

Survival Analysis II

Reading VGSM 6.2.5 - 6.2.13

Chiung-Yu Huang

April 6, 2021

ChiungYu.Huang@ucsf.edu

- Project description due today
- Hwk #1 due next Tuesday, 4/13/21 – upload via CLE
- Reading for next lecture VGSM 6.3 - 6.5
- Lab. 2 on web site - for this Thursday
- Lab quiz due by 5PM on Thursday

In this lecture

- Review of survival data and hazard function
- Cox model
 - Interpreting Cox model fits
 - Binary, categorical, and continuous predictors
 - Confounding and mediation
 - Interaction terms in the Cox model
- Adjusted survival curves

Survival review: key concepts

- Survival data: right censoring
- Linear/logistic regression sub-optimal
- Kaplan-Meier/logrank “raw” summary/test
- Hazard function: $\longrightarrow$ “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 $\longrightarrow$ $\log(\text{hazard ratio})$ 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

Binary Predictor (UNOS)

$$\mathbf{x} = \begin{cases} 0 & \text{Living} \\ 1 & \text{Cadaveric} \end{cases}$$

$$h(t \mid \mathbf{x}) = h_0(t) \exp(\beta \mathbf{x})$$

Binary Predictor (UNOS)

$$x = \begin{cases} 0 & \text{Living} \\ 1 & \text{Cadaveric} \end{cases}$$

$$h(t \mid x) = h_0(t) \exp(\beta x)$$

Hazard in cadaveric group

$$h(t \mid x=1) = h_0(t) \exp(\beta \times 1)$$

versus hazard in living donor group

$$h(t \mid x=0) = h_0(t) \exp(\beta \times 0)$$

$$\text{Ratio} = \frac{h_0(t) \exp(\beta)}{h_0(t) \exp(0)} = \exp(\beta)$$

$\exp(0) = 1$

Hazard Ratio (HR)

- β is the regression coefficient
- If a predictor variable has no effect then $\beta=0$
- HR for a unit increase in a predictor is $\exp(\beta)$ --
if predictor has no effect then $\exp(\beta) = \exp(0) = 1$
- Can readily switch from summarizing models in terms of coefficients, β , or in terms of hazard ratios, $\exp(\beta)$
 - HR's work better for interpretation, but math simpler based on coefficients

Cox Model - Wald test and CIs

- Key: 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
- Wald test: $Z = \hat{\beta} / SE(\hat{\beta})$ (i.e. assume approx. normal)
- 95% CI for **log HR** is $(\hat{\beta} - 1.96 \times SE(\hat{\beta}), \hat{\beta} + 1.96 \times SE(\hat{\beta}))$
- 95% CI for **HR** is

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

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

Lung Cancer Data

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

Univariate Cox regression

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

```
. stcox fdg25
```

No. of subjects =	40	Number of obs =	40
No. of failures =	12		
Time at risk =	1258.299998		
Log likelihood =	-31.394758	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	1.503172 92.1212

Wald test $Z = \hat{\beta} / SE(\hat{\beta})$

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

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

CIs are calculated from coefficients

-- Note: not HR/SE(HR)!
HR estimate is heavily skewed

Multiple predictors

```
stcox fdg25 tumorsize multifocal
```

test overall
predictor effect

No. of subjects = 40
No. of failures = 12
Time at risk = 1258.299998

Number of obs = 40

Log likelihood = -29.48613

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	.8903576	63.12297
tumorsize	1.249128	.1436471	1.93	0.053	.9970583	1.564924
multifocal	.296144	.3337985	-1.08	0.280	.0325141	2.697331

Binary predictor: two or more foci of cancer (Y/N)

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 =	1258.299998		
Log likelihood =	-29.48613	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	.8903576	63.12297
tumorsize	1.249128	.1436471	1.93	0.053	.9970583	1.564924
multifocal	.296144	.3337985	-1.08	0.280	.0325141	2.697331

fairly large HR

“non”-significant Wald test

Likelihood Ratio (LR) Tests

- Tests for effect of predictor(s) by comparing log-likelihood **between two models**
- Fit models **with and without** predictor(s) of interest
- -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)

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)

significant LR test

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

LR chi2(1) = 5.09
Prb > chi2 = 0.0240

Likelihood Ratio vs. Wald

- Two tests for the same null hypothesis
- Typically very close in results
- Will disagree when event number/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    = 1258.299998

Log likelihood   =   -35.78203          LR chi2(1)       =          1.26
                                          Prob > chi2       =          0.2623
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
over3cm	1.950839	1.196869	1.09	0.276	.5861334 6.493017

“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 a meaningful group
- Simplifies `lincom` 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 =	1258.299998		
Log likelihood =	-35.78203	LR chi2(1) =	1.26
		Prob > chi2 =	0.2623

	_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
	less3cm	.5125998	.3144878	-1.09	0.276	.1540116 1.706096

“less3cm” coded 0/1

0 = tumor greater than 3 cm

1 = tumor less than 3 cm

LR, Wald tests same. HR and it's 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

```
stcox fdg0
```

```
No. of subjects =          40
No. of failures =          12
Time at risk    = 1258.299998

Log likelihood   = -34.670661
```

LR test looks OK

```
Number of obs    =          40

LR chi2(1)        =          3.48
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: I = SUV > 0, 0 if SUV=0

stcox fdg0_gt0

No. of subjects = 40
No. of failures = 12
Time at risk = 1258.299998

Log likelihood = -34.670661

LR test is the same

Number of obs = 40

LR chi2(1) = 3.48
Prob > chi2 = 0.0621

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
fdg_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: age, gender, bilirubin, histologic stage, etc.

Cox Model

Is histology a significant predictor?

```
use pbc.dta
stset years, failure(status)
sts graph, by(histol)
```

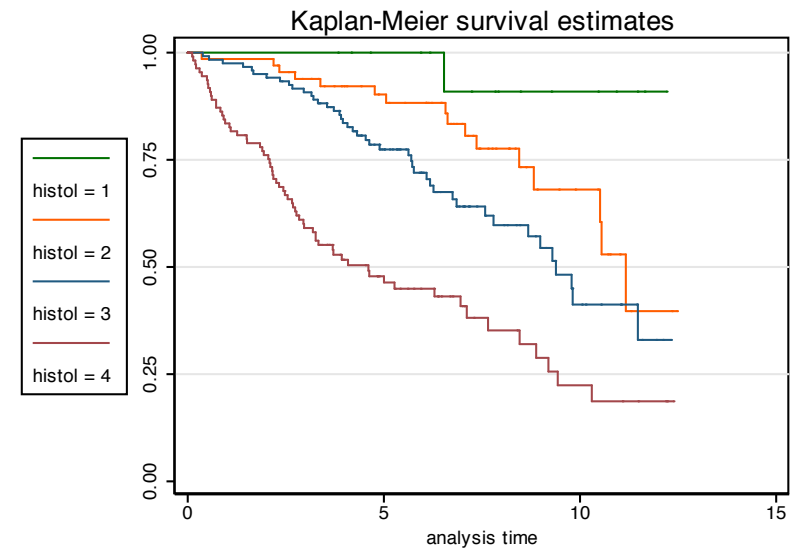
```
. stcox sex i.histol
```

```
No. of subjects =          312
No. of failures =          125
Time at risk    = 1713.853528
```

```
Log likelihood = -611.61794
```

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

_t		Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
sex histol		.6072455	.1433789	-2.11	0.035	.3822823	.9645939
	2	5.488862	5.667663	1.65	0.099	.7253584	41.53478
	3	9.459565	9.589963	2.22	0.027	1.296988	68.99321
	4	23.05048	23.28112	3.11	0.002	3.183916	166.8778



Stage 3 vs Stage 2

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

```
( 1)  - 2.histol + 3.histol = 0
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
3 vs 2	1.723411	.5056402	1.86	0.064	.9697295	3.06286

	sex	2.histol	3.histol	4.histol
Stage 3	x	0	1	0
Stage 2	x	1	0	0
diff	0	-1	1	0

Stage 4 vs Stage 3

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

```
( 1)  - 3.histol + 4.histol = 0
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
4 vs 3	2.436738	.4825026	4.50	0.000	1.652955	3.592168

	sex	2.histol	3.histol	4.histol
Stage 4	x	0	0	1
Stage 3	x	0	1	0
diff	0	0	-1	1

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 disease stage

- Overall Test (Wald test or LR test)

```
. testparm i.histol
```

```
( 1)  _Ihistol_2 = 0
```

```
( 2)  _Ihistol_3 = 0
```

```
( 3)  _Ihistol_4 = 0
```

```
chi2( 3) =    42.83
```

```
Prob > chi2 =    0.0000
```

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

```
. est store A
```

```
. stcox sex
```

```
. est store B
```

```
. lrtest A B
```

```
chi2( 3) =    52.95
```

```
Prob > chi2 =    0.0000
```

$p < 0.0001$, at least one group different

contrast

based on assigned 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 =	1713.853528		
Log likelihood =	-629.72592	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	1.000062 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    = 1713.853528
```

```
Log likelihood   = -629.72592          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	1.000062	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    = 1713.853528
Log likelihood   = -629.72592                LR chi2(1)           =          20.51
                                                Prob > chi2          =          0.0000
```

<u>_t</u>	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)
- Choose interpretable unit change in predictor
- Statistical significance is unaffected because SE is proportional to coefficient
- Can rescale by
 - (1) defining new variable
 - (2) using `lincom`

(I) Define new variable

*About 3650 days
per decade*



- 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

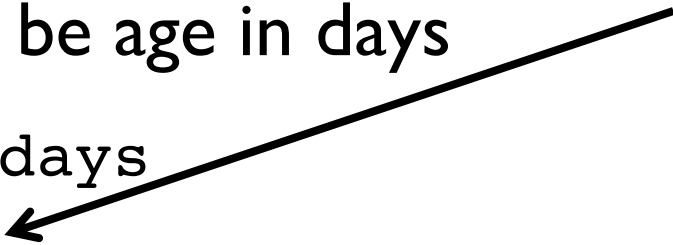
(2) Lincom

- Let `age_days` be age in days

- `stcox age_days`

- `lincom 3650*age_days, hr`

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



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*

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

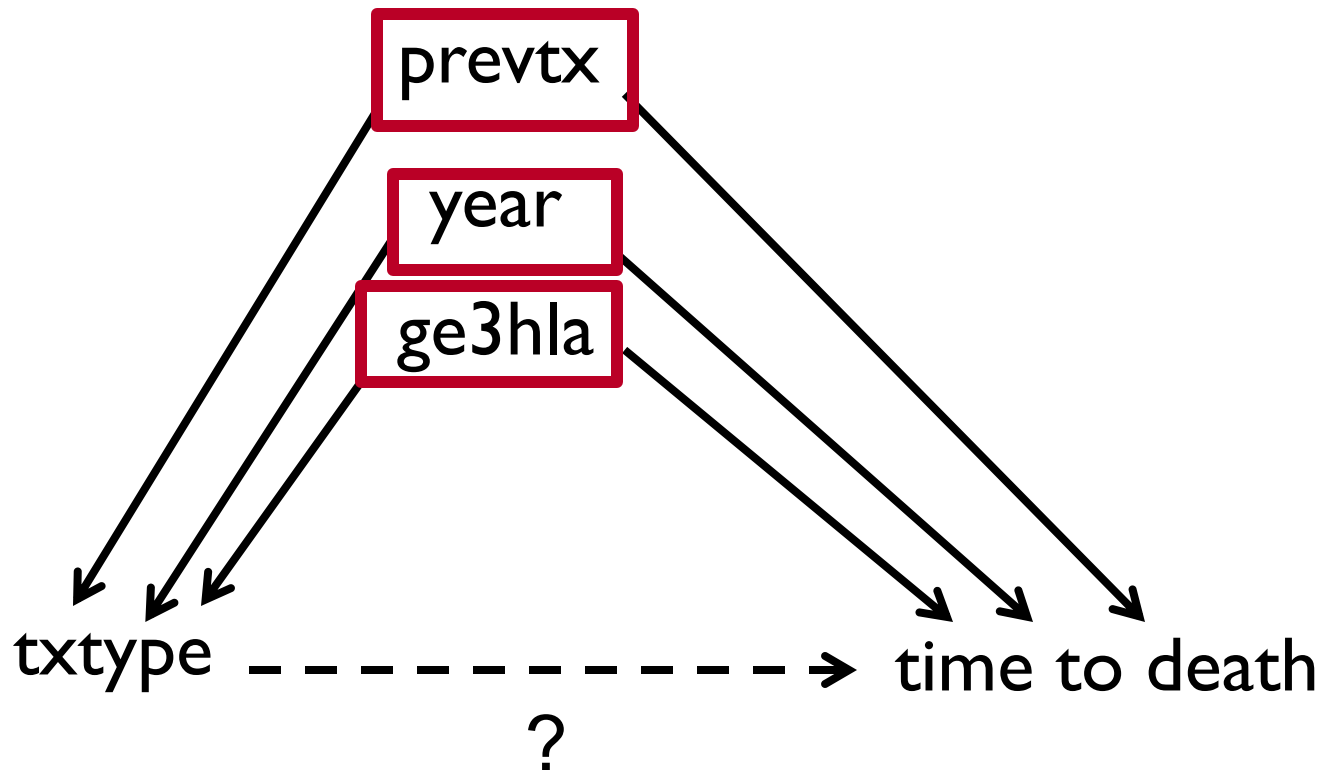
- Interest: How cadaveric donors recipients did compared to living kidney recipients
- Unadjusted HR = 1.97, 95% CI (1.63, 2.40) – i.e., when not correcting for other predictors
- What might vary between living/cadaveric recipients?

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	Number of obs =	9678
No. of failures =	407		
Time at risk =	38123.04385		
Log likelihood =	-3480.778	LR chi2(4) =	53.33
		Prob > chi2 =	0.0000

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
txtype	1.412006	.1868553	2.61	0.009	1.089417	1.830117
prevtx	1.316812	.1675536	2.16	0.031	1.026161	1.689788
year	.9456171	.0159334	-3.32	0.001	.9148981	.9773674
ge3hla	.7563095	.096678	-2.18	0.029	.5886967	.9716447

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”

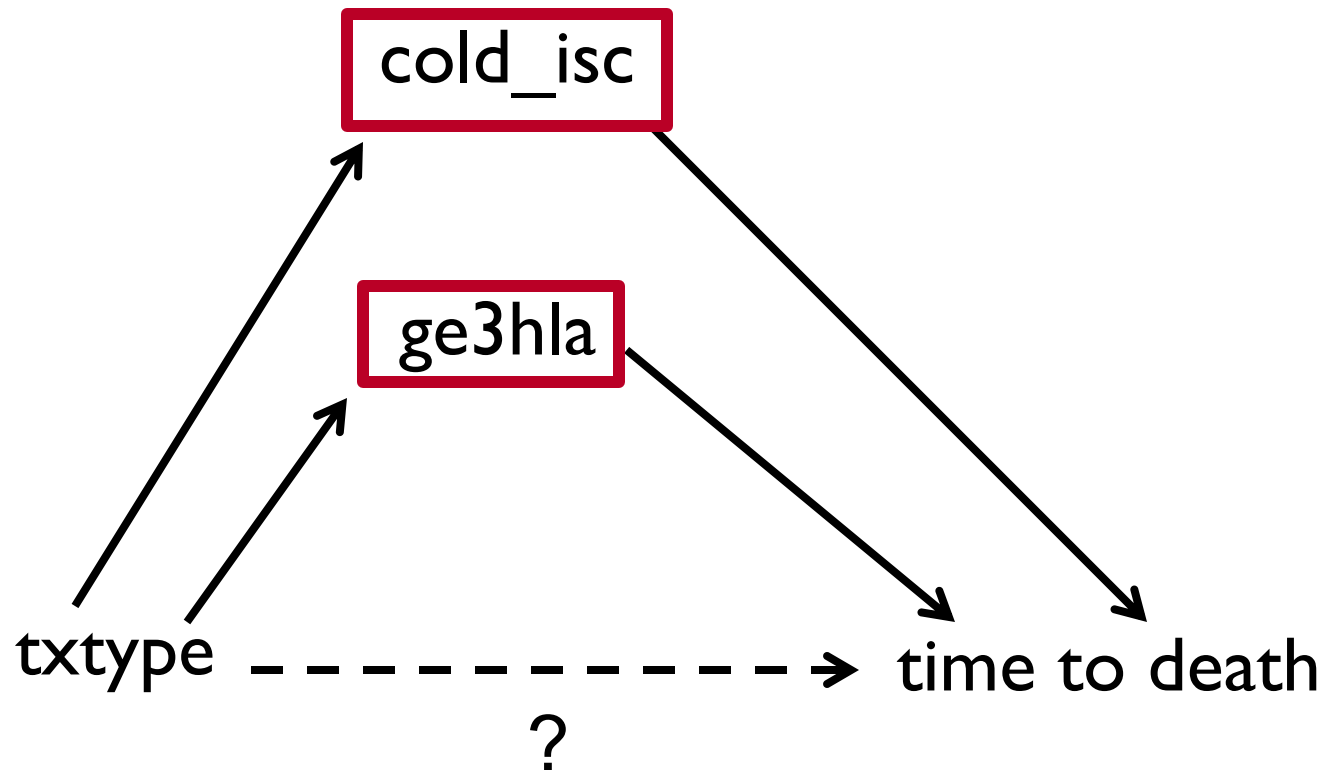
Is there confounding?

- Compare unadjusted and adjusted HR
- Screening of confounders based on association with mortality & txtype is too insensitive
(if highly predictive of mortality, but only slightly different between txtype, you will miss it but it's still an important confounder)
- Examination of the associations is a way of understanding potential confounding, not a screening method to rule out confounding
- Diff of 1.97 v. 1.41 -- 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 (`ge3h1a`) and less transport time for the donor organ (`cold_isc`)?
- A question of mediation
- To what extent do `ge3h1a` and `cold_isc` mediate the `txtype`/mortality relationship?

Directed Acyclic Graph (DAG)



After Adjustment

```
.stcox txttype ge3hla cold_isc
```

Log likelihood = -2776.2116 Prob > chi2 = 0.0000

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
txttype	1.463225	.2902374	1.92	0.055	.9919079	2.158494
ge3hla	.8178131	.112796	-1.46	0.145	.6240983	1.071655
cold_isc	1.005601	.0069314	0.81	0.418	.9921068	1.019278

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

Mediation Measure (Old Approach)

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

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

$$\beta_{\text{adj}} = \log(1.46) = 0.378$$

$$\frac{(0.678 - 0.378)}{0.678} \times 100\% = 44\%$$

“Approximately 44% of the mortality difference between living and cadaveric kidney recipients is explained by difference in HLA match and cold ischemia time” – Not recommended in modern mediation analysis

**Difference in coefficients under Cox model
does not give a causal interpretation!**

Causal Mediation

- Mediation from a counterfactual perspective
- Cox model fits poorly with mediation analysis
 - Cox model with and without mediator can not be valid at the same time
 - appropriate only when event is rare and no interaction between exposure and mediator
 - in general, can provide a valid test for whether there is mediation effect

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
```

```
Log likelihood = -20496.194 Prob > chi2 = 0.0000
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
prevtx	2.13531	.0999166	16.21	0.000	1.948188	2.340404
year	1.031579	.0080216	4.00	0.000	1.015976	1.047421

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

```
Log likelihood = -3496.171 Prob > chi2 = 0.0001
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
prevtx						
Y	8.39e+56	2.75e+58	4.00	0.000	1.14e+29	6.17e+84
year	1.047994	.0091017	5.40	0.000	1.030306	1.065986
prevtx#c.year						
Y	.9367223	.0153842	-3.98	0.000	.9070499	.9673652

Results

```
. stcox prevtx year
```

Log likelihood = -20496.194 Prob > chi2 = 0.0000

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
prevtx	2.13531	.0999166	16.21	0.000	1.948188	2.340404
year	1.031579	.0080216	4.00	0.000	1.015976	1.047421

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

huge HR for prevtx.

Log likelihood = -3496.171 Prob > chi2 = 0.0001

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
prevtx						
Y	8.39e+56	2.75e+58	4.00	0.000	1.14e+29	6.17e+84
year	1.047994	.0091017	5.40	0.000	1.030306	1.065986
prevtx#c.year						
Y	.9367223	.0153842	-3.98	0.000	.9070499	.9673652

interaction is significant

What gives?

- The 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 of prevtx in year 0

Centered Regression

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

Year centered at 1995

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]	
prevtx						
Y	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						
Y	.9367223	.0153842	-3.98	0.000	.9070499	.9673652

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

HR Interpretation

stcox prevtx##c.cyear

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
prevtx Y	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 Y	.9367223	.0153842	-3.98	0.000	.9070499	.9673652

- **prevtx Y**: HR of previous transplant in year 1995
- **year**: HR for +1 year when prevtx=0 (ref group)
- **prevtx#c.year Y**: 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 estimates may not correspond to something meaningful (sometimes yes, sometimes no)
- Instead: look at test for product term
- Followed by a series of `lincom` statements
 1. What is the effect of `prevtx` in 1990?
 2. What is the effect of `prevtx` in 1995?
 3. What is the effect of `prevtx` in 2000?

Effect of Previous Transplant in 1990

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

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


lincom (centered)

Effect of Previous transplant in 1990

```
. lincom 1.prevtx - 5 * 1.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 1.prevtx + 0 * 1.prevtx#c.cyear, 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 1.prevtx + 5 * 1.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

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.*

Interaction

- Same advice as in linear and logistic regression models: *create product, test product, calculate `lincom`*
- Centering gives same test for interaction and doesn't change `lincom`
- Centering 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

Adjusted Survival Curves

Effect of Sex: PBC data

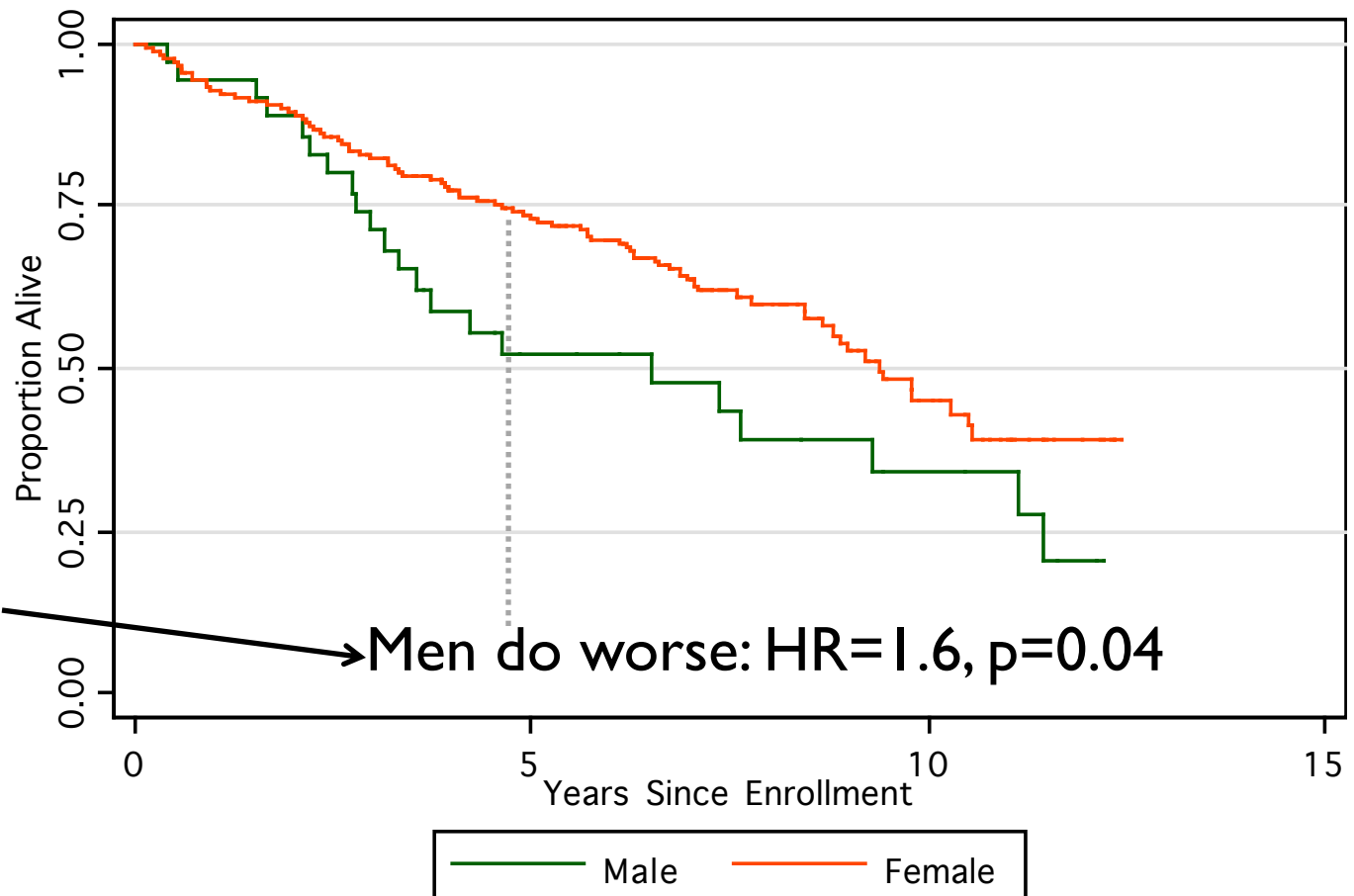
(unadjusted comparison)

```
use pbc.dat
```

```
stset years, failure(status)
```

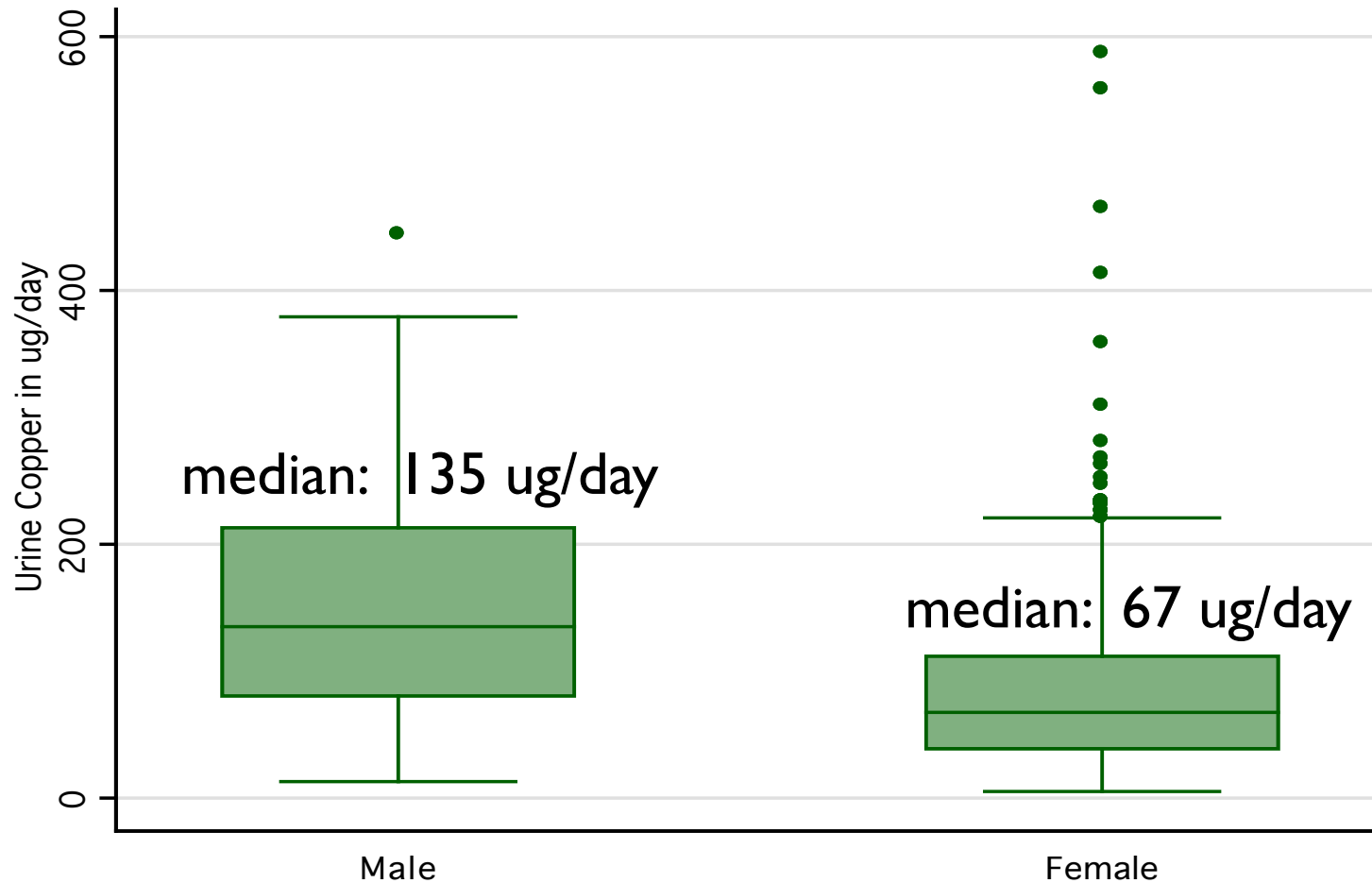
```
sts graph, by(sex)
```

stcox sex
(Cox model fit)



Men: 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 adjusted survival prediction based on a Cox model

Under the Cox Model

Consider $S(t|x)$: survival function (event-free proportion at time t) for someone with predictors x

$S_0(t)$:= baseline survival function

= survival function when all predictors equal zero

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

(β 's are the coefficients from the Cox model)

- In Cox model we see estimates of $\exp(\beta_p)$
- In background, Stata calculates estimates of $S_0(t)$

Adjusted Curves I

- Look at effect of x_1 (sex) adjusting for x_2 (copper)
- Create two curves with same value for x_2 (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 start with overall mean...

Adjusted Curves 2

```
. 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 copper=97.6) at2(sex=1 copper=97.6)
```

stcurve: gives predicted curves

survival: graph survival (not hazard default)

at1: (predictor values for curve 1)

at2: (predictor values for curve 2)

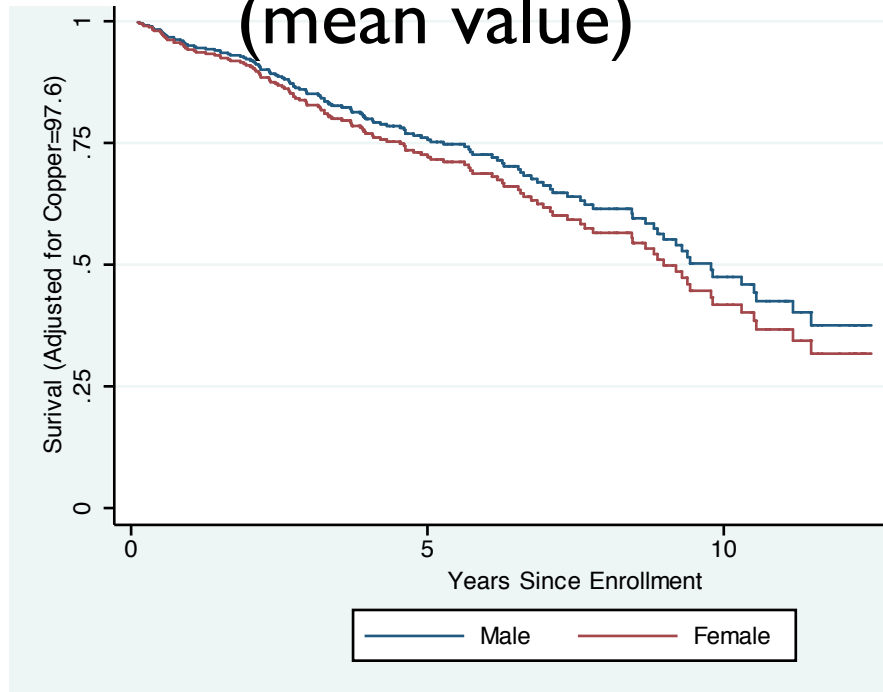
Note that the copper default is fixed at overall mean(=97.6)

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

Adjusted Curves 3

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

copper set to 97.6
(mean value)

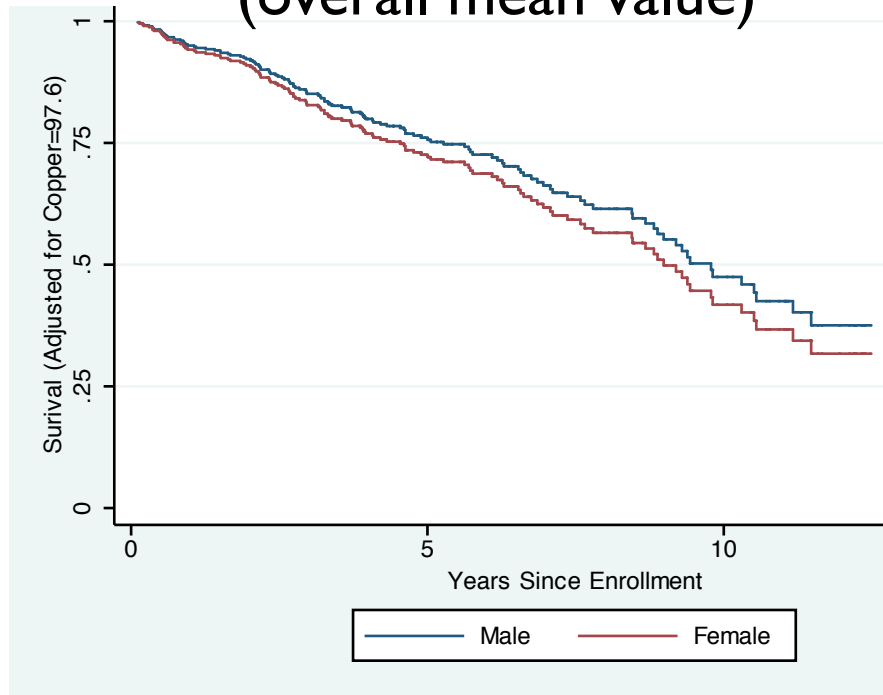


Adjusted Curves 4

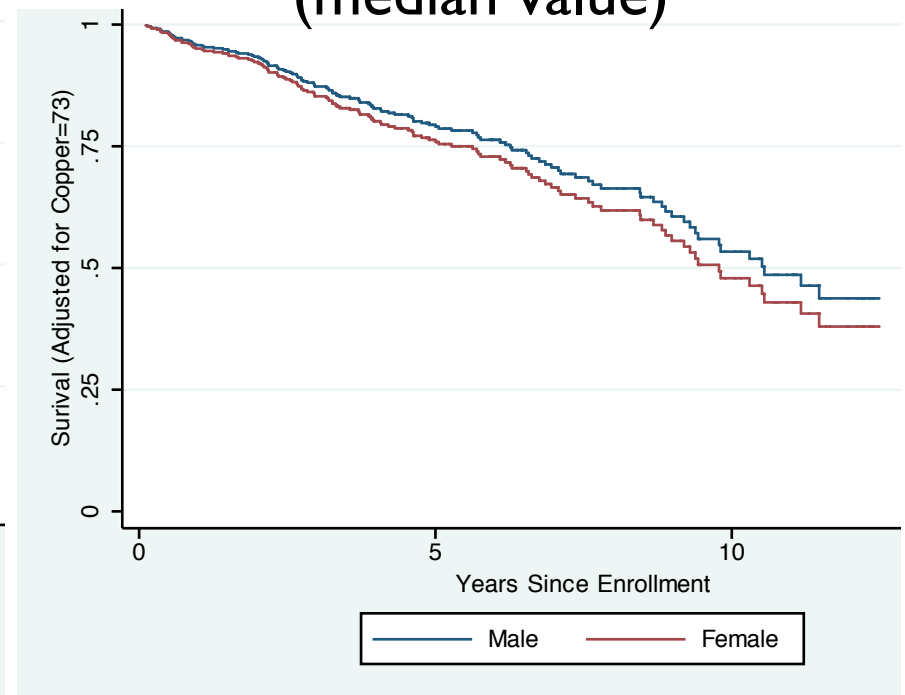
Let's compare mean vs. median copper as reference

```
stcurve, survival at1(sex=0 copper=97.6) at2(sex=1 copper=97.6)  
stcurve, survival at1(sex=0 copper=73) at2(sex=1 copper=73)
```

copper set to 97.6
(overall mean value)



copper set to 73
(median value)



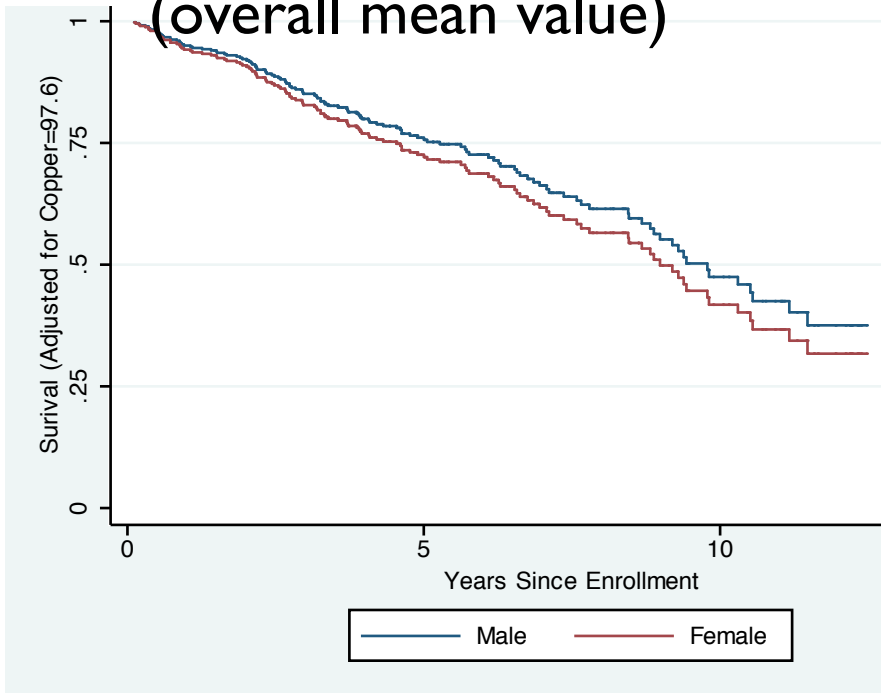
reference value for copper matters

Adjusted Curves 5

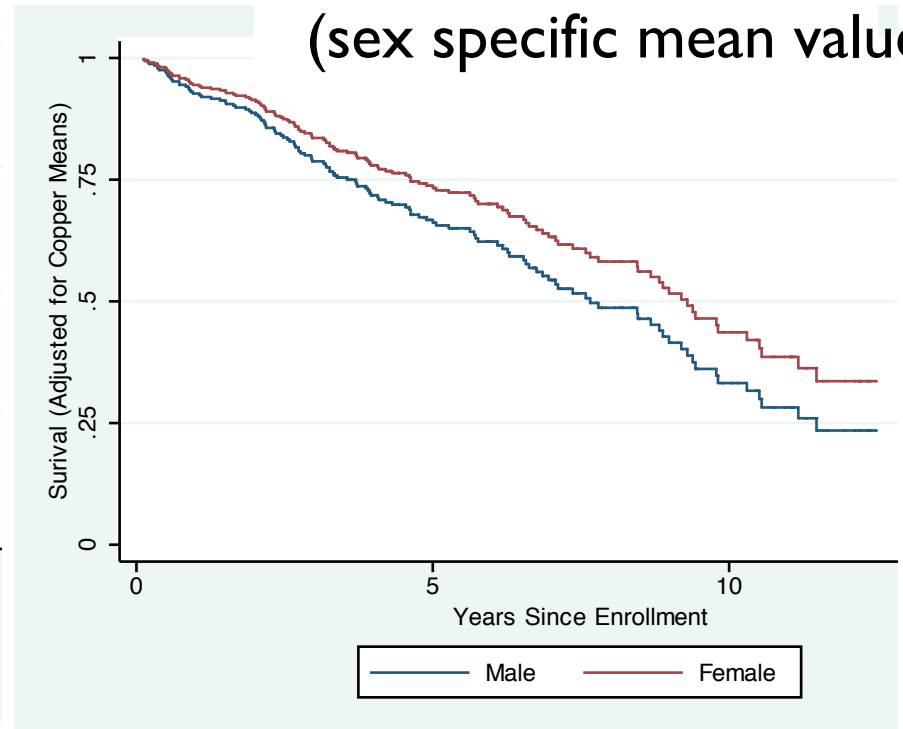
Now compare overall vs. group specific copper means as reference

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

- 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
- In Stata, `stcurve` is a flexible tool for creating adjusted survival curves
- Summary: Look at survival curves with `stcurve` while fixing other variables in the model

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

Don't forget...

- Project description and date limitations due today
- HW 1 due by start of lecture 4/13/21
- Lab 2 (available on syllabus page) session on Thursday

Causal Mediation

Lange and Hansen (2011) suggested using Aalen's additive hazard model (X: exposure, Z: covariate, M: mediator)

$$\lambda(t|X, Z, M) = \lambda_0(t) + \beta_1(t)X + \beta_2(t)Z + \theta(t)M$$

$$M = \alpha_0 + \alpha_1 x + \alpha_2 x + e$$

$e \sim$ normal random variable

Natural direct effect: $\beta_1(t)(x - x^*)$

Natural indirect effect: $\theta(t)\alpha_1(x - x^*)$

STATA: *stlh* command