L7: Comparing hazard functions

EPID 722
Spring 2021
UNC – Chapel Hill
jessedwards@unc.edu

An inflection point

So far, we have focused on estimating risk in a single sample.

Or comparing risk across samples using risk ratios and differences.

Today, we discuss various **regression models** for comparing groups with survival data

Why use a regression model?

Allows us to estimate different parameters of interest

- Hazard ratios
- Time ratios

“Easy” to account for confounding.

Today's plan

The purpose of today's lecture is to *introduce* you to **Cox proportional hazards models**, which are commonly used regression models in survival analysis.

Earlier this semester we also discussed **accelerated failure time models**, which are also regression models used in this context. (We will not cover those further here).

Chapter 1: Comparing hazards using Cox models

Recall, the hazard function

Discrete-time hazard

$$h(k) = y(k)/\{n(k)\Delta t\}$$

- k indexes (discrete) time
- $y(k) = \sum_{i=1}^{N} \delta_i(k)$
- $n(k)$: size of the risk set at time k

Continuous time hazard:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | W < t \leq T)}{\Delta t}$$

The components of a "survival analysis" we have discussed so far are still important

Origin/timescale

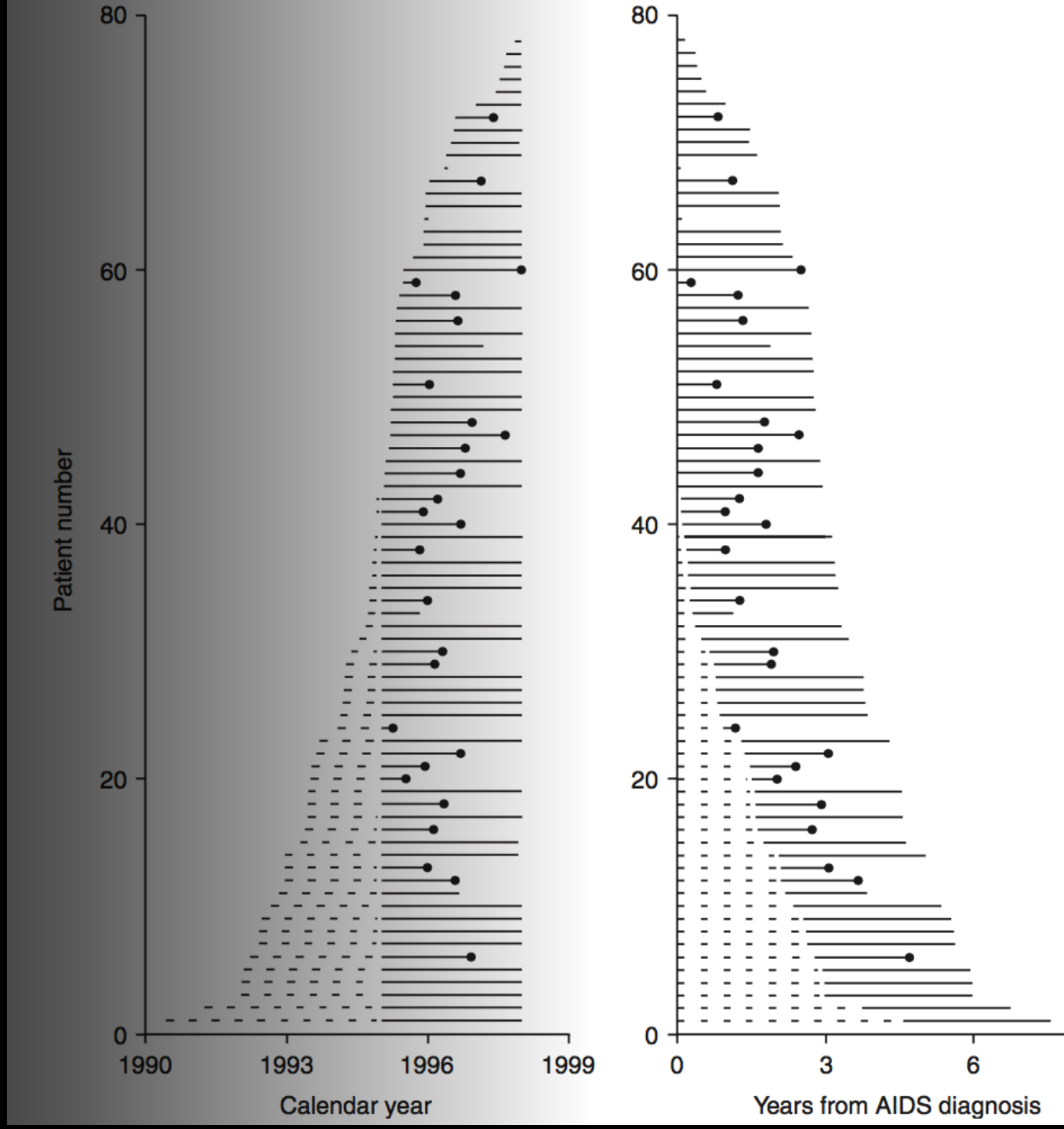
(t measures time since the origin)

Target population

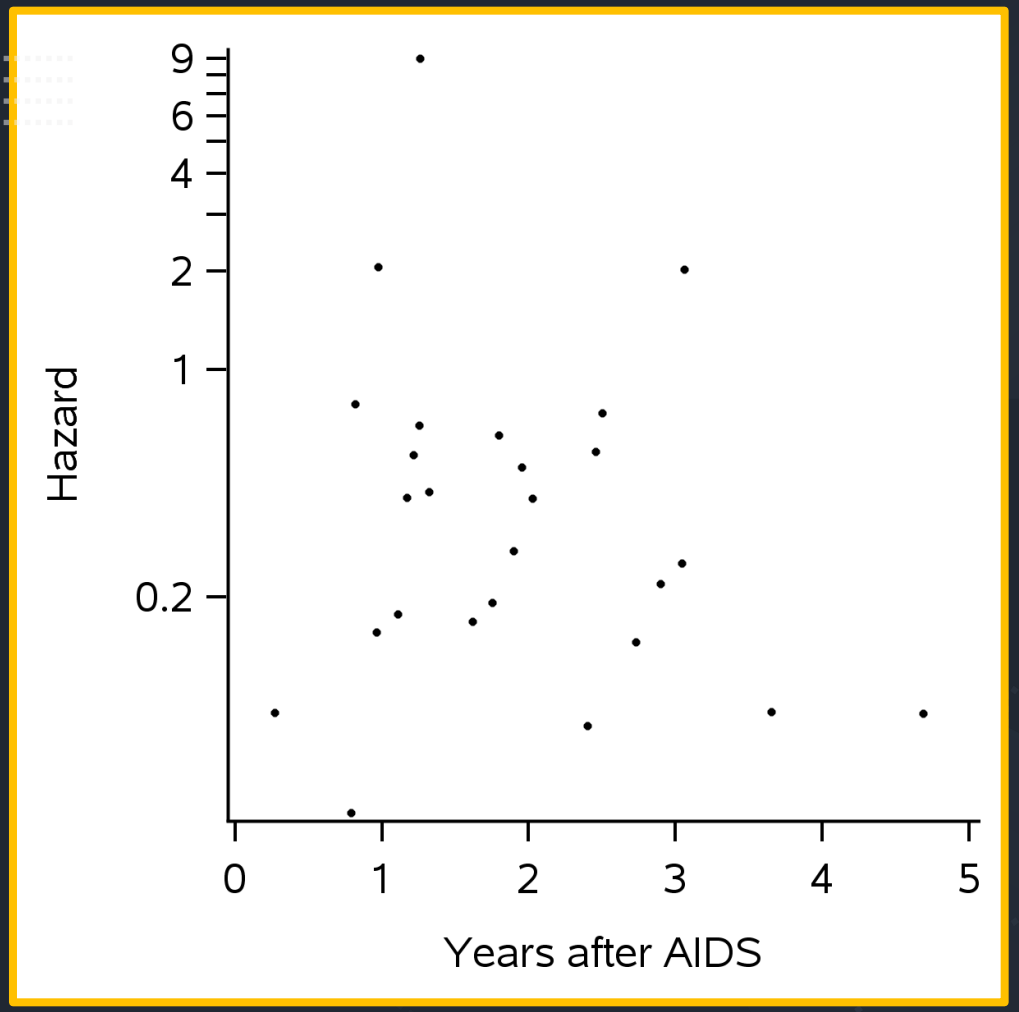
Event definition and timing

Group membership or actions

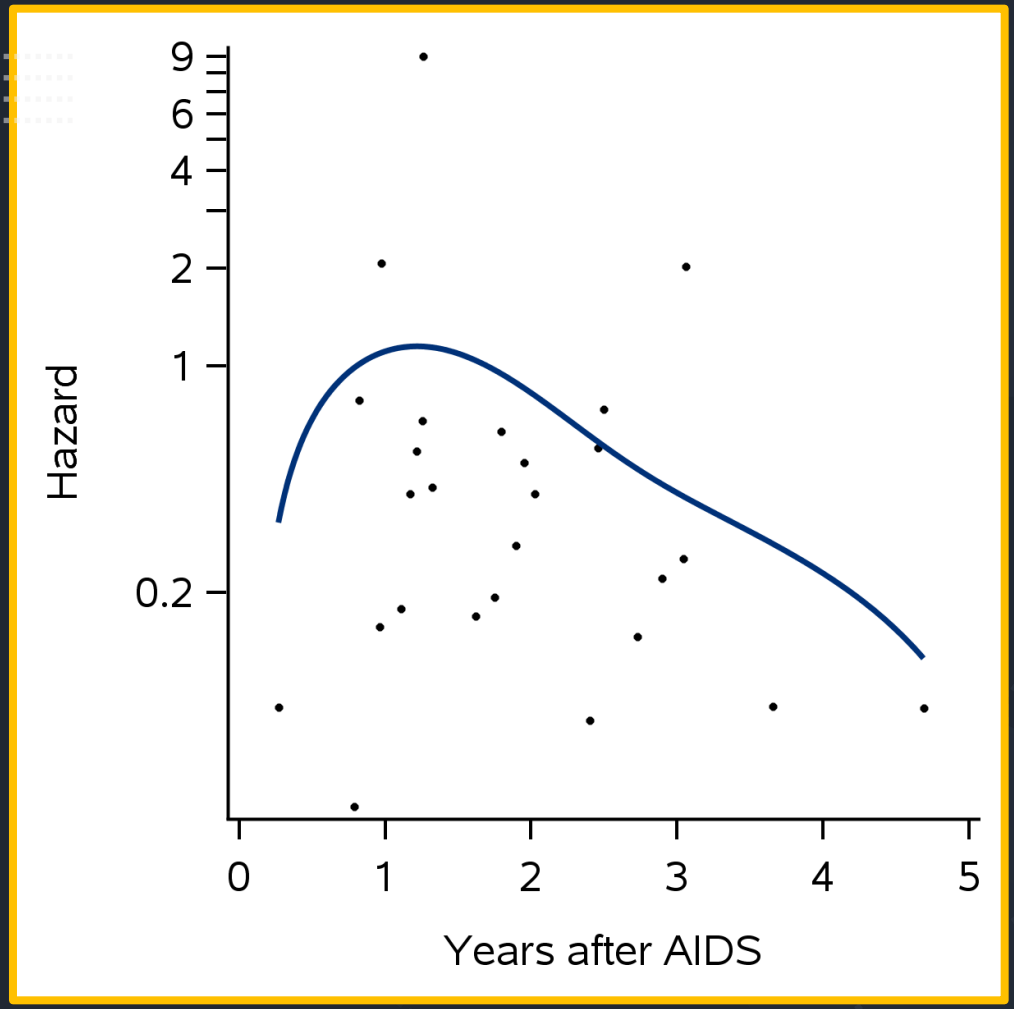
Recall, example



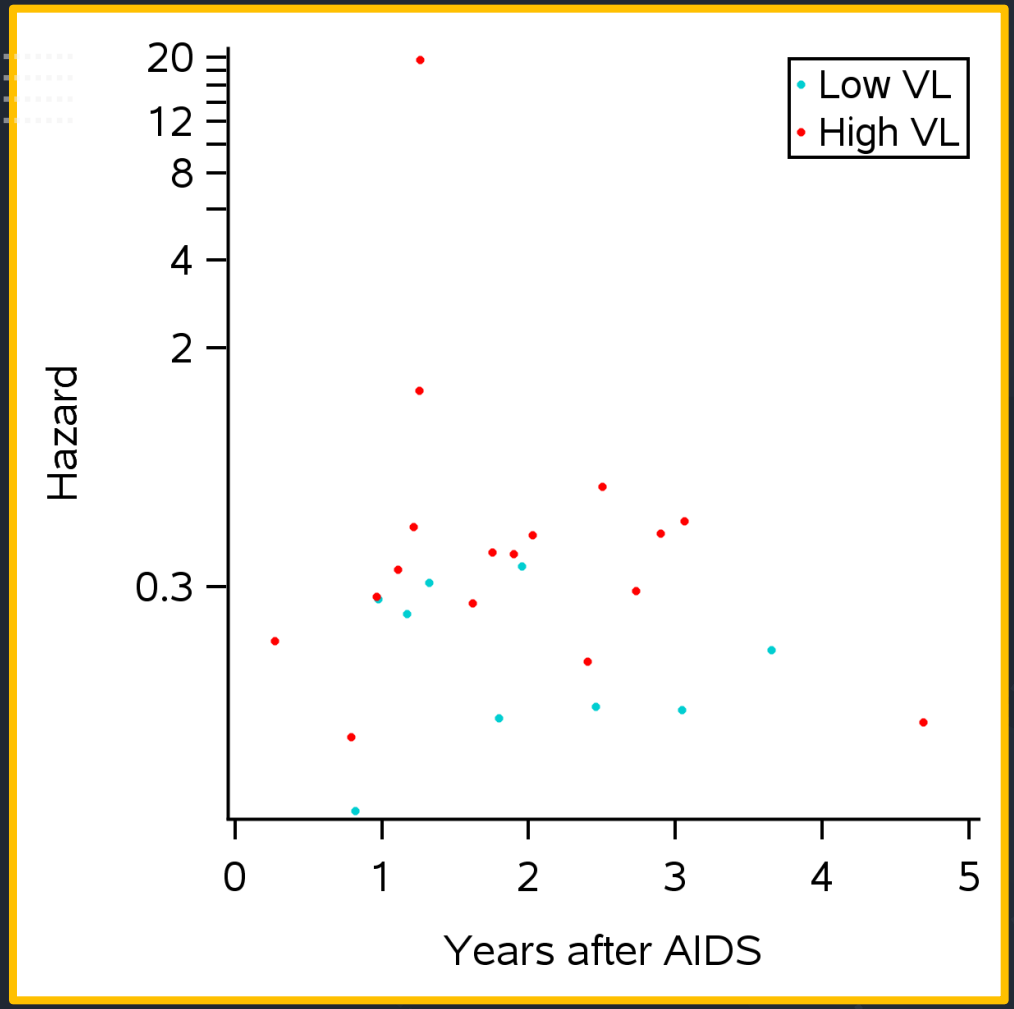
The hazard, illustrated



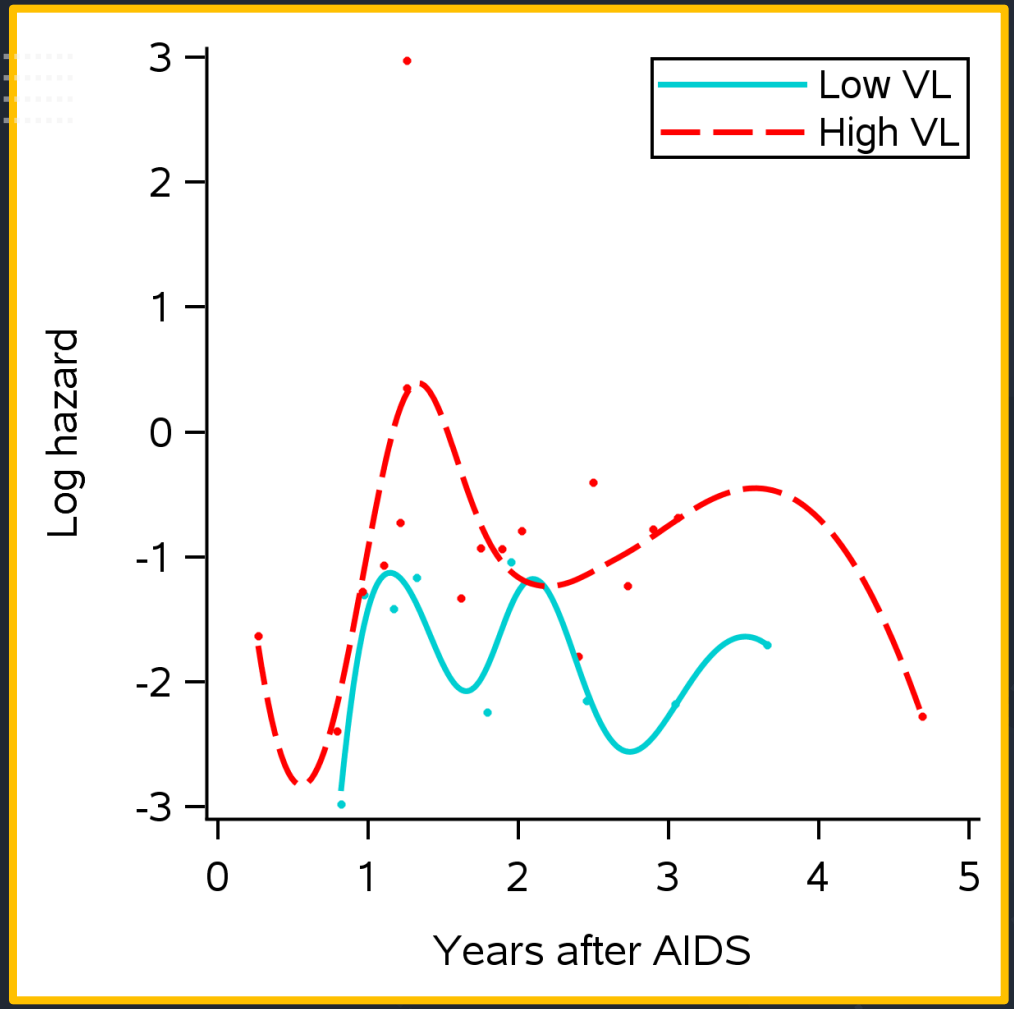
The hazard, illustrated



The hazard, by group (high vs low VL)



The hazard, by group (high vs low VL)



Intuition?

The hazard is the "instantaneous rate of the event"

To compare between groups:

$$h_{x=0}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | W < t \leq T, x = 0)}{\Delta t}$$

$$h_{x=1}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | W < t \leq T, x = 1)}{\Delta t}$$

Hazard ratio function: $HR(t) = h_1(t)/h_2(t)$

We can collapse $HR(t)$ to HR under the assumption that the HR is constant.

Cox model

Parameter of interest is $HR = \frac{h_1(t)}{h_0(t)}$

Encode parameter of interest into regression model:

$$HR = \exp(\beta) = h_1(t)/h_0(t)$$
$$h_1(t) = h_0(t)\exp(\beta)$$
$$h(t) = h_0(t)\exp(\beta x)$$

Hazard of $x = 1$ group is a multiple of $x = 0$ group.

Cox model intuition

β is the average log(HR)

β is the average $\log(h_1/h_0)$

β is the average $\log(h_1) - \log(h_0)$

β is the average difference in log hazards

$$HR = \exp\{\beta\}$$

The HR is the antilog of the average difference in log hazards.

t	$y_{x=1}$	$n_{x=1}$	$y_{x=0}$	$n_{x=0}$	$\boldsymbol{h_0(t)}$	$\boldsymbol{h_1(t)}$	$\boldsymbol{log(h_0(t))}$	$\boldsymbol{log(h_1(t))}$	$\boldsymbol{\Delta \log\{h_x(t)\}}$	**HR**
1	1	5	2	5						
2	1	5	2	5						
3	1	5	2	5						
4	1	5	2	5						
5	1	5	2	5						
6	1	5	2	5						

The "baseline" hazard function

An advantage of the cox model is that **we need not specify the "baseline" hazard function** $h_0(t)$.

Cox model: $h_x(t) = h_0(t)\exp(\beta x)$

To estimate the HR comparing $x = 1$ to $x = 0$:

$$\frac{h_1(t)}{h_0(t)} = \frac{h_0(t)\exp\{\beta(1)\}}{h_0(t)\exp\{\beta(0)\}} = \exp\{\beta\}$$

$h_0(\mathrm{t})$ cancels!

What is the baseline hazard function?
The hazard function among participants with 0 level of all variables included in the model.

We can learn about the ratio of hazard functions without learning about the hazard function in the reference group, just like we can learn about the difference between groups without learning about the level of groups in say the paired-difference t tests. [3]

[3] That we *can* do something is not warrant *for* doing it.

Partial likelihood

Functions as a standard likelihood; i.e., estimates are asymptotically consistent, asymptotically normal, and fully efficient given the model.

Factors out the reference hazard function $h_0(t)$, so that nuisance portion of the likelihood needn't be estimated (i.e., no intercept).

Because $h_0(t)$ is not estimated it is a semiparametric model.

See Appendix for details.

Ties

If there are no tied event times, the above PL is valid

When there are ties (why might this happen?)

- Breslow's method (SAS default)
- Efron's approximation (better)
- Exact method (best, but can be computationally intense)
- Discrete method (best if event times really are discrete)

CH5 of the Allison book covers these in detail

Proportional hazards

Standard Cox model assumes log hazard functions for X=1 and X=0 groups are equal distance over time.

Or, that we are interested in the average difference in log hazard functions over the study period of length τ.

We can check the proportional hazards assumption by plotting the cumulative hazard functions for each group and looking to see if they are parallel (see example).

Alternatives: product term with time, Schoenfeld residuals, others?

Relaxing the proportional hazards assumption

We can fit a model with product term between X and $g(t)$, where $g(.)$ is a user-specified function (say log), and t is a time-updated covariate (rather than T)

If HR is not constant, a refined choice of $g(.)$ may be the solution.

If HR is not constant (and we do not wish to average), we must report >1 HR. (see example).

Accounting for confounding in Cox models

Methods to account for confounding in Cox models

- Multivariable Cox models
- Stratified Cox models
- Inverse probability weighting

Say we wish to estimate the association between X and the hazard function, accounting for confounding by Z.

Multivariable Cox models

Recall: $h(t) = h_0(t)\exp(\beta x)$

Extend: $h(t) = h_0(t)\exp(\beta_1 x + \beta_2 z)$

Would you want to interpret $\exp(\beta_2)$?

Do you expect $\beta = \beta_1$ if Z is not associated with exposure?

Interpretation:

- $\exp(\beta_1)$is the HR for a unit difference in X holding Z constant at any level (calculated at each level, and information-weighted averaged over levels)
- $\exp(\beta_2)$ is the HR for unit difference in Z, likewise holding X constant at any level.

Stratified Cox models

Recall: $h(t) = h_0(t)\exp(\beta x)$

Extend: $h(t) = h_{0z}(t)\exp(\alpha x)$

$h_{0z}(t)$ represents the hazard function among the unexposed in stratum z of discrete nuisance variable Z.

More flexible than including Z in the model

- Does not constrain $h_{0z=1}$ and $h_{0z=0}$ to be proportional
- Implicitly allows for interactions between X and Z

Works by

- Constructing separate partial likelihood functions for each stratum of Z
- Multiplies functions together
- Chooses value of β that maximizes this function

Model form assumptions

While we can ignore the baseline hazard function, we must still model the parametric part of the model correctly

- Assess functional form between regressors and outcome
- Account for nonmultiplicativity using interaction terms
- Careful not to include too many regressors in small samples (rule of thumb: 10 events/regressor)

Chapter 2: Example

Recall, example

- Interested in describing survival after AIDS in MACS
- Origin = AIDS diagnosis, Event = all-cause mortality, Time scale = AIDS duration
- Enroll 42 men alive on 1 January 1995 with a prior clinical AIDS diagnosis and enroll 36 additional men with a clinical AIDS diagnosis between 1 January 1995 and 1 January 1998
- Follow all 78 (= 42 + 36) men for all-cause mortality through 1 January 1998, the date of study completion

Recall, example

Characteristics	Median (IQR)
Date, calendar year	‘94.89 (‘93.50, ‘95.48)
AIDS duration, years	0.11 (0, .5)
Age, years	40 (35, 44)
Nonwhite, %	14 (18%)
HIV viral load, log10 copies/ml (IQR)	5.0 (4.6, 5.2)
HIV viral load >10^5, n (%)	36 (46%)

Parameter of interest

Suppose we wish to estimate the association between high viral load at AIDS diagnosis and time to death using a hazard ratio.

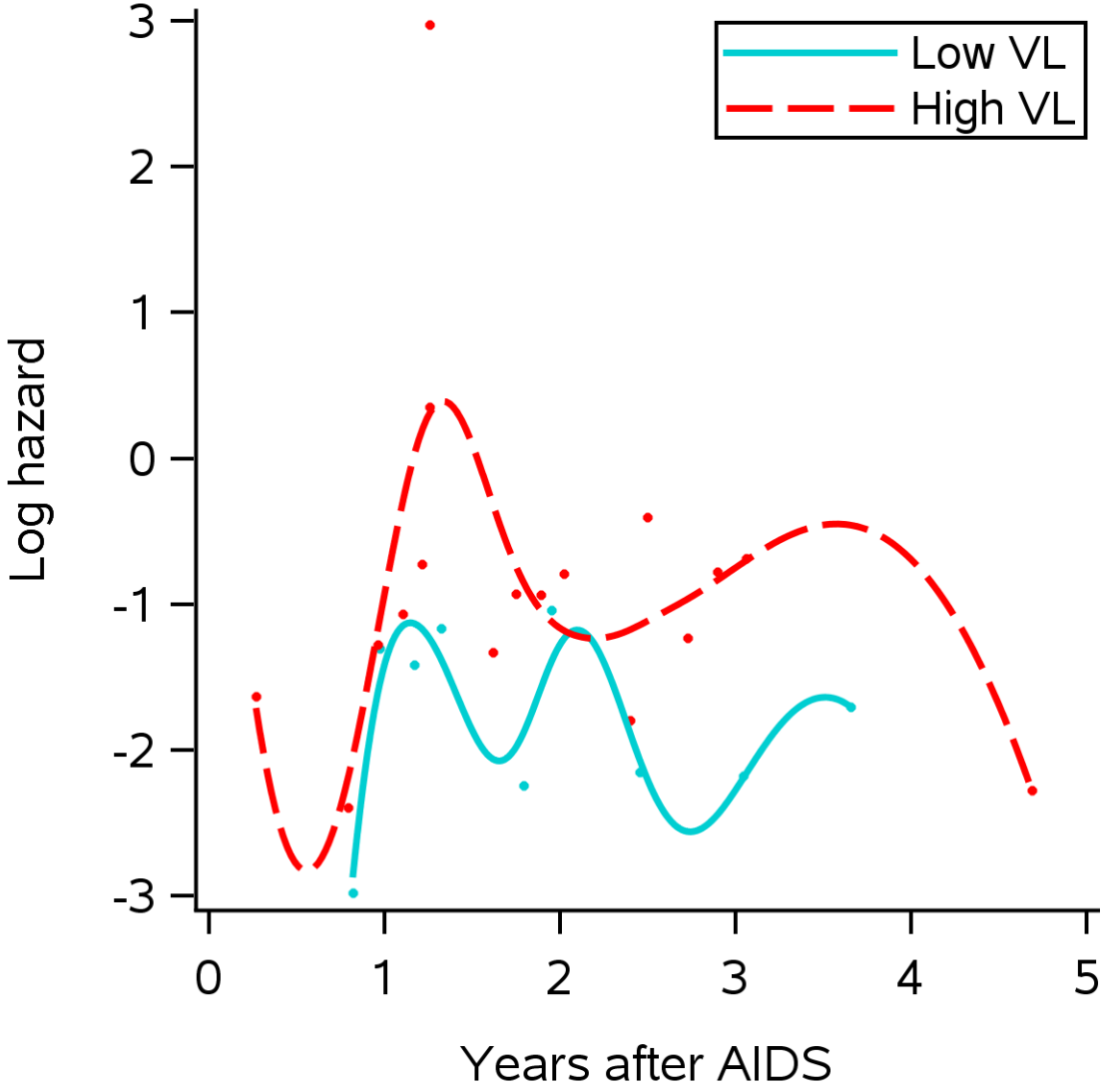
We wish to estimate this association both crude and accounting for confounding by age and race*.

* We control for race NOT because we think that race itself necessarily affects the exposure or outcome but because there are structural forces that are associated with both exposure and outcome, and reported ‘race’ is an imperfect proxy for identifying those most affected by these factors.

Takeaway messages: “controlling” for “race”

- Reported race is often an ascending or descending proxy for what we actually want to account for in our analyses
  - Might result in incomplete control for confounding
- How we categorize is important
  - In finite sample, we must often coarsen our categorization of confounders
  - Might result in incomplete control for confounding
- Think carefully about when you want to “control away” the effects of race vs when they are of interest
  - **This depends on your QUESTION**

The hazard, by group (high vs low VL)



Example: Crude Cox model

$$h(t) = h_0(t)\exp(\beta \times high\ VL)$$

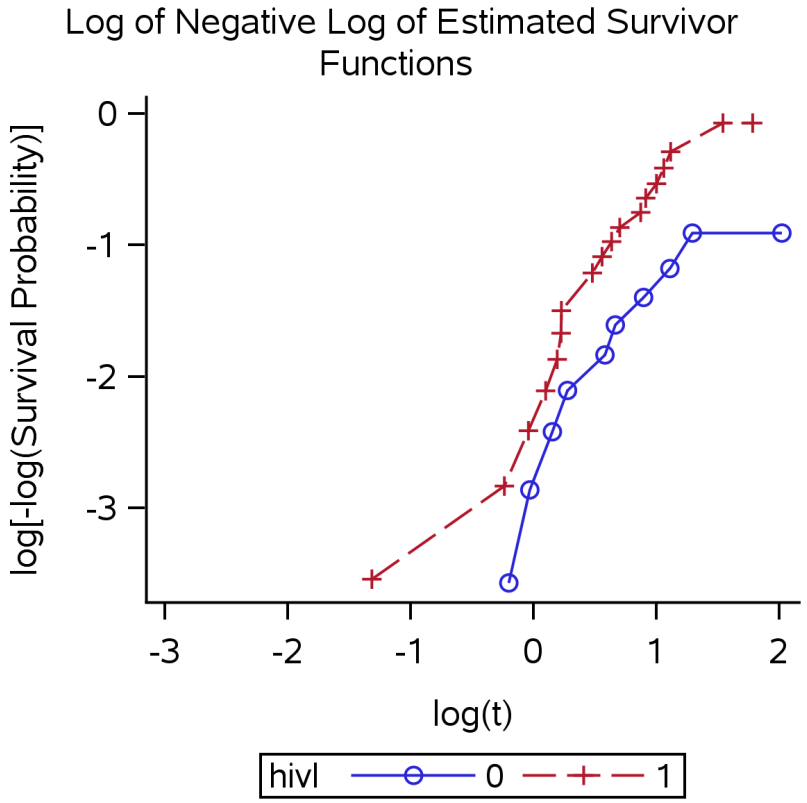
```
proc phreg data=a;
model t*d(0)=hivl/ties=efron rl;
run;
```

Type	HR	95% CI
Crude	2.16	0.97, 4.82

Example: Assessing PH graphically

Plot log(H(t)) = log(-log(S(t))) by group

```
proc lifetest data=a
      plots=(loglogs s) method=pl;
      time t*d(0);
      strata hivl;
run;
```



Example: Assessing PH (interactions with time)

```
proc phreg data=a;
     model t*d(0)=hivl tvl/ties=efron rl;
     tvl=t*hivl;
run;
```

Note that interaction programming statements come within model!

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
hivl	1	0.78490	0.89710	0.7655	0.3816	2.192
tvl	1	-0.00727	0.42018	0.0003	0.9862	0.993

Example: Assessing PH (interactions with time)

Dichotomous time (does HR vary before and after 2.5 years with AIDS?)

```
proc phreg data=a;
      model t*d(0)=hivl tvl/ties=efron rl;
      tvl=hivl*(t>2.5);
run;
```

HR before 2.5 years: 2.11

HR after 2.5 years = $\exp(\beta_1 + \beta_2) = 2.34$

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
hivl	1	0.74822	0.46301	2.6113	0.1061	2.113
tvl	1	0.10209	0.98448	0.0108	0.9174	1.107

Example: Multivariable model

Now fit multivariable model to account for confounding by age and race

$$h(t) = h_0(t)\exp(\beta_1 highvl + \beta_2 g(age) + \beta_3 nonwht)$$

Considerations:

How to model age?

How to categorize race?

Example: Results so far

Type	HR	95% CI
Crude	2.16	0.97, 4.82
Multivariable (race, loglinear age)	2.39	1.06, 5.36
Multivariable (race, curviloglinear age)	2.31	1.01, 5.27
Multivariable (race)	2.39	1.06, 5.36

Example: Stratified model

$$h(t) = h_{0z}(t)\exp(\beta_1 highvl)$$

```
proc phreg data=a ;
     strata nw;
     model t*d(0)=hivl/ties=efron rl;
run;
```

Type	HR	95% CI
Crude	2.16	0.97, 4.82
Multivariable (race, loglinear age)	2.39	1.06, 5.36
Multivariable (race)	2.39	1.06, 5.36
Stratified (race)	2.29	1.02, 5.13

Example: Interpretation

We estimated a series of HRs describing the association between high VL at baseline and mortality.

Why?

- Identify important risk factor for mortality among people with AIDS so we can better allocate resources?
- Estimate a causal effect of high VL at AIDS so that we can possibly intervene on high VL?

Chapter 3: Discussion

What are Dr. Hernán’s issues with the HR?

What does he propose instead?

Do you agree?

L7: Comparing hazard functions

EPID 722
Spring 2021
UNC – Chapel Hill
jessedwards@unc.edu

Appendix

Partial likelihood

The Cox model relies on a partial likelihood function

- The likelihood function factors into 2 parts (1 that depends on $h_0(t)$ and β and 1 that depends on β alone.
- The cox model discards the first part of the likelihood and maximizes only the second part, which does not depend on $h_0(t)$

Consequences:

- Robustness to shape of baseline hazard function
- Less precise than if we had used full likelihood (though loss of efficiency is small)

Partial likelihood (PL)

$$PL = \prod_{i=1}^{N} \left[\frac{\exp(\beta X_i)}{\sum_{j=1}^{N} R_j(T_i^*) \exp(\beta X_j)} \right]^{\delta_i}$$

When $\delta_i = 1$ (i.e., event) the PL equals the ratio of $\exp\{\beta X_i\}$ for the case i to same function summed over the entire risk set at time t. ($R_j(T_i^*)$ is an indicator that person j is in the risk set at event/censoring time T_i^*).

When $\delta_i = 0$ (i.e., nonevent), $PL = 1$ (meaning censored observations contribute only to denominator).