mathcal{L}(\texttt{FullModel}) \; - \; 2\log P\mathcal{L}(\texttt{ReducedModel})]$$

(Partial) Likelihood Ratio Tests

- Under H_0 (reduced is correct) then $X^2_{PLR} \sim \chi^2(\mathtt{df} = (\mathtt{k} - \mathtt{p}))$
- Degrees of freedom, $\mathtt{df} = (\mathtt{k} - \mathtt{p})$, equals the number of parameters set to 0 by the null hypothesis.
- Application is for situations where the models are "nested" – the reduced model is a special case of the full model.
- Also can use Wald tests, and/or score tests. The PLR (Partial Likelihood Ratio) test is particularly useful when df> 1.
- The PLR statistic is equivalent (using a "double negative") to:

$$X^2_{PLR} = \{[-2 \log P\mathcal{L}(\mathtt{ReducedModel})] \; - \; [-2 \log P\mathcal{L}(\mathtt{FullModel})]\}$$

Example of Cox Regression – Primary Billiary Cirrhosis (PBC)

- $\boxed{\textbf{Data:}}$
  - ▷ A randomized trial with $n = 312$ subjects.
  - ▷ Long-term follow-up (10 years!)

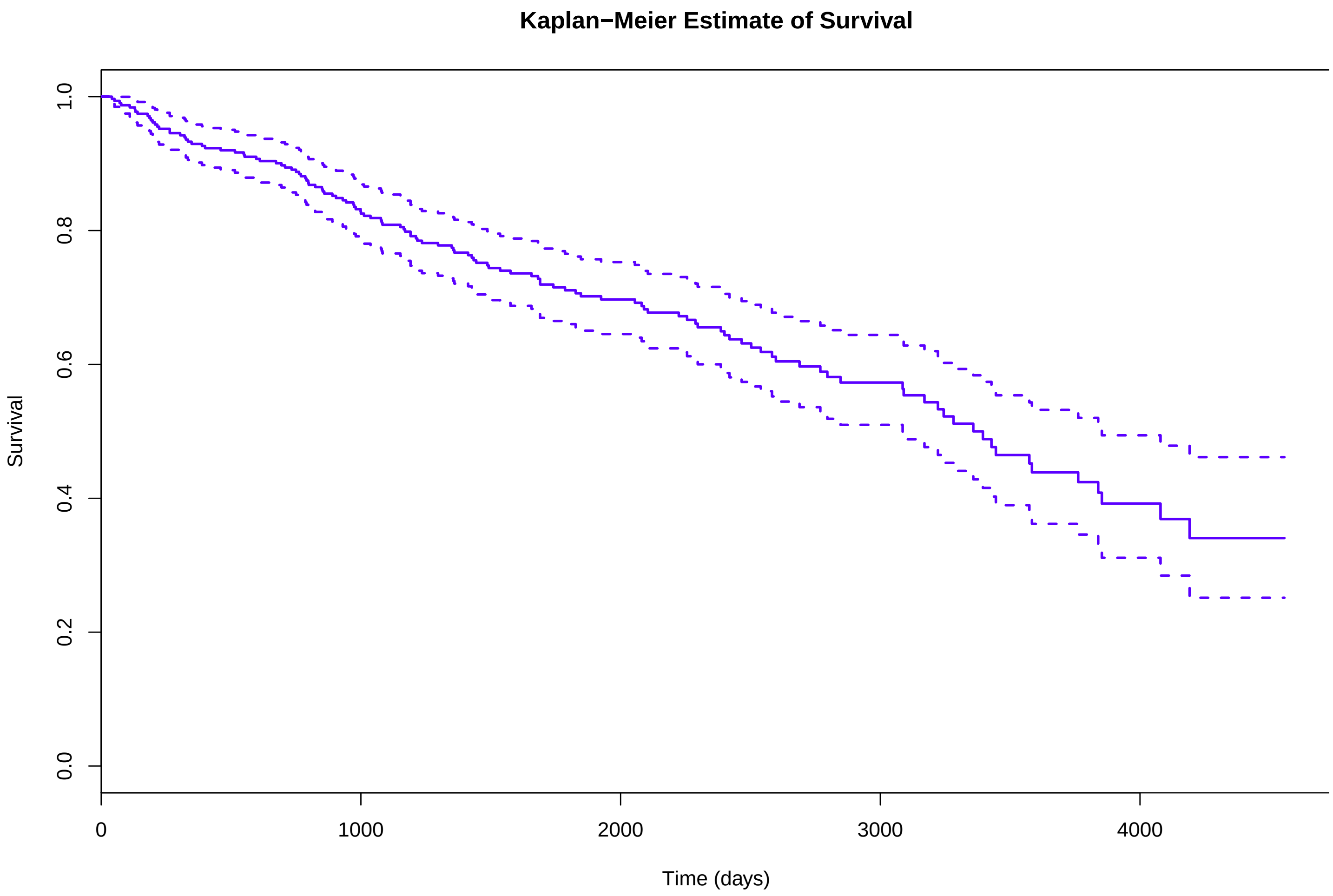
- $\boxed{\textbf{Objective:}}$
  - ▷ Could the available clinical information be used to construct a predictive model that could be used to guide medical decisions?

Example Analysis using R – PBC

```
pbc.data <- read.table( "pbc-data.txt", header=T )
#
pbc.data$survival.time <- pbc.data$fudays
pbc.data$survival.status <- as.integer( pbc.data$status==2 )
#
#####
library(survival)
#####
#
##### Kaplan-Meier curve
#
Sout <- survfit( Surv(survival.time,survival.status) ~ 1, data=pbc.data

plot( Sout, mark.time=F, col="blue", xlab="Time (days)", ylab="Survival"
      lwd=2  )
```

PBC Data



Kaplan−Meier Estimate of Survival
Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
Time (days)

Example of Cox Regression – PBC

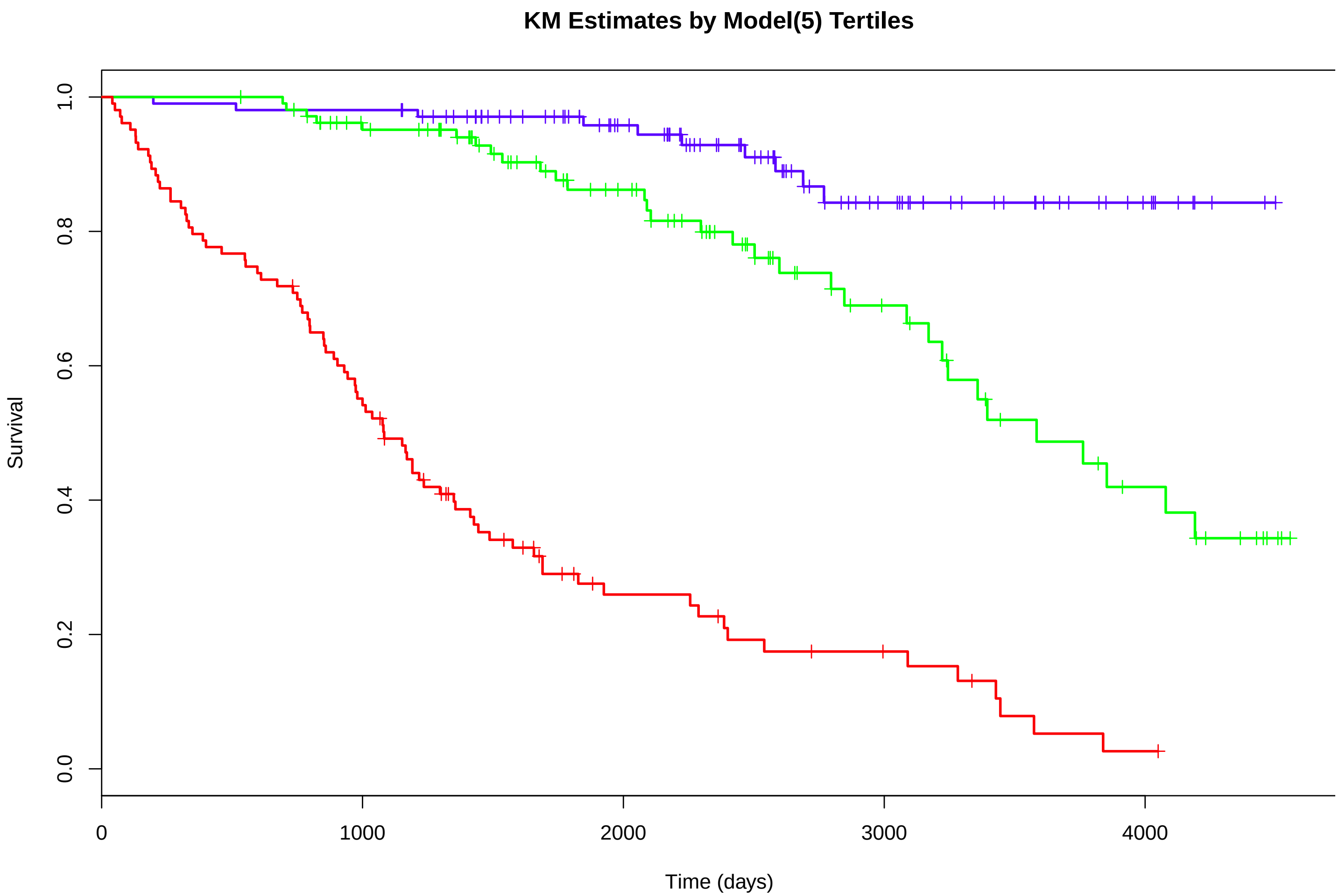
```
###### Cox model with log(bili), log(protime), edema, albumin, age
###### generate linear predictor for ordinary cox model: eta5
#
fit <- coxph( Surv(survival.time,survival.status) ~ log(bili) +
                                               log(protime) +
                                               edema +
                                               albumin +
                                               age,
              data=pbc.data )
summary( fit )
#
##### get the risk score
#
eta5 <- fit$linear.predictors
```

Example of Cox Regression – PBC

```
  n= 312, number of events= 125

                   coef  exp(coef)   se(coef)      z Pr(>|z|)
log(bili)     8.773e-01  2.404e+00  9.895e-02  8.866  < 2e-16 ***
log(protime)  3.013e+00  2.035e+01  1.025e+00  2.939 0.003296 **
edema         7.846e-01  2.192e+00  2.998e-01  2.617 0.008872 **
albumin      -9.445e-01  3.889e-01  2.370e-01 -3.985 6.73e-05 ***
age           9.169e-05  1.000e+00  2.363e-05  3.881 0.000104 ***
---
Rsquare= 0.471   (max possible= 0.983 )
Likelihood ratio test= 198.5  on 5 df,   p=0
```

Example of Cox Regression – PBC



Example of Cox Regression – PBC

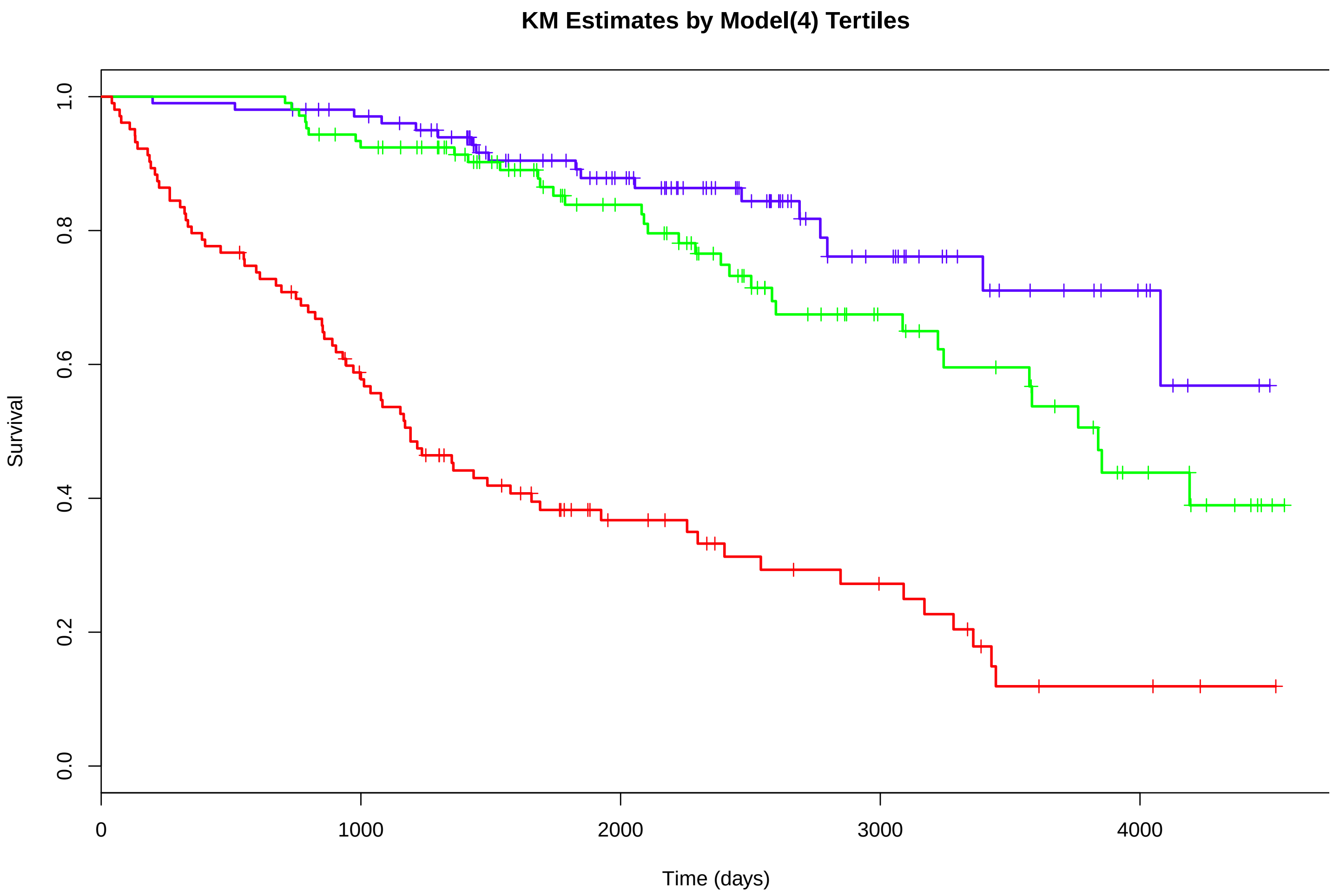
```
##### generate linear predictor for ordinary cox model: eta4
#
#
fit <- coxph( Surv(survival.time,survival.status) ~ log(protime) +
                                                     edema +
                                                     albumin +
                                                     age,
              data=pbc.data )
summary( fit )
#
# get the risk score
#
eta4 <- fit$linear.predictors
```

Example of Cox Regression – PBC

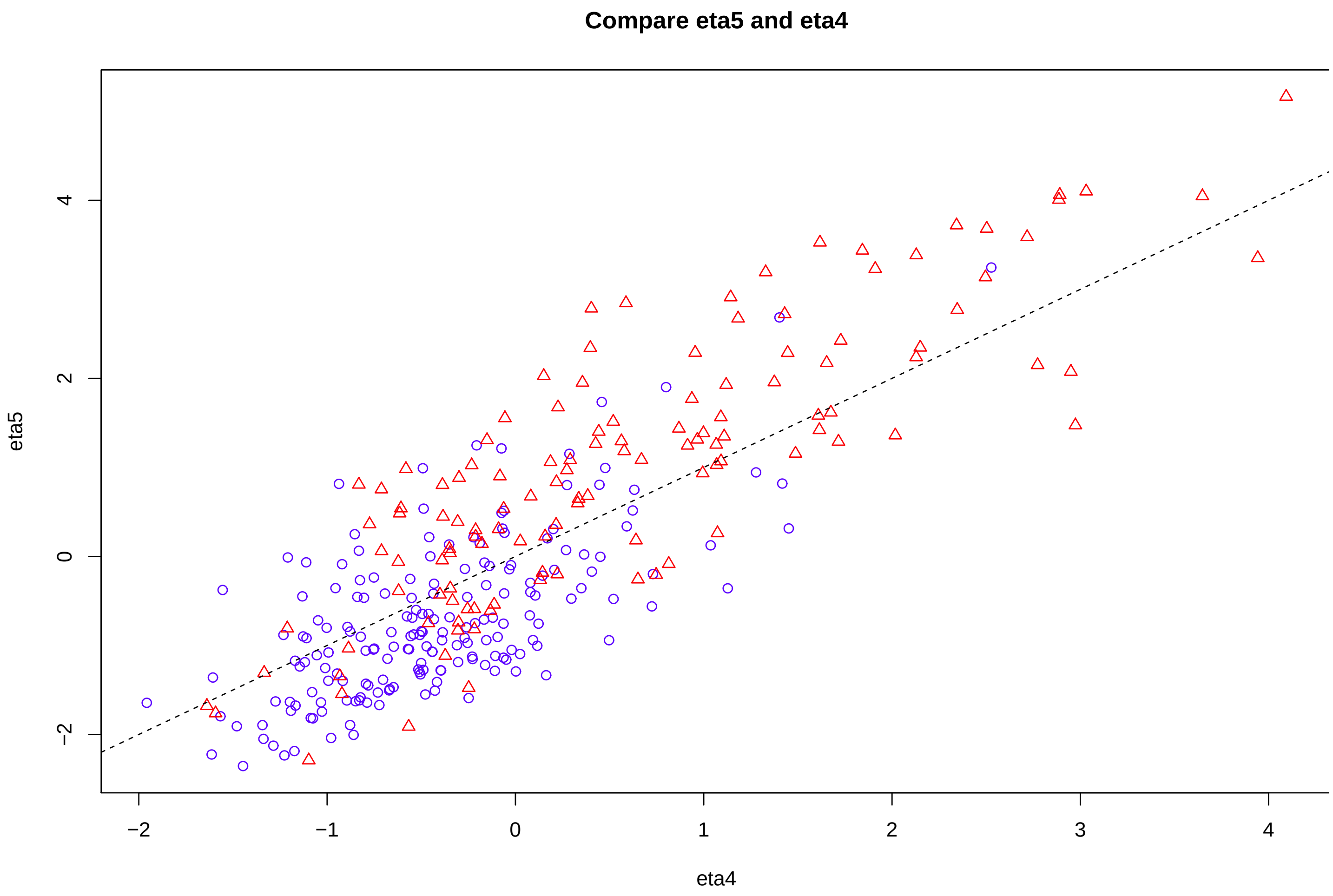
```
  n= 312, number of events= 125

                    coef  exp(coef)   se(coef)      z Pr(>|z|)
log(protime)  4.140e+00  6.283e+01  8.703e-01  4.758 1.96e-06 ***
edema         1.190e+00  3.288e+00  2.953e-01  4.031 5.56e-05 ***
albumin      -1.314e+00  2.687e-01  2.228e-01 -5.897 3.71e-09 ***
age           6.689e-05  1.000e+00  2.515e-05  2.660  0.00782 **
---
Rsquare= 0.32   (max possible= 0.983 )
Likelihood ratio test= 120.5  on 4 df,   p=0
```

Example of Cox Regression – PBC



Example of Cox Regression – PBC



Summary

- Cox regression is semi-parametric (and popular!)
- Covariates can be modeled in standard ways and inference performed using partial likelihood, score, and Wald tests.
- An important idea is the **risk set** at time t which usually includes a single CASE and multiple CONTROLS (at that time).
- **Q**: How well does the predictive model perform?
- **Q**: How to link the model to making medical decisions?