SISCER Mod 12

Survival Analysis

Lecture 4

Ying Qing Chen, Ph.D.

Department of Medicine
Stanford University

Department of Biostatistics
University of Washington

In Lecture 3

- What we discussed
  - Log-rank test
  - Weighted log-rank tests
  - Power and sample size calculation

More than two-sample

- Two-sample (multiple-samples)
  - Only gives hypothesis testing results
  - Does not provide exact effect size
  - Does not provide confidence intervals
- Covariates
  - Continuous factors
  - Multiple factors and interactions
  - Confounders
  - In particular to time-to-event outcomes: time-varying covariates
- Regression modeling
  - Parametric
  - Nonparametric
  - Semiparametric

Parametric regression modeling

- General paradigm
  - Set up underlying parametric distributions
  - Establish a relationship between outcomes and covariates by some parametric form
  - Apply MLE to estimate regression parameters
  - Make inferences
    - Wald's test
    - Score test
    - Likelihood ratio test

Parametric regression models

- Data (without censoring):

$$(T_i, Z_i), i = 1, 2, \ldots, n$$

- Likelihood function

$$\prod_{i=1}^{n} f(T_i | Z_i),$$

- Examples
  - A linear regression model:

$$-\log T_i = \beta_0 + \beta Z_i + \epsilon_i, \text{ where } \epsilon_i \sim N(0, \sigma^2)$$

Example. $T \sim \exp(\theta)$. The density function is $f(t;\theta) = \theta e^{-\theta t} I(t > 0)$.

Regression extension: Let $x_i = (x_{i1}, \ldots, x_{ip})$ be a $1 \times p$ vector of covariates and $\theta = (\theta_1, \ldots, \theta_p)^t$ a $p \times 1$ vector of parameters for subject i. Assume the hazard function is $\lambda(t; x_i, \theta) = x_i\theta = \sum_{j=1}^{p} x_{ij}\theta_j$. Assume T has the pdf $(x_i\theta)e^{-(x_i\theta)t_i}$. Based on $(x_1, t_1), \ldots, (x_n, t_n)$, the maximum likelihood techniques can still be applied to the likelihood function

$$L(\theta) = \prod_{i=1}^{n} (x_i\theta)e^{-(x_i\theta)t_i}$$

Exponential density

A constraint here is that the hazard $\lambda(t; x_i) = x_i\theta$ must be positive. To guarantee this, we sometimes use a positive-valued link function $\phi(\cdot)$ and assume the hazard $\lambda(t; x_i) = \phi(x_i\theta)$. For instance, $\phi(u) = u^2$ or $\phi(u) = e^u$. For the latter case, the likelihood becomes function

$$L(\theta) = \prod_{i=1}^{n} e^{x_i\theta} e^{-(e^{x_i\theta})t_i}$$

For censored time-to-event

- Recall
  - Log-rank test is mostly powerful to test the alternative when hazard functions are proportional
  - This shall serve as an important motivation for the so-called Cox proportional hazards model

Cox proportional hazards model

- 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: $\lambda(t)$
- 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.

Model specification

- Model:

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

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

Model interpretation

- Proportional Hazards:

$$\text{RR} = \frac{\lambda(t \mid X_1, X_2, \ldots, X_k)}{\lambda(t \mid X_1 = 0, X_2 = 0, \ldots, X_k = 0)}$$

$$= \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$$

- 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 $(0, 0, \ldots, 0)$."

- 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).

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

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

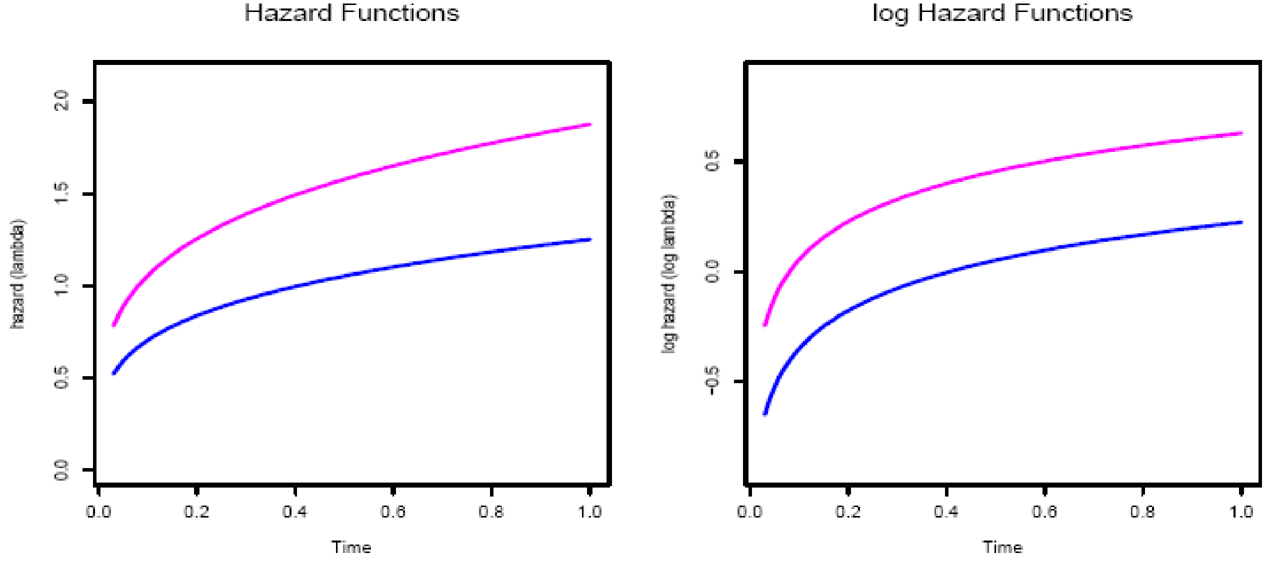
- The RR Comparing 2 Covariate Values (vectors):

  - RR comparing $(X_1, X_2, \ldots, X_k)$ to $(X_1', X_2', \ldots, X_k')$.

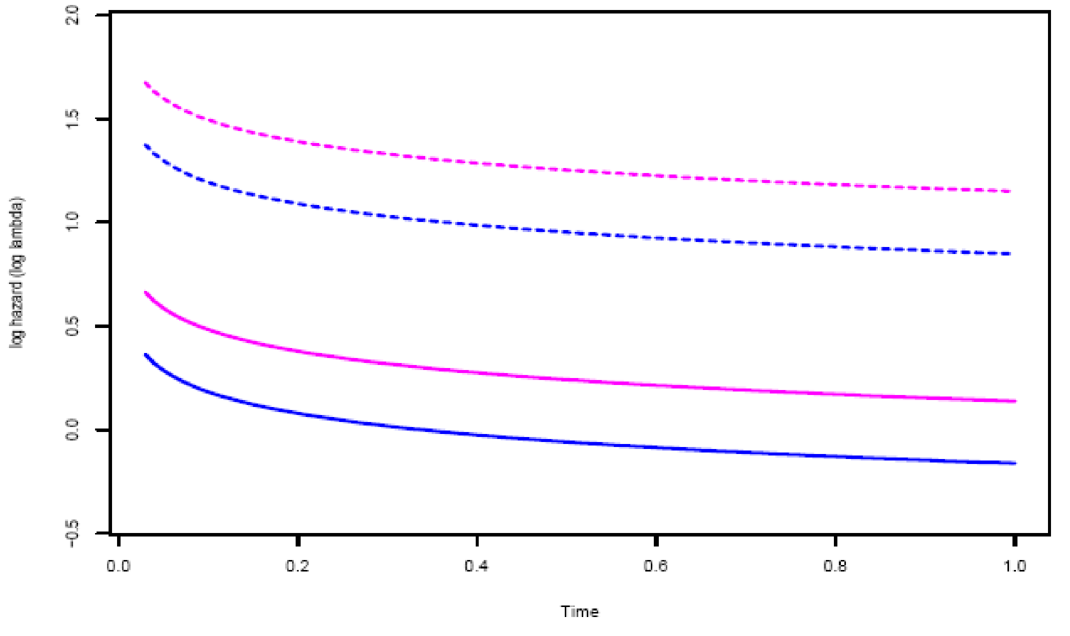
$$\begin{aligned} \mathrm{RR}(X \text{ vs. } X') &= \frac{\lambda(t \mid X_1, X_2, \ldots, X_k)}{\lambda(t \mid X_1', X_2' \ldots, X_k')} \\ &= \exp[\;\beta_1 \cdot (X_1 - X_1') + \\ &\qquad \beta_2 \cdot (X_2 - X_2') + \\ &\qquad \ldots + \\ &\qquad \beta_k \cdot (X_k - X_k')\;] \end{aligned}$$

Examples: Cox proportional hazards model

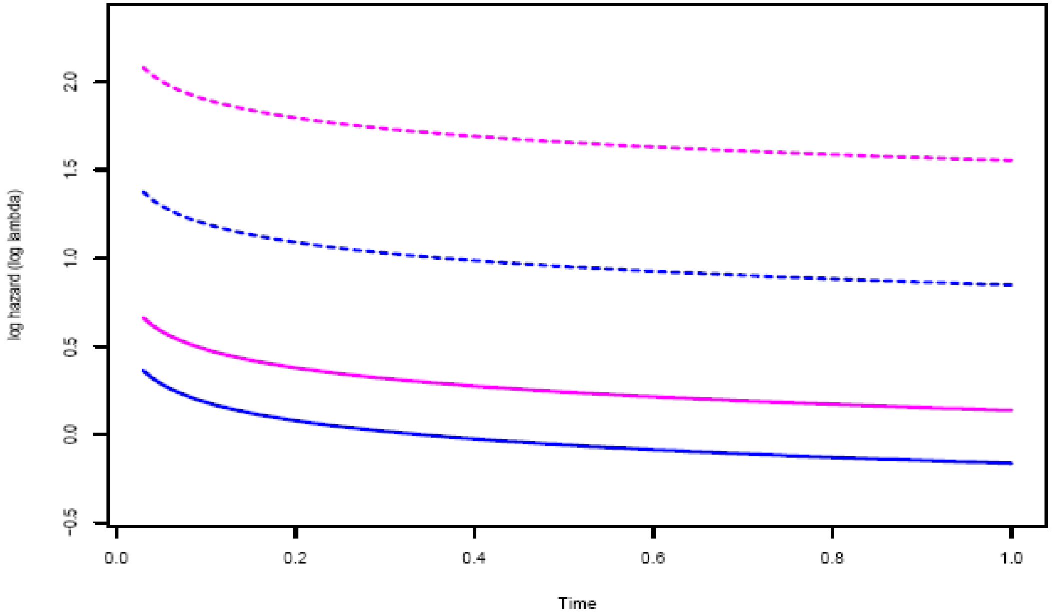
- **1**: One dichotomous covariate
  - ▷ $X_E = 1$ if exposed; $X_E = 0$ if not exposed.
  - ▷ $\lambda(t \mid X_E) = \lambda_0(t) \exp(\beta X_E)$



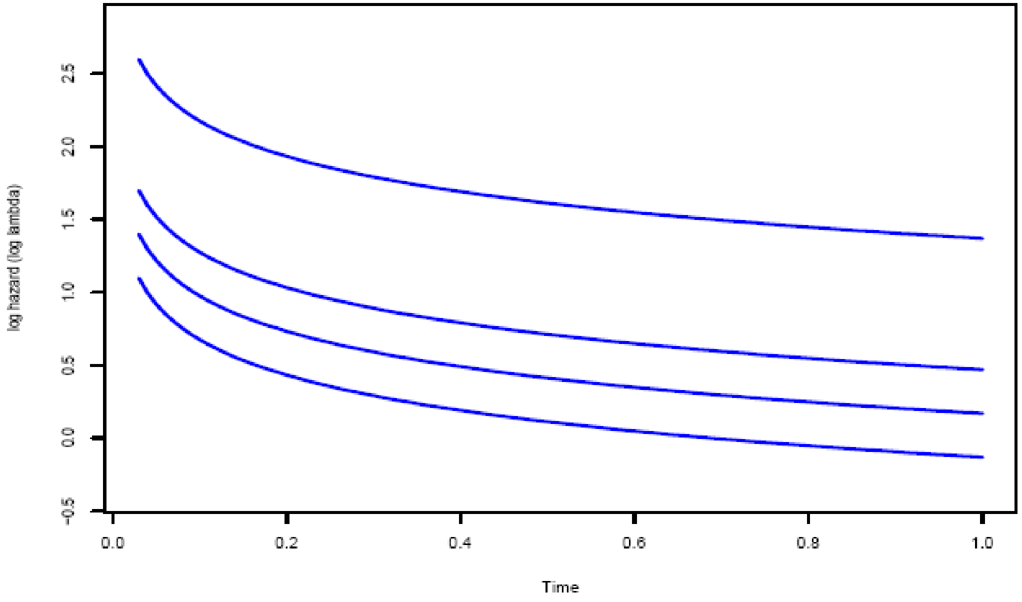
- **2**: Dichotomous covariate; Dichotomous confounder
  - ▷ $X_C = 1$ if level 2; $X_C = 0$ if level 1.
  - ▷ $\lambda(t \mid X_E, X_C) = \lambda_0(t)\exp(\beta_1 X_E + \beta_2 X_C)$



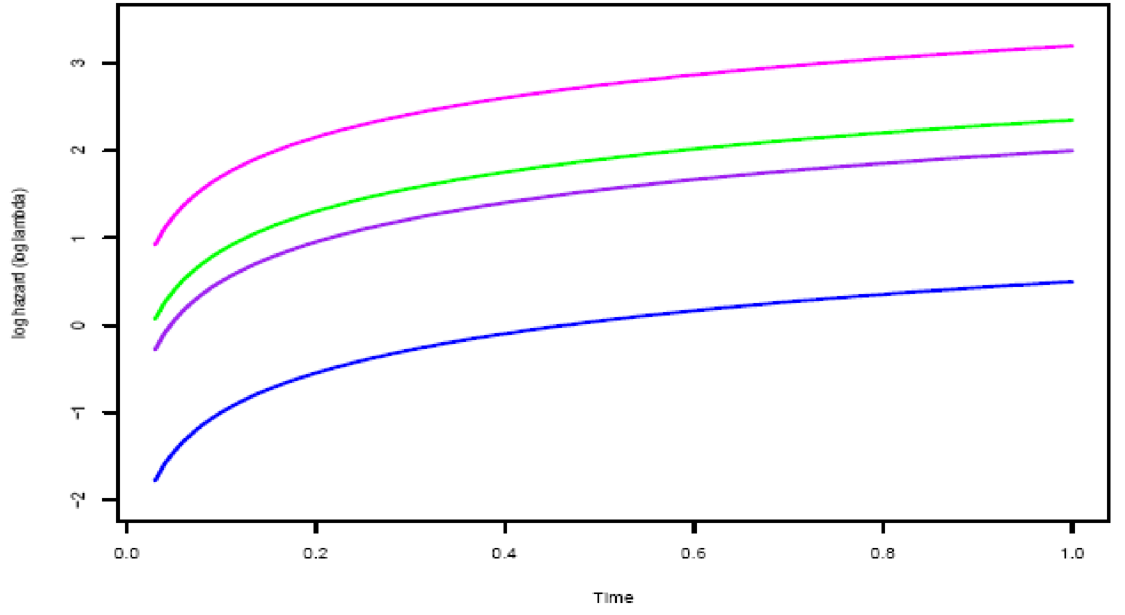
- **3**: Dichotomous covariate; confounder; (interaction)
  - ▷ With interaction
  - ▷ $\lambda(t \mid X_E, X_C) = \lambda_0(t)\exp(\beta_1 X_E + \beta_2 X_C + \beta_3 X_E X_C)$



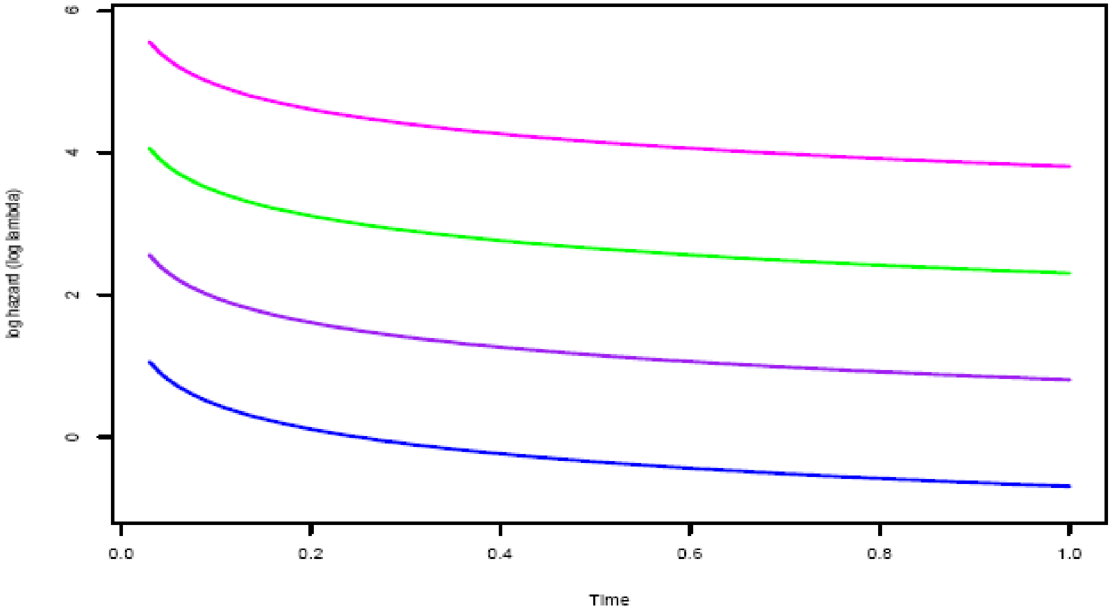
- **4**: One continuous covariate
  - $X_D = 1.0, 2.0, \ldots$
  - $\lambda(t \mid X_D) = \lambda_0(t) \exp(\beta_1 X_D)$



- **5**: K-sample Heterogeneity (K=4)
  - ▷ $X_j = \begin{cases} 1: \text{group } j \\ 0: \text{otherwise} \end{cases}$
  - ▷ $\lambda(t \mid X_2, X_3, X_4) = \lambda_0(t) \exp(\beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$



- **6**: K-sample Trend (K=4)
  - ▷ $X_D = \{ \; j : \text{group } j$
  - ▷ $\lambda(t \mid X_D) = \lambda_0(t) \exp(\beta X_D)$



About the Cox model

- 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 for example (5) with $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ is the same as the K-sample Heterogeneity test (generalization of LogRank).
- The score test for example (6) with $H_0 : \beta = 0$ is the same as Tarone's trend test.

Some history

- 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)
- 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: 19,502 citations (17 Jan 2005)
  - David R. Cox is knighted in 1985 in recognition of his scientific contributions.

How to estimate 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:

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

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

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

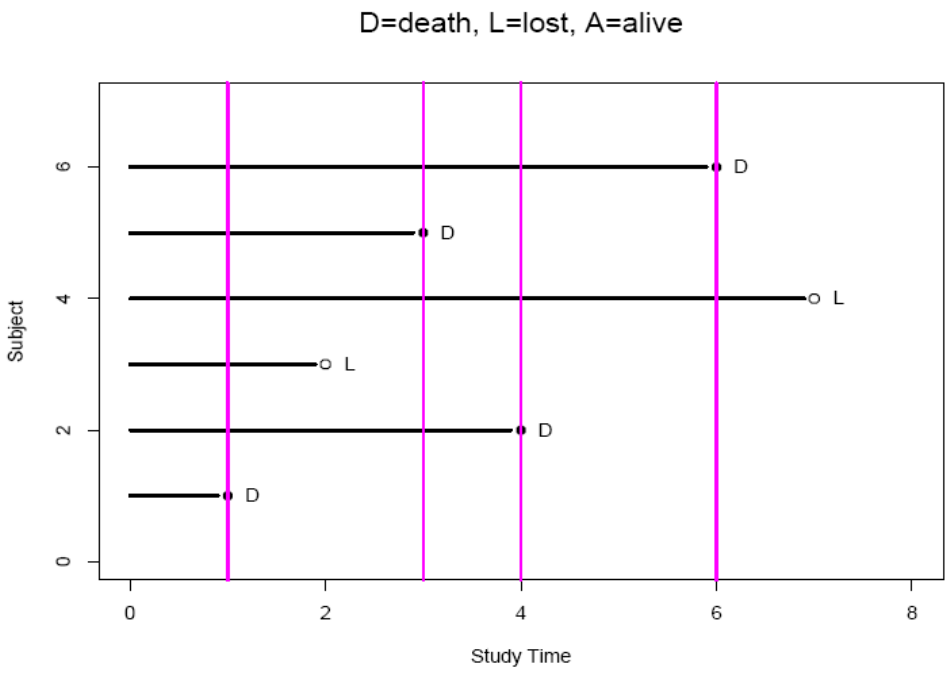
  - ▷ Wald test, score test, and likelihood ratio test similar to logistic regression. Now using the partial likelihood.

Partial likelihood

- Model: $\lambda(t \mid X_1, \ldots, X_k) = \lambda_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k)$
- 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



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

- **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?

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

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

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

- Probability One Death, Was (i) :

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

- Probability of One Death:

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

- Note: $(1 - P_j) \approx 1$ for small Δt.

- Now calculate the desired quantity:

$$\text{P(Observed Data} \mid \text{1 death)} = \frac{\text{P(Only (i) Dies)}}{\text{P(One Death)}}$$

$$= \frac{P_{(i)} \prod_{k \neq (i)} (1 - P_k)}{\sum_{j \in \mathcal{R}_i} P_j \prod_{k \neq j} (1 - P_k)}$$

$$\approx \frac{P_{(i)}}{\sum_{j \in \mathcal{R}_i} P_j}$$

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

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

- 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:

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

- Note: This likelihood contribution has the exact same form as a (matched) logistic regression conditional likelihood.

- Notice that our model is equivalent to

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

  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, as outlined in the notes on pages 86-87?

- **A**: To do so would require choosing a model for the baseline hazard, but we actually don't need to do that!

Handle 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.
- 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:

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

- 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.

- However, there is plenty of room between continuous and discrete.
  - ▷ Example: **USRDS Data** = 200,000 subjects.

    US Renal Data System

    - ∗ 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 test

- Full Model:

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

- 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

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

- Under H_0 (reduced is correct) then $X^2_{PLR} \sim \chi^2(\texttt{df} = (\texttt{k} - \texttt{p}))$
- Degrees of freedom, $\texttt{df} = (\texttt{k} - \texttt{p})$, 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:

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

STATA codes for Cox models

```
******************************
*   evaluate TX              *
******************************

stcox tx, nohr
est store LRmod1

xi: stcox i.group, nohr
est store LRmod2

xi: stcox tx i.group, nohr
est store LRmod3

lrtest LRmod3 LRmod2, stats
```

```
. xi: stcox i.group, nohr
Cox regression -- Breslow method for ties
No. of subjects =          456                  Number of obs    =       456
No. of failures =          374
Time at risk    =        46363
                                                LR chi2(2)       =     67.41
Log likelihood  =   -1986.2945                  Prob > chi2      =    0.0000

------------------------------------------------------------------------------
        _t |    Coef.    Std. Err.      z     P>|z|      [95% Conf. Interval]
-----------+------------------------------------------------------------------
_Igroup_2 | 1.14690    .1786005    6.42    0.000       .7968584    1.496959
_Igroup_3 | 1.51643    .2168077    6.99    0.000       1.091494    1.941365
------------------------------------------------------------------------------
```

```
. xi: stcox tx i.group, nohr
Cox regression -- Breslow method for ties
No. of subjects =          456                  Number of obs   =        456
No. of failures =          374
Time at risk    =        46363
                                                LR chi2(3)      =      68.49
Log likelihood  =   -1985.7542                  Prob > chi2     =     0.0000
------------------------------------------------------------------------------
          _t |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |   .111602   .1069722     1.04   0.297    -.0980588    .3212645
   _Igroup_2 |  1.171318   .1801767     6.50   0.000     .8181779    1.524457
   _Igroup_3 |  1.525078   .2170109     7.03   0.000     1.099745    1.950411
------------------------------------------------------------------------------
```

```
. lrtest LRmod3 LRmod2, stats

likelihood-ratio test                          LR chi2(1)  =      1.08
(Assumption: LRmod2 nested in LRmod3)          Prob > chi2 =    0.2986

------------------------------------------------------------------------
Model    |   nobs    ll(null)    ll(model)   df        AIC         BIC
---------+--------------------------------------------------------------
 LRmod2 |    456   -2019.999   -1986.294    2    3976.589    3984.834
 LRmod3 |    456   -2019.999   -1985.754    3    3977.508    3989.876
------------------------------------------------------------------------
```

Estimate baseline hazard function

- Recall: (math fact)

$$S(t) = \exp[-\int_0^t \lambda(s)ds] = \exp[-\Lambda(t)]$$

- Cox model:

$$\begin{aligned}
\lambda(t \mid X_1 \ldots X_k) &= \lambda_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k) \\
\Lambda(t \mid X_1 \ldots X_k) &= \Lambda_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k) \\
S(t \mid X_1 \ldots X_k) &= [S_0(t)]^{[\exp(\beta_1 X_1 + \ldots + \beta_k X_k)]}
\end{aligned}$$

- Therefore, in order to estimate the survival function, or the hazard for specific values of the covariates, $(X_1, X_2, \ldots, X_k)$ we need to estimate $\lambda_0(t), \Lambda_0(t)$, and/or $S_0(t)$.

- **Method 1:** Breslow Method (used in `STATA`)

$$\widehat{\Lambda}_0(t) = \sum_{i:t_{(i)} \le t} \frac{D_i}{\left[\sum_{j \in \mathcal{R}_i} \exp(\widehat{\beta}_1 X_{1j} + \ldots + \widehat{\beta}_k X_{kj})\right]}$$

- **Special Cases**

  ▷ **1** **One group, no covariates** — Nelson-Aalen Estimator

  This is like $(\widehat{\beta}_1 X_{1j} + \ldots + \widehat{\beta}_k X_{kj}) = 0$

$$\widehat{\Lambda}_0(t) = \sum_{i:t_{(i)} \le t} \frac{D_i}{\left[\sum_{j \in \mathcal{R}_i} \exp(0)\right]} = \sum_{i:t_{(i)} \le t} \frac{D_i}{N_i}$$

- Special Cases
  - 2 **Two groups: one dichotomous covariate**

$$X = \begin{cases} 0 & \text{group 1} \\ 1 & \text{group 2} \end{cases}, \qquad \lambda(t \mid X) = \lambda_0(t)\exp(\beta X).$$

$$\begin{aligned}
\widehat{\Lambda}_0(t) &= \sum_{i:t_{(i)}\le t} \frac{D_i}{\left[\sum_{j\in\mathcal{R}_i} \exp(\widehat{\beta}X_j)\right]} \\
&= \sum_{i:t_{(i)}\le t} \frac{D_i}{\left[\sum_{j\in\mathcal{R}_i,\ \text{group 1}} \exp(\widehat{\beta}X_j) + \sum_{j\in\mathcal{R}_i\ \text{group 2}} \exp(\widehat{\beta}X_j)\right]} \\
&= \sum_{i:t_{(i)}\le t} \frac{D_i}{\left[N_{1i} + \exp(\widehat{\beta})\cdot N_{2i}\right]}
\end{aligned}$$

- In this example we can consider $N_{1i} + \exp(\widehat{\beta})N_{2i}$ as the "effective risk set" at time $t_{(i)}$.
- The numerator, D_i, counts deaths equally from both group 1 and group 2.
- However, in order to represent cumulative hazard (risk) for group 1 some adjustment of the group 2 contributions is warranted.
- **Idea**: reweight the denominator
  - $\boxed{\widehat{\beta} > 0}$ more deaths in group 2, so effective risk set needs to be increased to estimate risk in group 1.
  - $\boxed{\widehat{\beta} < 0}$ fewer deaths in group 2, so effective risk set needs to be decreased to estimate risk in group 1.

- **3** In general, the denominator

$$\sum_{j \in \mathcal{R}_i} \exp(\widehat{\beta}_1 X_{1j} + \ldots + \widehat{\beta}_k X_{kj})$$

  - Is bigger than N_i when the average risk for a subject in $\mathcal{R}_i$ is greater than the risk for a subject with the reference value $(X_1 = 0, X_2 = 0, \ldots, X_k = 0)$.
  - Is smaller than N_i when the average risk for a subject in $\mathcal{R}_i$ is less than the risk for a subject with the reference value $(X_1 = 0, X_2 = 0, \ldots, X_k = 0)$.

- Survival

$$\widehat{S}_0(t) = \exp[-\widehat{\Lambda}_0(t)]$$

  - ▷ Not the default in STATA, but can be created.

- Hazard (similar to before)

$$\widehat{\lambda}_0(t) = \frac{1}{b} \cdot \sum_{j=1}^{J} K\left(\frac{t - t_{(j)}}{b}\right) \cdot \left\{ \frac{D_i}{\left[\sum_{j \in \mathcal{R}_i} \exp(\widehat{\beta} X_j)\right]} \right\}$$

  - ▷ Also not the default in STATA.

Alternative approach to estimate baseline survival function

- Kalbfleisch and Prentice (1973) discuss use of a discrete time model and use this to estimate the baseline survival.
- The PH model implies:

$$p_j(X_1,\ldots,X_k) = P[T \in [t_{j-1},t_j) \mid T \geq t_{j-1}, X_1,\ldots,X_k]$$

$$\begin{aligned} 1 - p_j(X_1,\ldots,X_k) &= \left[\frac{S_0(t_j)}{S_0(t_{j-1})}\right]^{\exp(\beta_1 X_1 + \ldots + \beta_k X_k)} \\ &= [\alpha_j]^{\exp(\beta_1 X_1 + \ldots + \beta_k X_k)} \end{aligned}$$

- K&P (1973) show that using such a discrete time approximation leads to a method to estimate these α_j. (see STATA manual p. 150 for further details)
- K&P (1973) are using maximum likelihood for the discrete model.

- Notice that once these estimates are obtained

$$S_0(t) = \left[\frac{S_0(t_1)}{1}\right] \times \left[\frac{S_0(t_2)}{S_0(t_1)}\right] \times \ldots \times \left[\frac{S_0(t_j)}{S_0(t_{j-1})}\right]$$

$$S_0(t) = \prod_{i:t_{(i)} \leq t} \alpha_i$$

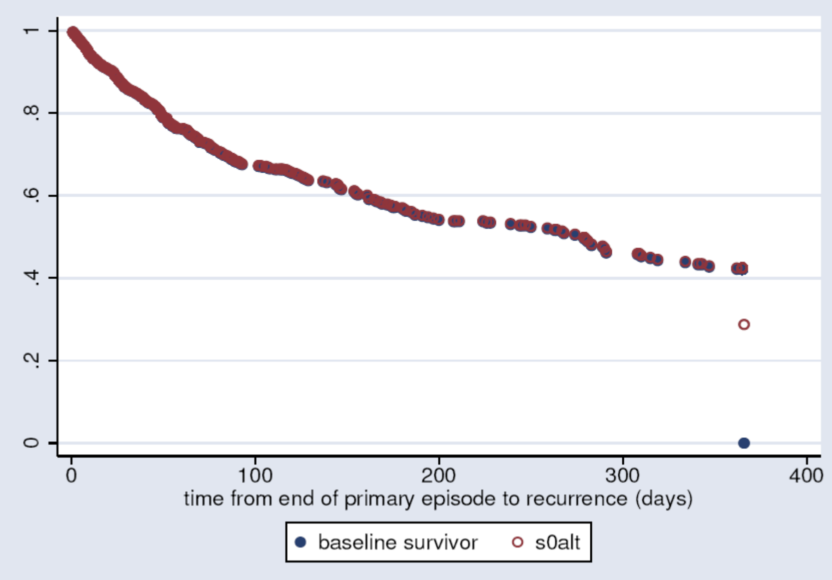
- This provides an estimate for the baseline survival function given as the default in STATA:

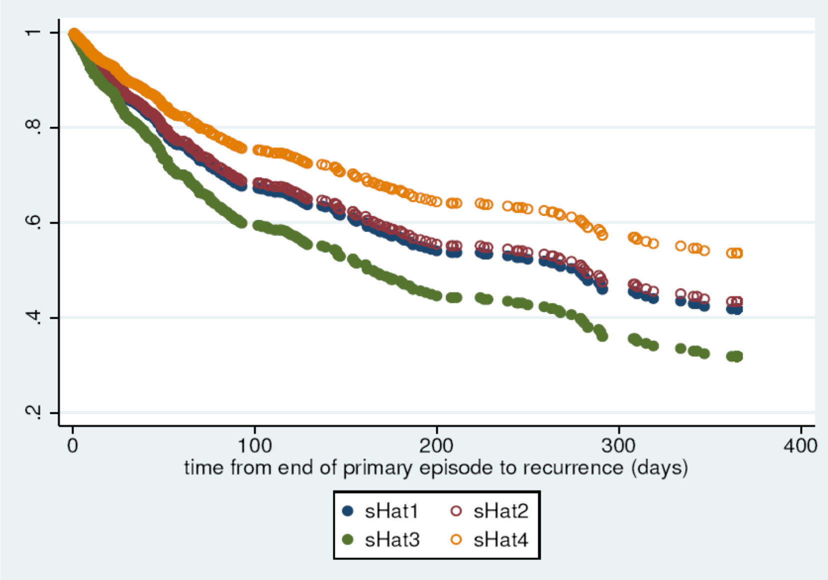
$$\widehat{S}_0(t) = \prod_{i:t_{(i)} \leq t} \widehat{\alpha}_i$$

- **Q**: How does this estimate compare to that obtained using the cumulative hazard?

STATA codes for baseline estimates

```
xi: stcox i.treat i.group age25 i.gender, basesurv( s0 ) basechazard( H0 )

gen s0alt = exp( -1 * H0 )

graph twoway (scatter s0 s0alt rectime )
```





Smoothed baseline hazard functions

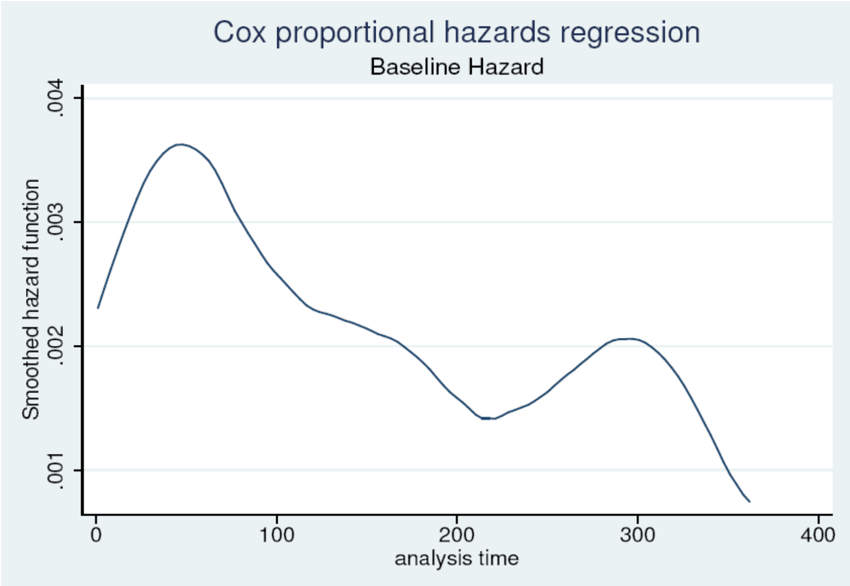
- Note: – with the estimates $\widehat{\alpha}_j$ we can also obtain estimates of the baseline hazard function:

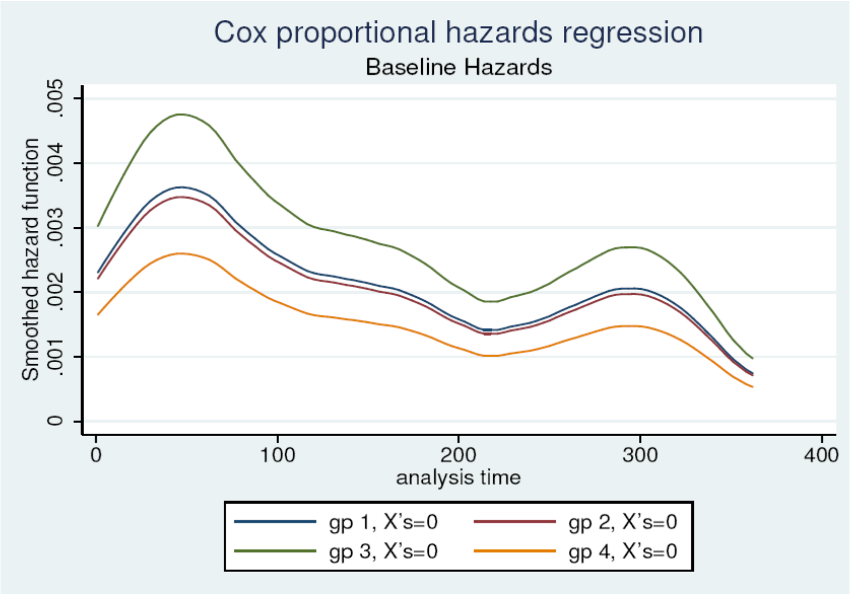
$$\widehat{\lambda}_0(t) = \frac{1}{b} \cdot \sum_{j=1}^{J} K\left(\frac{t - t_{(j)}}{b}\right) \cdot [(1 - \widehat{\alpha}_j)]$$

- STATA uses this method.

```
stcurve, hazard   at( _Itreat_1=0,
                      _Itreat_2=0,
                      _Itreat_3=0,
                      _Igroup_2=0,
                      _Igroup_3=0,
                      age25=0,
                      _Igender_2=0 ) subtitle("Baseline Hazard");
```

Examples: smoothed baseline hazard functions





Use of baseline estimates

- Uses:
  - ▷ Estimate survival or risk for specific sub-populations defined by a vector of covariate values.
  - ▷ Evaluate the shape of the estimated hazard as provided by the model. The model imposes constraints (e.g. PH).
  - ▷ To check the fit of the model, for example, by comparing the fitted survival curves for subsets to the survival curve estimated under the model.
  - ▷ Can be used to see whether different strata appear to satisfy PH after adjustment for key covariates (next!)

Stratification: use of dummy variables

- Suppose a confounder X_C has 3 levels on which we would like to stratify when comparing $X_E = 1$ to $X_E = 0$.
  - ▷ $\lambda(t \mid X_E, X_C)$
  - ▷ $\begin{cases} X_E = 1 & : \text{exposure} \\ X_E = 0 & : \text{no exposure} \end{cases}$
- 1 **"Dummy variables"**
  - ▷ $\begin{cases} X_j = 1 & : X_C = j \\ X_j = 0 & : X_C \neq j \end{cases}$
  - ▷ Model

$$\lambda(t \mid X_E, X_2, X_3) = \lambda_0(t) \exp(\beta_1 X_E + \beta_2 X_2 + \beta_3 X_3)$$

- Level 1 of X_C

$$\left.\begin{array}{ll} \text{exposed} & : \ \lambda_0(t)\exp(\beta_1) \\ \text{unexposed} & : \ \lambda_0(t) \end{array}\right\} \text{RR} = \exp(\beta_1)$$

- Level 2 of X_C

$$\left.\begin{array}{ll} \text{exposed} & : \ \lambda_0(t)\exp(\beta_1+\beta_2) \\ \text{unexposed} & : \ \lambda_0(t)\exp(\beta_2) \end{array}\right\} \text{RR} = \exp(\beta_1)$$

- Level 3 of X_C

$$\left.\begin{array}{ll} \text{exposed} & : \ \lambda_0(t)\exp(\beta_1+\beta_3) \\ \text{unexposed} & : \ \lambda_0(t)\exp(\beta_3) \end{array}\right\} \text{RR} = \exp(\beta_1)$$

Stratified Cox models

- In the previous approach each of the six groups has a log hazard that is "parallel" to any other group (e.g. one common curve characterizes time, $\log \lambda_0(t)$).
- More generally:
  - **Model**: $\lambda(t \mid X_E, X_C = j) = \lambda_{0,j}(t)\exp(\beta_1 X_E)$
  - $\lambda_{0,j}(t)$ represents an arbitrary function of time for the unexposed in strata $\{X_C = j\}$.
  - However, the comparison between exposed and unexposed within each strata is assumed to be constant [HR= $\exp(\beta_1)$].
- This approach is implicit in the stratified version of the LogRank test.
- "Stratified Cox Model"

- Level 1 of X_C

$$\left.\begin{array}{ll}\text{exposed} & : \ \lambda_{0,1}(t)\exp(\beta_1) \\ \text{unexposed} & : \ \lambda_{0,1}(t)\end{array}\right\} \text{RR} = \exp(\beta_1)$$

- Level 2 of X_C

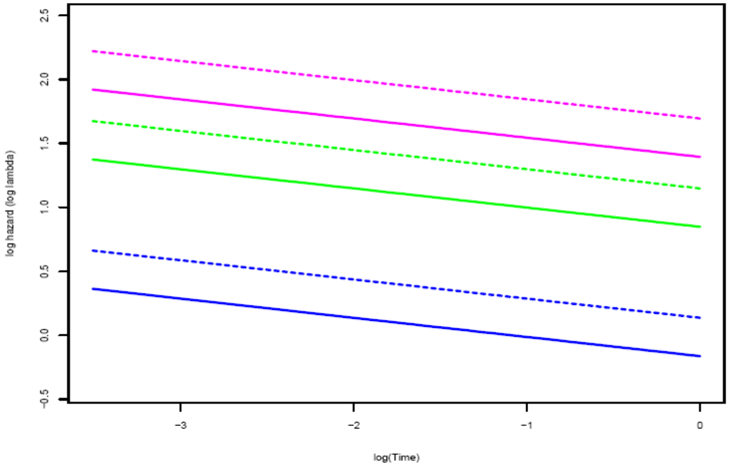
$$\left.\begin{array}{ll}\text{exposed} & : \ \lambda_{0,2}(t)\exp(\beta_1) \\ \text{unexposed} & : \ \lambda_{0,2}(t)\end{array}\right\} \text{RR} = \exp(\beta_1)$$

- Level 3 of X_C

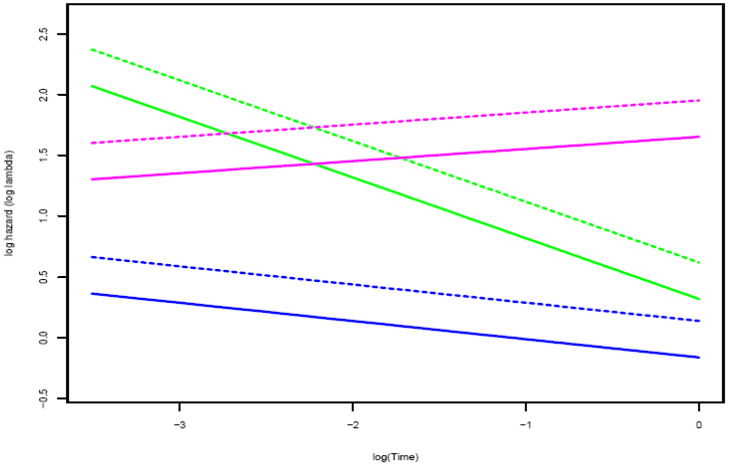
$$\left.\begin{array}{ll}\text{exposed} & : \ \lambda_{0,3}(t)\exp(\beta_1) \\ \text{unexposed} & : \ \lambda_{0,3}(t)\end{array}\right\} \text{RR} = \exp(\beta_1)$$

Comparison of two stratification methods

Adjustment Using Dummy Variables



Stratified Cox Model



- **Q**: When to choose separate baselines?
  - ▷ Dummy variables assume common time change across confounder groups. If not correct then X_C may be inadequately controlled, and may confound exposure evaluation.
  - ▷ PH can be checked using graphical methods of time-dependent covariates (later!).
  - ▷ True stratification is a more thorough adjustment when observations within each stratum are homogeneous. If X_C is measured as a continuous variable, and strata are formed by grouping its values then better control might be achieved with the original continuous variable (possibly with time-dependent) covariate adjustment.

- If X_C is controlled using true stratification then there is no single HR to report comparing the different levels of X_C. However, we can estimate baseline survival (hazard) within each level and can compare these curves.
- True stratification generally requires more data to obtain the same precision in coefficient estimates (a bias-variance trade-off).

STATA codes for stratification

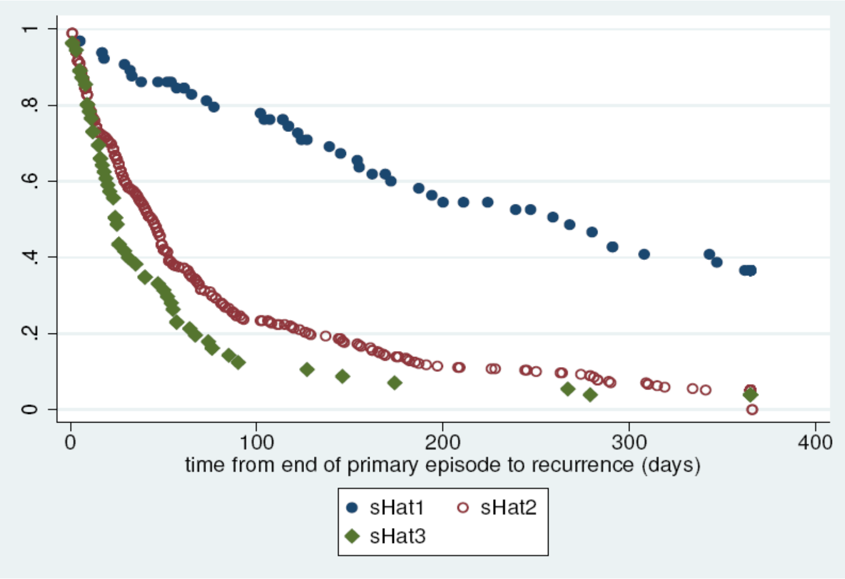
```
***
*** using dummy variables
***
xi: stcox i.treat i.group age25 i.gender


***
*** using stratified model
***
xi: stcox i.treat age25 i.gender, strata( group ) ///
    basesurv( s0 ) basehc( haz0 )
```

```
 xi: stcox i.treat i.group age25 i.gender
Cox regression -- Breslow method for ties
                                                  LR chi2(7)      =     86.54
Log likelihood  =   -1976.7301                    Prob > chi2     =    0.0000
----------------------------------------------------------------------
       _t | Haz. Ratio   Std. Err.    z    P>|z|    [95% Conf. Interval]
----------+-----------------------------------------------------------
 _Itreat_1 |   .98055   .1953991   -0.10   0.922     .663517    1.44909
 _Itreat_2 |  1.33508   .1593493    2.42   0.015    1.056606    1.68695
 _Itreat_3 |   .73497   .2392546   -0.95   0.344     .388313    1.39111
 _Igroup_2 |  3.55011   .6491291    6.93   0.000    2.480856    5.08021
 _Igroup_3 |  4.78591  1.050507     7.13   0.000    3.112625    7.35874
     age25 |   .97799   .0082657   -2.63   0.008     .961923     .99432
_Igender_2 |   .74549   .0849773   -2.58   0.010     .596231     .93211
----------------------------------------------------------------------
```

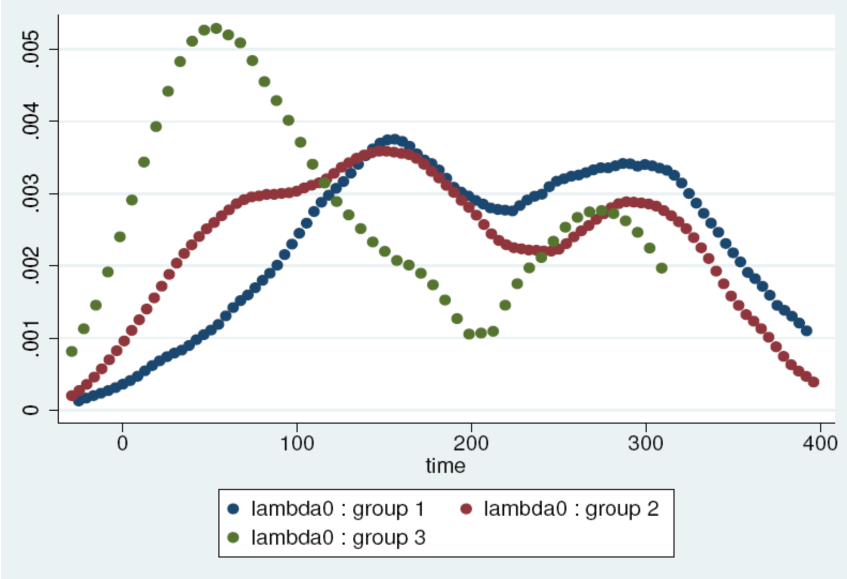
```
xi: stcox i.treat age25 i.gender, strata( group ) basesurv( s0 ) ///
    basehc( haz0 )
Stratified Cox regr. -- Breslow method for ties
                                             LR chi2(5)      =      16.94
Log likelihood  =   -1723.7986               Prob > chi2     =     0.0046
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_1 |    .958117    .1911902    -0.21   0.830      .647982   1.416688
   _Itreat_2 |   1.304738    .1562943     2.22   0.026     1.031712   1.650018
   _Itreat_3 |    .724621    .2358843    -0.99   0.322      .382843   1.371516
       age25 |    .980098    .0083365    -2.36   0.018      .963894    .996574
  _Igender_2 |    .755070    .0862966    -2.46   0.014      .603537    .944649
------------------------------------------------------------------------------
                                                          Stratified by group
```

Baseline functions

Separate S0 by Group



Separate $\lambda_{0,j}(t)$ by Group



Summary

- Cox Model parameters β_m are estimated using the partial likelihood. This focuses on the hazard ratios, HR or RR, and does not (directly) provide an estimate of the baseline hazard.
- Baseline hazard can be estimated using either the Breslow estimator of the cumulative hazard, or via a method introduced by Kalbfleisch & Prentice (`default in STATA`).
- The relationship among hazard, cumulative hazard, and survival functions allows estimation of one function to allow estimation of each of the other two functions:

$$\lambda(t \mid X) \Longleftrightarrow \Lambda(t \mid X) \Longleftrightarrow S(t \mid X)$$

- Stratified Cox models allow a more flexible adjustment for a stratifying variable. This is effectively allowing a separate baseline hazard for each level of the stratifying variable.
- No simple summary represents strata comparisons.
- Can be used to evaluate PH assumption relating strata after controlling for other covariates.