Class 7 Outline

1. Review of Cox proportional hazards regression model using a detailed example

2. Risk Factors for Infant Mortality in Sarlahi District, Nepal: A Survival Analysis of Data from the Nepal Nutrition Intervention Project (NNIPS-II)

1

0. Learning Objectives

- Explore a large survival data set, identifying data limitations, and data patterns useful in guiding decisions about a statistical analysis
- Use Kaplan-Meier survival curve estimates and crude survival rates to study dependence of survival on discrete characteristics of the mother and pregnancy
- Create predictor variables as candidates for survival regression models
- Choose, fit, and check Cox proportional hazards models using time independent predictor variables

2

1. Scientific Background

- NNIPS-II (Nepal Nutrition Intervention Project Sarlahi- II) was a randomized study of 15,987 infants born alive to 43,559 women who received Vitamin A, Beta-carotene, or a placebo prior to and during pregnancy

3

1.1 Objectives of Analysis

- Investigate effect of treatment on infant mortality
- Identify maternal and pregnancy characteristics that predict infant mortality
- Estimate shape of infant mortality rate curve as a function of gestational age and parity
- Estimate multivariable Cox model to predict survival as a function of treatment and other mother/child characteristics such as gestational age and parity

4

1.2 Data Exploration and "Cleaning"

- We are using a representative subset of 10,295 live births from the original data set
- Women/infant pairs followed for up to 6 months (180 days) post delivery

5

1.3a Key Variables

- Survival time (*stime*) in days
- Censoring indicator (*cens*) : 1 if death, 0 if censored
- Mother's parity (*parity*) : number of previous live births

6

1.3b Key Variables

- Night blindness (*nblind*): 1 if mother night blind during pregnancy; 0 if not
- Gestational Age (*gestage*): number of weeks from last completed menstrual cycle to birth
- Treatment (*treat*): 1 if Beta-carotene, 2 if placebo, 3 if Vitamin A
- Gender (*male*): 1 if male, 0 female

7

1.4a Codebook, Key Variables

```
. codebook stime cens
stime ------------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [0,918]                      units:  1
         unique values:  317                  coded missing:  0 / 10295

                  mean:    165.52
             std. dev:    58.2857

           percentiles:        10%       25%       50%       75%       90%
                               88       170       173       181       204


cens -------------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [0,1]                        units:  1
         unique values:  2                    coded missing:  0 / 10295

            tabulation:  Freq.  Value
                         9651  0
                          644  1
```

8

1.4b Codebook, Key Variables

```
. codebook parity nblind

parity ---------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [0,15]                       units:  1
         unique values:  15                   coded missing:  256 / 10295

                  mean:   2.32563
             std. dev:   2.14178

           percentiles:        10%       25%       50%       75%       90%
                                0         1         2         3         5


nblind ---------------------------------------------------------- (unlabeled)
                  type:  numeric (float)
                 label:  nb

                 range:  [0,1]                        units:  1
         unique values:  2                    coded missing:  0 / 10295

            tabulation:  Freq.   Numeric  Label
                         9372         0  No night blind
                          923         1  Night blind
```


1.4c Codebook, Key Variables

```
. codebook gestage treat

gestage --------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [28,46]                      units:  1
         unique values:  19                   coded missing:  612 / 10295

                  mean:    38.017
             std. dev:    3.7041

           percentiles:        10%       25%       50%       75%       90%
                               33        36        38        40        42


treat ----------------------------------------------------------- (unlabeled)
                  type:  numeric (float)
                 label:  allocl

                 range:  [1,3]                        units:  1
         unique values:  3                    coded missing:  0 / 10295

            tabulation:  Freq.   Numeric  Label
                         3265         1  Beta C
                         3387         2  Placebo
                         3643         3  Vit A
```

1.4d Codebook, Key Variables

```
. codebook male

male ------------------------------------------------------------- (unlabeled)
                  type:  numeric (float)
                 label:  sex

                 range:  [0,9]                        units:  1
         unique values:  3                    coded missing:  0 / 10295

            tabulation:  Freq.   Numeric  Label
                          4966         0  female
                          5195         1  male
                           134         9  Missing
```


2.0 Create “Survival” Data Set

```
. stset  stime, failure(cens=1)

     failure event:  cens == 1
obs. time interval:  (0, stime]
 exit on or before:  failure

--------------------------------------------------------------------------------
    10295  total obs.
      518  obs. end on or before enter()
--------------------------------------------------------------------------------
     9777  obs. remaining, representing
      559  failures in single record/single failure data
  1704029  total analysis time at risk, at risk from t =         0
                              earliest observed entry t =         0
                                   last observed exit t =       918
```

2.1a Why have Observations been Dropped?

- 518 observations have just been dropped by Stata because survival time (*stime*) recorded as 0 days

```
. tab cens if stime == 0

       cens |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        384       74.13       74.13
          1 |        134       25.87      100.00
------------+-----------------------------------
      Total |        518      100.00
```

- Infants who died on day 1

13

2.1b Why have Observations been Dropped?

- These 518 observations give information about survival experience!
- Remedy: add 1 to all survival times so that Stata will use these observations in its computations

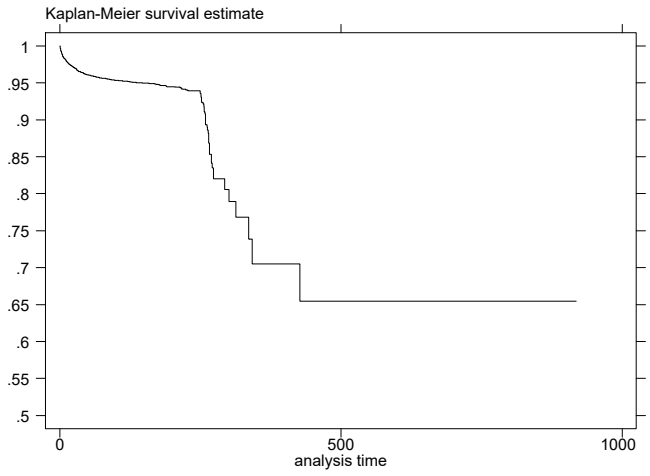
14

2.2 Remedy for Dropped Observations

```
. replace stime = stime + 1
(10295 real changes made)

. stset  stime, failure(cens=1)

     failure event:  cens == 1
obs. time interval:  (0, stime]
 exit on or before:  failure

--------------------------------------------------------------------------------
    10295  total obs.
        0  exclusions
--------------------------------------------------------------------------------
    10295  obs. remaining, representing
      693  failures in single record/single failure data
  1714324  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =       919
```

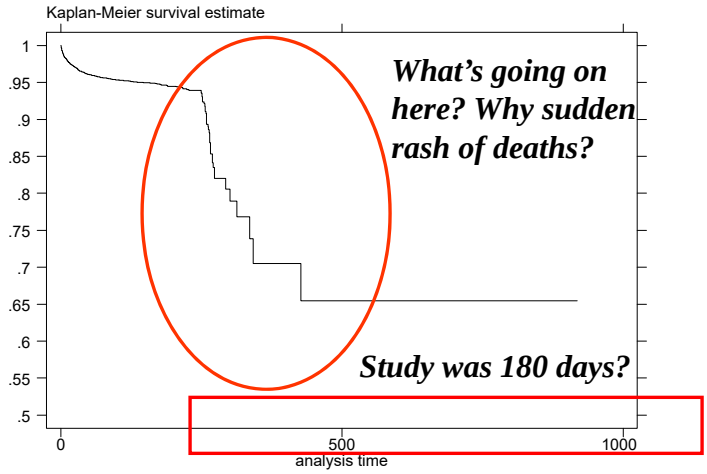

3.0a Overall View of Infant Survival

```
. sts graph, ylab(.85 (.05) 1)
```



3.0b Overall View of Infant Survival

```
. sts graph, ylab(.85 (.05) 1)
```



17

3.1 Checking Data Set, Again

```
. codebook stime


stime ------------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [1,919]                      units:  1
         unique values:  317                  coded missing:  0 / 10295

                  mean:    166.52
             std. dev:    58.2857

           percentiles:        10%       25%       50%       75%       90%
                                89       171       174       182       205
. tab cens if stime > 180, summ(stime)

           |         Summary of stime
      cens |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         0 |   210.18326   39.320311        2641
         1 |   260.28571   38.905334          49
-----------+------------------------------------
     Total |   211.09591   39.872823        2690
```

18

3.2a Follow-up With PIs

- According to the PIs, some women were followed beyond 180 days in “passive surveillance”
- Formal study (“active surveillance”) ended at 180 days even though additional data was collected
- Selection bias in using time to event information for this subset of women

19

3.2b Follow-up With PIs

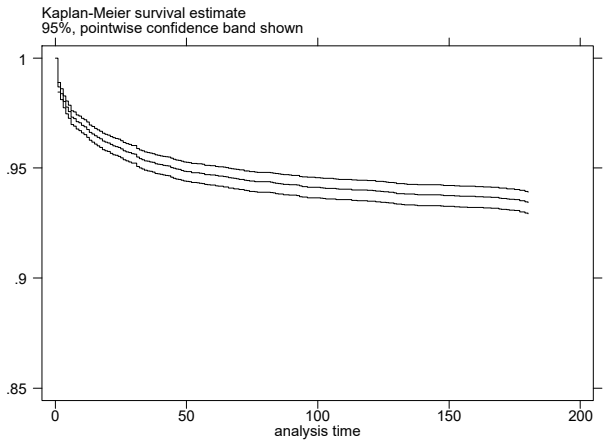
- “Remedy” – censor all observations at 180 days

```
. replace cens = 0 if stime > 180
(49 real changes made)

. replace stime = 180 if stime > 180
(2951 real changes made)

. stset stime, failure(cens=1)

     failure event:  cens == 1
obs. time interval:  (0, stime]
 exit on or before:  failure

--------------------------------------------------------------------------------
    10295  total obs.
        0  exclusions
--------------------------------------------------------------------------------
    10295  obs. remaining, representing
      644  failures in single record/single failure data
  1627725  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =       180
```

20

3.3 Kaplan-Meier Estimates For Analytic Data Set

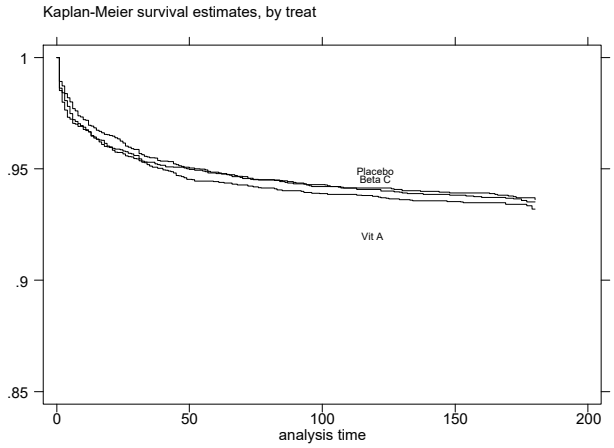
```
. sts graph, ylab(.85 (.05) 1) gwood
```



21

4.0 KM Curves, By Treatment Groups

```
. sts graph,by(treat) ylab(.85 (.05) 1)
```



22

4.1 Incidence Rates, By Treatment Groups

```
. stsum, by(treat)

        failure _d:  cens == 1
  analysis time _t:  stime

        |              incidence       no. of    |------ Survival time -----|
treat   | time at risk     rate        subjects        25%       50%       75%
--------+-------------------------------------------------------------------
  Beta C |       516692   .0003929          3265          .         .         .
 Placebo |       532438    .000385          3387          .         .         .
   Vit A |       578595   .0004079          3643          .         .         .
--------+-------------------------------------------------------------------
   total |      1627725   .0003956         10295          .         .         .
```

23

4.2 Hazard Ratios for Treatment Groups

```
. stcox i.treat
i.treat           _Itreat_1-3         (naturally coded; _Itreat_1 omitted)
        failure _d:  cens == 1
  analysis time _t:  stime
Iteration 0:   log likelihood = -5902.8126
Iteration 1:   log likelihood = -5902.5964
Iteration 2:   log likelihood = -5902.5964
Refining estimates:
Iteration 0:   log likelihood = -5902.5964
Cox regression -- Breslow method for ties
No. of subjects =        10295                     Number of obs   =     10295
No. of failures =          644
Time at risk    =      1627725
                                                   LR chi2(2)      =      0.43
Log likelihood  =   -5902.5964                     Prob > chi2     =    0.8056
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_2 |   .9770902   .0967476    -0.23   0.815     .8047334    1.186362
   _Itreat_3 |   1.039412   .0994985     0.40   0.686     .8616002     1.25392
```

24

5.0 Gestational Age

First, create some categories for *gestage*

```
.gen ga_cat=1 if gestage < 36 & gestage ~=.
.replace ga_cat=2 if gestage >=36 & gestgage < 38
.replace ga_cat=3 if gestage >=38 & gestgage < 39
.replace ga_cat=4 if gestage >=39 & gestgage < 41
.replace ga_cat=5 if gestage >=41

.label define gestcat 1 “< 36 wks” 2 “36-38 weeks” 3 “38-39
   weeks” 4 “39-41 weeks” 5 “41+ weeks”

.label values ga_cat gestcat
```

25

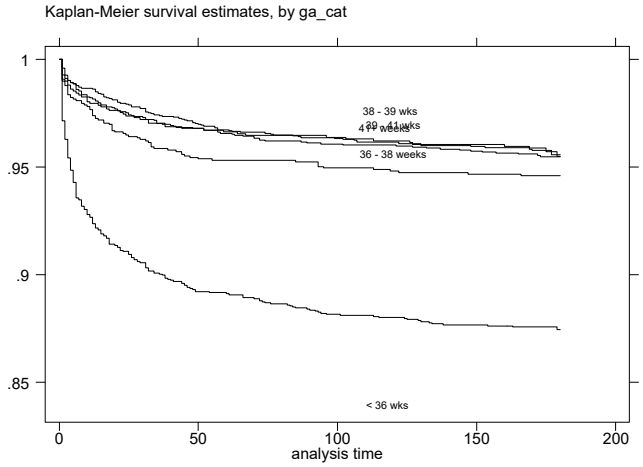
5.1 Distribution of Gestational Age Categories

```
. tab ga_cat

       ga_cat |      Freq.     Percent        Cum.
--------------+-----------------------------------
     < 36 wks |       2176       22.47       22.47
36 - 38 weeks |       1493       15.42       37.89
  38 - 39 wks |       1245       12.86       50.75
  39 - 41 wks |       2499       25.81       76.56
    41+ weeks |       2270       23.44      100.00
--------------+-----------------------------------
        Total |       9683      100.00
```

26

5.2 Kaplan-Meier Curves by Gestational Age Categories

```
. sts graph,by(ga_cat) ylab(.85 (.05) 1)
```



5.3 Incidence Rates, By Gestational Age Categories

```
. stsum, by(ga_cat)

        failure _d:  cens == 1
  analysis time _t:  stime

         |              incidence       no. of    |------ Survival time -----|
ga_cat   | time at risk     rate        subjects        25%       50%       75%
---------+--------------------------------------------------------------------
 < 36 wk |       323590   .0008128          2176          .         .         .
 36 - 38 |       238819   .0003266          1493          .         .         .
 38 - 39 |       203632   .0002554          1245          .         .         .
 39 - 41 |       409594   .0002564          2499          .         .         .
 41+ wk |       367150   .0002696          2270          .         .         .
---------+--------------------------------------------------------------------
   total |      1542785    .000387          9683          .         .         .
```

5.4 Incidence Rates by Gestational Age (With Missing Values)

```
. stsum, by(ga_cat_miss)

        failure _d:  cens == 1
  analysis time _t:  stime

         |              incidence       no. of    |------ Survival time -----|
ga_cat~s | time at risk     rate        subjects        25%       50%       75%
---------+--------------------------------------------------------------------
 < 36 wk |       323590   .0008128          2176          .         .         .
 36 - 38 |       238819   .0003266          1493          .         .         .
 38 - 39 |       203632   .0002554          1245          .         .         .
 39 - 41 |       409594   .0002564          2499          .         .         .
  41+ wk |       367150   .0002696          2270          .         .         .
 Missing |        84940   .0005533           612          .         .         .
---------+--------------------------------------------------------------------
   total |      1627725   .0003956         10295          .         .         .
```

Note high mortality among those missing gestational age!

29

6.0 Mother's Parity Categories

First create categories for *parity*

```
.gen par_cat=0 if parity ==0
.replace par_cat=1 if parity==1
.replace par_cat=2 if parity >=2 & parity < =4
.replace par_cat=3 if parity >=5 & parity < =8
.replace par_cat=4 if parity > 8 & parity~=.


.label define par 0 "No prev child" 1 "1 prev child" 2 "2-4
   prev child" 3 "5-8 prev child" 4 "8+ prev child"

.label values par_cat par
```

30

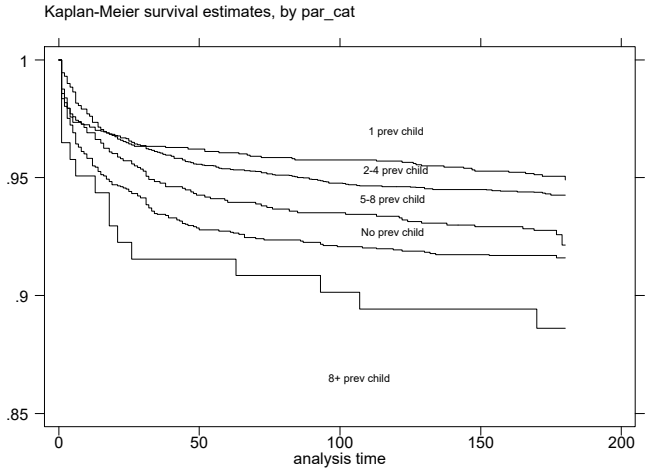
6.1 Distribution of Parity Categories

```
. tab par_cat

       par_cat |      Freq.     Percent        Cum.
---------------+-----------------------------------
 No prev child |       2254       22.45       22.45
  1 prev child |       2018       20.10       42.55
2-4 prev child |       4262       42.45       85.01
5-8 prev child |       1363       13.58       98.59
 8+ prev child |        142        1.41      100.00
---------------+-----------------------------------
         Total |      10039      100.00
```

31

6.2 Kaplan-Meier Curves by Parity Categories

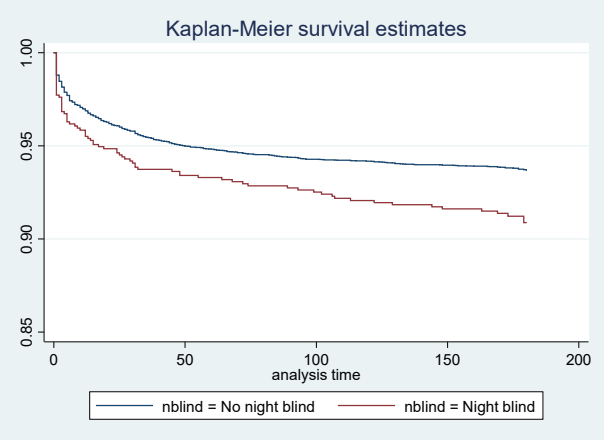
```
. sts graph,by(par_cat) ylab(.85 (.05) 1)
```



32

7.0 Kaplan-Meier Curves, By Mother's Night Blindness

```
. sts graph,by(nblind) ylab(.85 (.05) 1)
```



33

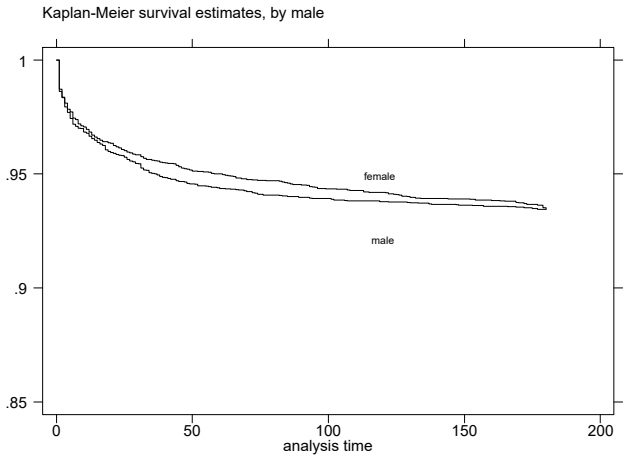
7.1 Incidence Rates, By Mother's Night Blindness

```
. stsum, by(nblind)

        failure _d:  cens == 1
  analysis time _t:  stime

         |              incidence       no. of    |------ Survival time -----|
nblind   | time at risk     rate        subjects       25%       50%       75%
---------+--------------------------------------------------------------------
No night |      1482249   .0003805          9372         .         .         .
Night bl |       145476   .0005499           923         .         .         .
---------+--------------------------------------------------------------------
   total |      1627725   .0003956         10295         .         .         .
```

34

8.0 Kaplan-Meier Curves, By Gender
(excluding missing)

```
. sts graph, by(male) ylab(.85 (.05) 1)
```



35

8.1 Incidence Rates, By Gender

```
. stsum, by(male)

         failure _d:  cens
   analysis time _t:  stime

         |              incidence       no. of    |------ Survival time -----|
male     | time at risk     rate        subjects        25%       50%       75%
---------+--------------------------------------------------------------------
  female |       840324   .0003689          4966          .         .         .
    male |       872832   .0003792          5195          .         .         .
---------+--------------------------------------------------------------------
   total |      1713156   .0003742         10161          .         .         .
```

36

9.0a Cox Proportional Hazards Model

- h(t,X) = hazard for people at risk with predictor values X = (X_1,X_2, …..X_p)
- We can write

$$h(t,X) = h_0(t) \cdot e^{\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p}$$

- Similarly,

  $\log[h(t,\mathbf{x})] = \log[h_o(t)] + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_p X_p$

37

9.0b Cox Proportional Hazards Model

- Hazard Ratio (relative hazard) interpretation of the β's

$$e^{\beta_1} = \frac{h(t,X_1+1,X_2,...X_p)}{h(t,X_1,X_2,...X_p)}$$

  = relative hazard by one unit increase in X_1, but have same values for X_2, …. X_p (at any fixed time t)

38

9.1 Interpretation of β's

- Why do we have a hazard ratio interpretation?

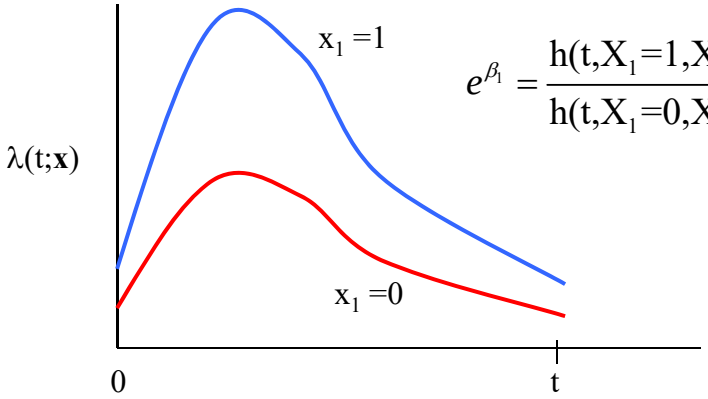
$\log[h(t,X_1+1,X_2,..X_p)] = \log[h_o(t)]+\beta_1(X_1-1)+\beta_2X_2+\ldots\beta_pX_p$
$- \log[h(t,X_1,X_2,..X_p)] = \log[h_o(t)]+\beta_1X_1+\beta_2X_2+\ldots\beta_pX_p$

$$\beta_1 = \log\left[\frac{h(t,X_1+1,X_2,...X_p)}{h(t,X_1,X_2,...X_p)}\right]$$

39

9.2 Cox Proportional Hazards Model Assumption

- Assumes proportional hazards over time

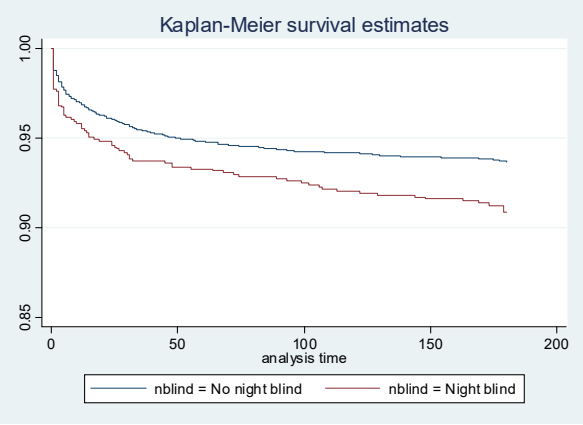


$$e^{\beta_1} = \frac{h(t,X_1=1,X_2,...X_p)}{h(t,X_1=0,X_2,...X_p)}$$

40

9.3 Example: Survival by Maternal Night Blindness

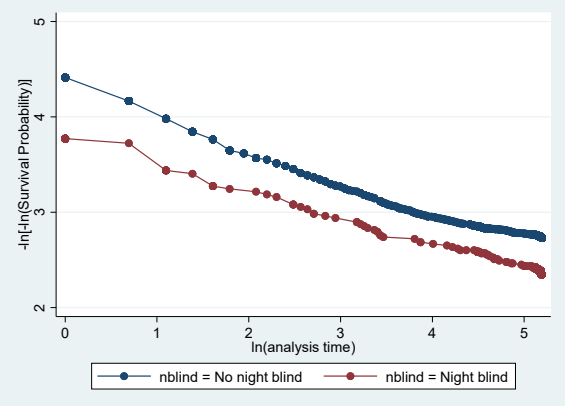
```
. sts graph, by(nblind) ylab(0.85(0.05)1)
```



41

9.4 Proportional Hazards Check

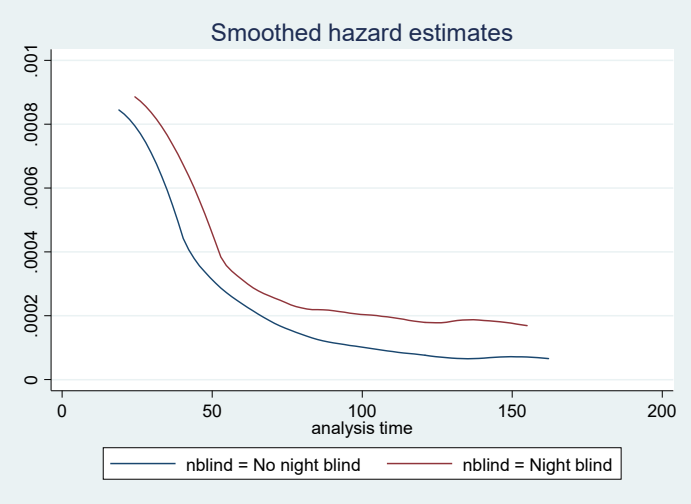
```
.stphplot, by(nblind)ylab(2(1) 5) xlab(0 (1) 5)
```



42

9.5 Graph of Estimated Hazards over Time

```
.sts graph, hazard by(nblind)
```



43

9.6 Corresponding Cox Estimate

```
. stcox nblind

        failure _d:  cens == 1
  analysis time _t:  stime

Iteration 0:   log likelihood = -5902.8126
Iteration 1:   log likelihood = -5898.7079
Iteration 2:   log likelihood = -5898.5873
Iteration 3:   log likelihood = -5898.5873
Refining estimates:
Iteration 0:   log likelihood = -5898.5873

Cox regression -- Breslow method for ties

No. of subjects =        10295                  Number of obs   =     10295
No. of failures =          644
Time at risk    =      1627725
                                                LR chi2(1)      =      8.45
Log likelihood  =   -5898.5873                  Prob > chi2     =    0.0036

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      nblind |   1.438342   .1718403     3.04   0.002     1.138068    1.817842
------------------------------------------------------------------------------
```

44

9.7 Interpretation

- Children of night-blind mothers are estimated to be at higher risk of death than children of non-night-blind mothers, averaged over the study duration
- The estimated relative hazard of death for children of night-blind versus non-night-blind mothers is 1.44 (95% CI 1.14 – 1.82) at any time in the 6-month follow-up period

45

10. Analysis Plan

- Using Cox Regression
  - First, look at gestational age and parity individually to choose a model form
  - Second, build a multivariable model for treatment including gestational age and parity, as well as other predictors (gender, mother's night blindness)

46

10.1 Gestational Age, Categorical

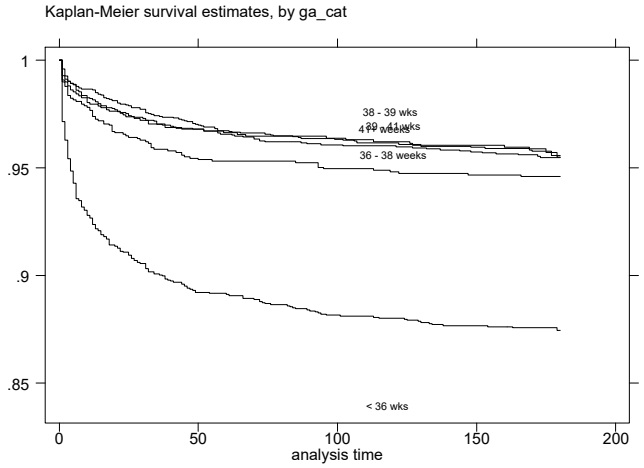
```
. tab ga_cat

       ga_cat |      Freq.     Percent        Cum.
--------------+-----------------------------------
     < 36 wks |       2176       22.47       22.47
36 - 38 weeks |       1493       15.42       37.89
  38 - 39 wks |       1245       12.86       50.75
  39 - 41 wks |       2499       25.81       76.56
    41+ weeks |       2270       23.44      100.00
--------------+-----------------------------------
        Total |       9683      100.00
```

Recall, total n = 10,295: 612 observations missing gestational age

47

10.2 Kaplan-Meier Curves by Gestational Age Categories

```
. sts graph, by(ga_cat) ylab(.85 (.05) 1)
```



48

10.3 Cox Results with Hazard Ratios (HRs)

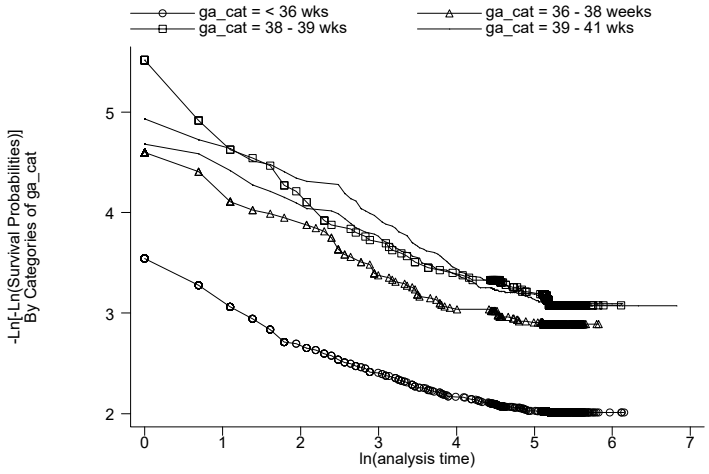
```
. stcox i.ga_cat
i.ga_cat          _Iga_cat_1-5       (naturally coded; _Iga_cat_1 omitted)

        failure _d:  cens
  analysis time _t:  stime

Iteration 0:   log likelihood = -5439.3955
Iteration 1:   log likelihood = -5368.3146
Iteration 2:   log likelihood = -5362.5375
Iteration 3:   log likelihood = -5362.5354
Refining estimates:
Iteration 0:   log likelihood = -5362.5354

Cox regression -- Breslow method for ties

No. of subjects =         9683                     Number of obs   =      9683
No. of failures =          597
Time at risk    =      1619695
                                                   LR chi2(4)      =    153.72
Log likelihood  =   -5362.5354                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |
          _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
36-38 wks _Iga_cat_2 |    .413508   .0533154    -6.85   0.000     .3211698     .532394
38-39 wks _Iga_cat_3 |   .3251237   .0493452    -7.40   0.000     .2414675    .4377626
39-41 wks _Iga_cat_4 |   .3264775   .0376917    -9.70   0.000      .260365    .4093776
 > 41 wks _Iga_cat_5 |   .3422158   .0403542    -9.09   0.000     .2715977    .4311952
------------------------------------------------------------------------------
```


10.4 Cox Results with log HRs

```
. stcox i.ga_cat, nohr
i.ga_cat          _Iga_cat_1-5       (naturally coded; _Iga_cat_1 omitted)

        failure _d:  cens == 1
  analysis time _t:  stime

Iteration 0:   log likelihood = -5439.3955
Iteration 1:   log likelihood = -5368.3146
Iteration 2:   log likelihood = -5362.5375
Iteration 3:   log likelihood = -5362.5354
Refining estimates:
Iteration 0:   log likelihood = -5362.5354

Cox regression -- Breslow method for ties

No. of subjects =         9683                     Number of obs   =      9683
No. of failures =          597
Time at risk    =      1542785
                                                   LR chi2(4)      =    153.72
Log likelihood  =   -5362.5354                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |
          _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
36-38 wks _Iga_cat_2 |  -.8830784   .1289344    -6.85   0.000    -1.135785   -.6303715
38-39 wks _Iga_cat_3 |  -1.123549   .1517737    -7.40   0.000     -1.42102   -.8260785
39-41 wks _Iga_cat_4 |  -1.119394   .1154495    -9.70   0.000    -1.345671   -.8931173
 > 41 wks _Iga_cat_5 |  -1.072314   .1179203    -9.09   0.000    -1.303433   -.8411944
------------------------------------------------------------------------------
```

10.5 Proportional Hazards Check

```
. stphplot, by(ga_cat) ylab(2 (1) 5) xlab(0 (1) 7)
```



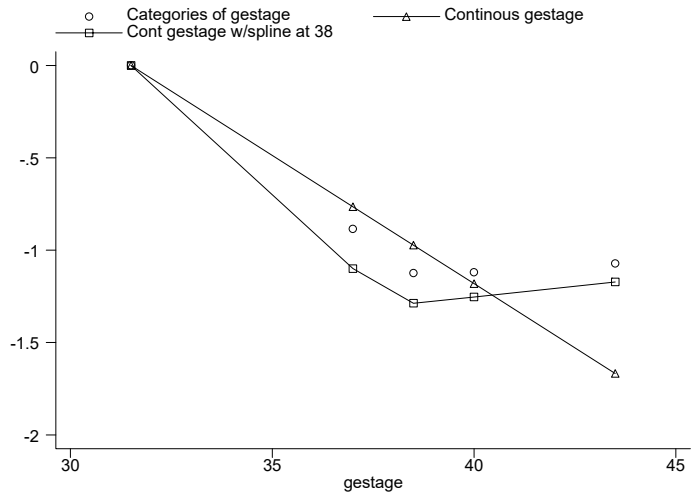
10.6 Gestational Age, Continuous (Centered at 38 Weeks)

```
.stcox gestage_c
        failure _d:  cens
  analysis time _t:  stime

Iteration 0:   log likelihood = -5439.3955
Iteration 1:   log likelihood = -5359.6802
Iteration 2:   log likelihood = -5358.7331
Iteration 3:   log likelihood = -5358.7331
Refining estimates:
Iteration 0:   log likelihood = -5358.7331

Cox regression -- Breslow method for ties

No. of subjects =         9683                     Number of obs   =      9683
No. of failures =          597
Time at risk    =      1619695
                                                   LR chi2(1)      =    161.32
Log likelihood  =   -5358.7331                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
         _t |
         _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
   gestage_c |   .8747498   .0089821   -13.03   0.000     .8573212    .8925327
------------------------------------------------------------------------------
```

10.7 Add a Spline Term for Gestational Age at 38 Weeks

```
.stcox gestage_c ga_sp38
        failure _d:  cens == 1
  analysis time _t:  stime

Iteration 0:   log likelihood = -5439.3955
Iteration 1:   log likelihood =  -5375.191
Iteration 2:   log likelihood =  -5338.943
Iteration 3:   log likelihood = -5338.7324
Iteration 4:   log likelihood = -5338.7324
Refining estimates:
Iteration 0:   log likelihood = -5338.7324

Cox regression -- Breslow method for ties

No. of subjects =         9683                     Number of obs   =      9683
No. of failures =          597
Time at risk    =      1542785
                                                   LR chi2(2)      =    201.33
Log likelihood  =   -5338.7324                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
         _t |
         _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
   gestage_c |   .8185724   .0115865   -14.14   0.000     .7961754    .8415995
     ga_sp38 |   1.258339   .0440248     6.57   0.000     1.174944    1.347654
------------------------------------------------------------------------------
```


10.8a Graph: Predicted Log HR At Midpoint of Gestage Categories



10.8b Notes on Graph of Predicted log(HR)

- Estimated log HR of death for children in each gestational age group as compared to children in lowest gestational age group
- Both the spline model and the model with gestational in categories yield similar estimates, and same trend

55

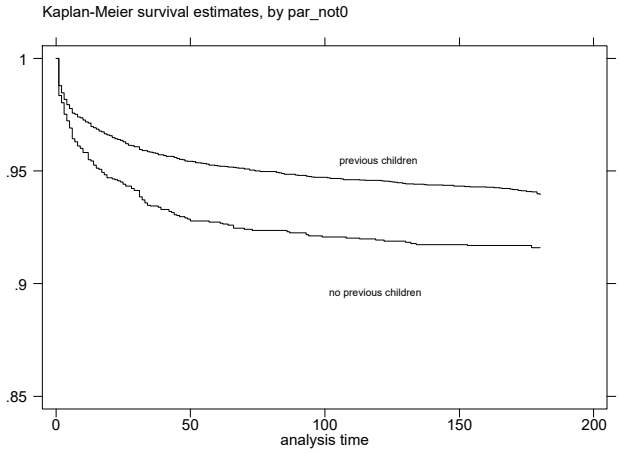
11.0 Parity, As Binary Categorical

```
. codebook parity

type:  numeric (float)
                 range:  [0,15]                       units:  1
         unique values:  15                       missing .:  256/10,295

                  mean:   2.32563
              std. dev:   2.14178

           percentiles:        10%       25%       50%       75%       90%
                                 0         1         2         3         5

. gen par_not0=1 if parity >0 & parity < 16
(2,510 missing values generated)
. replace par_not0=0 if parity==0
(2,254 real changes made)

. tab par_not0

   par_not0 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,254       22.45       22.45
          1 |      7,785       77.55      100.00
------------+-----------------------------------
      Total |     10,039      100.00
```

56

11.1 Kaplan-Meier Curves by Parity (Binary)

```
. sts graph, by(par_not0) ylab(.85 (.05) 1)
```



57

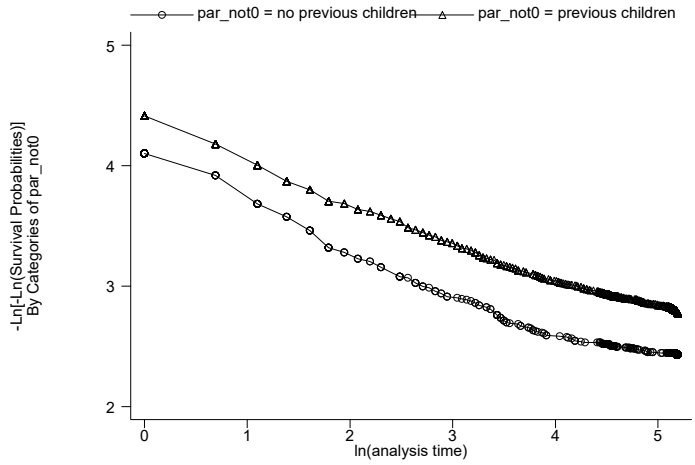
11.2 Cox Results with Parity, As Binary Categorical

```
.stcox par_not0

        failure _d:  cens
  analysis time _t:  stime

Iteration 0:   log likelihood =  -5781.046
Iteration 1:   log likelihood =  -5773.349
Iteration 2:   log likelihood =  -5773.268
Iteration 3:   log likelihood =  -5773.268
Refining estimates:
Iteration 0:   log likelihood =  -5773.268

Cox regression -- Breslow method for ties

No. of subjects =        10039                     Number of obs   =     10039
No. of failures =          632
Time at risk    =      1683476
                                                   LR chi2(1)      =     15.56
Log likelihood  =    -5773.268                     Prob > chi2     =    0.0001

------------------------------------------------------------------------------
         _t |
         _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
    par_not0 |   .6995863   .0616632    -4.05   0.000      .588592    .8315114
------------------------------------------------------------------------------
```

58

11.3 Proportional Hazards Check

```
. stphplot, by(par_not0) ylab(2 (1) 5) xlab(0 (1) 5)
```



59

11.4 Parity, as Multi-Categorical

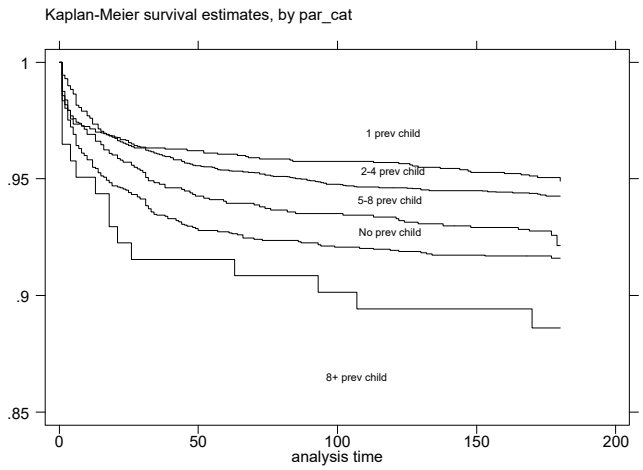
```
. tab par_cat
```

par_cat	Freq.	Percent	Cum.
No prev child	2254	22.45	22.45
1 prev child	2018	20.10	42.55
2-4 prev child	4262	42.45	85.01
5-8 prev child	1363	13.58	98.59
8+ prev child	142	1.41	100.00
Total	10039	100.00	

60

11.5 Kaplan-Meier Curves by Parity (Multi-Categorical)

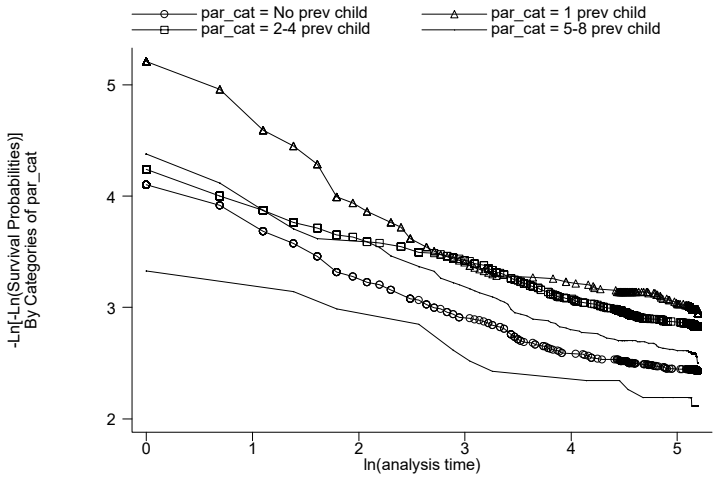
```
. sts graph, by(par_cat) ylab(.85 (.05) 1)
```



11.6 Parity, as Multi-Categorical

```
.stcox i.par_cat
i.par_cat         _Ipar_cat_0-4       (naturally coded; _Ipar_cat_0 omitted)

        failure _d:  cens
  analysis time _t:  stime

Iteration 0:   log likelihood =  -5781.046
Iteration 1:   log likelihood =  -5766.178
Iteration 2:   log likelihood = -5765.8311
Iteration 3:   log likelihood = -5765.8293
Iteration 4:   log likelihood = -5765.8293
Refining estimates:
Iteration 0:   log likelihood = -5765.8293

Cox regression -- Breslow method for ties

No. of subjects =        10039                     Number of obs   =     10039
No. of failures =          632
Time at risk    =      1683476
                                                   LR chi2(4)      =     30.43
Log likelihood  =   -5765.8293                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |
          _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Ipar_cat_1 |   .5819646   .0733035    -4.30   0.000     .4546533    .7449254
 _Ipar_cat_2 |   .6729974   .0664822    -4.01   0.000     .5545328    .8167695
 _Ipar_cat_3 |   .8852947   .1100691    -0.98   0.327     .6938372    1.129583
 _Ipar_cat_4 |   1.368832   .3570949     1.20   0.229     .8209036    2.282484
------------------------------------------------------------------------------
```

11.7 Proportional Hazards Check



63

11.8 How to Model the Mortality and Parity Relationship?

- Both the binary and categorical modeling of parity showed strong parity/mortality association
- Categorical approach allowed for better modeling of relationship
- We could have also used a linear predictor approach (and also examined spline terms)

64

12.0 A Multivariable Model for Describing Child Mortality

- Two purposes
  - Estimating adjusted relationships between hazard of mortality and each predictor
  - Building a model to estimate hazard (and subsequently survival) of children with certain child and/or maternal characteristics

65

12.1a Cox Regression Results (HRs)

```
.stcox i.ga_cat i.par_cat i.male i.nblind i.treat if male~=9

i.ga_cat          _Iga_cat_1-5        (naturally coded; _Iga_cat_1 omitted)
i.par_cat         _Ipar_cat_0-4       (naturally coded; _Ipar_cat_0 omitted)
i.male            _Imale_0-9          (naturally coded; _Imale_0 omitted)
i.nblind          _Inblind_0-1        (naturally coded; _Inblind_0 omitted)
i.treat           _Itreat_1-3         (naturally coded; _Itreat_1 omitted)
         failure _d:  cens == 1
   analysis time _t:  stime
note: _Imale_9 dropped due to collinearity
Iteration 0:   log likelihood = -5312.0553
Iteration 1:   log likelihood = -5222.0508
Iteration 2:   log likelihood = -5215.5804
Iteration 3:   log likelihood =  -5215.576
Iteration 4:   log likelihood =  -5215.576
Refining estimates:
Iteration 0:   log likelihood =  -5215.576
Cox regression -- Breslow method for ties
No. of subjects =         9443                     Number of obs   =      9443
No. of failures =          584
Time at risk    =      1523838
                                                   LR chi2(12)     =    192.96
Log likelihood  =    -5215.576                     Prob > chi2     =    0.0000
```

66

12.1b Cox Regression Results (HRs)

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Iga_cat_2	.4118807	.0537904	-6.79	0.000	.3188651	.5320296
_Iga_cat_3	.3230919	.0495576	-7.37	0.000	.2392017	.4364033
_Iga_cat_4	.3212073	.0377706	-9.66	0.000	.2550898	.4044619
_Iga_cat_5	.3469316	.0413817	-8.88	0.000	.2746082	.438303
_Ipar_cat_1	.5396267	.0716611	-4.65	0.000	.4159641	.7000531
_Ipar_cat_2	.640755	.0654042	-4.36	0.000	.524574	.7826674
_Ipar_cat_3	.7880633	.1013753	-1.85	0.064	.6124403	1.014048
_Ipar_cat_4	1.179086	.3282661	0.59	0.554	.683227	2.034821
_Imale_1	1.008702	.0836094	0.10	0.917	.8574488	1.186635
_Inblind_1	1.407019	.1767372	2.72	0.007	1.099966	1.799783
_Itreat_2	.9556154	.0988269	-0.44	0.661	.7802871	1.170339
_Itreat_3	.9609895	.0968532	-0.39	0.693	.7887337	1.170865

67

12.2 Summary of Multivariable Model Results

- Notice that hazard ratio estimates for both gestational age and parity categories are similar to the estimates obtained in the simple models (Why is this?)
- Each hazard ratio estimate is interpretable as an adjusted hazard ratio – interpretation depends on coding of predictor

68

12.3a Summary of Multivariable Model Results

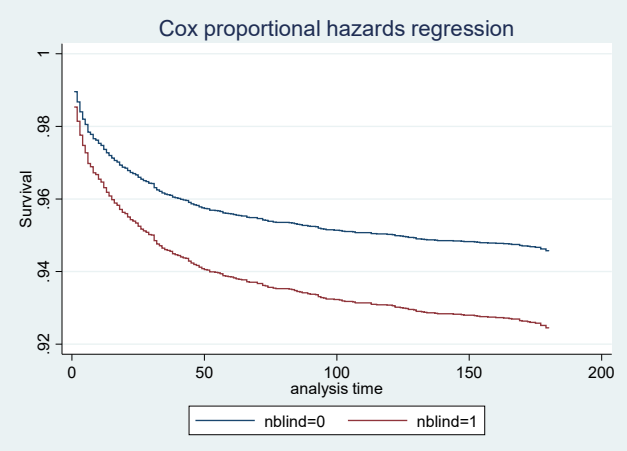
- Example: Night blindness
  - Coefficient estimates the adjusted log HR for children with night-blind mothers compared to children with non-night-blind mothers

```
------------------------------------------------------------------------------
          _t |  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_  _Inblind_1 |    1.407019   .1767372     2.72   0.007     1.099966    1.799783
```

69

12.4 Adjusted Kaplan-Meier Curves by Maternal Night Blindness

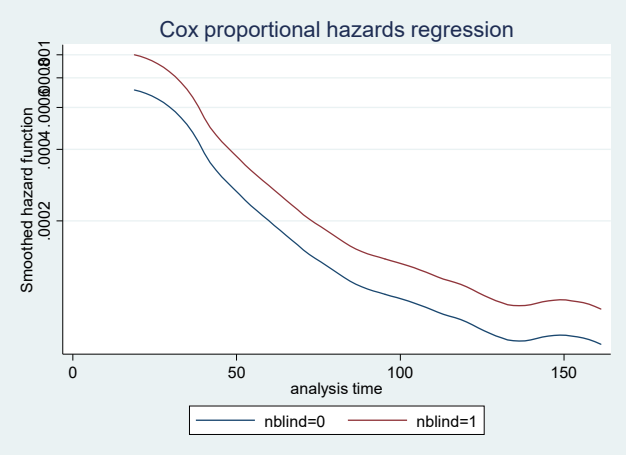
```
After .stcox for the multivariable model:
.stcurve, survival at1(nblind=0) at2(nblind=1) ylab(0.85 (0.05) 1)
```



70

12.5 Adjusted Hazard Curves by Maternal Night Blindness

```
After .stcox for the multivariable model:
. stcurve, hazard at1(nblind=0) at2(nblind=1)yscale(log)
```



71

12.3b Summary of Multivariable Model Results

- Model estimates children with night-blind mothers have 41% increased adjusted hazard of death than children with mothers who are not night-blind.
- Conclusions?

  The hazard "risk" of death is estimated to be 41% greater for an infant whose mother is night-blind (95% CI: 10 to 80% higher).

72