Lecture 5

Monday, October 2, 2023 10:01

BIOS522_Slides5

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | Department of Biostatistics and Bioinformatics

BIOS 522: Survival Analysis Methods

Lecture 5:

Introduction to the Cox model

Previously

- Estimated the survival function using the Kaplan-Meier estimator
- Compared survival functions using the log-rank test
- Adjusted for categorical covariates using the stratified log-rank test
- Defined the hazard function and its relationship to the survival function

2

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer

The primary predictor of interest is trial arm (with or without trastuzumab). Researchers also considered the study (trial B-31 vs. trial N9831), intended chemotherapy schedule (every three weeks vs. weekly), nodal status (0, 1-3, 4-9, or 10+ positive nodes), and hormone-receptor status (estrogen- or progesterone-receptor positive vs. estrogen- and progesterone-receptor negative).

Comparison of the two groups was based on a log-rank test, stratified according to the study trial, intended chemotherapy schedule, nodal status, and hormone-receptor status. Thus, there were 2x2x4x2=32 strata

The stratified log-rank test p-value is <0.0001, indicating significant improvement in disease-free survival in the trastuzumab group adjusting for key covariates.

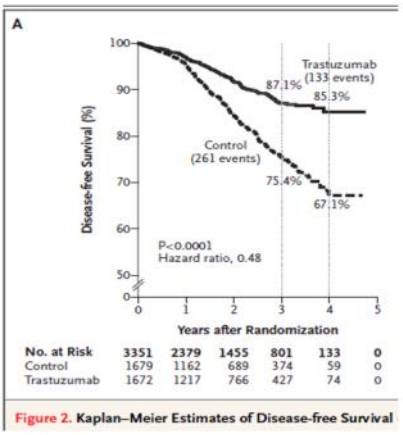


Figure 2. Kaplan–Meier Estimates of Disease-free Survival

3

Stratified log-rank test

But this required separating 32x !

- The stratified log-rank test stratified over 32 (!) unique strata
  - Women in trial B-31 on weekly chemo with 0-3 positive lymph nodes with hormone receptor negative cancer
  - Women in trial B-31 on weekly chemo with 4-9 positive lymph nodes with hormone receptor positive cancer (and so on…)
- We compare patients with and without trastuzumab in each strata
- In this way, the stratified log-rank test adjusts for the stratifying variables

But ....

4

Limitations of the stratified log-rank test

- It does not allow us to measure the simultaneous impact of these variables on survival
- It does not borrow information across similar groupings
  - Same trial, hormone receptor status, chemo frequency, but different # of nodes
- We can only stratify on categorical covariates
  - We may wish to model the number of positive lymph nodes as a continuous variable
- It does not provide a summary statistic for the effect size → can do p-value but doesn't provide estimate of effect
- *Thus, we may prefer a regression framework…*

5

Regression models

Dependent variable = function of independent variables

Linear regression: mean of continuous var

$$E[Y_i] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

Logistic regression: log odds

$$\log\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

Intercept that characterizes the dependent variable in the reference group

Coefficients that characterize the effect on the dependent variable of a one-unit change in the independent variable

6

Cox proportional hazards regression

Dependent variable = function of independent variables

Cox model:

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

Baseline hazard function that characterizes the dependent variable in the reference group

Coefficients that characterize the effect on the dependent variable of a one-unit change in the independent variable

this is a FUNCTION of t, no longer just mean or log odds

hazard in group w/ all covariates = 0

Assuming a common shape, which is shared across groups

covariates are multiplied (multiplicative effect)

7

Coefficients of the Cox model

- β_j is the **log hazard ratio** for the jth covariate
- $\exp(\beta_j)$ is the **hazard ratio** for the jth covariate

- For a **binary covariate**, this is the hazard ratio comparing the group with $X_{ij} = 1$ vs. the group with $X_{ij} = 0$
  - (*holding all other covariates constant*)
- For a **continuous covariate**, this is the hazard ratio for a one-unit increase in X_{ij}, e.g. $X_{ij} = 10$ vs. $X_{ij} = 9$
  - (*holding all other covariates constant*)

8

Ratio of hazards

holding other covariates constant

$$\frac{h_i'(t)}{h_i(t)} = \frac{h_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_j(X_{ij}+1) + \cdots + \beta_k X_{ik})}{h_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_j X_{ij} + \cdots + \beta_k X_{ik})}$$

$$= \exp[(\beta_1 X_{i1} + \cdots + \beta_j(X_{ij}+1) + \cdots + \beta_k X_{ik}) - (\beta_1 X_{i1} + \cdots + \beta_j X_{ij} + \cdots + \beta_k X_{ik})]$$

$$= \exp[\beta_j(X_{ij}+1) - \beta_j X_{ij}]$$

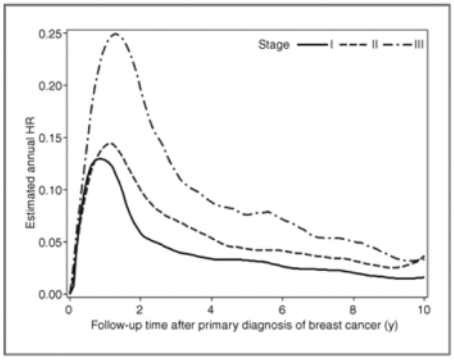
$$= \exp(\beta_j)$$

$$\log\left[\frac{h_i'(t)}{h_i(t)}\right] = \beta_j$$

log hazard ratio

Hazard ratio for one unit change in a covariate holding others zero

Example: Breast cancer recurrence



- Stage III vs. Stage I [ref]
  - Hazard ratio ≈ 2 → Stage III is 2x than stage I
  - Coefficient ≈ 0.69 → about the log of 2 if this is the only covariate
- Stage I vs. Stage III [ref]
  - Hazard ratio ≈ 0.5 → stage I is 0.5x than stage III
  - Coefficient ≈ −0.69

reciprocal

Source: Cheng et al. (2012) Cancer, Epi, Biomarkers & Prevention 10.1158/1055-9965.EPI-11-1089

Example: Lung cancer mortality

- Goal: Study risk factors associated with 3-year overall survival after lung cancer surgery
- Population: 601 lung cancer patients who underwent surgical treatment
- Outcome variable: Time from surgery to death by any cause
- Predictor variables: Age, gender, forced expiratory volume (FEV1), cancer stage, neoadjuvant therapy, and decortication

Source: Brembilla et al. (2018) J Thorac Dis 10.21037/jtd.2018.06.15

Example: Lung caner mortality

Covariates	Coefficient	Standard error	P value	HR	95% CI	
					Lower	Upper
Age (0: ≤70, 1: >70)	0.584	0.191	0.002	1.793	1.232	2.610
Gender (0: M, 1: F)	0.276	0.230	0.229	1.318	0.840	2.068
FEV1 (0: <80, 1: ≥80)	−0.197	0.165	0.233	0.821	0.594	1.135
Simplified cancer stage						
Stage 2	1.509	0.528	0.004	4.523	1.609	12.720
Stage 3	1.616	0.511	0.002	5.031	1.848	13.700
Stage 4	1.808	0.619	0.004	6.097	1.811	20.528
Neoadjuvant therapy (0: no, 1: yes)	0.055	0.188	0.769	1.057	0.731	1.529
Decortication (0: no, 1: yes)	0.471	0.232	0.043	1.602	1.016	2.527

FEV1, forced expiratory volume in 1 second.

Source: Brembilla et al. (2018) J Thorac Dis 10.21037/jtd.2018.06.15

12

(Proportional hazards)

PH with multiple covariates

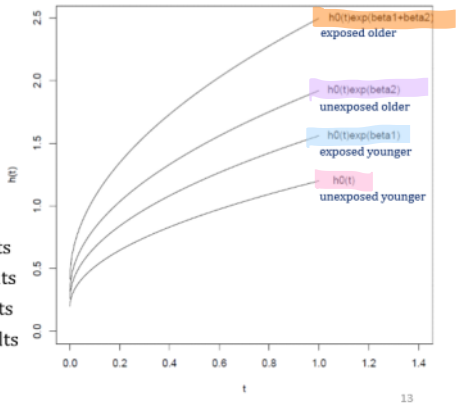
Reference group: Unexposed, young adults

- Say we have two covariates
  - $X_{i1} = 1$ if exposed, 0 if unexposed
  - $X_{i2} = 1$ if older, 0 if younger

- Fit a Cox model

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2})$$

$h_i(t) = h_0(t)$ for unexposed younger adults
$h_i(t) = h_0(t)e^{\beta_1}$ for exposed younger adults
$h_i(t) = h_0(t)e^{\beta_2}$ for unexposed older adults
$h_i(t) = h_0(t)e^{\beta_1+\beta_2}$ for exposed older adults



13

PH with multiple covariates

- Say we have two covariates
  - $X_{i1} = 1$ if exposed, 0 if unexposed
  - $X_{i2} = 1$ if older, 0 if younger

So the log hazard will just be increased (shifted up) by the β's

- Fit a Cox model

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2})$$

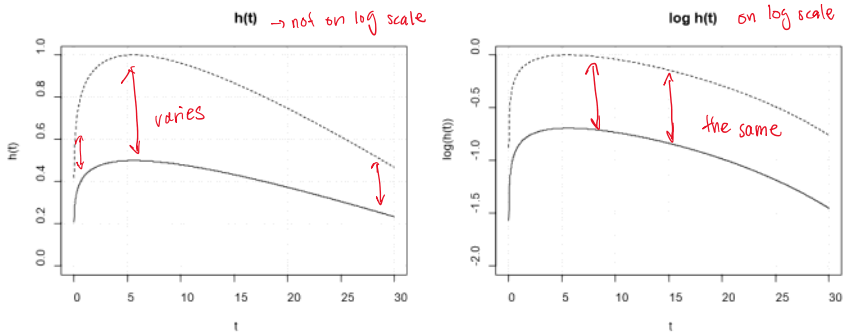
- Equivalently

$$\rightarrow \quad \log(h_i(t)) = \log(h_0(t)) + \beta_1 X_{i1} + \beta_2 X_{i2}$$

$h_i(t) = h_0(t)$ for unexposed younger adults → $\log(h_i(t)) = \log(h_0(t))$
$h_i(t) = h_0(t)e^{\beta_1}$ for exposed younger adults → $\log(h_i(t)) = \log(h_0(t)) + \beta_1$
$h_i(t) = h_0(t)e^{\beta_2}$ for unexposed older adults → $\log(h_i(t)) = \log(h_0(t)) + \beta_2$
$h_i(t) = h_0(t)e^{\beta_1+\beta_2}$ for exposed older adults → $\log(h_i(t)) = \log(h_0(t)) + \beta_1 + \beta_2$

14

PH on the log scale



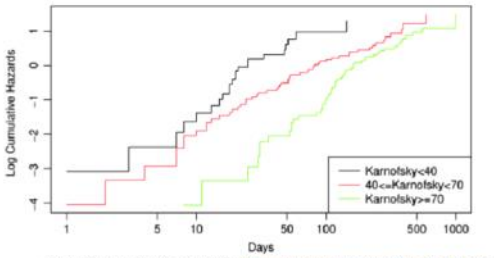
Suggests that one way to examine the proportional hazards assumption is to check that the log hazard functions are parallel

15

PH on the log cumulative hazard scale

$$H_1(t) = \exp(\beta)H_0(t)$$

$$\log\big(H_1(t)\big) = \beta + \log\big(H_0(t)\big)$$



Source: Wang and Li (2017) Quant Biosci https://doi.org/10.22283%2Fqbs.2017.36.2.85

Suggests that another way to check the proportional hazards assumption is by seeing if the log cumulative hazard functions are parallel

16

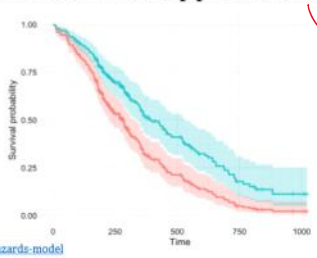
PH on the survival scale

$$S_1(t) = [S_0(t)]^{\exp(\beta)}$$

because there is a steady divergence

- Under the proportional hazards assumption, the survival curves diverge over time (until survival approaches 0)

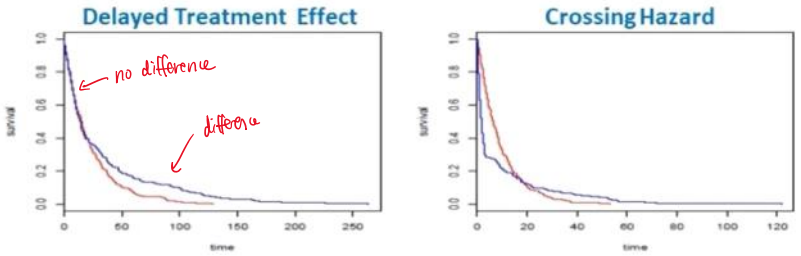
↳ you run out of people



Source: http://www.sthda.com/english/wiki/cox-proportional-hazards-model

17

When proportional hazards is violated



Relation between two groups is not the same over time → can't fit one value if the relationship changes over time

Source: http://onbiostatistics.blogspot.com/2020/10/visual-inspection-and-statistical-tests.html

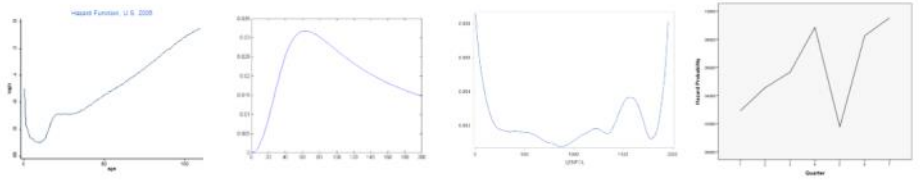
18

Understanding the Cox model

Assuming this basic format

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

- Instead of an intercept, the baseline hazard is an entire FUNCTION
- This function can take any of an infinite number of shapes



19

Semiparametrics

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

- The baseline hazard function $h_0(t)$ is **nonparametric** → no assumption about distr. of fxn
- The hazard ratio $\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$ is **parametric** → we ARE assuming for hazard ratio
  - We assume that the effect of each covariate is to multiply the hazard
  - We assume that the hazards are proportional (constant ratio for all t)

- The Cox model is **semiparametric**

20

Inference on the Cox model

- We are most interested in estimating the terms $\beta_1, \ldots, \beta_k$
- Fitting the nonparametric baseline hazard is more complex
  - We might consider $h_0(\cdot)$ to be a "nuisance"

- It is possible to estimate $\beta_1, \ldots, \beta_k$ without estimating $h_0(\cdot)$ or making distributional assumption

$$L(\beta_1, \ldots, \beta_k) L(h_0(\cdot), \beta_1, \ldots, \beta_k)$$

only coefficients; hazard fxn; likelihood breaks into 2 terms

- The left-hand term is known as the **Cox partial likelihood**
- We can find the values $\hat{\beta}_1, \ldots, \hat{\beta}_k$ that maximize the partial likelihood
  - These are known as *maximum partial likelihood estimators*

21

Cox partial likelihood → aka $L(\beta_1, \ldots, \beta_k)$ from previous slide

- Consider unique failure times $T_1, \ldots, T_m$
- At time T_i, individual i fails; the risk set at that time is $R(T_i)$ — who is at risk (haven't failed/been censored yet)
- The partial likelihood is the product over all m unique failure times:

$$\prod_{i=1}^{m} \frac{\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})}{\sum_{j \in R(T_i)} \exp(\beta_1 X_{j1} + \cdots + \beta_k X_{jk})}$$

hazard ratio for person who DID fail vs. hazard of everyone still at risk

- The contribution at each failure time is the conditional probability that individual i was the one to fail out of risk set $R(T_i)$

22

Notes on the partial likelihood

you can rescale, etc.

$$\prod_{i=1}^{m} \frac{\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})}{\sum_{j \in R(T_i)} \exp(\beta_1 X_{j1} + \cdots + \beta_k X_{jk})}$$

- Note that the partial likelihood does not contain $h_0(t)$
- Note that this calculation assumes all failure times are unique; modifications are required to accommodate tied times

need to be able to assign who failed first

23

Methods for handling ties

Method	Approach	Advantages/Disadvantages
Exact method	Considers all $d_i!$ possible combinations of breaking ties	(++) Most accurate (−) Computationally complex
Efron approximation	Approximates the mean denominator of the Exact method	(+) Accurate (+) Computationally efficient
Breslow approximation	Ignores the fact that failures are removed from the risk set	(+) Simplest (−) Least accurate

Prediction

- Given a final model, we are usually most interested in reporting which covariates increase/decrease survival
  - Hazard ratios = $\exp(\hat{\beta})$
  - Hypothesis tests
- Less frequently, we also wish to generate predictions of survival based on our semiparametric regression model
  - Survival curve, median survival, survival at a landmark time
  - Calculated for a particular set of covariates

Prediction

- The proportional hazards model gives us a framework for making predictions

hazard function

$$\hat{h}_i(t) = \hat{h}_0(t)\exp(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})$$

cumulative hazard function

$$\hat{H}_i(t) = \hat{H}_0(t)\exp(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})$$

$$\hat{S}_i(t) = \exp\left(-\hat{H}_i(t)\right)$$

Survival

- Using partial likelihood estimation, we estimate $\hat{\beta}_1, \ldots, \hat{\beta}_k$ but not the flexible baseline hazard or cumulative hazard functions

Baseline cumulative hazard

- **Breslow estimator** of the baseline cumulative hazard function:

$$\tilde{H}_0(t) = \sum_{j:t_j \le t} \frac{d_j}{\sum_{i=1}^{n_j} \exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)}$$

- Related to the Nelson-Aalen cumulative hazard estimator:

$$\tilde{H}(t) = \sum_{j:t_j \le t} \frac{d_j}{n_j}$$

instead, sum each individual's hazard ratio

27

Prediction

- We can predict survival at time t for a person with covariates $X_{i1}, \ldots, X_{ik}$ as:

$$\tilde{S}_i(t) = \exp\left(-\tilde{H}_0(t)\exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)\right)$$

$$= \left[\tilde{S}_0(t)\right]^{\exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)}$$

$$= \left[e^{-\tilde{H}_0(t)}\right]^{\exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)}$$

28

Example: Liver disease

- Goal: An existing **prognostic model** for **end stage liver disease** (MELD) has been validated for short-term and intermediate-term mortality outcomes for patients with cirrhosis and end-stage liver disease. This model is valuable for transplant planning. Researchers sought to **extend this model** for predicting outcomes for patients with early-stage cirrhosis and alcoholic hepatitis.

Good for transplant stuff seeing who will respond best

29

Example: Liver disease

Journal of Hepatology 40 (2004) 897–903

Journal of Hepatology
www.elsevier.com/locate/jhep

Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease

Adnan Said[1], John Williams[1], Jeremy Holden[1], Patrick Remington[2], Ronald Gangnon[3], Alexandru Musat[1], Michael R. Lucey[1,*]

- Population: They constructed a retrospective cohort using data on **1611 patients with chronic liver disease** from the hepatology clinics and inpatient service of a university hospital between January 1994 and December 2001.

- Outcome variable: Survival was calculated from the **date of first clinical contact to date of death**, abstracted from hospital records and the national social security death index. Survival was censored if the patient received a liver transplant.

Example: Liver disease

Journal of Hepatology 40 (2004) 897–903

Journal of Hepatology
www.elsevier.com/locate/jhep

Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease

Adnan Said[1], John Williams[1], Jeremy Holden[1], Patrick Remington[2], Ronald Gangnon[3], Alexandru Musat[1], Michael R. Lucey[1,*]

- Predictor variables: The primary predictor of interest was **MELD score**, which was calculated based on serum creatinine, bilirubin, and INR (international normalized ratio, measure of liver function) measured at the first visit. The continuous MELD score was then categorized (MELD categories 1-4, scores ≤9, 10-19, 20-29, ≥30, respectively). Higher categories indicate poorer liver function.
- **Other prognostic variables** considered included age, gender, race, etiology of chronic liver disease (alcoholic liver disease, hepatitis C virus, other), diagnosis of end-stage liver disease, and CTP score (another prognostic measure).

Example: Liver disease

Journal of Hepatology 40 (2004) 897–903

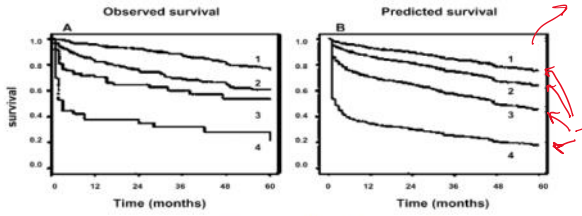
Journal of Hepatology
www.elsevier.com/locate/jhep

Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease

Adnan Said[1], John Williams[1], Jeremy Holden[1], Patrick Remington[2], Ronald Gangnon[3], Alexandru Musat[1], Michael R. Lucey[1,*]

- Statistical analysis: Researchers fit a **Cox proportional hazards model** to identify covariates that were predictive of survival in a model that already includes MELD score. **Survival probabilities** for each MELD category were predicted by Kaplan-Meier and by Cox regression (after adjusting for other covariates).

Results: Survival curves are plotted in Figure 1 (Kaplan-Meier, left; Cox model, right).



Why is this smoother? proportional hazards model

the predicted curves are fit using ALL of the data, people just have different weights

And multiplying by proportional terms

Fig. 1. Survival of cohort with chronic liver disease. Observed survival curves (Kaplan–Meier estimates) and predicted (Cox regression) survival curves predict increasing mortality with higher MELD categories. Categories: 1 = MELD ≤ 9, 2 = MELD 10 to 19, 3 = MELD 20 to 29, 4 = MELD ≥ 30. Pairwise comparisons: 1 vs 2, p<0.0001; 1 vs 3, p<0.0001; 1 vs 4, p<0.0001; 2 vs 3, p = 0.004; 2 vs 4, p<0.0001; 3 vs 4, p<0.0001. Comparison of observed (A) and predicted (B) curves: Category 1 (≤ 9), p = 0.95; Category 2 (10–19), p = 0.05; Category 3 (20–29), p = 0.34; Category 4 (≥ 30), p = 0.51; Overall (non-categorized MELD), p = 0.18.

Said et al. (2004)

Example: Liver disease

ELSEVIER
Journal of Hepatology

Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease

Adnan Said[1], John Williams[1], Jeremy Holden[1], Patrick Remington[2], Ronald Gangnon[3], Alexandru Musat[1], Michael R. Lucey[1,*]

- Results, continued: **MELD score is a significant predictor of survival**, capturing large differences in short-term survival between the risk categories.
- There is **good concordance** between observed and predicted survival. Note that the curves generated by the Cox method are smoother, as they borrow information across all participants. They also share the same basic shape, because their hazard functions have the same underlying shape.

Looking ahead

Lecture 6
- Interpreting hazard ratios for binary and continuous data
- Hypothesis testing
- Interactions, transformations, polynomials, splines

Lecture 7
- Examining the proportional hazards assumption
- Residuals
- Goodness of fit
- Visual diagnostics
- Model selection

Lecture 8
- Time-dependent covariates
- Time-varying effects
- Stratified Cox model