Survival Analysis III

John Kornak
April 14, 2015
john.kornak@ucsf.edu

Reading VGSM 6.3 - 6.5

- Homework #*1* due Today in class
- Homework Q/A after class today *12-1*
- Lab 3 on Thursday 10.30 – MH-1400
- Homework #2 due next Tuesday (4/21) in class

So far…

- Survival data and censoring
- Reviewed Kaplan-Meier survival curves and Logrank test
- Hazard function and hazard ratio (HR)
- Proportional hazards model
- Cox Model (no baseline hazard model)
- Binary, categorical and continuous predictors
- Wald and likelihood ratio tests
- Zero/infinite HR
- Confounding, mediation, adjusting for other variables
- Interactions and `lincom` statements (danger of extrapolation)

In this lecture
(extensions to the Cox model)

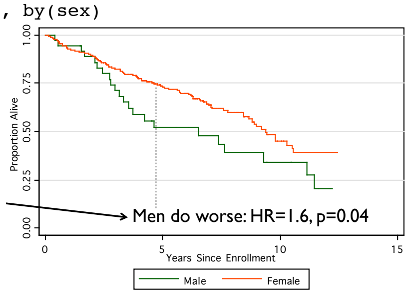
- Adjusted survival curves (for Cox model)
- Time-dependent covariates
- Diagnostics (model checking) - proportional hazards?
- Non-proportional Hazards: Stratification
- Non-proportional Hazards: generate time-dependent covariates trick
- Diagnostics (model checking) – linearity of predictors?
- Other methods: Clustered data, Left-Truncation, Interval-censoring

3

Evaluating Cox model fits: Adjusted Survival Curves

Effect of Sex: PBC data
(unadjusted comparison)

```
use pbc.dat
stset years,failure(status)
sts graph, by(sex)
```

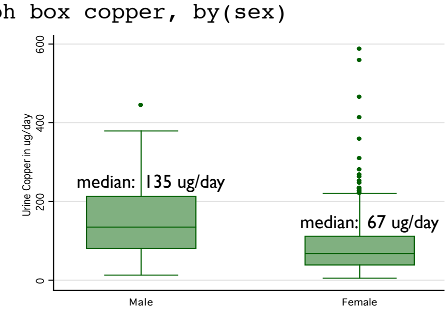


`stcox sex`
(Cox model fit)

Men do worse: HR=1.6, p=0.04

Men: Higher Copper

```
graph box copper, by(sex)
```



Adjusted Survival Curves (for Cox model)

- Would like to visualize the adjusted effects of variables (e.g., controlling for copper)
- Can make survival prediction based on a Cox model
- $S(t|x)$: survivor function (event-free proportion at time t) for someone with predictors x

Under the Cox Model

$$S(t|x) = S_0(t)^{\exp(\beta_1 x_1 + \ldots + \beta_p x_p)}$$

β's are the coefficients from the Cox model

$S_0(t)$:= baseline survivor function

= survivor function when all predictors equal zero

In Cox model we see estimates of $\exp(\beta_p)$

In background, Stata calculates estimates of $S_0(t)$

Adjusted Curve

- Look at effect of x_1 (sex) adjusting for x_2 (copper)
- Create two curves with same value for x_2 *(we are <u>not really adjusting</u> for copper, we <u>are examining</u> the effect of sex with copper held constant)*
- But copper differs by sex! – so should we use sex-specific adjustments?
- So what value for x_2?
  the choice of value will affect the curves
- Let's use overall mean or median

Compare Adjusted Curves 1

```
. stcox sex copper
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   1.171796   .2996835     0.62   0.535     .7098385    1.934391
      copper |   1.006935   .0008328     8.36   0.000     1.005304    1.008569
------------------------------------------------------------------------------

. stcurve, survival at1(sex=0) at2(sex=1)
```

stcurve: gives predicted curves
survival: graph survival (not hazard default)
at1: (value for curve 1)
at2: (value for curve 2)
Note that the copper default is fixed at overall mean(=97.6)

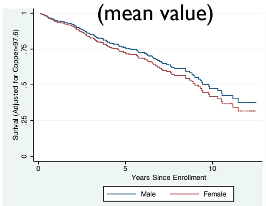
```
. stcurve, survival at1(sex=0 copper=97.6) at2(sex=1 copper=97.6)
```

gives same result

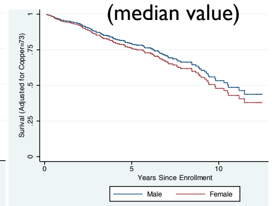
Compare Adjusted Curves 2

```
stcurve, survival at1(sex=0) at2(sex=1)
stcurve, survival at1(sex=0 copper=73) at2(sex=1 copper=73)
```

copper set to 97.6 (mean value)



copper set to 73 (median value)

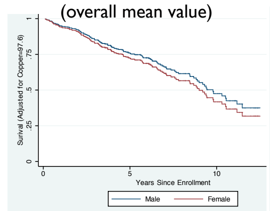


reference value for copper matters

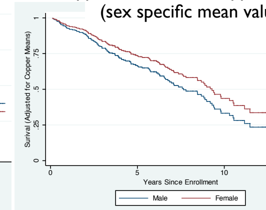
Compare Adjusted Curves 3

```
stcurve, survival at1(sex=0) at2(sex=1)
stcurve, survival at1(sex=0 copper=154) at2(sex=1 copper=90)
```

male and females= 97.6 (overall mean value)



male copper=154, female copper=90 (sex specific mean values)



adjusting for sex differences in copper matters

Adjusted/Predicted Curves

- Can be useful for visualizing effect of predictor
- Must choose reference values for confounders
  - often choose mean for continuous variable
  - most common category for categorical
- `stcurve` is a flexible tool for creating adjusted or predicted survival curves
- Summary: Look at survival curves with `stcurve` while fixing other variables in the model

Time Dependent Covariates

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

… and is measured (or evaluated) at multiple times during the study

14

Example

- Risk factors for pregnancy in a cohort of HIV infected women in Uganda
- Is the development of pregnancy affected by CD4 cell counts?
- We could consider only <u>baseline</u> CD4 count as a predictor (i.e. CD4 value at study onset)
- *But, CD4 cell count measured throughout the study!*
- Multiple measures of CD4 during study could provide additional prognostic information

15

Example

E.g., Patient #24901:

CD4 at baseline: 143
CD4 at day 123: 202
CD4 at day 216: 344
CD4 at day 284: 373
Pregnant on day 380

16

Data

multiple records per subject

	idno	t_from	t_to	cd4	prg
218.	24901	0	123	143	0
219.	24901	123	216	202	0
220.	24901	216	284	344	0
221.	24901	284	380	373	1
229.	25601	0	117	112	0
230.	25601	117	216	304	0
231.	25601	216	293	319	0
232.	25601	293	379	297	0
233.	25601	379	468	302	0
234.	25601	468	560	264	0
235.	25601	560	574	277	0
236.	25601	574	651	277	0
237.	25601	651	738	268	0

- `idno`: subject id #
- `t_from`: start of interval
- `t_to`: end of interval
- `cd4`: cd4 cell count during interval
- `prg`: pregnancy (1 = yes, 0 = no)

Stata syntax to define dataset:

```
stset t_to, failure(prg) id(idno)
```

17

TDC Cox model

```
gen cd4_50 = cd4/50
stcox cd4_50


No. of subjects =          702                   Number of obs   =       4935
No. of failures =           85
Time at risk    =       448321
                                                 LR chi2(1)      =     129.91
Log likelihood  =   -485.32684                   Prob > chi2     =     0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cd4_50 |   .5456291   .0344751    -9.59   0.000     .4820756    .6175612
------------------------------------------------------------------------------
```

Interpretation: a 50 cell increase in CD4 cell count (at any time point) is associated with a 45% reduction in the rate of pregnancy, 95% CI (-52% to -38%), p < 0.001

18

A different TDC example…

- Does lung transplant extend life of patients with Cystic Fibrosis?
- Outcome: Time from listing to death or censoring
- Predictor: Received lung transplant (yes or no)

Possible problem?

A different TDC example…

- Does lung transplant extend life of patients with Cystic Fibrosis?
- Outcome: Time from listing to death or censoring
- Predictor: Received lung transplant (yes or no)
- Bias: waiting list mortality!

Short-term survivors unlikely to get a transplant!

Solution

Treat transplant as a time-dependent covariate

$$tx(t) = \begin{cases} 0 & \text{before transplantation} \\ 1 & \text{after transplantation} \end{cases}$$

$$h(t|tx) = \begin{cases} h_0(t) & \text{before transplantation} \\ \exp(\beta)\, h_0(t) & \text{after transplantation} \end{cases}$$

group membership changes over time

Summary
TDC Cox Model

- TD Covariates useful when values of predictors change
- Key is to set up dataset properly
- Straightforward fitting the Cox model
- We will look at another way to use TDC to accommodate non-proportional hazards later…

Diagnostics for model checking: *testing the proportional hazards assumption*

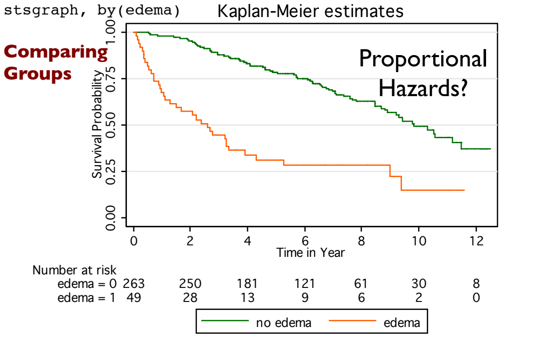
"all models are wrong, but some are useful…"

George Box - statistician (1987)

Model Checking PBC Data

`stsgraph, by(edema)`

Comparing Groups

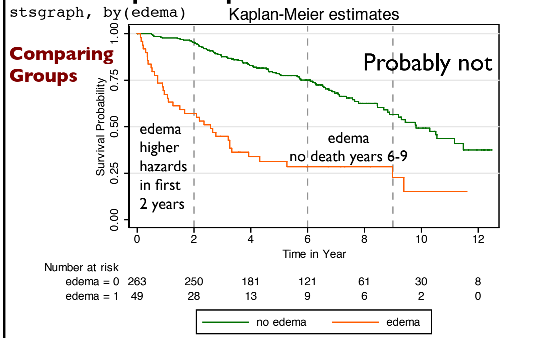


25

Group Proportional Hazards?

`stsgraph, by(edema)`

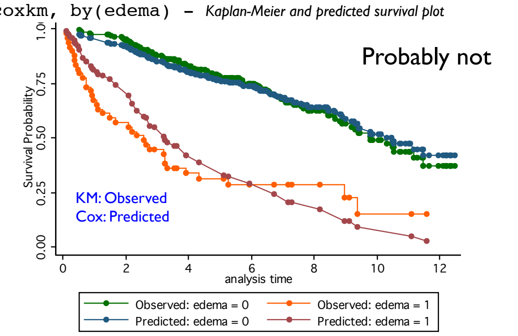
Comparing Groups



26

Group Proportional Hazards?

`stcoxkm, by(edema)` – *Kaplan-Meier and predicted survival plot*



27

Log minus log plots for categorical predictors

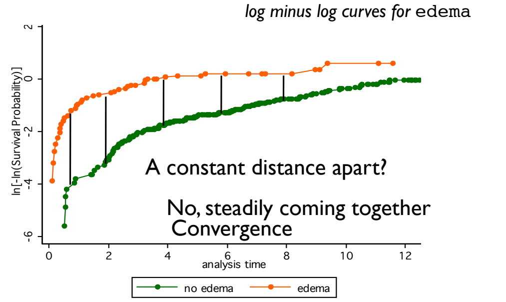
Under the Cox model:

- $\log(-\log(S_1(t))) = \beta + \log(-\log(S_0(t)))$
- Estimate survival curves, transform them by: (1) taking log, (2) multiplying by *-1*, then (3) taking log again
- Therefore the curves $\log(-\log(S_1(t)))$ and $\log(-\log(S_0(t)))$ should be a constant distance apart

28

Log minus log plot: edema

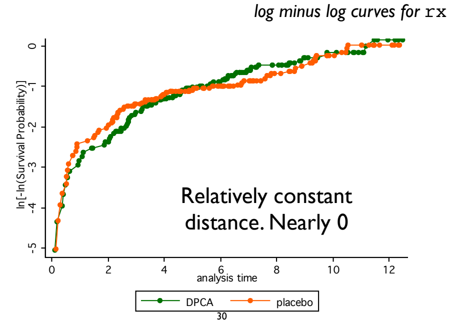
`stphplot, by(edema) nonegative nolntime` – *log minus log curves for* `edema`



29

Log minus log plot: rx

`stphplot, by(rx) nonegative nolntime` – *log minus log curves for* `rx`



30

Interpreting Curves

- Easily calculated (pro)
- Naturally subjective (con)
  - Not so easy to interpret
  - Look for pronounced convergence/divergence, or marked crossing
- Only works for categorical variables (con)
- Multiple crossing is evidence of a lack of overall effect (i.e., difference=0, HR=1)

31

Smoothed Hazard Ratio

(for continuous and categorical predictors)

- Possible to use "residuals" to estimate shape of hazard ratio over time
- HR(t): hazard ratio at time t
  - If HR(t) is reasonably constant: prop. hazards
  - If not, gives description of shape of HR
- The method estimates log(HR(t)) = β(t)

32

How does it work?

- Fit Cox model with relevant predictors
- Obtain "scaled Schoenfeld residuals"
  complex formula to generate
  residuals for each predictor & time point
- LOWESS: smooth residuals vs. time
- Plot the smooth curve estimates of β(t)
- *Note that estimated curves may change with bandwidth selection*

33

Stata

```
gen age10 = age/10

stcox edema age10, scaledsch(r_edema r_age)
```

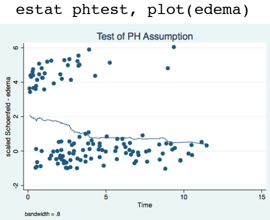
=> *saves residuals* `r_edema` *for* edema, `r_age` *for* `age10`

```
No. of subjects =          312                  Number of obs   =        312
Log likelihood  =    -614.3788                  Prob > chi2     =     0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio  Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       edema |   3.471158  .7099928     6.08   0.000      2.324711    5.182981
       age10 |   1.355471  .1166134     3.54   0.000      1.145143    1.604429
------------------------------------------------------------------------------
```

34

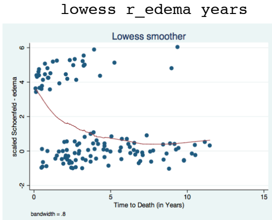
β(t) for edema

`estat phtest, plot(edema)`



running mean smoother `r_edema` *vs.* `time` *to estimate* β(t)

`lowess r_edema years`



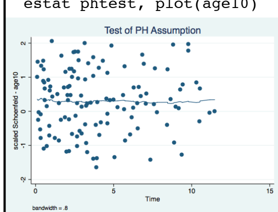
lowess smoother `r_edema` *vs.* `time` *to estimate* β(t)

flat line? *line is not flat, HR is not constant*

35

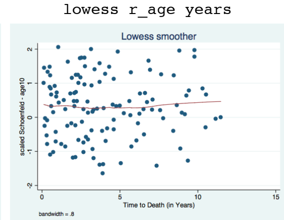
β(t) for age10

`estat phtest, plot(age10)`



running mean smoother `r_age` *vs.* `time` *to estimate* β(t)

`lowess r_age years`



lowess smoother `r_age` *vs.* `time` *to estimate* β(t)

flat line? *line is approximately flat, HR is relatively constant*

36

Smoothing Hazard Ratio

- Present the smoothed curves as a summary
- Augment it with the table to explain the HR
- Get those values by typing

```
lowess r_edema years, gen(smloghr) nogr
gen smhr=exp(smloghr)
sort years
list years sm* if status==1
```

Lowess $\beta(t)$ values for edema

Years	Log HR	HR
1	2.5	12.2
2	1.6	5.1
4	0.86	2.4
6	0.58	1.8

37

Test of Proportional Hazards

- Null hypothesis: Hazards are proportional
  i.e., $\beta(t)$ is constant over time
  i.e., no association between residuals & time
- Alternative: Hazards are not proportional
  i.e., $\beta(t)$ changes with time
  i.e., association between residuals & time

Idea is to look at correlation between residuals and time?

38

Schoenfeld Test

- A test for non-proportional hazards:
  correlation between residual and time
  `stcox edema age10`
  `estat phtest, detail`
- Small p-value means
  proportional hazards is rejected –
  the proportional hazards assumption can be shown false

39

Schoenfeld Test

```
stcox edema age10
estat phtest, detail
```

```
Test of proportional-hazards assumption

Time:  Time
----------------------------------------------------------------
            |       rho            chi2       df       Prob>chi2
------------+---------------------------------------------------
edema       |   -0.33749          13.09        1          0.0003
age10       |    0.01747           0.03        1          0.8540
------------+---------------------------------------------------
global test |                     13.52        2          0.0012
----------------------------------------------------------------
```

rho is correlation between residuals and time

We see that edema is statistically significant
= *non-proportional hazards*

40

Scaled Schoenfeld Residuals (plot & test)

- Technical/subjective, so hard to explain (con)
- Poor for multilevel categorical variables
  would need a plot for each level of category (con)
- Handles continuous variables well (pro)
- Can display effects on HR over time
- Note that different time-scaling functions can be used with `estat phtest` - can be important if there are outliers

41

Graphs vs. Tests

- Graphs and tests are complementary
- Need to look at whether the graph shows evidence of important violation
- Test helps as an objective assessment of graph
- However, tests have low power when *n* is small (and "too much" power when *n* is large)
- Graphs can show problem with test
  e.g., single outlier can affect test

42

Handling Non-Proportionality

- Stratification (categorical predictors)
- Time Dependent Covariates (continuous predictors)

43

Dealing with non-proportionality: Stratification

Stratified Cox Model

PBC data

- We have seen that baseline edema does not obey proportional hazards, but age does… so model,
- $h(t|edema=1,age) = h_{01}(t)\ \exp(\beta \times age)$

  $h(t|edema=0,age) = h_{00}(t)\ \exp(\beta \times age)$
- Models two separate baseline reference groups
- Proportional within edema but not across:
  relative effect of a 1-unit change in age on hazard is the same for edema = 1 or edema = 0;
  implicitly assumes no interaction between edema and age

45

Stratification Approach

- Fit a Cox model with terms for proportional variable and stratify by non-proportional variable

  `stcox age10, strata(edema)`
  (proportional) *(non-proportional)*
- Use adjusted survival curves to present the effect of edema

46

Stratified Cox Model

- Easily implemented in Stata:
  - Proportional hazards model for `age`
  - Stratified by `edema`

```
. stcox age10, strata(edema)

Stratified Cox regr. -- Breslow method for ties
No. of subjects =          312                  Number of obs   =        312
No. of failures =          125
Time at risk    =  1713.853528
                                                LR chi2(1)      =      11.60
Log likelihood  =   -546.68714                  Prob > chi2     =     0.0007
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       age10 |   1.342085   .1162448     3.40   0.001     1.132539    1.590402
------------------------------------------------------------------------------
                                                             Stratified by edema
```

No p-value, HR etc. for `edema`

47

Interpretation

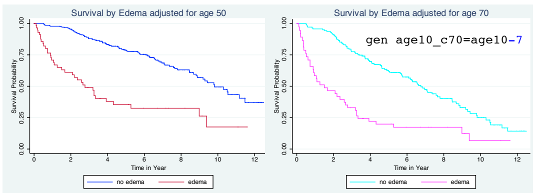
For each 10 years increase in age, there is a 34% increase in the hazard of death after adjusting for edema via stratification, 95% CI (13% increase to 59% increase)

Could mention you did a stratified model in the method section, rather than in the results.

48

Effect of Edema

- `gen age10_c50=age10-5`
  should center first, graph sets adjusted variables to 0
- `sts graph, by(edema) adjustfor(age10_c50)`



effect of edema is fading in time

49

Stratification Pros/Cons

- Fairly simple and non-technical approach
- What if the non-proportional variable is continuous?
- What if more than one non-proportional variable?

50

Summary

Stratified Cox Model

- Stratification requires multiple baseline hazards
- Stratification of a continuous variable (e.g., bilirubin) requires cutting it into categories
- Need to be at least 3-5 events per stratum
- Can use stratification as a way to adjust for non-proportional variable or to *avoid* proportional hazard assumption
- Gives no summary of the effect of stratum, but *adjusted survival curves can show strata effects*

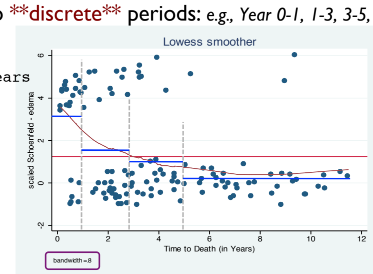
51

Dealing with non-proportionality 2: The time-dependent covariates "trick" (continuous and categorical predictors)

Time-Dep Cov Approach

Divide time into **discrete** periods: *e.g., Year 0-1, 1-3, 3-5, 5+*

Edema residuals
`lowess junk_e years`



The trick here is that the time interval itself becomes the time varying covariate!

53

Time-Dep Cov Approach

- Divide time into a series of periods (*e.g., Year 0-1, 1-3, 3-5, 5+*)
- Estimate HR for edema for each period
- Acheived by creating a series of TD covariates:
  `edema01, edema13, edema35, edema5p`
  that separately give the effect of edema in each period

```
stset years, failure(status) id(number)  // generates _t0, _t, _d

stsplit grp, at(1 3 5)  // split time variable (years) at these times into groups ≤ years (grp)
// that is, generate multiple rows for each subject; one for each timepoint up to and including
// the time of censoring or time of death

recode status .=0  // recodes all newly generated rows to "censored" status

gen edema01=edema*(grp==0)  // This set of commands generates 4 separate
gen edema13=edema*(grp==1)  // edema variables specific to each time interval;
gen edema35=edema*(grp==3)  // that is, edemaXX only equals 1 if the patient has edema
gen edema5p=edema*(grp==5)  // AND the dataset row corresponds to period XX
```

54

TD Cov Set-Up

```
. list number _t0 _t status edema grp edema01 edema13 edema35 edema5p in 1/12, sepby(number)
// lists values of the variables: number _t0 _t status edema grp edema01 edema13 edema 35 edema5p;
// 'in 1/12' restricts to first 12 rows; 'sepby(number)' draws line between each subject::
```

	number	_t0	_t	status	edema	grp	edema01	edema13	edema35	edema5p
1.	1	0	1	Censored	1	0	1	0	0	0
2.	1	1	1.0951403	Dead	1	1	0	1	0	0
3.	2	0	1	Censored	0	0	0	0	0	0
4.	2	1	3	Censored	0	1	0	0	0	0
5.	2	3	5	Censored	0	3	0	0	0	0
6.	2	5	12.320329	Censored	0	5	0	0	0	0
7.	3	0	1	Censored	1	0	1	0	0	0
8.	3	1	2.770705	Dead	1	1	0	1	0	0
9.	4	0	1	Censored	1	0	1	0	0	0
10.	4	1	3	Censored	1	1	0	1	0	0
11.	4	3	5	Censored	1	3	0	0	1	0
12.	4	5	5.2703629	Dead	1	5	0	0	0	1

A separate edema variable is set up for each time period so that we can have hazard ratio estimates for edema specific to each time interval!

55

Output

```
gen age10=age/10
stcox edema?? age10
```

```
No. of subjects =          312                  Number of obs   =       1001
No. of failures =          125
Time at risk    =  1713.853528
                                                LR chi2(5)      =      69.23
Log likelihood  =   -605.36554                  Prob > chi2     =     0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
edema01	**14.45344**	**6.974774**	5.53	0.000	5.613169	37.21639
edema13	3.423855	1.241054	3.40	0.001	1.682588	6.967111
edema35	3.187902	1.495416	2.47	0.013	1.2712	7.994587
edema5p	.8742166	.526164	-0.22	0.823	.2687244	2.844009
age10	1.33777	.1153185	3.38	0.001	1.129812	1.584006

HR declines with time,
does not statistically significantly differ from 1 after year 5

56

Interpretation

"Adjusted for age, during the first year of follow-up, subjects with edema at baseline have about 14-fold (5.6-37) higher hazard of death. During years 1-3 and 3-5, it is 3.4-fold (1.7, 7.0) and 3.2-fold (1.3, 8.0) higher, respectively compared to those with no edema. After year 5, the relative hazard is 0.87 (0.3, 2.8), not statistically significantly different from 1.0 (although the confidence interval is wide enough to include important differences)."

57

Output


```
gen age10=age/10
stcox edema?? age10
```


Similar?

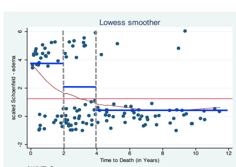

```
No. of subjects =          312                  Number of obs   =       1001
No. of failures =          125
Time at risk    =  1713.853528
                                                LR chi2(5)      =      69.23
Log likelihood  =   -605.36554                  Prob > chi2     =     0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
edema01	**14.45344**	**6.974774**	5.53	0.000	5.613169	37.21639
edema13	$\boxed{3.423855}$	1.241054	3.40	0.001	1.682588	6.967111
edema35	$\boxed{3.187902}$	1.495416	2.47	0.013	1.2712	7.994587
edema5p	.8742166	.526164	-0.22	0.823	.2687244	2.844009
age10	1.33777	.1153185	3.38	0.001	1.129812	1.584006

58

Output

- What if divide time into
  Year 0-2, 2-4, 4+?



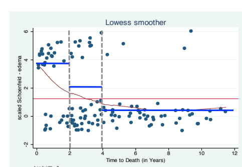

```
No. of subjects =          312                  Number of obs   =        784
Log likelihood  =   -604.93786                  Prob > chi2     =     0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
edema02	10.54145	3.869923	6.42	0.000	5.133481	21.64657
edema24	3.2131	1.138287	3.29	0.001	1.604626	6.433907
edema4p	.9417363	.4928127	-0.11	0.909	.3376709	2.626425
age10	1.342123	.1157712	3.41	0.001	1.13336	1.58934

59

Output

- What if divide time into
  Year 0-2, 2-4, 4+?




```
No. of subjects =          312                  Number of obs   =        784
Log likelihood  =   -604.93786                  Prob > chi2     =     0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
edema02	10.54145	3.869923	6.42	0.000	5.133481	21.64657
edema24	$\boxed{3.2131}$	1.138287	3.29	0.001	1.604626	6.433907
edema4p	.9417363	.4928127	-0.11	0.909	.3376709	2.626425
age10	1.342123	.1157712	3.41	0.001	1.13336	1.58934

60

What about Follow Up Edema Values?

- Recall, edema only codes for baseline edema
- Subjects with edema die off fairly fast
- PBC is a progressive disease
- Subjects are developing edema over time
- If used the yearly information on edema as a TD covariate, the effect may not fade with time

TDC Pros/Cons

- Con: Bit of programming to set up
- Con: Somewhat artificial on choice of cut-points
- Pro: Works for continuous and categorical predictors
- Pro: Estimates time-varying HRs and 95% CIs
- Pro: Clinicians love cutpoints
  e.g. *might say "edema doesn't matter after 4-5 yrs"*

Alternative time dependent covariates approach: *continuous* interaction with time

- Allows effect of predictor to vary continuously with time
- `stcox` with `tvc()` and maybe `texp(fn(_t))`, or `stsplit, at(failures) --` [see Sections 6.4.2.6 and 6.9 of VGSM plus time-varying subsection of `help stcox` and `help tvc_note`: especially see hyperlink to Cox regression with continuous time-varying covariates]
- The continuous interaction with time model is more realistic than splitting at specific times (pro)
- Typically requires transformation so that hazard ratio is a linear (or other "smooth") function of the interaction with time (con)

Linearity in Predictors?

continuous predictors

- We considered the proportional hazard assumption
- But what about the linearity of predictors assumption?
- $\log(h(t|x)/h_0(t)) = \beta_1 x_1 + \ldots + \beta_p x_p$

Linearity in Predictors?

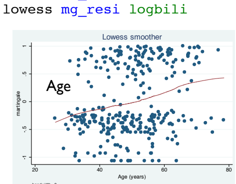
continuous predictors

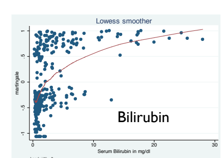
- Possible to use "martingale residuals" to discover a suitable functional form for a *predictor* (possibly adjusted for others)
- Fit Cox model with no predictor, and give a name "*mg_resi*" to the martingale residuals
  `. stcox, mgale(mg_resi) estimate`
- LOWESS: smooth *mg_resi* vs. *predictor*
- Does it appear to be approx. a straight line?

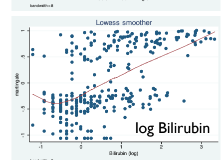
Model Checking: PBC

Linearity in predictors?

```
. use pbc.dta
. stset years, failure(status)
. stcox, mgale(mg_resi) estimate
. lowess mg_resi age
. lowess mg_resi bili
. lowess mg_resi logbili
```







Similar plots can be made while adjusting for other covariates by `stcox covariates, mgale(mg_rescov)` and then, e.g., `lowess mg_rescov age`

Other Survival Topics

- **Clustered data**
  - Dr. McCulloch's lectures have a strong focus on the analysis of clustered data and you will see many examples (but not survival)
  - Clustered survival data consist of e.g., multiple subjects clustered by center, or multiple events (multiple failure times) on a subject (see Dr. McCulloch's lectures)
  - use `shared(cluster_id), cluster(id) or vce(cluster id)` in Stata with `stcox`
  - The key is to appropriately set up the data for the type of clustering – ordered vs. non-ordered events, single or multiple event types etc.
  - See http://www.stata.com/support/faqs/statistics/multiple-failure-time-data

Other Survival Topics

- **Interval Censoring**
  - Conventional survival data is right-censored only
  - Often only know event happens in an interval (e.g., between scheduled hospital visits).
  - Regular intervals can use pooled logistic regression (e.g., all patients visit at time 0, 12 months, 24 months and 36 months
  - Non-regular intervals need non-parametric Kaplan-Meier-Turnbull or parametric modeling (i.e. with a model for baseline hazard or survival distribution)
  - Use `intcens` in STATA from SSC for parametric models type `ssc install intcens` in STATA then look at: `help intcens`

Other Survival Topics

- **Left-Truncation**
  - PBC used time from enrolment for time 0, but makes more sense to use diagnosis time
  - Why LT a problem? subjects with early death less likely to be enrolled – if not accounted for will underestimate early deaths
  - `stset years_since_diag, failure(status) enter(disease_dur)`
    `disease_dur = truncation times (time from Dx to enrollment)`
  - Survival function drops faster early on after accounting for left-truncation
  - See section 6.6.6 in VGSM

Summary

- Time dependent covariates
- Testing proportional hazards: graphs and test
- Non-proportional hazards solutions
  1) Stratified Cox
  2) Time dependent covariate trick (discrete or continuous)

Other extensions: clustering, left-truncation, interval censoring – competing risks (later…)

Don't forget…

- HW 1 discussion will follow now
- Next lecture: "Common Biostatistical Problems" 4/21
- Upload HW 2 by start of 4/21 lecture.
- HW 2 discussion will follow lecture on 4/21
- **The homework of 4/21 will be due by the lecture on 4/23 -- only a two day window!!!!**