SESSION 1:
REVIEW AND COX MODEL
FOR ADJUSTMENT AND INTERACTION

Module 13: Survival Analysis for Observational Data
Summer Institute in Statistics for Clinical Research
University of Washington
July, 2019

Ying Qing Chen, Ph.D.
Affiliate Professor
Department of Biostatistics
University of Washington

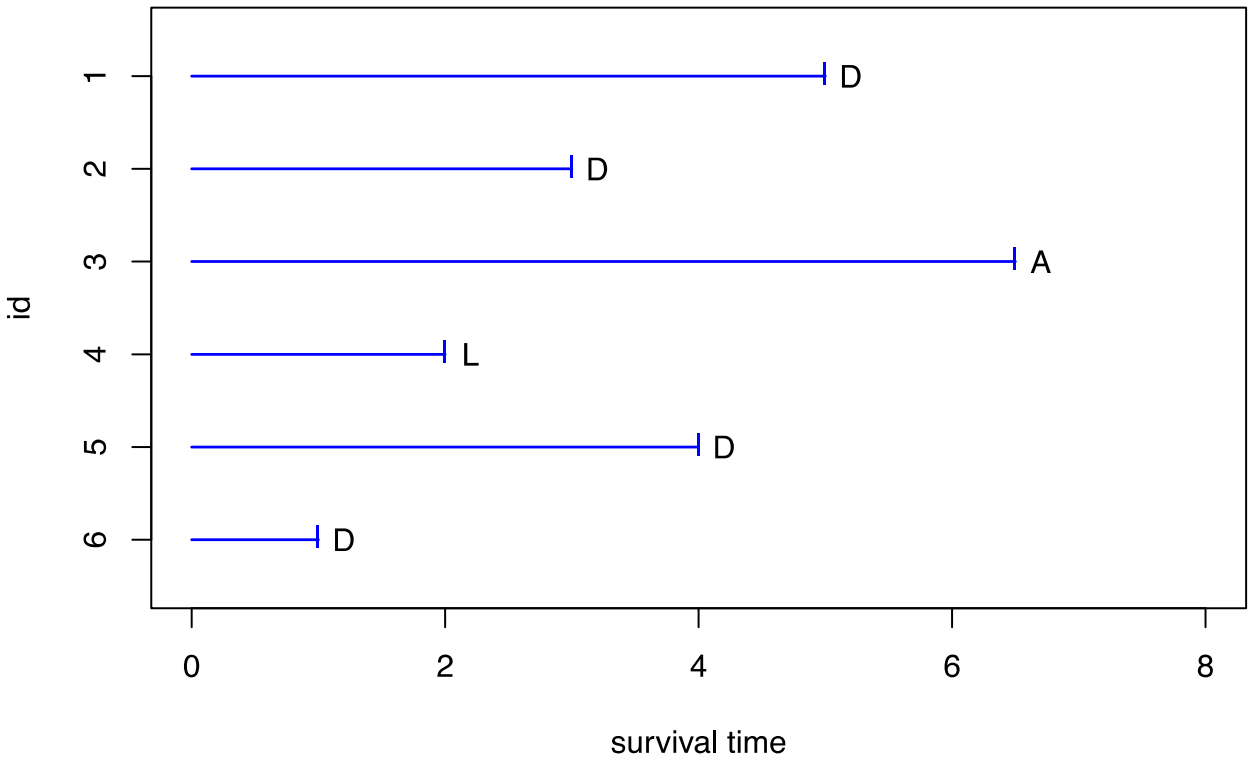
OVERVIEW

- Session 1
  - Quick review of introductory material
  - Adjustment in the Cox model: confounding and precision
  - Effect modification in the Cox model
- Session 2
  - Nonparametric hazard function estimation
  - Competing risks
  - Cumulative Incidence estimation
- Session 3
  - Left entry and left truncation
  - Choice of the time variable
  - Interactions with functions of time
- Session 4
  - Immortal time bias
  - Time-dependent covariates

OUTLINE

- Review of censored data, KM estimation, logrank test and Cox model basics
- Covariate adjustment in Cox model
- Stratification adjustment in Cox model
- Interaction (Effect Modification) in Cox Model
- Precision in Cox model

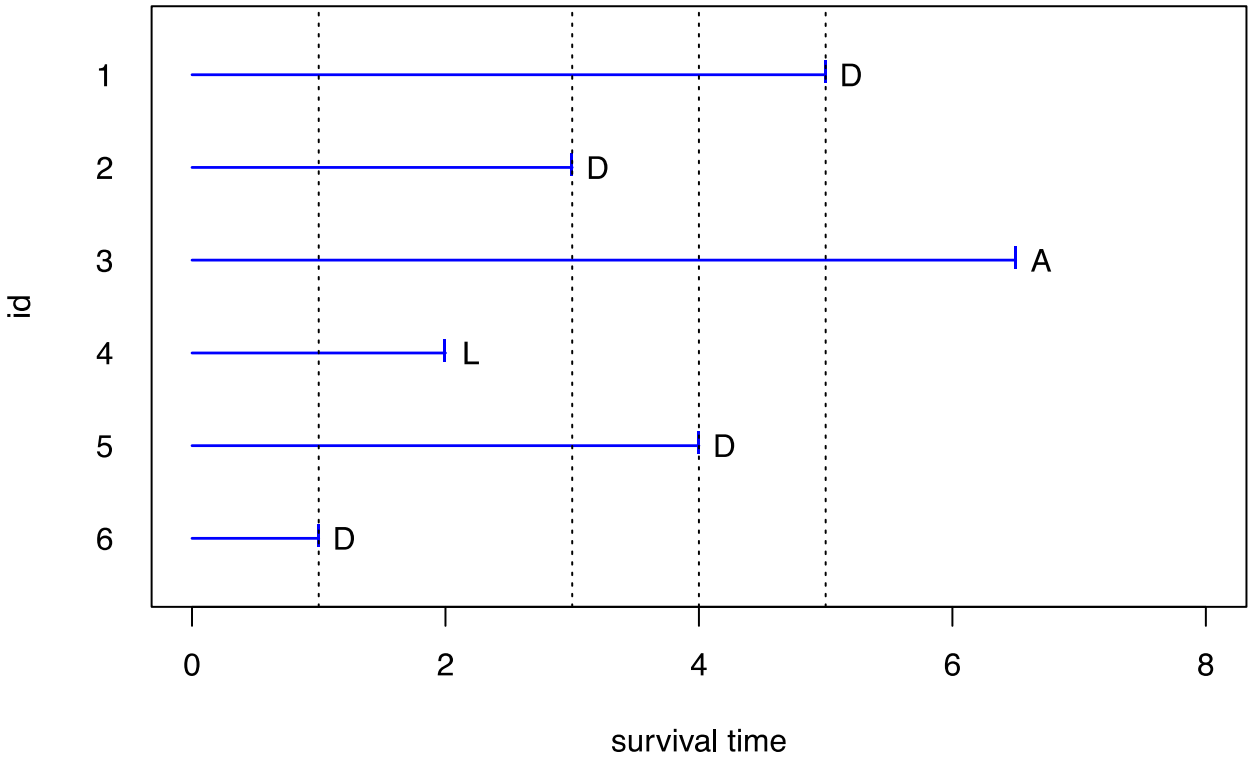
CENSORED DATA



id	Y	δ
1	5	1
2	3	1
3	6.5	0
4	2	0
5	4	1
6	1	1

"Censored" observations give some information about their survival time.

RISK SETS



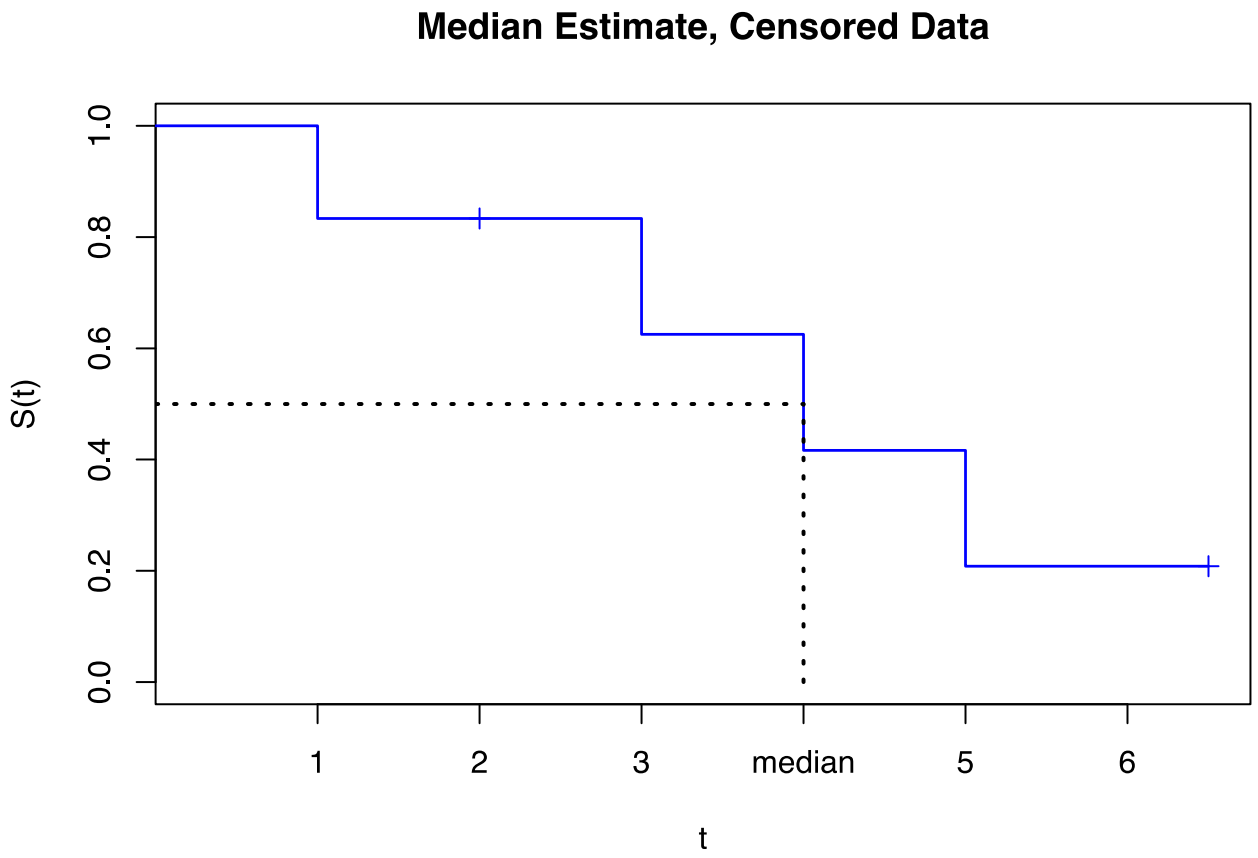
R_1 R_2 R_3 R_4

{1,2,3,4,5,6} {1,2,3,5} {1,3,5} {1,3}

CENSORED DATA ASSUMPTION

- Important assumption: subjects who are censored at time t are at the same risk of dying at t as those at risk but not censored at time t.

MEDIAN & SURVIVAL CENSORED DATA



EQUIVALENT CHARACTERIZATIONS

- Any <u>one</u> of the density function(f(t)), the survival function(S(t)) or the hazard function(λ(t)) is enough to determine the survival distribution.
- They are each functions of each other:
  - $S(t) = \int_t^\infty f(s)ds = e^{-\int_0^t \lambda(s)ds}$
  - $f(t) = -\frac{d}{dt}S(t) = \lambda(t)e^{-\int_0^t \lambda(s)ds}$
  - $\lambda(t) = \frac{f(t)}{S(t)}$

LOGRANK TEST

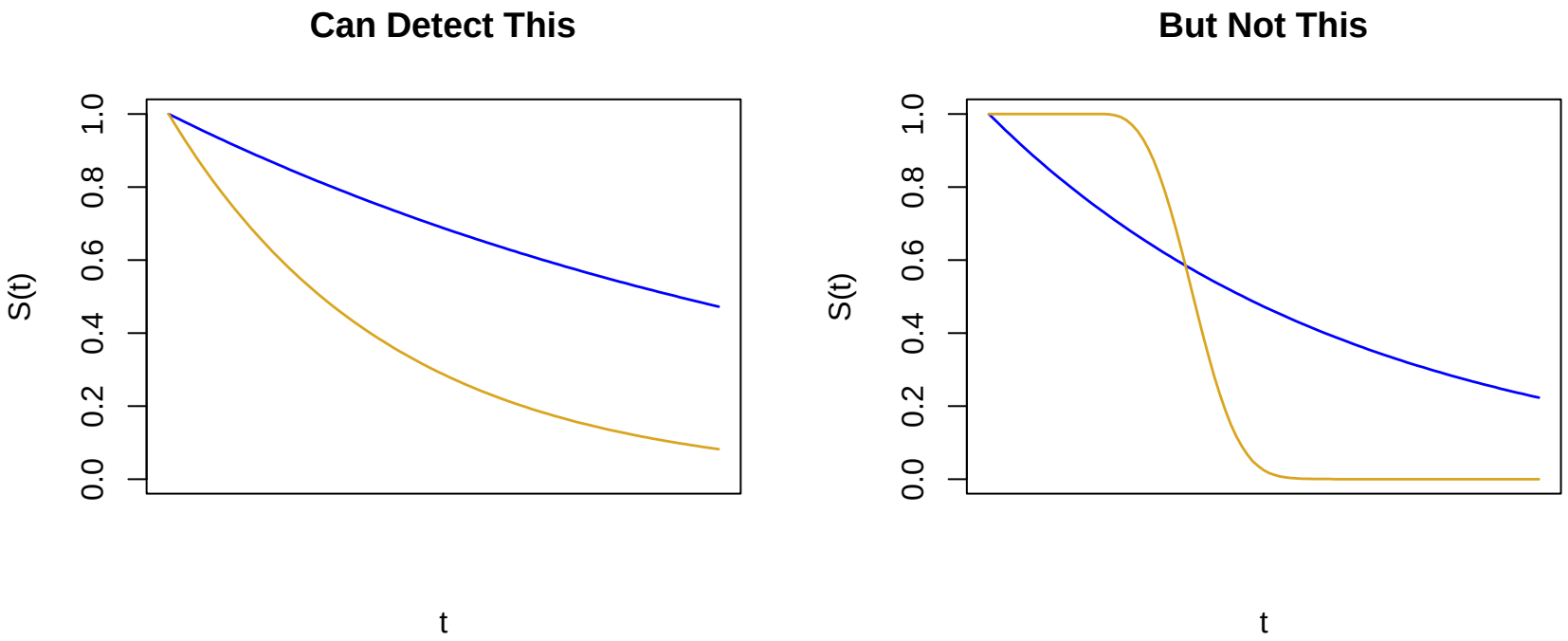
- The test is based on a 2x2 table of group by current status at each observed failure time (ie for each risk set)
- $T_{(j)}$, $j=1,\ldots m$, as shown in the Table below.

Event/Group	1	2	Total
Die	$d_{1(j)}$	$d_{2(j)}$	$D_{(j)}$
Survive	$n_{1(j)}-d_{1(j)}= s_{1(j)}$	$n_{2(j)}-d_{2(j)} = s_{2(j)}$	$N_{(j)}-D_{(j)} = S_{(j)}$
At Risk	$n_{1(j)}$	$n_{2(j)}$	$N_{(j)}$

LOGRANK TEST

- Detects consistent differences between survival curves over time.
- Best power when:
  - $H_0: S_1(t) = S_2(t)$ for all t vs $H_A: S_1(t) = [S_2(t)]^c$, or
  - $H_0: \lambda_1(t) = \lambda_2(t)$ for all t vs $H_A: \lambda_1(t) = c\,\lambda_2(t)$
- Good power whenever survival curve difference is in consistent direction

LOGRANK TEST



Other tests (generalized Wilcoxon and others) can give more weight to early or late differences.

COX REGRESSION MODEL

- Usually written in terms of the hazard function
- As a function of independent variables $x_1, x_2, \ldots x_k$,

$$\lambda(t) = \lambda_0(t) e^{\beta_1 x_1 + \cdots + \beta_k x_k}$$

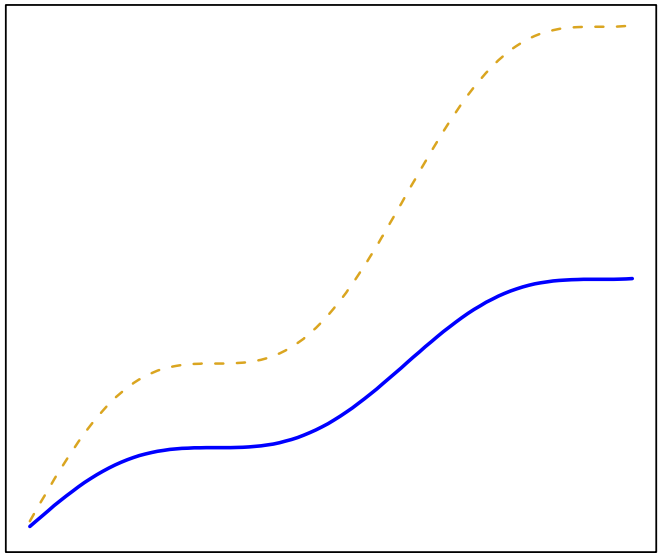
↑ relative risk / hazard ratio

$$\log \lambda(t) = \log \lambda_0(t) + \beta_1 x_1 + \cdots + \beta_k x_k$$

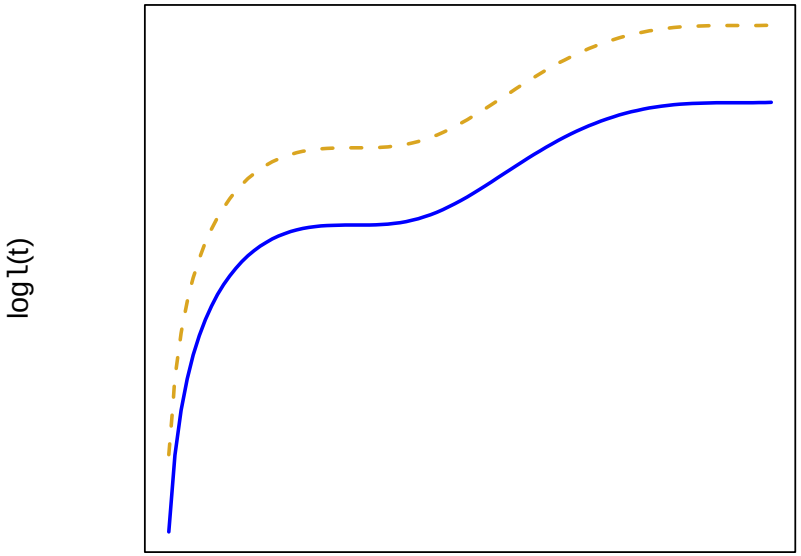
↑ intercept

EXAMPLE

Proportional Hazards



Parallel Log Hazards



RELATIONSHIP TO SURVIVAL FUNCTION

Single binary x:

$$x = \begin{cases} 1 & \text{Test treatment} \\ 0 & \text{Standard treatment} \end{cases}$$

$$\lambda(t) = \lambda_0(t)e^{\beta x} \implies S(t) = [S_0(t)]^{e^{\beta x}}$$

In terms of $S_0(t)$:

$S(t)$ for $x = 1$: $\quad [S_0(t)]^{e^{\beta \cdot 1}} = [S_0(t)]^{e^{\beta}}$

$S(t)$ for $x = 0$: $\quad [S_0(t)]^{e^{\beta \cdot 0}} = [S_0(t)]^{1} = S_0(t)$

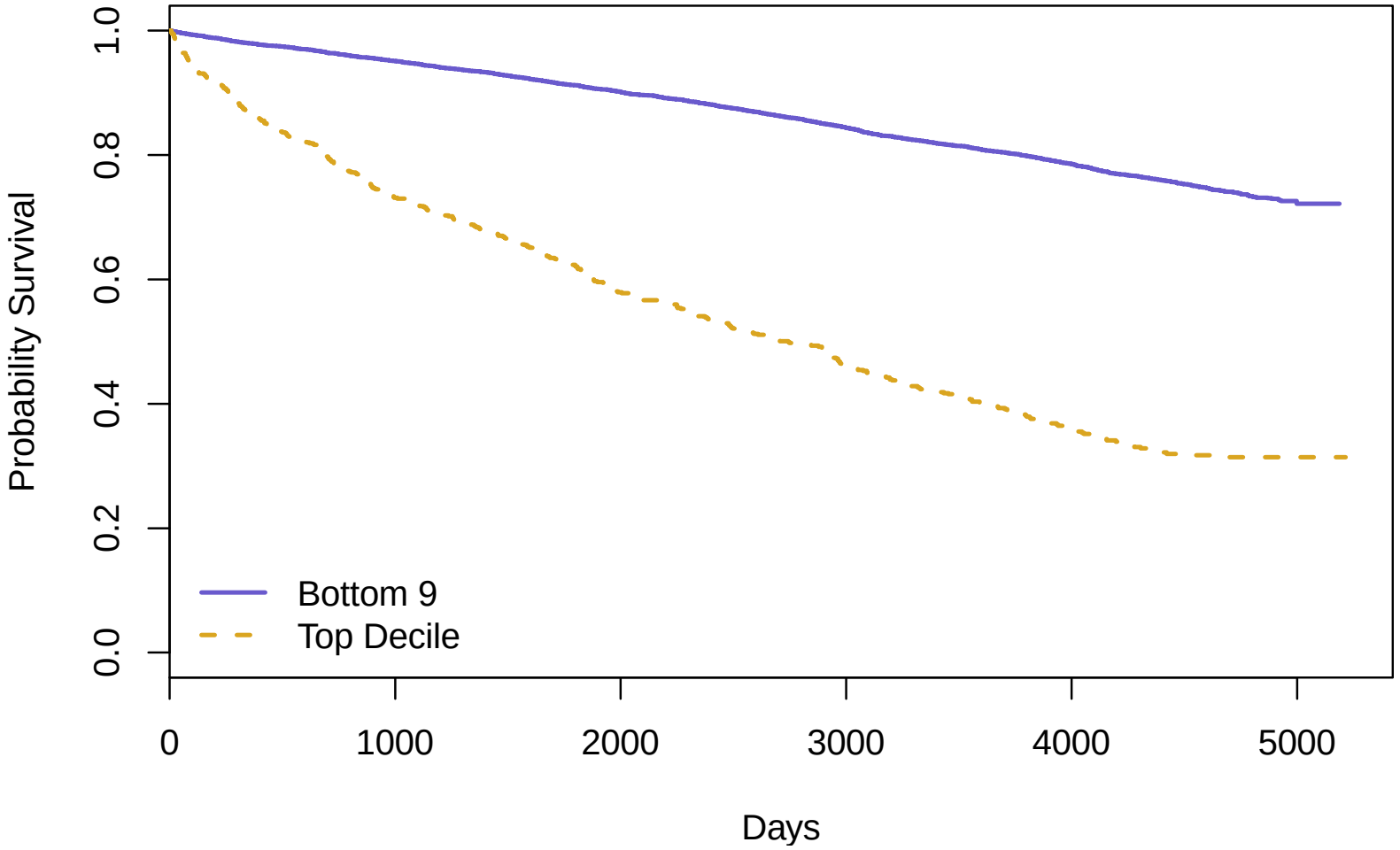
CONFOUNDING

- Observational data: sometimes observed associations between an explanatory variable and outcome can be due to their joint association with another variable.
  - Age related to both sex and risk of death.
  - Age related to immunoglobulin levels and risk of death (example next)

SURVIVAL AND IG

- Random subset of the data from A. Dispenzieri, J. Katzmann, R. Kyle, D. Larson, T. Therneau, C. Colby, R. Clark, G. Mead, S. Kumar, L.J. Melton III, and S.V. Rajkumar. Use of monoclonal serum immunoglobulin (ig) free light chains (flc) to predict overall survival in the general population. Mayo Clinic Proc, 87:512–523, 2012.
- Are high free-chain ig levels associated with survival?
  - Population-based Olmstead County example
  - Men and women 50+ years of age

TOP DECILE FLC



COX REGRESSION

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
topdecileTRUE	1.452639	4.274378	0.0523126	27.7684	0

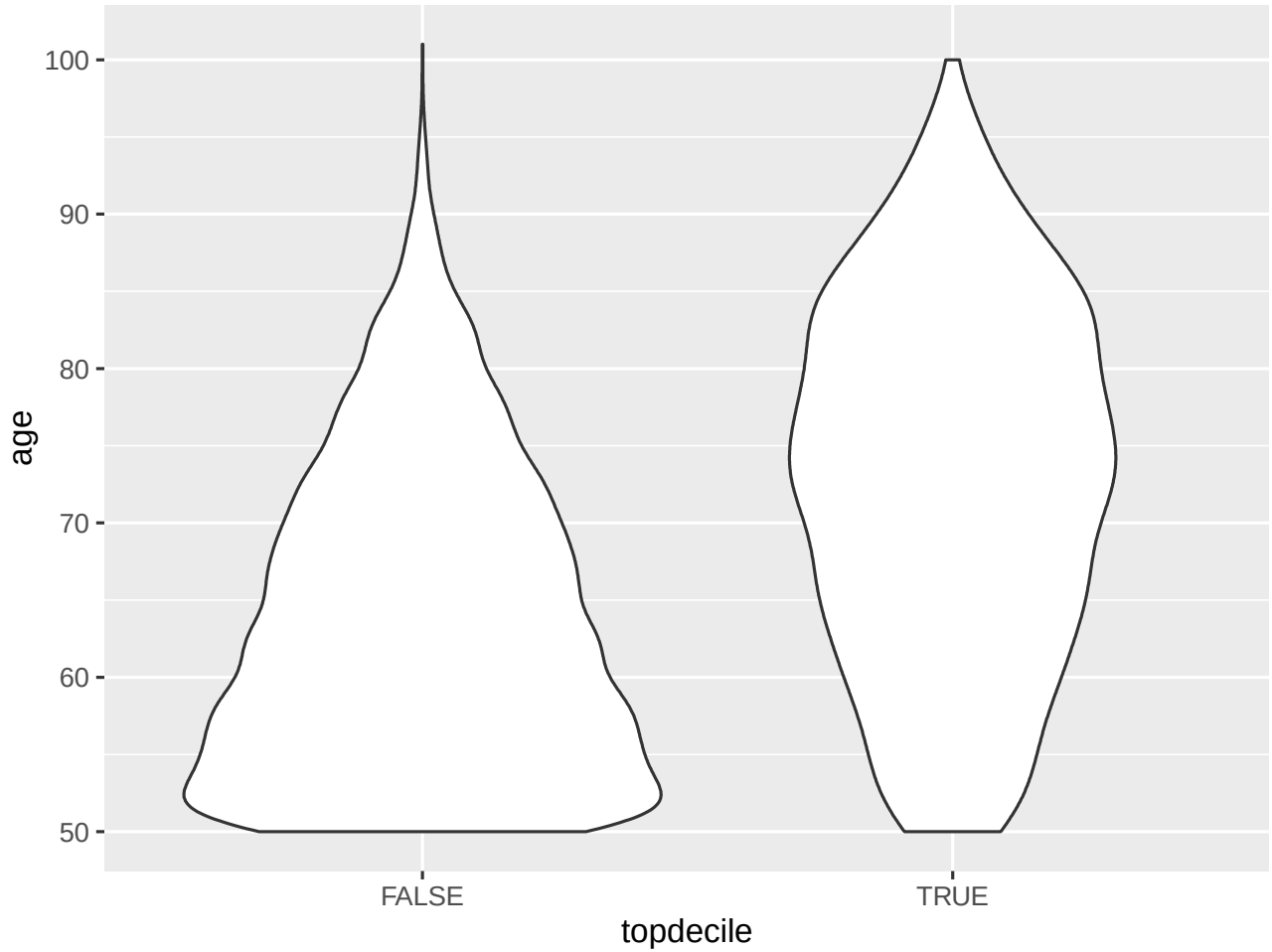
	2.5 %	97.5 %
topdecileTRUE	3.857841	4.735889

ADJUSTED COX REGRESSION

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
topdecileTRUE	0.8012613	2.228350	0.0543721	14.73663	0
age	0.1018649	1.107234	0.0022780	44.71700	0

	2.5 %	97.5 %
topdecileTRUE	2.003096	2.478934
age	1.102301	1.112189

WHY?



ADJUSTMENT MODEL

One binary variable, x_1, with continuous adjustment variable x_2:

$$x_1 = \begin{cases} 1 & \text{Top decile FLC} \\ 0 & \text{Otherwise} \end{cases}$$

$x_2 =$ Age in years

$$\lambda(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2}$$

Interpretation of e^{β_1}:

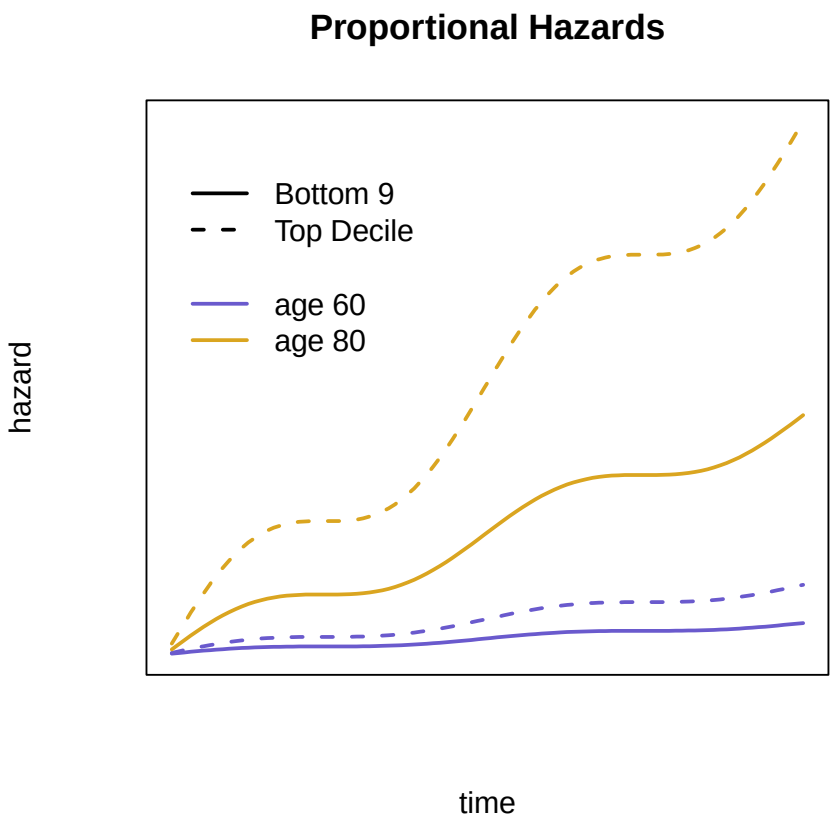
"Relative risk (or hazard ratio) comparing top decile FLC to the rest, among those of the same age".

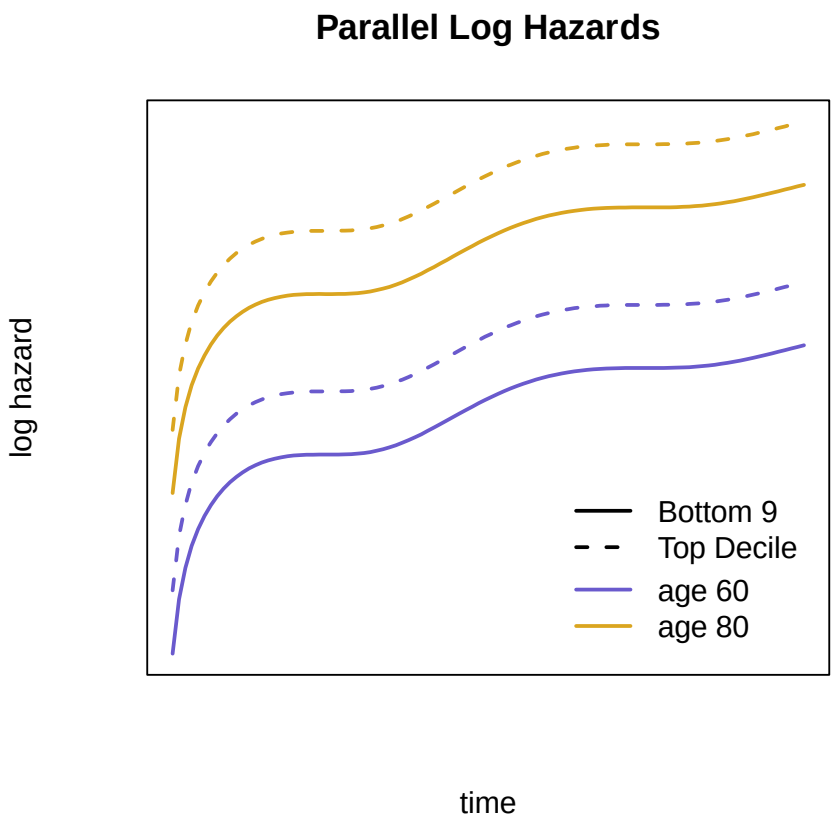
$\lambda(t)$ for $x_1 = 1$ and x_2: $\lambda_0(t)e^{\beta_1 \cdot 1 + \beta_2 x_2}$

$\lambda(t)$ for $x_1 = 0$ and x_2: $\lambda_0(t)e^{\beta_1 \cdot 0 + \beta_2 x_2}$

ratio: $e^{\beta_1(1-0)+\beta_2(x_2-x_2)} = e^{\beta_1}$

ADJUSTMENT


Proportional Hazards
hazard
Bottom 9
Top Decile
age 60
age 80
time


Parallel Log Hazards
log hazard
Bottom 9
Top Decile
age 60
age 80
time

RESULTS

- “We found strong evidence that adjusted for age, free light chain(FLC) values in the top decile were associated with the risk of death ($P < .0001$). Among individuals of the same age, we estimate that having an FLC value in the top decile is associated with 2.23 times the hazard of death (95% CI: 2.00, 2.48).”

STRATIFICATION ADJUSTMENT

One binary variable, x_1, with grouped adjustment variable x_2:

$$x_1 = \begin{cases} 1 & \text{Top decile FLC} \\ 0 & \text{Otherwise} \end{cases}$$

$$x_2 = \begin{cases} 0 & \text{age 50-59} \\ 1 & \text{age 60-69} \\ 2 & \text{age 70-79} \\ 3 & \text{age 80-89} \\ 4 & \text{age 90+} \end{cases}$$

$$\lambda(t) = \lambda_{0x_2}(t)e^{\beta_1 x_1}$$

STRATIFICATION ADJUSTMENT

$$\lambda(t) = \lambda_{0x_2}(t)e^{\beta_1 x_1}$$

Interpretation of e^{β_1}:

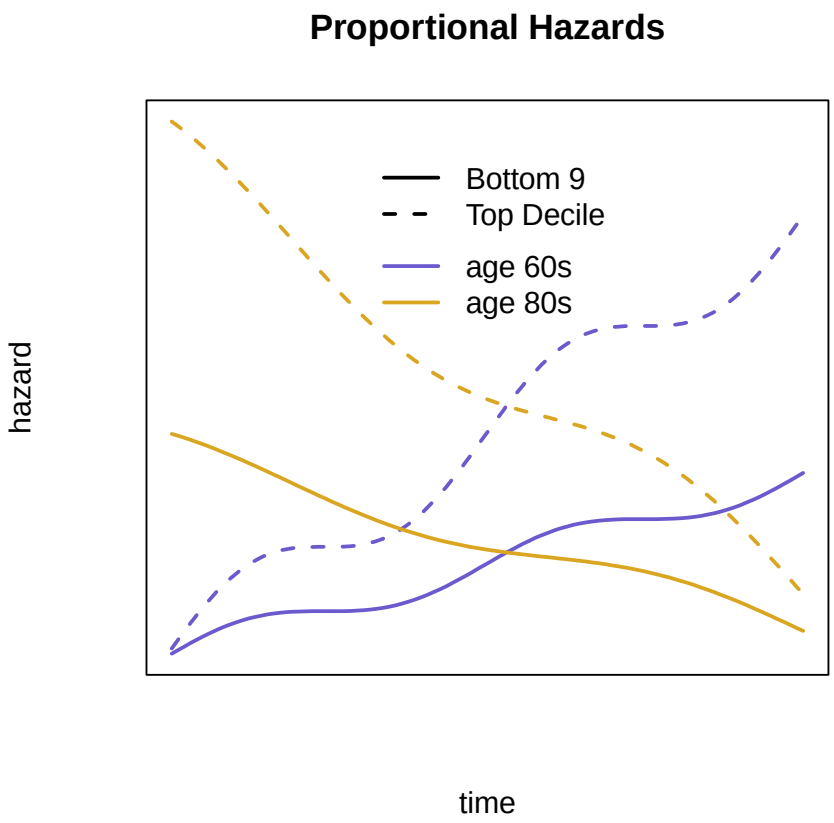
"Relative risk (or hazard ratio) comparing top decile FLC to the rest, among those in the same age group".

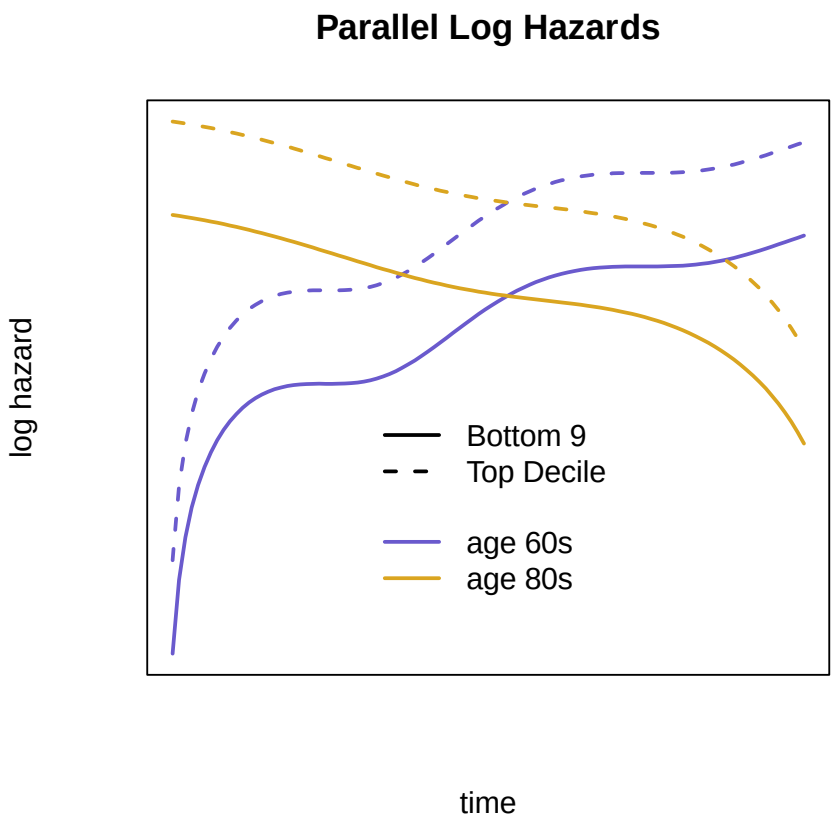
$\lambda(t)$ for $x_1 = 1$ and x_2: $\lambda_{0x_2}(t)e^{\beta_1 \cdot 1}$

$\lambda(t)$ for $x_1 = 0$ and x_2: $\lambda_{0x_2}(t)e^{\beta_1 \cdot 0}$

ratio: $\frac{\lambda_{0x_2}(t)}{\lambda_{0x_2}(t)}e^{\beta_1(1-0)} = e^{\beta_1}$

STRATIFICATION ADJUSTMENT





INTERACTION

One binary variable with continuous linear interaction, x_1 and x_2

$$x_1 = \begin{cases} 1 & \text{Top Decile FLC} \\ 0 & \text{Otherwise} \end{cases}$$

$x_2 =$ Age in years

$$\lambda(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2}$$

Interpretation of e^{β_1}:

"Relative risk (or hazard ratio) comparing top decile FLC to the rest among those with age (= x_2) = zero".

Interpretation of $e^{\beta_1 + x_2\beta_3}$:

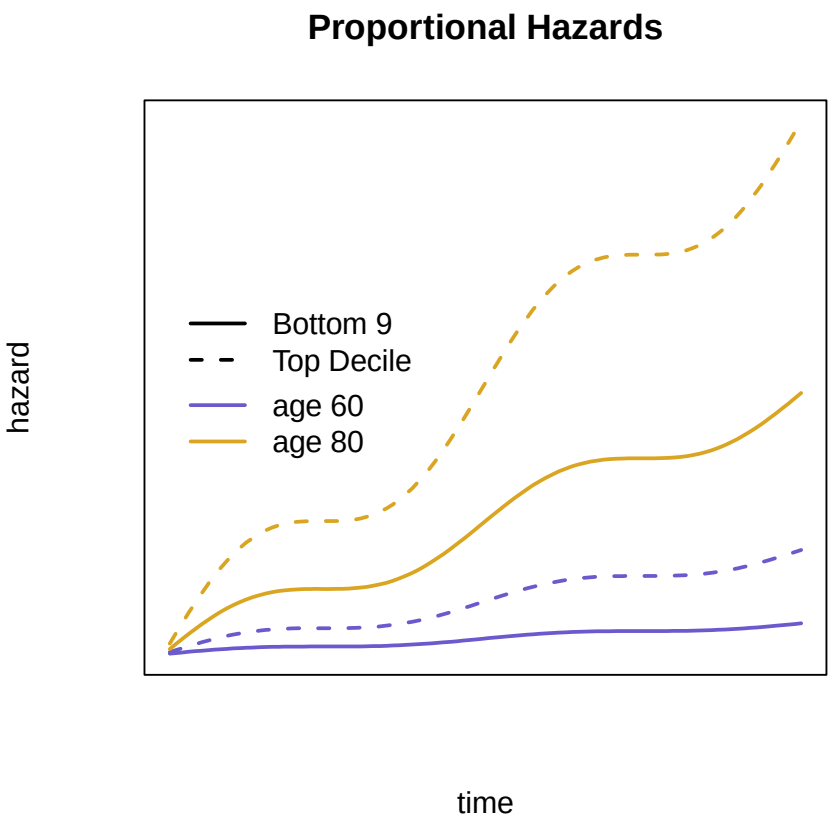
"Relative risk (or hazard ratio) comparing top decile FLC to the rest among those with age = x_2".

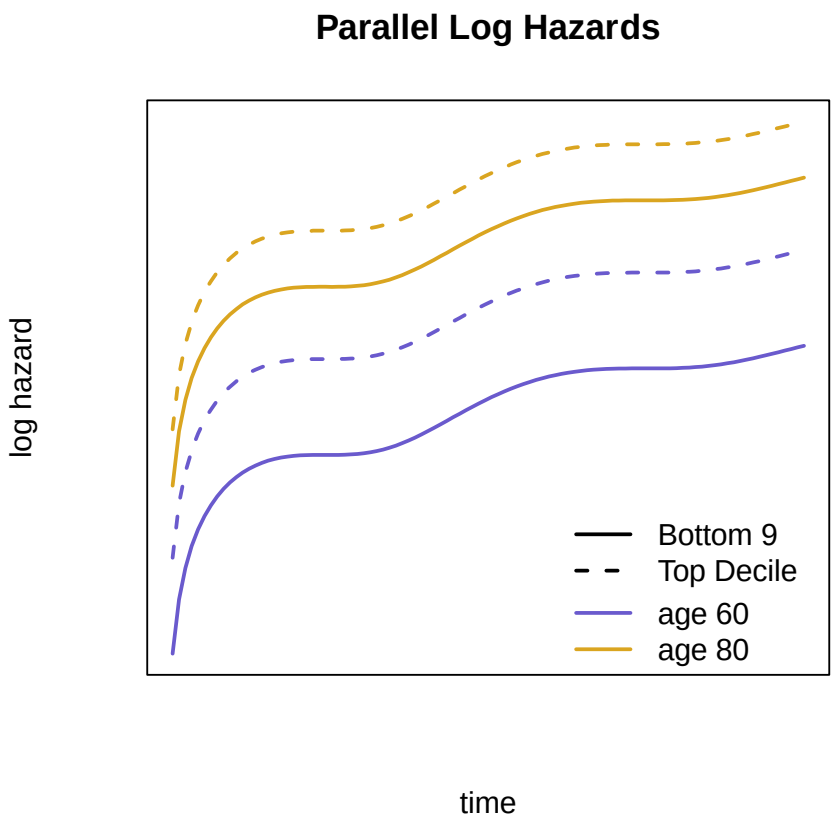
$\lambda(t)$ for $x_1 = 1$ and $x_2 = 0$: $\lambda_0(t)e^{\beta_1 \cdot 1}$ $\lambda(t)$ for $x_1 = 1$ and $x_2 \neq 0$: $\lambda_0(t)e^{\beta_1 \cdot 1 + \beta_2 \cdot x_2 + \beta_3 \cdot 1 \cdot x_2}$

$\lambda(t)$ for $x_1 = 0$ and $x_2 = 0$: $\lambda_0(t)e^{\beta_1 \cdot 0}$ $\lambda(t)$ for $x_1 = 0$ and $x_2 \neq 0$: $\lambda_0(t)e^{\beta_1 \cdot 0 + \beta_2 \cdot x_2 + \beta_3 \cdot 0}$

ratio: $e^{\beta_1(1-0)} = e^{\beta_1}$ ratio: $e^{\beta_1(1-0)+\beta_3(x_2-0)} = e^{\beta_1 + x_2\beta_3}$

INTERACTION





INTERACTION

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
topdecileTRUE	2.7312322	15.3517922	0.4154009	6.574930	0.0e+00
age	0.1067648	1.1126726	0.0025185	42.392311	0.0e+00
topdecileTRUE:age	-0.0252304	0.9750852	0.0054342	-4.642936	3.4e-06

	2.5 %	97.5 %
topdecileTRUE	6.8009436	34.6536508
age	1.1071938	1.1181785
topdecileTRUE:age	0.9647549	0.9855261

TOP DECILE HR BY AGE

age	exp(coef)	z	Pr(>|z|)	2.5 %	97.5 %
50	3.897886	8.499784	0.00e+00	2.848328	5.334189
60	3.077554	10.309487	0.00e+00	2.485373	3.810831
70	2.429865	13.162515	0.00e+00	2.128957	2.773302
80	1.918486	10.861243	0.00e+00	1.705679	2.157843
90	1.514729	4.368336	1.25e-05	1.257254	1.824932

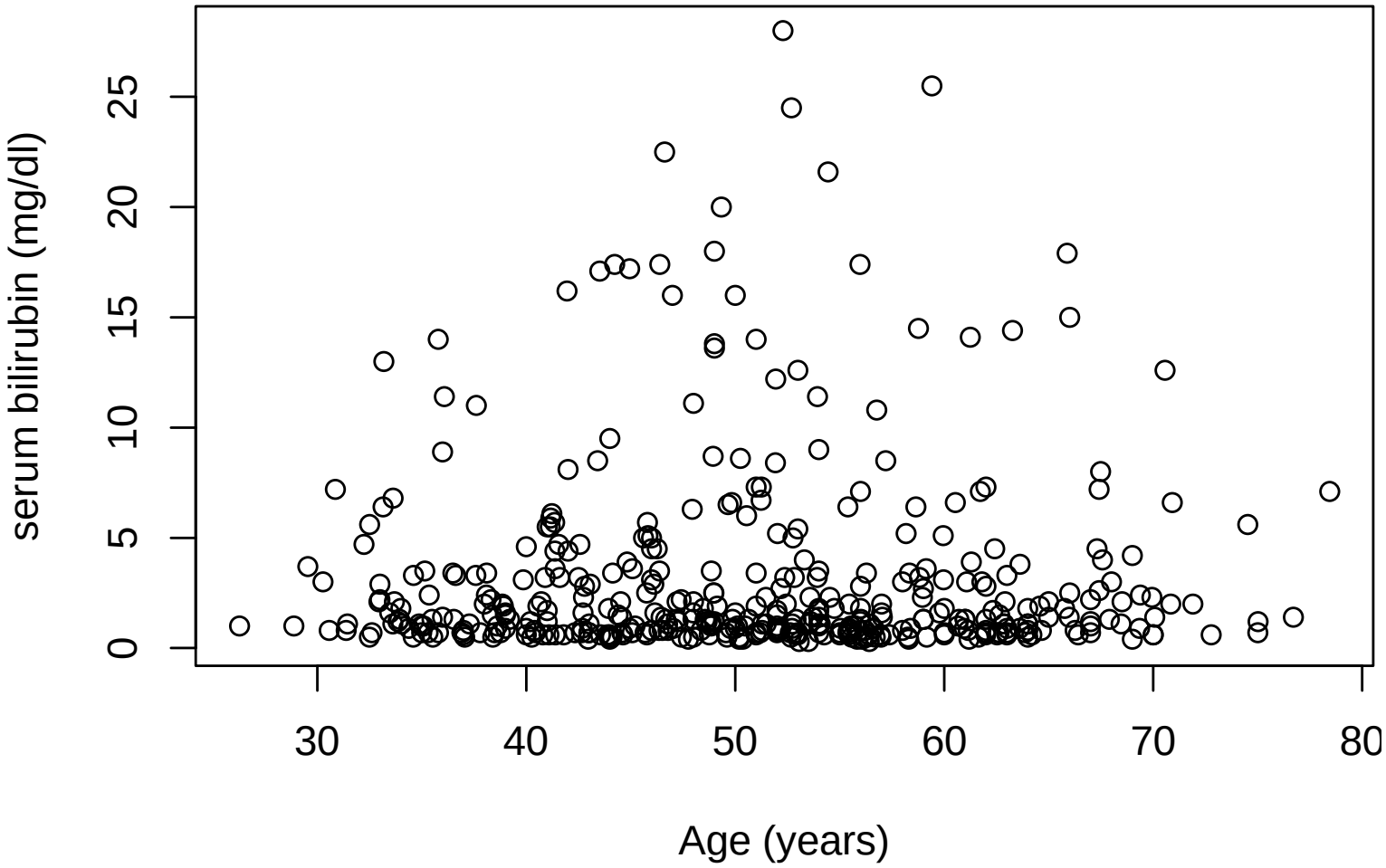
ADJUSTMENT AND PRECISION

- In Cox regression, addition of variables to a model that are associated only with the outcome can improve power.
- There is little effect on the coefficient estimate for other variables (eg treatment) or their standard errors, except when the association between outcome and the added variable is very strong.
- When there is an effect of adding a predictive variable, this is what happens to inference for the treatment variable or other variable of interest:
  - The standard error of its coefficient increases
  - The estimate of the coefficient moves farther from zero
  - The test of whether the coefficient is zero has more power.

PRIMARY BILIARY CIRRHOSIS

- Clinical trial with virtually no treatment effect
- Conducted before widespread use of immune suppressive therapies
- Good data for examining prognostic factors in PBC
- Some patients received liver transplant—treated as censored here
- Serum bilirubin associated with survival
- Treating age as a “precision variable”

AGE-BILIRUBIN ASSOCIATION



PRECISION

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
bili	0.1418533	1.152408	0.0115685	12.26201	0

	2.5 %	97.5 %
bili	1.126572	1.178836

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
bili	0.1436238	1.154450	0.0114189	12.577714	0e+00
age	0.0431303	1.044074	0.0080554	5.354198	1e-07

	2.5 %	97.5 %
bili	1.128899	1.180578
age	1.027719	1.060689

TO WATCH OUT FOR:

- Coefficients in Cox regression are positively associated with risk, not survival.
  - Positive β means large values of x are associated with shorter survival.
- Without certain types of time-dependent covariates (more later), Cox regression does not depend on the actual times, just their order.
  - Can add a constant to all times to remove zeros (which are removed by some software) without changing inference
- For LRT, nested models must be compared based on same subjects.
  - If some values of variables in larger model are missing, these subjects must be removed from fit of smaller model.
- Coefficient interpretation depends on what other variables are in the model and how they are coded (ie. interaction terms, 0/1 vs 1/-1 etc.)