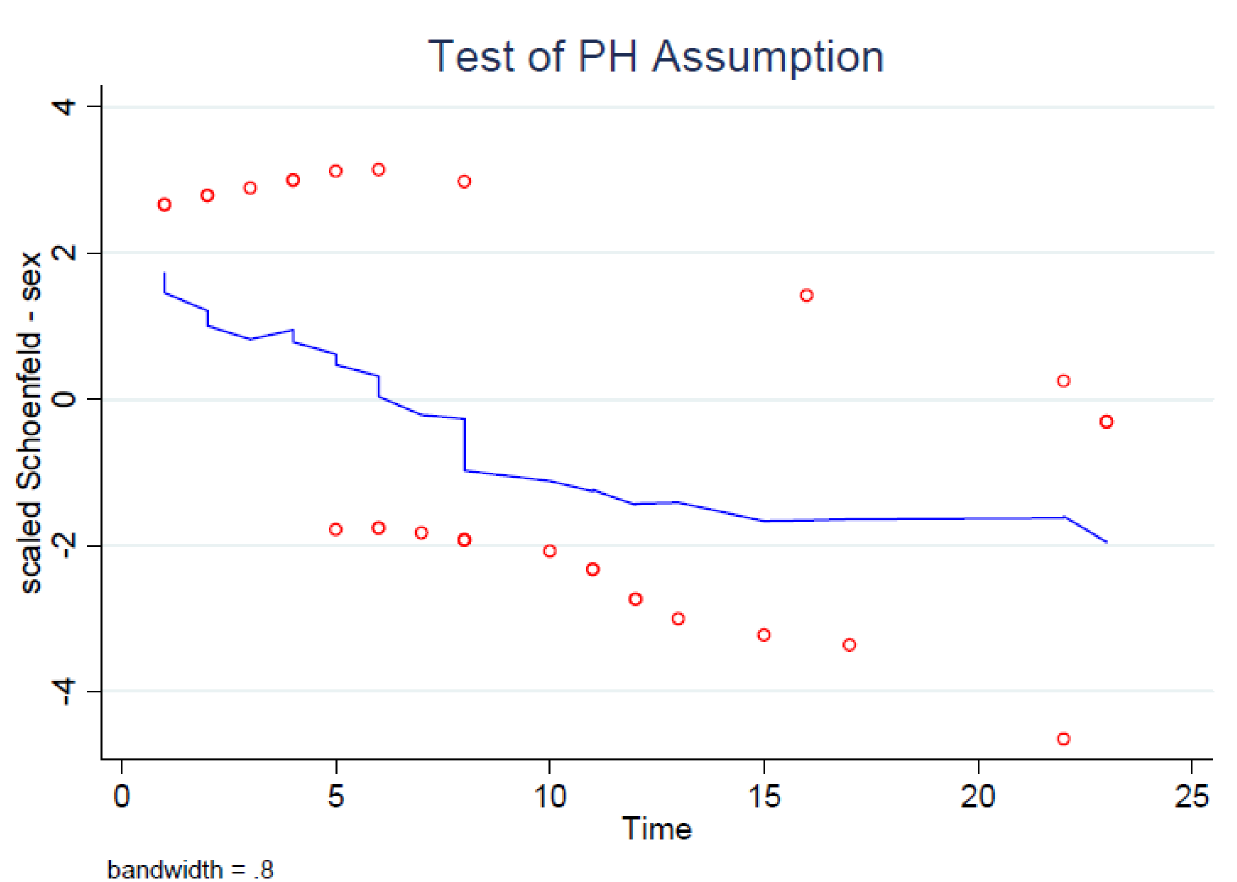
Example: Remission Duration

```
. stcox sex, nohr efron

         failure _d:  status
   analysis time _t:  time

Cox regression -- Efron method for ties

No. of subjects =           42                     Number of obs   =        42
No. of failures =           30
Time at risk    =          541
                                                   LR chi2(1)      =      0.60
Log likelihood  =   -92.885605                     Prob > chi2     =    0.4396

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.3101396   .4042745    -0.77   0.443    -1.102503    .4822238
------------------------------------------------------------------------------

. estat phtest

      Test of proportional-hazards assumption

      Time:  Time
      ----------------------------------------------------------------
                  |                     chi2       df       Prob>chi2
      ------------+---------------------------------------------------
      global test |                    10.90        1         0.0010
      ----------------------------------------------------------------
```

Example: Remission Duration

```
. estat phtest, plot(sex)
```

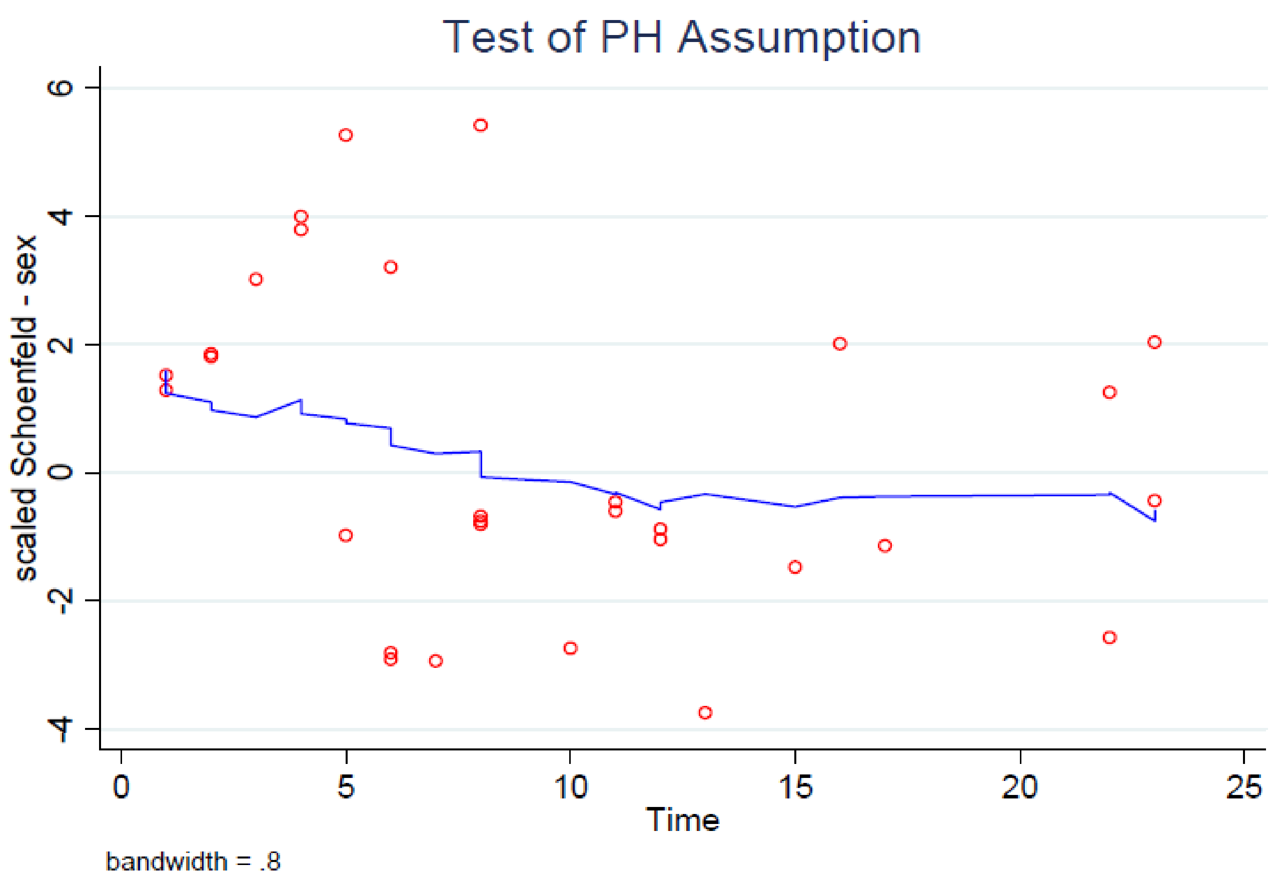


Example: Remission Duration

```
. stcox tx logwbc3 sex, nohr efron

         failure _d:  status
   analysis time _t:  time

Cox regression -- Efron method for ties

No. of subjects =           42                 Number of obs   =         42
No. of failures =           30
Time at risk    =          541
                                                LR chi2(3)      =      47.19
Log likelihood  =   -69.590483                 Prob > chi2     =     0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |  -1.503591   .4615127    -3.26   0.001    -2.408139   -.5990429
     logwbc3 |   1.681942   .3365836     5.00   0.000     1.022251    2.341634
         sex |    .314678   .4545115     0.69   0.489    -.5761482    1.205504
------------------------------------------------------------------------------

. estat phtest, detail

      Test of proportional-hazards assumption

      Time:  Time
      ----------------------------------------------------------------
                  |       rho            chi2       df       Prob>chi2
      ------------+---------------------------------------------------
      tx          |      0.05700         0.11        1         0.7425
      logwbc3     |      0.06864         0.21        1         0.6438
      sex         |     -0.29209         2.56        1         0.1094
      ------------+---------------------------------------------------
      global test |                      2.82        3         0.4206
      ----------------------------------------------------------------
```

Example: Remission Duration

```
. estat phtest, plot(sex)
```



Example: Remission Duration

Is the relationship with logwbc linear?

```
. gen lwbc2=logwbc3^2

. stcox group logwbc3 lwbc2, nohr

LR chi2(3)      =     44.50
Log likelihood  =    -71.73582                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -1.366605   .4303963    -3.18   0.001    -2.210167   -.5230442
     logwbc3 |   1.510339   .3221063     4.69   0.000     .8790224    2.141656
       lwbc2 |   .2710913   .2558792     1.06   0.289    -.2304227    .7726052
------------------------------------------------------------------------------
```

Model Checking Summary
Cox PH Model

Assumption	Technique	Stata	Remarks
PH	-log-log plot	`stphplot`	graphical, cat. covars only
	obs v pred KM	`stcoxkm`	graphical, cat. covars only
	H_0: $\beta(t) = \beta$	`estat phtest`	test each cov., any kind of cov.

What to do if PH fails?

Time dependent covariates:

- Interaction between covariate and t or log t
- smoothed estimates of $\beta(t)$
- See Biost 537 and other advanced courses

Stratification:

- Discrete (grouped) covariates only
- Separate baseline hazard for each covariate class
- graphical analysis – no quantitative estimate of HR

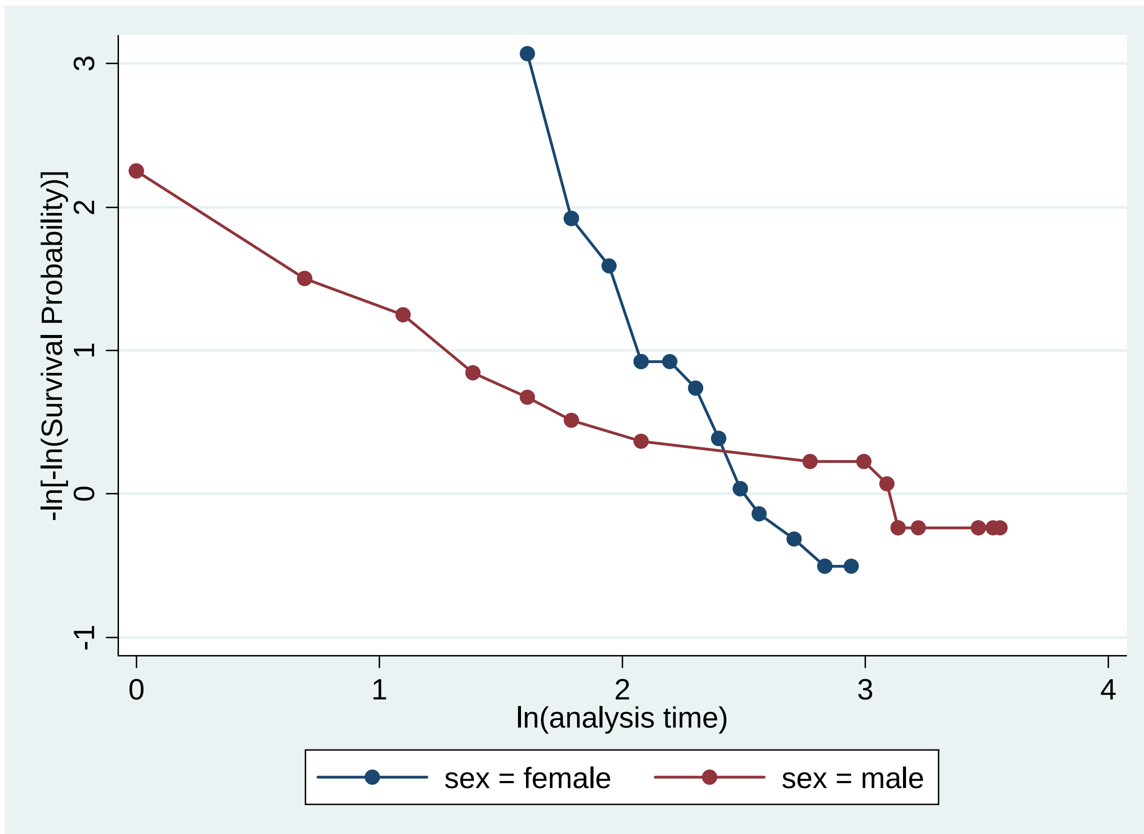
Stratified Cox Model

Remission duration study also includes information on gender

```
. tab sex

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
     female |         22       52.38       52.38
       male |         20       47.62      100.00
------------+-----------------------------------
      Total |         42      100.00

. stcox group logwbc3 sex, nohr efron scaledsch(SCA*)

LR chi2(3)      =     47.19
Log likelihood  =   -69.590483                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group | -1.503591   .4615127    -3.26   0.001    -2.408139   -.5990429
     logwbc3 |  1.681942   .3365836     5.00   0.000     1.022251    2.341634
         sex |   .314678   .4545115     0.69   0.489    -.5761482    1.205504
------------------------------------------------------------------------------
. estat phtest, detail log

      Test of proportional-hazards assumption

      Time:  Log(t)
      ----------------------------------------------------------------
                  |       rho            chi2       df       Prob>chi2
      ------------+---------------------------------------------------
      group       |      0.13307         0.59        1         0.4431
      logwbc3     |      0.08070         0.30        1         0.5867
      sex         |     -0.36205         3.94        1         0.0472
      ------------+---------------------------------------------------
```

Stratified Cox Model

```
. stphplot, by(sex)
```



Stratified Cox Model

Q: Evidence is against PH assumption for sex (though less strong in multivariate model). If we don't accept PH assumption on *sex*, can we still make PH inference about *group* and *logwbc* even if *sex* does not satisfy the PH assumption?

A: Yes. In order to do this we can perform a "stratified" analysis. This is different from using dummy variables, and is different from using separate analyses by gender.

Stratified Cox Model

Idea:

We can use a model where, within each gender, we have the same PH model, but we allow men and women to have *different baseline hazards*:

$$\text{women} : h(t, X) = h_{0,0}(t)\exp(\beta_1 \cdot \text{group} + \beta_2 \cdot \text{logwbc3})$$

$$\text{men} : h(t, X) = h_{0,1}(t)\exp(\beta_1 \cdot \text{group} + \beta_2 \cdot \text{logwbc3})$$

Such a model is said to be "stratified on sex". It is analogous to allowing interactions between sex and time, but we still have additive terms for the other covariates (no dependence on time).

Stratified Cox Model

Proportional Hazards Model:

$$\log[\mathrm{h}(\mathrm{t},\mathrm{X})] = \log[h_0(t)] + \beta_1 \mathrm{group} + \beta_2 \mathrm{logwbc3} + \beta_3 \mathrm{sex}$$

Stratified Cox Model:

$$\log[\mathrm{h}(\mathrm{t},\mathrm{X})] = \log[h_{0,sex}(t)] + \beta_1 \mathrm{group} + \beta_2 \mathrm{logwbc3}$$

Q: What's the interpretation of β_1 in each model?

Stratified Cox Model

Proportional Hazards:

$$F: h(t, X) = h_0(t)\exp(\beta_1\ \text{group} + \beta_2\ \text{logwbc3})$$

$$M: h(t, X) = h_0(t)\exp(\beta_1\ \text{group} + \beta_2\ \text{logwbc3} + \beta_3)$$

Separate Models:

$$F: h(t, X) = h_{0,1}(t)\exp(\beta_1^{(1)}\text{group} + \beta_2^{(1)}\text{logwbc3})$$

$$M: h(t, X) = h_{0,2}(t)\exp(\beta_1^{(2)}\text{group} + \beta_2^{(2)}\text{logwbc3})$$

Stratified Model #1:

$$F: h(t, X) = h_{0,F}(t)\exp(\beta_1\text{group} + \beta_2\text{logwbc3})$$

$$M: h(t, X) = h_{0,M}(t)\exp(\beta_1\text{group} + \beta_2\text{logwbc3})$$

Stratified Cox Model

Stratified Model #2:

$$\mathrm{h(t,X)} = h_{0,sex}(t)\,\mathrm{exp}(\beta_1\mathrm{group} + \beta_2\mathrm{logwbc3} + \beta_3\mathrm{group} \times \mathrm{sex})$$

$$\mathrm{F : h(t,X)} = h_{0,F}(t)\,\mathrm{exp}(\beta_1\mathrm{group} + \beta_2\mathrm{logwbc3})$$

$$\mathrm{M : h(t,X)} = h_{0,M}(t)\,\mathrm{exp}(\beta_1\mathrm{group} + \beta_2\mathrm{logwbc3} + \beta_3\mathrm{group})$$

Example: Remission Duration

Proportional hazards model

```
. stcox group logwbc3 sex, efron nohr

No. of subjects =           42                     Number of obs   =        42
No. of failures =           30
Time at risk    =          541
                                                   LR chi2(3)      =     47.19
Log likelihood  =   -69.590483                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group | -1.503591   .4615127    -3.26   0.001    -2.408139   -.5990429
     logwbc3 |   1.681942   .3365836     5.00   0.000     1.022251    2.341634
         sex |    .314678   .4545115     0.69   0.489    -.5761482    1.205504
------------------------------------------------------------------------------
```

Example: Remission Duration

Separate Models

Females

```
. stcox group logwbc3 if sex==0, nohr efron

                                                  LR chi2(2)      =      6.65
Log likelihood  =   -33.090979                    Prob > chi2     =    0.0361

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group | -.3112706   .5635539    -0.55   0.581    -1.415816    .7932747
     logwbc3 |  1.206146   .5034893     2.40   0.017     .2193255    2.192967
------------------------------------------------------------------------------
```

Males

```
. stcox group logwbc3 if sex==1, nohr efron

                                                  LR chi2(2)      =     29.18
Log likelihood  =   -20.760908                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group | -1.977887    .739202    -2.68   0.007    -3.426697   -.5290782
     logwbc3 |  1.742777   .5357723     3.25   0.001     .6926825    2.792871
------------------------------------------------------------------------------
```

Stratified Cox Model – Remission Duration

Stratified #1

```
. stcox group logwbc3, strata(sex) efron nohr bases(S) basech(H)
                                                LR chi2(2)      =     32.06
Log likelihood  =   -55.734815                  Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.9981037   .4735546    -2.11   0.035    -1.926254   -.0699538
     logwbc3 |   1.453654   .3440687     4.22   0.000     .7792919    2.128017
------------------------------------------------------------------------------
                                                             Stratified by sex
. estimates store model1
```

Stratified #2

```
. gen txsex=group*sex
. stcox group logwbc3 txsex, strata(sex) nohr efron
                                                LR chi2(3)      =     35.28
Log likelihood  =   -54.126889                  Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.2865729   .5685327    -0.50   0.614    -1.400877    .8277308
     logwbc3 |   1.472627   .3517843     4.19   0.000     .7831426    2.162112
       txsex |  -1.642102   .9140899    -1.80   0.072    -3.433685    .1494813
------------------------------------------------------------------------------
                                                             Stratified by sex
. est stor model2
. lrtest model1 model2

Likelihood-ratio test                                 LR chi2(1)  =      3.22
(Assumption: model1 nested in model2)                 Prob > chi2 =    0.0729
```

Time Varying Covariates

- Stratification allows us to make inferences about everything except the stratification variable. In effect, we have adjusted away the effect of sex.
- Suppose we wanted to make inferences on sex. Is it possible to model the time dependent HR?
- Yes! Include an interaction between sex and time.

Note: A different, but related, situation arises when we have a covariate for which we have measured changes over time e.g. sexual behavior, CD4 count

Time Varying Covariates

Time-independent model:

$$\log h(t;X) = \log h_0(t) + \beta_1 sex$$

Time-dependent model:

$$\log h(t;X) = \log h_0(t) + \beta_1 sex + \beta_2 sex * t$$

(note: no main effect for t – it is part of log h_0(t))

What is the HR for sex?

$$\frac{h(t;sex=1)}{h(t;sex=0)} = \frac{h_0(t)\exp(\beta_1 + \beta_2 t)}{h_0(t)\exp(0)}$$

$$= \exp(\beta_1 + \beta_2 t)$$

β_2 positive $\Rightarrow$ HR increases over time

β_2 negative $\Rightarrow$ HR decreases over time

Stratified Cox Model – Remission Data

```
. stcox tx logwbc sex, nohr efron tvc(sex) texp(_t)

Cox regression -- Efron method for ties

No. of subjects =           42                     Number of obs   =        42
No. of failures =           30
Time at risk    =          541
                                                   LR chi2(4)      =     51.09
Log likelihood  =   -67.641012                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
main         |
          tx | -1.186389   .4828406    -2.46   0.014    -2.132739   -.2400388
      logwbc |  1.535217   .3405832     4.51   0.000     .8676864    2.202748
         sex |  1.727716   .8966549     1.93   0.054    -.0296948    3.485128
-------------+----------------------------------------------------------------
tvc          |
         sex | -.1655309   .0904991    -1.83   0.067    -.3429058     .011844
------------------------------------------------------------------------------
Note: variables in tvc equation interacted with _t
```

Stratified Cox Model – Remission Data

$Hr_{sex}(t = 5) =$

```
lincom [main]sex + 5*[tvc]sex, hr
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
(1)	2.459755	1.397205	1.58	0.113	.8079549	7.488532

$Hr_{sex}(t = 25) =$

```
lincom [main]sex + 25*[tvc]sex, hr
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
(1)	.0897653	.1417142	-1.53	0.127	.0040673	1.981143

Overall LR test for sex:

```
stcox tx logwbc, nohr efron
estimates store model0
stcox tx logwbc sex, nohr efron tvc(sex)
estimates store model1

lrtest model0 model1

Likelihood-ratio test                                 LR chi2(2)  =      4.37
(Assumption: model0 nested in model1)                 Prob > chi2 =    0.1122
```

Survival Analysis and Sample Size

Q: What are the considerations for determining the sample size necessary when the study endpoint is a time-until-event?

Planned Analysis:

- Assessment of percent surviving beyond *t**.
  - Need to know α, power, S(*t**;X=0), S(*t**;X=1)
  - Comparison of proportions (see STATA sampsi!)
- Assessment of survival function and/or hazard ratio.
  - Log-rank / Cox regression.
  - Power depends on events only! $L = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{\log(HR)/2} \right)^2$
    - *L* is number of events (i.e. deaths)
    - *HR* is relative hazard.
  - N/arm based on expected event rate, duration of follow-up and *HR*.

Sample Size Example

HPTN 039

Assume:

- 2 treatment arms with *N* subjects each
- $h_0 = 0.035$/yr with $HR = 0.5$
- $f = 1$ year follow-up
- $\alpha = 0.05$, Power $= 1 - \beta = 0.90$

Compute:

- $L = [(1.96+1.28)/(-0.3466)]^2 =$ **87.4**
- $N = L/[\, h_0 * f * (1 + HR)\,] = 87.4/[(0.035)*(1)*(1.5)] =$ **1665**
- Inflate *N* further for expected lost to follow-up
- Compare to 1865/arm (sampsi) for $p_0 = .035$ to $p_1 = 0.0175$ at 1 year.
- Less censoring >> greater savings in sample size