

Session 8 - Survival Analysis III

Levi Waldron

Welcome and outline - session 8

Learning Objectives:

- ▶ Check model assumptions and fit of the Cox model
 - ▶ residuals analysis
 - ▶ log-minus-log plot
- ▶ Fit and interpret multivariate Cox models
 - ▶ perform tests for trend
 - ▶ predict survival for specific covariate patterns
 - ▶ predict survival for adjusted coefficients
- ▶ Explain and perform stratified analysis and its use
- ▶ Explain time-dependent covariates and when to use them
- ▶ Vittinghoff sections 6.2-6.4

Cox proportional hazards model

- ▶ Cox proportional hazard regression assesses the relationship between a right-censored, time-to-event outcome and multiple predictors:
 - ▶ categorical variables (e.g., treatment groups)
 - ▶ continuous variables

$$\log(HR(x_i)) = \log \frac{h(t|x_i)}{h_0(t)} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

- ▶ $HR(x_i)$ is the hazard of patient i relative to baseline
- ▶ $h(t|x_i)$ is the time-dependent hazard function $h(t)$ for patient i
- ▶ $h_0(t)$ is the *baseline hazard function*, and is the negative of the slope of the $S_0(t)$, the baseline *survival* function.
- ▶ Multiplicative model

Caveats and Assumptions

- ▶ Categories with no events
 - ▶ can occur when the group is small or its risk is low
 - ▶ HRs with respect to such a reference group are infinite
 - ▶ hypothesis tests and CIs are difficult / impossible to interpret
- ▶ Assumptions of Cox PH model
 - ▶ Constant hazard ratio over time (proportional hazards)
 - ▶ Linear association between $\log(\text{HR})$ and predictors (log-linearity)
/ multiplicative relationship between hazard and predictors
 - ▶ Independence of survival times between individuals in the sample

Residuals analysis

- ▶ Residuals are used to investigate the lack of fit of a model to a given subject.
- ▶ For Cox regression, there's no easy analog to the usual “observed minus predicted” residual

```
library(pensim); set.seed(1)
mydat <- create.data(nvars=c(1, 1), nsamples=500,
  cors=c(0, 0), associations=c(0.5, 0.5),
  firstonly=c(TRUE, TRUE), censoring=c(0, 8.5))$data
```

```
## Rename variables of simulated data, and make one variable categorical
colnames(mydat)[1:2] <- c("Var1", "Var2")
mydat$Var1 <- cut(mydat$Var1, breaks=2, labels=c("low", "high"))
mydat$time <- ceiling(mydat$time*1000)
```

Simulated data to test residuals methods

```
summary(mydat)
```

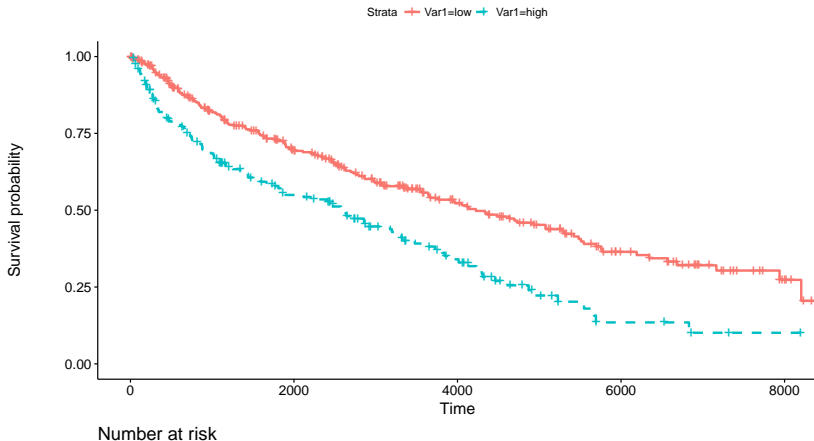
##	Var1	Var2	time	cens
##	low :323	Min. :-2.99695	Min. : 5	Min. :0.000
##	high:177	1st Qu.: -0.79008	1st Qu.: 691	1st Qu.:0.000
##		Median : -0.02126	Median :1970	Median :1.000
##		Mean :-0.04594	Mean :2529	Mean :0.526
##		3rd Qu.: 0.68933	3rd Qu.:3874	3rd Qu.:1.000
##		Max. : 3.05574	Max. :8481	Max. :1.000

Kaplan-Meier plot of simulated data, stratified by Var1

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

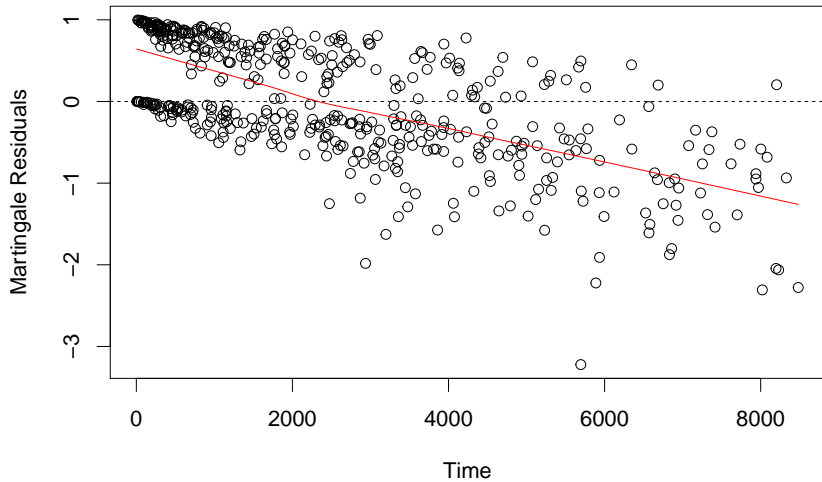
```
## Loading required package: magrittr
```



Martingale residuals

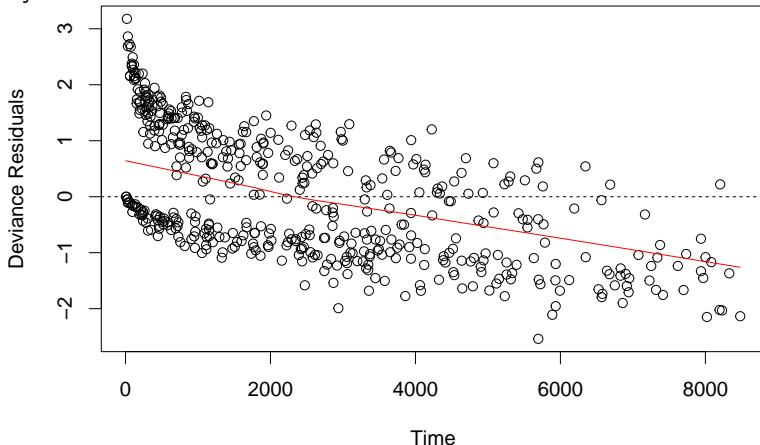
- ▶ censoring variable c_i (1 if event, 0 if censored) minus the estimated cumulative hazard function $H(t_i, X_i, \beta_i)$ (1 - survival function)
 - ▶ E.g., for a subject censored at 1 year ($c_i = 0$), whose predicted cumulative hazard at 1 year was 30%, Martingale = $0 - 0.30 = -0.30$.
 - ▶ E.g. for a subject who had an event at 6 months, and whose predicted cumulative hazard at 6 months was 80%, Martingale = $1 - 0.8 = 0.2$.
- ▶ Problem: not symmetrically distributed, even when model fits the data well

Martingale residuals in simulated data



Deviance residuals in simulated data

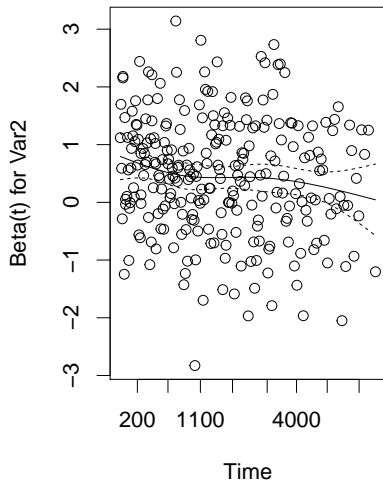
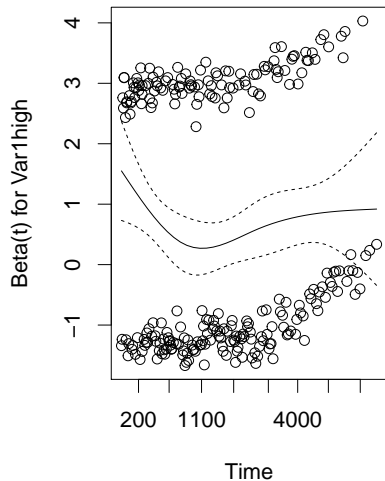
- ▶ Deviance residuals are scaled Martingale residuals
- ▶ Should be more symmetrically distributed about zero?
- ▶ Observations with large deviance residuals are poorly predicted by the model



Schoenfeld residuals

- ▶ technical definition: contribution of a covariate at each event time to the partial derivative of the log-likelihood
- ▶ intuitive interpretation: the observed minus the expected values of the covariates at each event time.
- ▶ a random (unsystematic) pattern across event times gives evidence the covariate effect is not changing with respect to time
- ▶ If it is systematic, it suggests that as time passes, the covariate effect is changing.

Schoenfeld residuals for simulated data

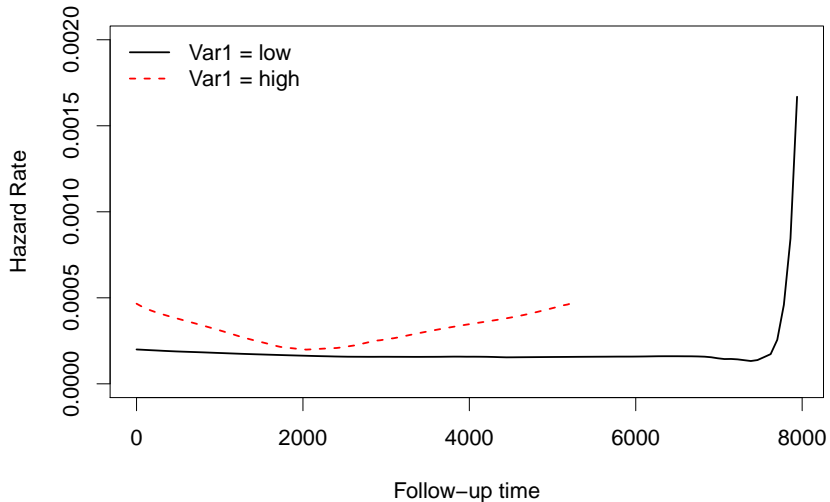


Schoenfeld test for proportional hazards

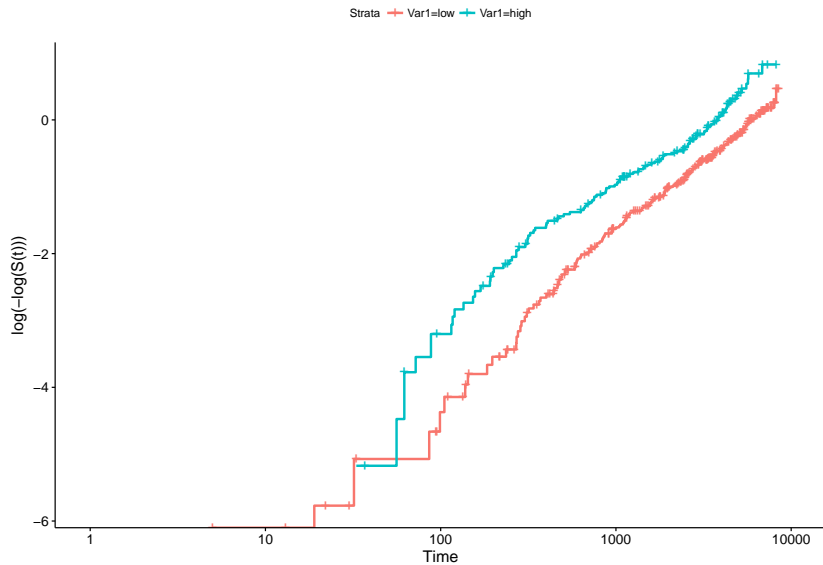
- ▶ Tests correlation between scaled Schoenfeld residuals and time
- ▶ Equivalent to fitting a simple linear regression model with time as the predictor and residuals as the outcome
- ▶ Parametric analog of smoothing the residuals against time using LOWESS
- ▶ If the hazard ratio is constant, correlation should be zero.
 - ▶ Positive values of the correlation suggest that the log-hazard ratio increases with time.

##		rho	chisq	p
##	Var1high	-0.0185	0.0903	0.7638
##	Var2	-0.1315	4.6360	0.0313
##	GLOBAL	NA	4.6438	0.0981

The hazard function $h(t)$, stratified by Var1



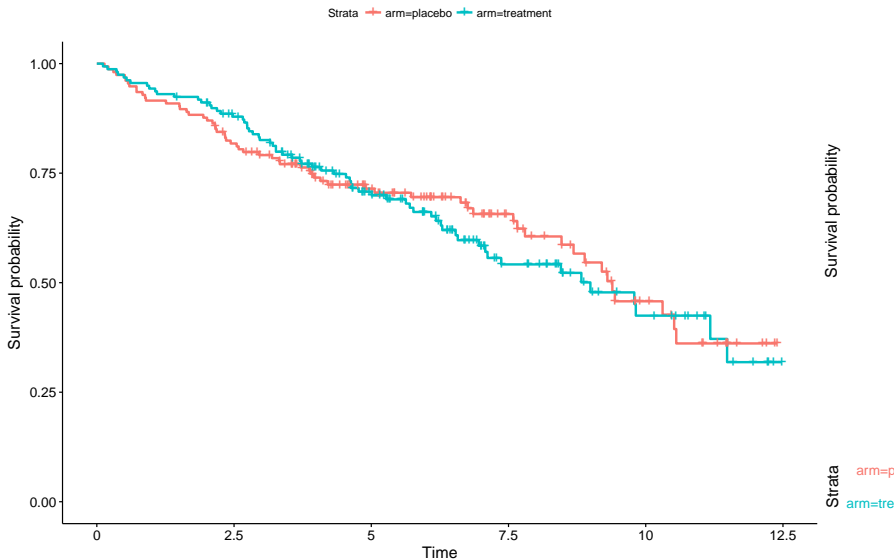
Log-minus-log plot



Motivating Example: Primary Biliary Cirrhosis (PBC)

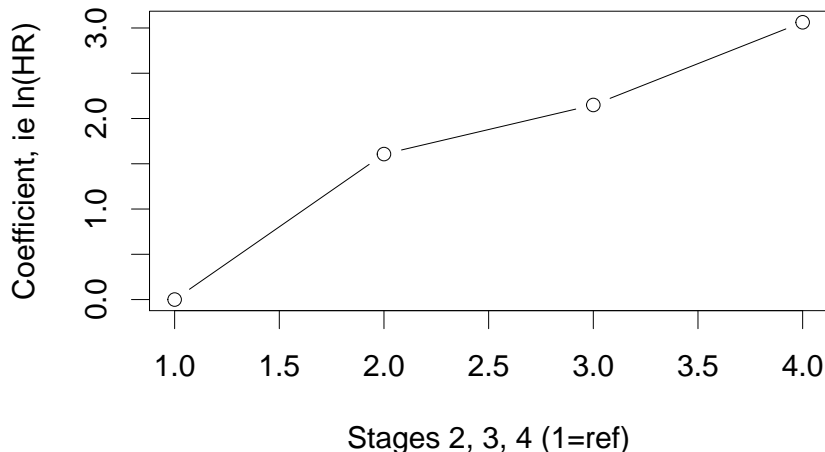
- ▶ Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984, $n=424$ patients.
- ▶ randomized placebo controlled trial of the drug D-penicillamine.
 - ▶ 312 cases from RCT, plus additional 112 not from RCT.
- ▶ Primary outcome is (censored) time to death

Kaplan-Meier plot of treatment and placebo arms



Tests for trend

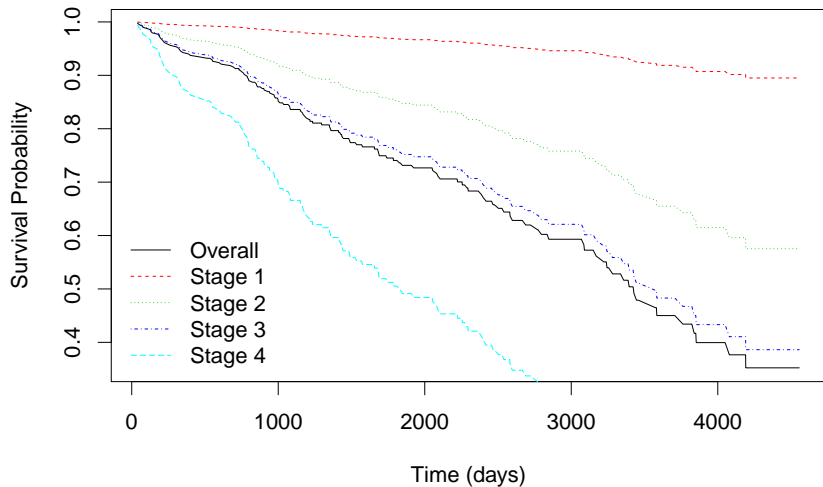
- ▶ For ordinal variables like stage (1, 2, 3, 4)
 - ▶ This is a test for linear / quadratic / cubic relationship between coefficients and their index
 - ▶ Model selection by LRT or Wald Test



Predicted survival for specific covariate patterns

- ▶ The Cox model is a *relative* risk model
 - ▶ only predicts relative risks between pairs of subjects
- ▶ Key is to calculate the overall $S(t)$, then multiply it by the relative hazard for the specific covariate pattern.
- ▶ In this example we plot the baseline survival for all stages together, then for stages 1-4 separately.

Predicted survival for specific covariate patterns



Multivariate regression

- ▶ Same coding and objectives as for `lm()` and `glm()`
 - ▶ controlling for confounding
 - ▶ testing for mediation
 - ▶ testing for interaction

Multivariate regression

```
fit <- coxph(Surv(time, os) ~ age + sex + edema
              + stage + arm, data=pbcc.os)
summary(fit)
```

```
## Call:
## coxph(formula = Surv(time, os) ~ age + sex + edema + stage +
##       arm, data = pbcc.os)
##
##      n= 312, number of events= 125
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## age           0.027618   1.028003   0.009362  2.950  0.00318 **
## sexm           0.317540   1.373744   0.248839  1.276  0.20193
## edema0.5       0.538715   1.713804   0.275287  1.957  0.05036 .
## edema1         2.080422   8.007845   0.276959  7.512 5.84e-14 ***
## stage2         1.535263   4.642546   1.034854  1.484  0.13793
## stage3         1.998217   7.375893   1.016097  1.967  0.04923 *
## stage4         2.666263  14.386101   1.016234  2.624  0.00870 **
## armtreatment   0.057946   1.059658   0.189200  0.306  0.75940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Predicted survival for adjusted coefficients

- ▶ Can create Kaplan-Meier curves for crude or unadjusted coefficients
 - ▶ Section 6.3.2.3 in Vittinghoff
- ▶ Idea is to estimate hazard ratio in an unadjusted model:

```
unadjfit <- coxph(Surv(time, os) ~ stage, data=dbc.os)  
coef(unadjfit)
```

```
##      stage2      stage3      stage4  
## 1.607014 2.149500 3.062775
```

Predicted survival for adjusted coefficients

- and in an adjusted model:

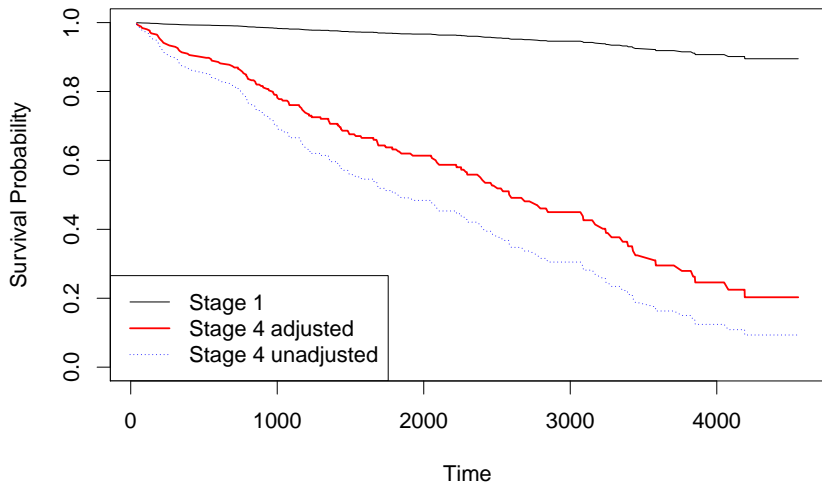
```
adjfit <- coxph(Surv(time, os) ~ age + sex + edema  
               + stage + arm, data=pbcc.os)  
coef(adjfit)
```

##	age	sexm	edema0.5	edema1	stage2
##	0.0276179	0.3175396	0.5387152	2.0804217	1.5352629
##	stage3	stage4	armtreatment		
##	1.9982170	2.6662626	0.0579460		

Predicted survival for adjusted coefficients (cont'd)

- The survival function will be calculated for a “baseline” group, say stage 1, then exponentiated with the adjusted coefficient, e.g.:

$$[S_{stage=1}(t)]^{\exp(\beta_{stage=4})}$$



Stratification

- ▶ Vittinghoff 6.3.2
- ▶ Separates the analysis into strata
 - ▶ must have an adequate number of events in each stratum (at least 5 to 7)
 - ▶ can be used to adjust for variables with strong impact on survival
 - ▶ can help solve proportional hazards violations
- ▶ Strata have different baseline hazards
- ▶ Coefficients / Hazard Ratios are calculated within stratum then combined.

Stratification

Example - in R, strata() can be added to any model formula

```
mycox <- coxph(Surv(time, os) ~ trt + strata(stage), data=pbpc.os)  
summary(mycoc)
```

```
## Call:  
## coxph(formula = Surv(time, os) ~ trt + strata(stage), data = pbpc.os)  
##  
##    n= 312, number of events= 125  
##  
##           coef exp(coef) se(coef)      z Pr(>|z|)  
## trt -0.1063      0.8992   0.1814 -0.586   0.558  
##  
##      exp(coef) exp(-coef) lower .95 upper .95  
## trt    0.8992      1.112   0.6302    1.283  
##  
## Concordance= 0.494  (se = 0.041 )  
## Rsquare= 0.001    (max possible= 0.958 )  
## Likelihood ratio test= 0.34  on 1 df,   p=0.5576  
## Wald test            = 0.34  on 1 df,   p=0.5578  
## Score (logrank) test = 0.34  on 1 df,   p=0.5576
```

Time-dependent covariates

Definition: A time-dependent covariate in a Cox model is a predictor whose values may vary with time.

- ▶ For example, Aurora et al. (1999) followed 124 patients to study the effect of lung transplantation on survival in children with cystic fibrosis. * The natural time origin in this study is the time of listing for transplantation, not transplantation itself, because the children are most comparable at that point.
 - ▶ However, waiting times for a suitable transplant can be long, and there is considerable mortality among children on the waiting list.

Time-dependent covariates

- ▶ Lung transplantation *must* be treated as a TDC.
 - ▶ To see this, consider the alternative in which transplantation is modeled as a fixed binary covariate
 - ▶ this compares mortality risk in the group of children who undergo transplantation during the study to risk among those who do not.
- ▶ *Immortal time bias* can then make transplantation look more protective than it really is