

Survival Analysis II

Reading VGSM 6.2.5 - 6.2.13

John Kornak
April 5, 2016

john.kornak@ucsf.edu

- Project description due today
- Hwk #1 due next Tuesday, 4/12/16 – upload via syllabus
- Reading for next lecture VGSM 6.3 - 6.5
- Lab. 2 on web site - for this Thursday Rm MH 1400
- Check and use Biostat 209 discussion forum

In this lecture

- Review Survival data and the Cox model
- Interpreting Cox model fits
- Binary, categorical and continuous predictors
- Confounding and mediation
- Review of interaction effects
- Interaction terms in the Cox model
- Adjusted survival curves (for Cox model)

Survival review: key concepts

- Survival data: right censoring
- Linear/logistic regression sub-optimal
- Kaplan-Meier/logrank “raw” summary/test
- Hazard function: $\rightarrow$ “instantaneous risk”
- Proportional hazards assumption: ratio of hazards is constant over time
- Cox model - no baseline hazard model
- Effect of 1 unit change in a predictor on Survival, given in terms of “hazard ratio”: the *relative hazard*

Cox regression review

- Proportional hazards model $\rightarrow$ $\log(\text{hazard ratio})$ depends linearly on regression coefficients
- $h(t|x) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p)$
- $\log(h(t|x)/h_0(t)) = \beta_1 x_1 + \dots + \beta_p x_p$
- C.f. *log-odds* in logistic regression and *outcome* in linear regression - each depends linearly on regression coefficients

Review Cox Model

- Assumes Proportional Hazards
- Can summarize predictor effects based on coefficients, β , or in terms of hazard ratios, $\exp(\beta)$
- Hazard ratios work better for interpretation
- Math simpler based on coefficients (easy to go back and forth)

Cox Model - Wald test and CIs

- Confidence intervals and Wald tests are based on the fact that $\hat{\beta}$ has an approximate normal distribution (rule of thumb: at least 15-25 **events**)
- Test and confidence interval are based on estimators $\hat{\beta}$ for coefficients β
- 95% CI for HR is
Upper limit: $\exp(\hat{\beta} + 1.96 \times \text{SE}(\hat{\beta}))$
Lower limit: $\exp(\hat{\beta} - 1.96 \times \text{SE}(\hat{\beta}))$
- Wald test: $Z = \hat{\beta} / \text{SE}(\hat{\beta})$ (i.e. assume approx. normal)

Lung Cancer Data

- 40 subjects with Bronchioloalveolar Carcinoma (BAC / lung cancer)
- Each subject underwent a Positron Emission Tomography (PET) scan
- Determined uptake of Fludeoxyglucose, (18F FDG) in standard units: (variable `fdg25`, if tumor Standard Uptake Value (SUV) > 2.5, Y/N)
- 12 subjects died during follow-up

Wald test and CI

```
load lung.dta
stset time, failure(event)
. stcox fdg25
```

No. of subjects	=	40	Number of obs	=	40
No. of failures	=	12			
Time at risk	=	1258.299 9 98			
Log likelihood	=	-31.3947 58	LR chi2(1)	=	10.03
			Prob > chi2	=	0.0015

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
fdg25	11.7675	12.35468	2.35	0.019	1.50317 2 92.1212

Upper limit: $\exp(\hat{\beta} + 1.96 \times SE(\hat{\beta}))$

Lower limit: $\exp(\hat{\beta} - 1.96 \times SE(\hat{\beta}))$

Wald test: $Z = \hat{\beta} / SE(\hat{\beta})$ -- Note: not HR/SE(HR)

CI's are calculated from coefficients not hazard directly

Normal approx. for coeffs.

Multiple predictors

```
stcox fdg25 tumorsize multifocal
```

No. of subjects	=	40	Number of obs	=	40
No. of failures	=	12			
Time at risk	=	1258.299 9 98			
Log likelihood	=	-29.486 13	LR chi2(3)	=	13.85
			Prob > chi2	=	0.0031

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
fdg25	7.4968	8.149509	1.85	0.064	.890357 6 63.1229 7
tumorsize	1.24912 8	.1436471	1.93	0.053	.997058 3 1.56492 4
multifocal	.296144	.3337985	-1.08	0.280	.032514 1 2.69733 1

Likelihood Ratio (LR) Tests

- Tests for effect of predictor(s) by comparing log-likelihood **between two models**
- Fit models **with and without** predictor(s) to be tested
- -2 times difference in log-likelihoods is compared to a chi-square distribution
- Important to use when **number of failures** is small and the HR is far from 1 (strong effect)

Likelihood Ratio vs. Wald

```
stcox fdg25 tumorsize multifocal
```

few failures

No. of subjects	=	40	Number of obs	=	40
No. of failures	=	12			
Time at risk	=	1258.299 9 98			
Log likelihood	=	-29.486 13	LR chi2(3)	=	13.85
			Prob > chi2	=	0.0031

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
fdg25	7.4968	8.149509	1.85	0.064	.890357 6 63.1229 7
tumorsize	1.24912 8	.1436471	1.93	0.053	.997058 3 1.56492 4
multifocal	.296144	.3337985	-1.08	0.280	.032514 1 2.69733 1

fairly large HR

"non"-significant Wald test

Performing corresponding Likelihood Ratio test for fdg25

- `stcox fdg25 tumorsize multifocal`
`est store A`
fitted the model with all predictors (the reference model), and then asks Stata to save log-likelihood for above model, call it "A"

- `stcox tumorsize multifocal`
`est store B`
fits model leaving out `fdg25`

- `lrtest A B` (or `lrtest A`)
compare log-likelihoods (defaults to comparing with current model)

Likelihood-ratio test
(Assumption: B nested in A)

LR chi2(1) = 5.09
Prb > chi2 = 0.024

significant LR test

Likelihood Ratio vs. Wald

- Two tests for the same null hypothesis
- Typically very close in results
- Will disagree when sample size small and HR are far from 1 or/and if colinearity is present (strong correlations between predictors)
- When they disagree, the likelihood ratio test is more reliable.
- LR test always better -- just less convenient to compute

Binary Predictors

```
.stcox over3cm
No. of subjects = 40          Number of obs = 40
No. of failures = 12
Time at risk = 1258.299 9 98
Log likelihood = -35.782 03    LR chi2(1) = 1.26
                                Prob > chi2 = 0.2623
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
over3cm	1.95083 9	1.19686 9	1.09	0.276	.586133 4 6.49301 7

"over3cm" coded 0/1

0 = tumor less than 3 cm

1 = tumor greater than 3 cm

relative hazard for ≥ 3 cm compared to < 3 cm = 1.95

Hazard is about double!

Binary Predictors

Suggest 0/1 coding:

- One-point change is easy to interpret
- Makes the baseline hazard an identifiable group e.g., those with tumors < 3 cm
- Simplifies lincoms when we consider interactions (we will model interactions soon...)

- Get same answer if coded 10/11

- Get same p-value but different HR if coded 0/2

Reversed Coding

```
. recode over3cm 0=1 1=0, gen(less3cm)
. stcox less3cm
No. of subjects = 40          Number of obs = 40
No. of failures = 12
Time at risk = 1258.299 9 98
Log likelihood = -35.782 03    LR chi2(1) = 1.26
                                Prob > chi2 = 0.2623
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
less3cm	.512599 8	.314487 8	-1.09	0.276	.154011 6 1.70609 6

"less3cm" coded 0/1

0 = tumor greater than 3 cm

1 = tumor less than 3 cm

LR, Wald tests same. HR and its CI are reciprocals

.5125998 = 1 / 1.950839

Potential Pitfall: Zero Hazard Ratio

Define fdg0 based on SUV>0

	Alive	Dead
Tumor SUV=0	4	0
Tumor SUV>0	24	12

No Deaths in Those with Tumor SUV=0

LR test looks OK

```
.stcox fdg0
No. of subjects = 40          Number of obs = 40
No. of failures = 12
Time at risk = 1258.299 9 98
Log likelihood = -34.6706 61    LR chi2(1) = 48
                                Prob > chi2 = 0.0621
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
fdg0	6.53e+17	5.87e+09	-0.00	1.000	0

Hazard Ratio equals zero

Wald test and CI's have broken down

Reverse the Reference?

fdg_gt0: 1 = SUV > 0, 0 if SUV=0

LR test is the same

```
.stcox fdg0_gt0
No. of subjects = 40          Number of obs = 40
No. of failures = 12
Time at risk = 1258.299 9 98
Log likelihood = -34.6706 61    LR chi2(1) = 48
                                Prob > chi2 = 0.0621
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
fdg0_gt0	2.07e+15	6.95e+22	0.00	1.000	0

Hazard Ratio equals ∞

Wald test and CI's still don't work

Interpretation

“Zero of four subjects with a SUV of 0 died while 12/36 subjects with SUV > 0 died (estimated hazard ratio = 0); the effect did not reach statistical significance ($p=0.06$)”

Zero/Infinite HR

- Two sides of the same coin
(depends on reference)
- Category has either 0% or 100% events
(often happens with lots of categories)
- Use likelihood ratio tests: they're fine
(Wald test performs poorly)
- Confidence intervals: see statistician
(to calculate likelihood ratio based CI)
- Sometimes can consolidate categories to handle the issue

Categorical Predictors

- Fit in Stata: `stcox i.categoricalpredictor`
- Many different possible tests and comparisons
 - Overall versus trend tests (unordered vs. ordered)
 - Making pairwise comparisons
- Can also use the “`i.`” syntax when binary predictor is not coded as 0/1

PBC Data

- 312 patients: Primary Biliary Cirrhosis (PBC)
- Randomized trial: DPCA vs. Placebo
- 125 subjects died
- 15 predictors: hepatomegaly, bilirubin, etc.

Cox Model

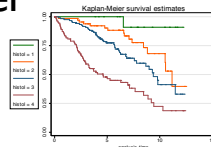
Is histology a significant predictor?

```
use pbc.dta
stset years, failure(status)
```

```
. stcox sex i.histol
```

```
No. of subjects = 312
No. of failures = 125
Time at risk = 1713.853 528
```

```
Log likelihood = -611.617 94
```



```
LR chi2(4) = 56.72
Prob > chi2 = 0.0000
```

	_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
sex	1	.607245 5	.1433789	-.2.11	0.035	.382282 3	.964593 9
histol	2	5.48886 2	5.667663	1.65	0.099	.725358 4	41.5347 8
	3	9.45956 5	9.589963	2.22	0.027	1.29698 8	68.9932 1
	4	23.0504 8	23.28112	3.11	0.002	3.18391 6	166.877 8

```
. lincom 3.histol-2.histol, hr
```

```
. lincom 4.histol-3.histol, hr
```

3 vs 2	1.72341 1	.5056402	1.86	0.064	.969729 5	3.06286
4 vs 3	2.43673 8	.4825026	4.50	0.000	1.65295 5	3.59216 8

Overall vs. Trend Tests for Multiple Categories

- Both have same null hypothesis:
(no difference in event risks between the groups)
- But different alternative hypothesis:
overall: at least one group is different
trend: there is a trend across the groups
- Use trend tests only for **ordered** predictors
(no trend test for ethnicity)
- When trends exist, a trend test is (typically) more powerful
- For ordinal predictors it is (typically) more interpretable

Trend vs. Overall Tests

- Trend Test (Wald test)

```
contrast p.histol
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	chi2	P>chi2
histol (linear)	1	10.69	0.0011

$p = 0.001$, there is a survival trend with pathology grade

- Overall Test (Wald test or LR test)

```
. testparm i.histol          . est store A
( 1) _Ihistol_2 = 0          . stcox sex
( 2) _Ihistol_3 = 0          . est store B
( 3) _Ihistol_4 = 0          . lrtest A B
chi2( 3) = 42.83             chi2( 3) = 52.95
Prob > chi2 = 0.0000         Prob > chi2 = 0.0000
```

$p < 0.0001$, at least one group different

contrast

based on group values based on order of group values
contrast p.histol (or contrast q.histol)

```
contrast p.histol
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	chi2	P>chi2
histol (linear)	1	10.69	0.0011
quadratic	1	0.60	0.4400
cubic	1	1.26	0.2608
Joint	3	42.83	0.0000

Note: fitting histol as categorical followed by contrast is better than adding histol as a continuous predictor, because initial fit not constrained to be linear

	Contrast	Std. Err.	[95% Conf. Interval]
histol (linear)	1.113267	.3404304	.4460361 1.780499
quadratic	-.2030152	.2629276	-.7183439 .3123136
cubic	.1682384	.1496035	-.1249791 .4614559

Continuous Predictors PBC data: age (in days) as predictor – what's going on?

```
.stcox age_days
No. of subjects = 312          Number of obs = 312
No. of failures = 125
Time at risk = 1713.853 528
Log likelihood = -629.725 92
LR chi2(1) = 20.51
Prob > chi2 = 0.0000
```

t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age_days	1.00011	.0000241	4.54	0.000	1.00006	1.000157

HR is nearly one
Wald and LR tests highly significant

** Pitfall – Poor Scaling of Continuous Predictors - PBC data: age (in days) as predictor

```
.stcox age_days
No. of subjects = 312          Number of obs = 312
No. of failures = 125
Time at risk = 1713.853 528
Log likelihood = -629.725 92
LR chi2(1) = 20.51
Prob > chi2 = 0.0000
```

t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age_days	1.00011	.0000241	4.54	0.000	1.00006	1.000157

HR is nearly one
Wald and LR tests highly significant

PBC data: age (decades) as predictor

```
.stcox age_decades
No. of subjects = 312          Number of obs = 312
No. of failures = 125
Time at risk = 1713.853 528
Log likelihood = -629.725 92
LR chi2(1) = 20.51
Prob > chi2 = 0.0000
```

t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age_decades	1.491811	.1314533	4.54	0.000	1.255188	1.773041

HR is greater!
Wald and LR tests exactly the same
(HR per year is about 1.04)

Continuous Predictor Scaling

- HR greatly affected by the scale of measurement (e.g., age in decades, years or days)
- Statistical significance is unaffected because SE is proportional to coefficient
- Choose interpretable unit change in predictor
- Can rescale by
 - (1) defining new variable
 - (2) using `lincom`
 - (3) direct calculation

(1) Define new variable

- Let `age_days` be age in days
 - `gen age_decades=age_days/(3650)`
 - `stcox age_decades` gives HR for one-decade older
- Works for every regression -- always
- Dividing by 3650: effect of one decade younger
- The most simple method

About 3650 days
per decade

(2) Lincom

- Let `age_days` be age in days
 - `stcox age_days`
 - `lincom 3650*age_days, hr`
 gives the effect of a decade (being 3650 days older)
- The HR option is important
otherwise get coefficient not the HR
- Less effort to implement than redefining variables
(especially for one-off calculations) but easier to make mistakes

A 1-unit change
in decade is
3650 unit change
in days

(3) Direct Calculation

- Let `HRage_d` be the HR for age in days
 - HR for decade = $(\text{HRage_days})^{3650}$
 - HR for confidence limits: also raised to 3650
 - test, p-values exactly the same
- HR for k -days = $(\text{HRage_days})^k$
 k is any arbitrary number, *could even be negative*
- Rarely need to use this method
but useful to know what calculations are going on

(3) Direct Calculation

	HR	Lower 95% CI	Upper 95% CI	Wald p- value	LR p-value
Day	1.00011	1.000062	1.00016	<0.0001	<0.0001
Decade	1.00011^{3650} = 1.49	1.000062^{3650} = 1.26	1.00016^{3650} = 1.79	same as above	same as above

Confounding in the Cox model

- Handled the same way as other regression models
- Confounders added into model
- Interpretation: HR of a 1-unit change holding all other predictors constant

UNOS Kidney Example

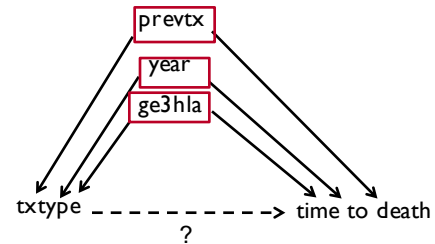
- Interest: How recipients from cadaveric donors do compared to living kidney recipients
- Unadjusted HR = 1.97, 95% CI (1.63, 2.40) – i.e., when not correcting for other predictors
- What might vary between living/cadaveric recipients?

UNOS Kidney Example

- Interest: How recipients from cadaveric donors do compared to living kidney recipients
- Unadjusted HR = **1.97**, 95% CI (1.63, 2.40) – i.e., when not correcting for other predictors
- What might vary between living/cadaveric recipients? **previous transplant, year of transplant, HLA match (0-2 loci vs. 3+)**
- Not accounting for these differences could lead to inflated unadjusted HR

Directed Acyclic Graph (DAG)

Potential confounding:



Adjusted Model

	stcox	txtype	prevtx	year	ge3hla	
No. of subjects	=					9678
No. of failures	=					407
Time at risk	=					38123.04 3 85
Log likelihood	=					-3480.7 78
						LR chi2(4) = 53.33
						Prob > chi2 = 0.0000

	_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
txtype		1.412006	.1866553	2.61	0.009	1.08941 7 1.83011 7
prevtx		.1675536	.216	0.031	1.02616 1	1.68978 8
year		.945617 1	.0159334	-3.32	0.001	.914898 1 .977367 4
ge3hla		.756309 5	.096678	-2.18	0.029	.588696 7 .971644 7

Attenuated HR for transplant type
vs

Unadjusted HR = **1.97**

Interpretation

“The hazard ratio of mortality for the recipient of a cadaveric kidney is 1.41 compared to living kidney (p=0.01), adjusting for year of transplant, history of previous transplants and degree of HLA compatibility. The 95% CI for the hazard ratio is 1.09 to 1.83”

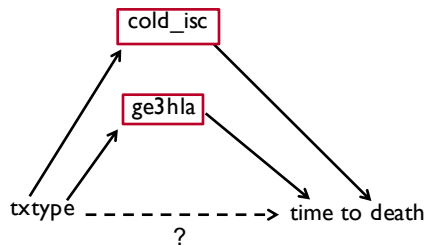
Is there confounding?

- Compare unadjusted and adjusted HR
- Screening of confounders based on association with mortality & txtype is too insensitive (if very predictive of mortality, but only slightly different between txtype then can still be important confounder)
- Examination of the associations is a way of understanding potential confounding, not a screening method for confounding
- Diff of 2.0 v. 1.4 -- clinically important? yes, the txtype association is confounded

Mediation

- How much of the effect of better prognosis of living recipients is explained by closer HLA match (ge3hla) and less transport time for the donor organ (cold_isc)?
- A question of mediation
- To what extent do ge3hla and cold_isc mediate the txtype/mortality relationship?

Directed Acyclic Graph (DAG)



After Adjustment

```
.stcox txtype ge3hla cold_isc
```

Log likelihood	=	-2776.2116	Prob > chi2	=	0.0000
----------------	---	------------	-------------	---	--------

t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
txtype	1.46322 5		.2902374	1.92	0.055	.991907 9 2.15849 4
ge3hla	.817813 1		.112796	-1.46	0.145	.624098 3 1.07165 5
cold_isc	1.00560 1		.0069314	0.81	0.418	.992106 8 1.01927 8

Reduction in txtype HR due to HLA and cold ischemia time is evidence of mediation

Mediation Measure

$$\% \text{ mediation} = \frac{\beta_{\text{unadj}} - \beta_{\text{adj}}}{\beta_{\text{unadj}}} 100\%$$

$$\beta_{\text{unadj}} = \log(1.97) = 0.678$$

$$\beta_{\text{adj}} = \log(1.46) = 0.378 \quad \frac{(0.678 - 0.378) \times 100}{0.678} = 44\%$$

"Approximately 44% of the mortality difference between living and cadaveric kidney recipients is explained by difference in HLA match and cold ischemia time"

See Sect 4.5 of VGSM for details

Aside: interactions in regression

Binary Variable Interactions

Outcome = Systolic BP

No interaction

	Drink = 0	Drink = 1
Smoke = 1	a	a + b
Smoke = 0	0	b

a = Smoking effect, b = Drinking effect

Binary Variable Interactions

Outcome = Systolic BP

Interaction

	Drink = 0	Drink = 1
Smoke = 1	a	a + b + c
Smoke = 0	0	b

a = Smoking effect, b = Drinking effect, c = interaction

Binary Variable Interaction with a Continuous Variable

$$y = \alpha + x_{age}\beta_{age} + x_{Dx}\beta_{Dx} + x_{age}x_{diag}\beta_{inter} + \varepsilon$$

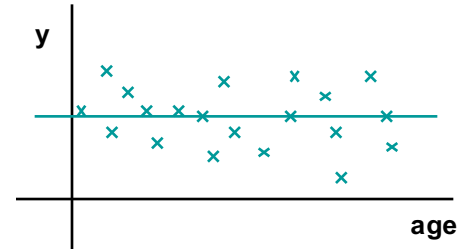
y = cognition score - outcome

$Dx = 1$, AD (Alzheimer's)

$Dx = 0$, HC (Healthy Control)

Case 1

$$\beta_{age} = 0, \beta_{Dx} = 0, \beta_{inter} = 0$$



$$y = \alpha + x_{age}\beta_{age} + x_{Dx}\beta_{Dx} + x_{age}x_{diag}\beta_{inter} + \varepsilon$$

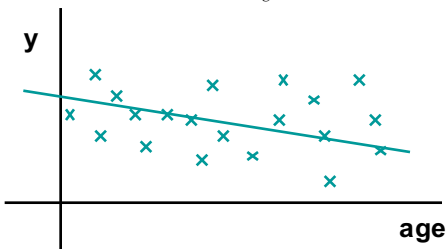
y = cognition score - outcome

$Dx = 1$, AD (Alzheimer's)

$Dx = 0$, HC (Healthy Control)

Case 2

$$\beta_{age} \neq 0, \beta_{Dx} = 0, \beta_{inter} = 0$$



$$y = \alpha + x_{age}\beta_{age} + x_{Dx}\beta_{Dx} + x_{age}x_{diag}\beta_{inter} + \varepsilon$$

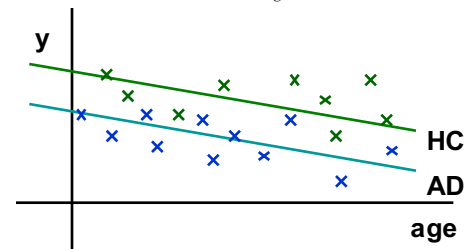
y = cognition score - outcome

$Dx = 1$, AD (Alzheimer's)

$Dx = 0$, HC (Healthy Control)

Case 3

$$\beta_{age} \neq 0, \beta_{Dx} \neq 0, \beta_{inter} = 0$$



$$y = \alpha + x_{age}\beta_{age} + x_{Dx}\beta_{Dx} + x_{age}x_{diag}\beta_{inter} + \varepsilon$$

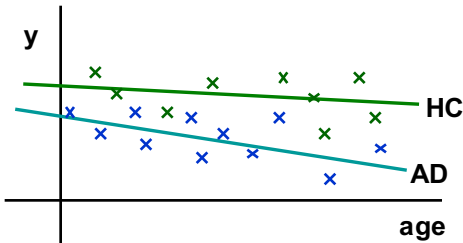
y = cognition score - outcome

$Dx = 1$, AD (Alzheimer's)

$Dx = 0$, HC (Healthy Control)

Case 4

$$\beta_{age} \neq 0, \beta_{Dx} \neq 0, \beta_{inter} \neq 0$$



$$y = \alpha + x_{age}\beta_{age} + x_{Dx}\beta_{Dx} + x_{age}x_{diag}\beta_{inter} + \varepsilon$$

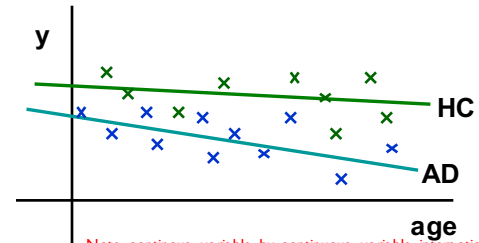
y = cognition score - outcome

$Dx = 1$, AD (Alzheimer's)

$Dx = 0$, HC (Healthy Control)

Case 4

$$\beta_{age} \neq 0, \beta_{Dx} \neq 0, \beta_{inter} \neq 0$$



Note, continuous variable by continuous variable interaction
even trickier to interpret -- we won't go there

Interaction

- Addressed the same way across all regression methods
 - Create product terms (use # or ## Stata 11 on)
 - Test of product terms reveals interaction
 - Understand interaction through series of `lincom` commands
- Predictors of **graft failure** in UNOS
- Is there an interaction between previous transplant and year of transplant?

Results

```
. stcox prevtx year
```

	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
prevtx	2.13531	.0999166	16.21	0.000	1.94818 8 2.34040 4
year	1.03157 9	.0080216	4.00	0.000	1.01597 6 1.04742 1

```
. stcox prevtx#c.year
```

	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
2.prevtx	6.39e+5 6	2.75e+58	4.00	0.000	1.14e+2 9 6.17e+8 4	
year	1.04799 4	.0091017	5.40	0.000	1.03030 6 1.06598 6	
prevtx#c.year	2	.936722 3	.0153842	-3.98	0.000	.907049 9 .967365 2

What gives?

- There is a huge HR for `prevtx`. Is this an example of colinearity?

What gives?

- There is a huge HR for `prevtx`. Is this an example of colinearity?
- There might be some colinearity, but it is a minor issue here
- The big issue: the HR gives the effect of `prevtx` when all other predictors are equal to zero (including `year`!)
It's huge because it's a meaningless extrapolation!

HR Interpretation

	_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval
	2.prevtx	8.39e+5 6	2.75e+58	4.00	0.000	1.14e+2 9 6.17e+8 4
	year	1.04799 4	.0091017	5.40	0.000	1.03030 6 1.06598 6
prevtx#c.year	2	.936722 3	.0153842	-3.98	0.000	.907049 9 .967365 2

- 2.prevtx**: HR of previous transplant in year 0
- year**: HR "year of tx" with no `prevtx` effect of +1 year when `prevtx=1` (ref group)
- prevtx#c.year 2**: HR more difficult to interpret: corresponds to an effect modifier – change in effect of year when had previous transplant (or change in effect of previous tx for each additional year)

Advice

- Don't fixate on the model HRs
- HRs may not correspond to something meaningful (sometimes yes, sometimes no)
- Instead: look at test for product term
- Followed by a series of `lincom` statements
 - What is the effect of `prevtx` in 1990?
 - What is the effect of `prevtx` in 1995?
 - What is the effect of `prevtx` in 2000?

Effect of Previous Transplant in 1990

	2.prevtx	year	product= 2.prevtx#c. year
previous transp.	1	1990	1990
no prev tx	0	1990	0
diff	1	0	1990

```
lincom 2.prevtx + 1990* 2.prevtx#c.year, hr
```

lincom

Effect of Previous transplant in 1990

```
. lincom 2.prevtx + 1990* 2.prevtx#c.year, hr
```

_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	2.686016		.1950779	13.60	0.000	2.329637 3.096914

Effect of Previous transplant in 1995

```
. lincom 2.prevtx + 1995* 2.prevtx#c.year, hr
```

_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	1.937148		.1054406	12.15	0.000	1.74113 2.155234

Effect of Previous transplant in 2000

```
. lincom 2.prevtx + 2000* 2.prevtx#c.year, hr
```

_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	1.397066		.1661109	2.81	0.005	1.106648 1.7637

Interpretation

The effect of previous transplant on risk of graft failure *varies* by year of transplant ($p < 0.0001$).

The relative hazards (and 95% Conf. Int.) for the previous transplant are 2.7 (2.3 - 3.1), 1.9 (1.7 - 2.2) and 1.4 (1.1 - 1.8), in the years 1990, 1995 and 2000, respectively.

Centered Regression

Year centered at 1995

```
. gen cyear=year-1995
      stcox prevtx#c.cyear
```

No. of subjects	=	9678	Number of obs	=	9678
No. of failures	=	2501			
Time at risk	=	38123.04385			
Log likelihood	=	-20488.046	LR chi2(3)	=	252.04
			Prob > chi2	=	0.0000

_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
2.prevtx	1.937148		.1054406	12.15	0.000	1.74113 2.155234
cyear	1.047994		.0091017	5.40	0.000	1.030306 1.065986
prevtx#c.cyear						
2		.9367223	.0153842	-3.98	0.000	.9070499 .9673652

Only 2.prevtx has changed: corresponds to effect of prev transplant in 1995

Effect of Previous Transplant in 1990

	2.prevtx	year	product= 2.prevtx#c. year
previous transp.	1	-5	-5
no prev tx	0	-5	0
diff	1	0	-5

```
lincom 2.prevtx - 5 * 2.prevtx#c.cyear, hr
```

lincom (centered)

Effect of Previous transplant in 1990

```
. lincom 2.prevtx - 5 * 2.prevtx#c.cyear, hr
```

_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	2.686016		.1950779	13.60	0.000	2.329637 3.096914

Effect of Previous transplant in 1995

```
. lincom 2.prevtx + 0 * 2.prevtx#c.cyear, hr
```

Effect of Previous transplant in 2000

```
. lincom 2.prevtx + 5 * 2.prevtx#c.cyear, hr
```

_t	Haz.	Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	1.397066		.1661109	2.81	0.005	1.106648 1.7637

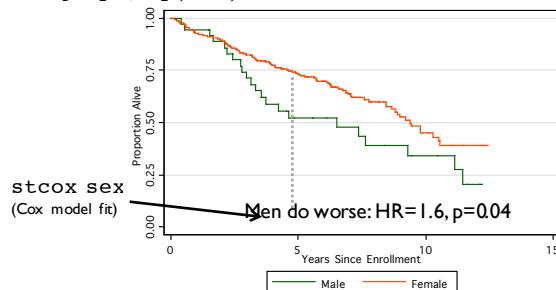
Interaction

- Same advice as in linear and logistic regression models: *create product, test product, calculate lincom*
- Colinearity is not the problem
- Centering gives same test for interaction and doesn't change `lincom`
- Center makes some coefficients more interpretable, but can make use of `lincom` subject to error (if forget the variable is centered)
- Don't forget "hr" in those `lincom` commands

Evaluating Cox model fits: Adjusted Survival Curves

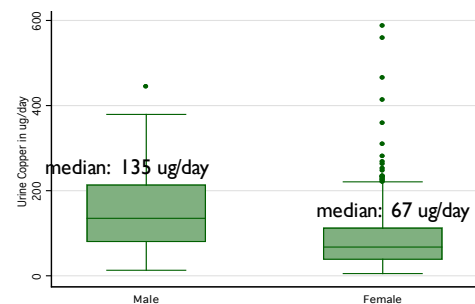
Effect of Sex: PBC data

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



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

$$S(t|x) = S_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p)$$

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 not really adjusting for copper, we are examining 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 I

```
. stcox sex copper
```

	_t	Haz.	Ratio	Std.	Err.	z	P> z	[95% Conf. Interval]
sex			1.17179 6		.2996835	0.62	0.535	.709838 5 1.93439 1
copper			1.00693 5		.0008328	8.36	0.000	1.00530 4 1.00856 9

```
. 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)

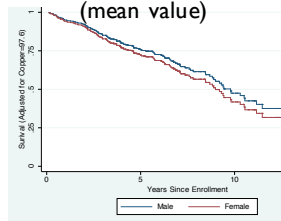
```
. stcurve, survival at1(sex=0 copper=97.6) at2(sex=1 copper=97.6)
```

gives same result

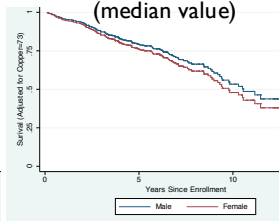
Compare Adjusted Curves 2

```
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)

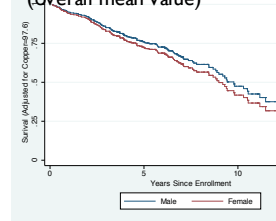


reference value for copper matters

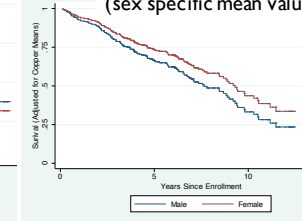
Compare Adjusted Curves 3

```
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

Summary of this lecture...

- Wald and likelihood ratio tests
- Zero/infinite HR
- Confounding, mediation, adjusting for other variables
- Interactions and lincom statements (danger of extrapolation)
- Useful tool: adjusted survival curves

Next Lecture

(extensions to the Cox model)

- Reading VGSM, Ch. 6.3 - 6.5
- Time-dependent covariates
- Diagnostics (model checking) - proportional hazards?
- Non-proportional Hazards: Stratification
- Non-proportional Hazards: generate time-dependent covariates trick
- Followed by homework session

79

Don't forget...

- Project description and date limitations due today
- HW 1 due by start of lecture 4/12/16
- Lab 2 (available on syllabus page) session on Thursday MH 1400
- Reading for next lecture VGSM 6.3 - 6.5