Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Introduction to Models That Relax The Proportional Hazards Assumption

Stratification
We now consider models that allow for varying [and nonproportional] baseline hazards. Perhaps we have s strata. We are concerned that the baseline hazard will vary in nonproportional ways across these strata. We anticipate using a model like:

$$\log h(t) = \log h_{0i}(t) + \sum_{j=1}^{k} \beta_j x_j \quad \text{for i = 1 to s}$$

One can use 'parametric' forms for the stratum specific baseline hazards $\log h_{0i}(t)$ or use the Cox approach that then constructs the regression coefficient estimates admitting stratum-specific hazards without specifying their forms. Notice, that, in both approaches, the proportional hazards assumption is partially maintained but is now within strata.

A worthy example comes from addiction research. [Caplehorn & Bell(1991)]. They studied heroin addicts receiving methadone maintenance treatment to help them overcome their addiction. Early dropout is an important issue with this treatment. We will consider the time from admission to termination of treatment (in days). Status refers to dropout (1) or end of study (0). Possible explanatory variables are maximum methadone dose (dose), prison record (prison). Participants came from two different clinics.
The investigators were concerned that the baseline hazard for the two clinics might be nonproportional. We fit a fairly elaborate model that allows for clinic specific baseline hazard and explanatory variables that may depend on clinic as well. Then we view the two clinic specific log baseline cumulative hazards versus time.

```
use caplehorn.dta
gen d60=dose-60
gen cl=clinic-1
gen pc=prison*cl
gen dp=d60*prison
gen dpc=dp*cl
gen dc=d60*cl
stcox d60 prison dp dc pc dpc,strata(clinic) basech(bch) nohr
gen lbch=log(bch)
twoway (line lbch time if cl==0,connect(stairstep))(line lbch time if
cl==1,connect(stairstep))
```

The two curves look to be about the same until about the 400 day mark and then the curves separate from one another. This offers a visual cue that these 2 curves are not separated by a constant vertical distance. Strong visual evidence that the two hazards are not proportional.

The modeling process could then proceed as usual but keeping the two clinic specific baseline hazards in each model. One might then arrive at:

```
. stcox d60 prison,strata(clinic) nohr

Stratified Cox regr. -- Breslow method for ties

No. of subjects =          238                     Number of obs   =       238
No. of failures =          150
Time at risk    =        95812
```

```
                                                  LR chi2(2)      =     33.94
Log likelihood  =     -597.714                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d60 |  -.0351449    .006465    -5.44   0.000    -.0478162   -.0224737
      prison |   .3887882   .1689154     2.30   0.021     .0577201    .7198563
------------------------------------------------------------------------------
                                                        Stratified by clinic
```

We can then explore whether there are meaningful differences between the regression coefficients with and without the separate baseline hazards.

```
. stcox d60 prison,nohr

Cox regression -- Breslow method for ties

No. of subjects =          238                     Number of obs   =       238
No. of failures =          150
Time at risk    =        95812
                                                  LR chi2(2)      =     38.22
Log likelihood  =   -686.55176                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d60 |  -.0360866    .006001    -6.01   0.000    -.0478484   -.0243248
      prison |   .1897446   .1642743     1.16   0.248    -.1322272    .5117163
------------------------------------------------------------------------------
```

With regard to dose, the assumed common slope estimates are similar but the impact of a prison record (assumed common to dose) is seen more clearly with the 'stratified' model.

Time Varying Explanatory Variables

Now we relax the proportional hazards assumption in a quite different way. We now consider circumstances where some of the explanatory variables may change during the period of follow up.

$\log h(t) = \log h_0(t) + \sum_{j=1}^{k_1} \beta_j x_j + \sum_{j=k_1+1}^{k_2} \beta_j x_j(t)$ where the second set of explanatory variables are functions of time $x_j(t)$ for $j = k_1 + 1$ to k_2

Time Varying Explanatory Variables : Indicators

The simplest form of time varying explanatory variable is an indicator variable. We will use as an illustration the [rather infamous] Stanford Heart Transplant study. The outcome is time until death (in days from entry into the trial) and the exposure is heart transplant. Age, year of entry and previous surgery are possible confounders or modifiers. A fragment of this dataset looks like:

```
. list id transplant start stop event in 1/25,sep(25)

     +---------------------------------------+
     | id   transp~t   start   stop   event |
     |---------------------------------------|
  1. |  1          0       0     50       1 |
  2. |  2          0       0      6       1 |
  3. |  3          0       0      1       0 |
  4. |  3          1       1     16       1 |
  5. |  4          0       0     36       0 |
  6. |  4          1      36     39       1 |
  7. |  5          0       0     18       1 |
```

8.	6	0	0	3	1
9.	7	0	0	51	0
10.	7	1	51	675	1
11.	8	0	0	40	1
12.	9	0	0	85	1
13.	10	0	0	12	0
14.	10	1	12	58	1
15.	11	0	0	26	0
16.	11	1	26	153	1
17.	12	0	0	8	1
18.	13	0	0	17	0
19.	13	1	17	81	1
20.	14	0	0	37	0
21.	14	1	37	1387	1
22.	15	0	0	1	1
23.	16	0	0	28	0
24.	16	1	28	308	1
25.	17	0	0	36	1

Notice that the first 2 IDs (patients) have one row while IDs 3, 4 and 7 have 2 rows each. IDs 3, 4 and 7 had transplants while IDs 1, 2, 5, 6 and 8 did not have a transplant. It is acknowledged that the change in transplant status may change the hazard and, for any given patient, the change in status may or may not happen during the trial time and, if the transplant occurs may occur at varying times through the course of the study. Hence we say that 'transplant' E(t)) is a time varying.

E(t) is either always zero (no transplant) or it is a step function with a single step from zero to one at the time of the transplant. Patient age, year accepted into the trial and previous surgery are measured at 'baseline' and are said to be time fixed.

Lets start with a 'simple' model (ignoring age (A), year (Y) and surgery (S))

$$\log(h(t))=\log(h_0(t))+\beta_1 E(t)$$

With this model, all patients that did not have a transplant (E(t)=0 for all time t) have the same hazard function (here, the baseline hazard function). Consider patient 7 (transplant at day 51). For this patient, this model presents the same baseline log of hazard until time 51. At time 51, the log hazard changes by the amount β_1 and then follows the same shape as the baseline log hazard but vertically shifted by the amount β_1 .

Now consider a model with age (A) included:

$$\log(h(t))=\log(h_0(t))+\beta_1 A+\beta_2 E(t)+\beta_3 AE(t)$$

[Age has been centred at age 48]. For a patient of [baseline] age 48 (A=0), we get the same interpretation as above. Now consider a patient of age 49 (A=1) that did not have a transplant. The log of the hazard is the baseline log hazard shifted by β_1 . For a 49 year old patient that did have a transplant at day 51, the log hazard follows the same log hazard as the untransplanted 49 year old until time 51 at which time the log hazard shifts by the amount $\beta_2+\beta_3$. More generally, for a patient of age A receiving a transplant at time T. The log of hazard is the baseline log of hazard shifted by $\beta_1 A$ until time T at which time the log hazard shifts by $\beta_2+\beta_3 A$ If β_3 is zero, then we can see that the impact of transplant does not depend on [baseline] age.

It is worth emphasizing that baseline age is time fixed. One might think that one could add a patient's actual age as time varying. This action serves no purpose since a patient's age and time are merely shifted versions of each other. Adding actual age as explanatory variable serves to change the baseline hazard [and its interpretation] but has no impact on the regression coefficients. Adding time varying

functions of age is another [and complicating] matter.
Models can be built that include time varying variables, time fixed variables and interactions between any 2 (or more) of such (whether time varying or time fixed).

Here are some analyses of the Stanford Heart Transplant dataset. These analyses essentially reproduce the results from pg. 139 of Kalbfleisch & Prentice (1980 & 2002). [A 'corrected' version of this dataset (from http://lib.stat.cmu.edu/datasets/) was used] It is instructive [and challenging!] to read the interpretations given in K&P and to consider possible next steps in analysis.

```
(Stanford Heart Transplant Data)
. stset
-> stset stop, id(id) failure(event)

                id:  id
     failure event:  event != 0 & event < .
obs. time interval:  (stop[_n-1], stop]
 exit on or before:  failure

------------------------------------------------------------------------------
      172  total obs.
        0  exclusions
------------------------------------------------------------------------------
      172  obs. remaining, representing
      103  subjects
       75  failures in single failure-per-subject data
    31954  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      1800

. stvary

         failure _d:  event
   analysis time _t:  stop
                 id:  id

               subjects for whom the variable is
                                                      never    always   sometimes
    variable |  constant    varying             missing   missing   missing
-------------+----------------------------------------------------------------
  transplant |        34         69                 103         0         0
       start |        34         69                 103         0         0
         age |       103          0                 103         0         0
        year |       103          0                 103         0         0
     surgery |       103          0                 103         0         0
          ta |        34         69                 103         0         0
          ts |        90         13                 103         0         0
          ty |        34         69                 103         0         0


. stcox transplant age surgery ta ts,nohr

Cox regression -- Breslow method for ties
                                                LR chi2(5)      =     12.45
Log likelihood  =   -292.09953                  Prob > chi2     =    0.0291

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |   .1181019   .3276902     0.36   0.719    -.5241591    .7603629
         age |   .0138234   .0181253     0.76   0.446    -.0217015    .0493482
     surgery |  -.5457023     .61091    -0.89   0.372    -1.743064    .6516592
          ta |   .0347659   .0272543     1.28   0.202    -.0186515    .0881833
          ts |  -.2916307   .7581853    -0.38   0.701    -1.777647    1.194385
------------------------------------------------------------------------------

. stcox transplant year ty,nohr

Cox regression -- Breslow method for ties
                                                LR chi2(3)      =      8.61
Log likelihood  =   -294.02098                  Prob > chi2     =    0.0350
```

```
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |  -.2822266   .5142526    -0.55   0.583    -1.290143    .7256901
        year |   -.264717   .1051108    -2.52   0.012    -.4707305   -.0587036
          ty |   .1362093   .1409024     0.97   0.334    -.1399543    .4123729
------------------------------------------------------------------------------

. stcox transplant age year ta ty,nohr

Cox regression -- Breslow method for ties
                                                LR chi2(5)      =     14.83
Log likelihood  =   -290.90889                  Prob > chi2     =    0.0111

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |  -.5883537   .5427372    -1.08   0.278    -1.652099    .4753917
         age |    .015514   .0173418     0.89   0.371    -.0184754    .0495033
        year |  -.2735364   .1058311    -2.58   0.010    -.4809614   -.0661113
          ta |   .0338558   .0279495     1.21   0.226    -.0209242    .0886359
          ty |    .201259   .1424636     1.41   0.158    -.0779645    .4804825
------------------------------------------------------------------------------
```

Refer to Crowley & Hu(1977) and K&P(2002) for their interpretations.

year is in fact Year(19XX) + [the YY day of the year]/365.25 – [1967 +275/365.25]

eg) November 15, 1967 is 1967+320/365.25 – [1967+275/365.25] = 45/365.25 = 0.12320329

```
. stcox transplant year surgery ty ts,nohr

Cox regression -- Breslow method for ties
                                                LR chi2(5)      =     12.35
Log likelihood  =   -292.14897                  Prob > chi2     =    0.0303

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |  -.2920895   .5059006    -0.58   0.564    -1.283636    .6994575
        year |  -.2536811   .1076625    -2.36   0.018    -.4646958   -.0426664
     surgery |  -.2361504   .6281973    -0.38   0.707    -1.467395    .9950937
          ty |   .1644914   .1416135     1.16   0.245    -.1130661    .4420488
          ts |  -.5504738   .7758498    -0.71   0.478    -2.071111    .9701638
------------------------------------------------------------------------------

. stcox transplant age year surgery ta ts,nohr

Cox regression -- Breslow method for ties
                                                LR chi2(6)      =     16.21
Log likelihood  =   -290.22034                  Prob > chi2     =    0.0127

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |   .0771957   .3316176     0.23   0.816    -.5727628    .7271541
         age |   .0149866   .0176007     0.85   0.395    -.0195102    .0494834
        year |  -.1363152   .0709655    -1.92   0.055     -.275405    .0027746
     surgery |  -.4191803   .6156507    -0.68   0.496    -1.625833    .7874728
          ta |   .0269781   .0271197     0.99   0.320    -.0261756    .0801318
          ts |   -.298129   .7580001    -0.39   0.694    -1.783782    1.187524
------------------------------------------------------------------------------

. stcox transplant age year surgery ta,nohr

Cox regression -- Breslow method for ties
                                                LR chi2(5)      =     16.06
Log likelihood  =   -290.29562                  Prob > chi2     =    0.0067

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
  transplant |   .0474531   .3221818     0.15   0.883    -.5840116    .6789178
         age |   .0152199   .0175019     0.87   0.385    -.0190832    .0495229
        year |  -.1360785   .0708987    -1.92   0.055    -.2750373    .0028803
     surgery |  -.6211691   .3678687    -1.69   0.091    -1.342178    .0998403
          ta |   .0270955   .0271401     1.00   0.318    -.0260981    .0802892
------------------------------------------------------------------------------
```

Time Varying Variables With A Specific Functional Form

More elaborate time varying variables can be considered.

$\log(h(t)) = \log(h_0(t)) + \sum_{j=1}^{k} \beta_j x_j + g(t) \sum_{l=1}^{m} \gamma_l z_l$ Where g(t) is some chosen function. With such choices of the function g (other than unit steps) , the models are not invariant to changes in analysis time (even with Cox models). A direct assessment requires a dataset that records, for each subject, a separate row of data for each distinct failure time if g(t) changes in any given time interval. This enables the consideration of 'continuous' time varying variables as the model fitting process is only dependent on the values of such functions at the distinct failure times. In principle, more than one g function could be considered. Stata has option [called tvc] that can handle some of the dataset matters for you.

Now let us consider the data from a study of recovery time from walking pneumonia in pneumonia.dta Two drugs (Type =0 or 1) are being compared. The patient 's [baseline] age (Age) is also involved. Suppose we know that the actual level of either drug in the body has a half-life of 2 days so that level is proportional to $e^{-0.35t}$

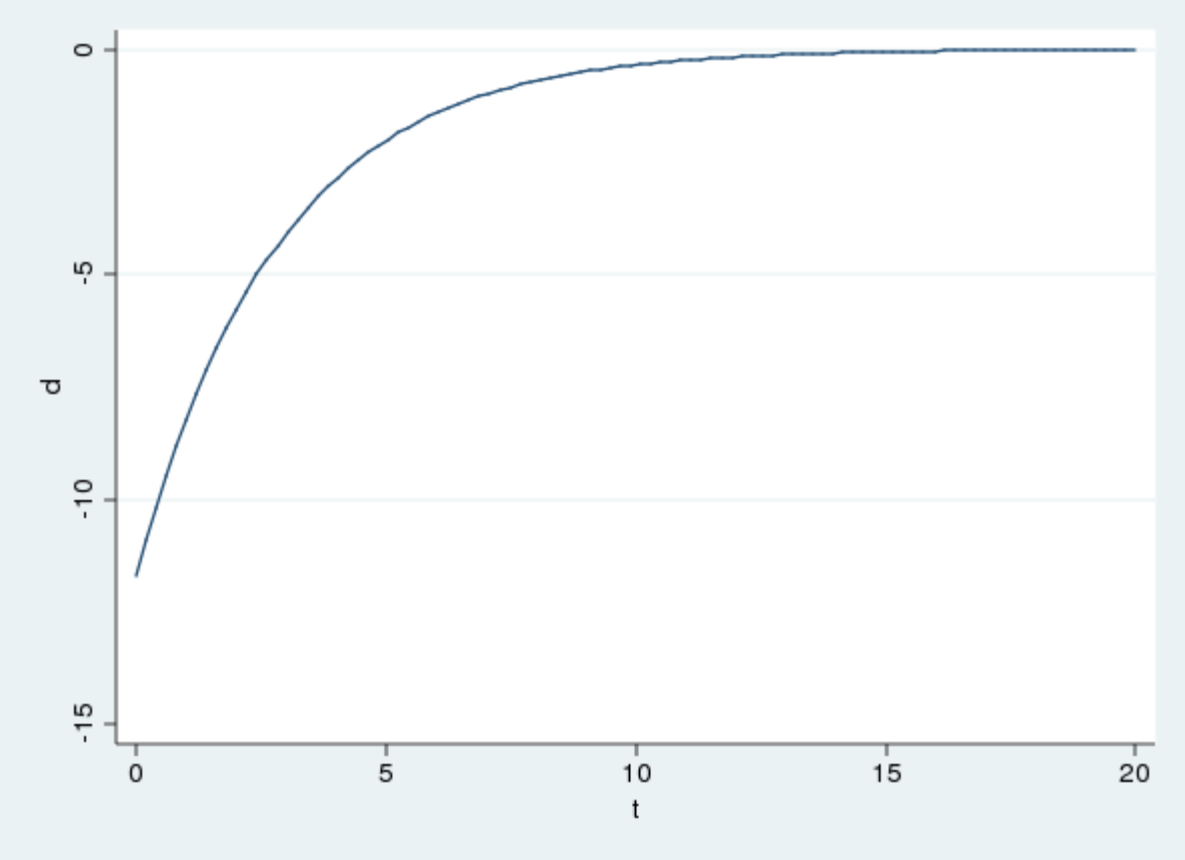
$$\log h(t) = \log h_0(t) + \beta_1 \text{Age} + e^{-0.35t} \gamma_1 \text{Type}$$

```
. stcox age type,nohr

        failure _d:  cured
  analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           45                     Number of obs   =        45
No. of failures =           36
Time at risk    =  677.9000034
                                                   LR chi2(2)      =     27.28
Log likelihood  =   -102.90267                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.1275093   .0280598    -4.54   0.000    -.1825055    -.072513
        type |  -.7711755   .3563117    -2.16   0.030    -1.469534   -.0728173
------------------------------------------------------------------------------

. stcox age, tvc(type) texp(exp(-0.35*_t)) nohr

        failure _d:  cured
  analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           45                     Number of obs   =        45
No. of failures =           36
Time at risk    =  677.9000034
                                                   LR chi2(2)      =     28.06
Log likelihood  =   -102.51376                     Prob > chi2     =    0.0000
```

```
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rh           |
         age |  -.1306558   .0299297    -4.37   0.000     -.189317   -.0719946
-------------+----------------------------------------------------------------
t            |
        type |  -11.72067   5.184889    -2.26   0.024    -21.88286   -1.558474
------------------------------------------------------------------------------

Note: second equation contains variables that continuously vary with respect
      to time; variables are interacted with current values of exp(-0.35*_t).

. disp exp(-0.35)
.70468809

. disp exp(-0.35*2)
.4965853
```

$$\log \hat{h}(t) = \log \hat{h}_0(t) - 0.1307\,\text{Age} - 11.7207\, e^{-0.35t}\,\text{Type}$$

So for any given age, the difference between log of the hazard with drug2 (type=1) and the log of the hazard with drug1 (type=0) is:



This model necessarily requires that the impact of the drug difference eventually goes to zero. We could check if there is a lasting effect by including Type as a 'Time Invariant' covariate.

$$\log h(t) = \log h_0(t) + \beta_1 \text{Age} + \beta_2 \text{Type} + e^{-0.35t}\gamma_1 \text{Type}$$

If the sign of β_2 and the sign of γ_1 are the same, then β_2 records the lasting effect initially detailed by $\beta_2+\gamma_1$ at time =0. To see this, notice that $e^{-0.35t}$ is one when t=0 while $e^{-0.35t}$ is near zero when t is 'large'.

If the sign of β_2 and the sign of γ_1 are not the same, then we get other scientifically interesting scenarios. As always, a careful graphing of the situation enables an appropriate interpretation.

To explore the circumstance here, we can consider:

```
. stcox age type, tvc(type) texp(exp(-0.35*_t)) nohr

Cox regression -- Breslow method for ties

No. of subjects =           45                     Number of obs   =        45
No. of failures =           36
Time at risk    =  677.9000034
                                                   LR chi2(3)      =     28.37
Log likelihood  =   -102.36053                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rh           |
         age |  -.1298054   .0292817    -4.43   0.000    -.1871964   -.0724143
        type |  -.3283862   .5795018    -0.57   0.571    -1.464189    .8074163
-------------+----------------------------------------------------------------
t            |
        type |  -8.202869   8.054333    -1.02   0.308    -23.98907    7.583333
------------------------------------------------------------------------------

Note: second equation contains variables that continuously vary with respect
      to time; variables are interacted with current values of exp(-0.35*_t).
```

Apparently, there is no lasting effect.

In principle, models could be contemplated that allow for the estimates of the exponent (instead of assuming the value -0.35) Such model fitting is not directly available with Stata [as of the Fall 2015].

Time Varying Explanatory Variables : Discrete Time-To-Event
Once one has the dataset in person period format and one adds the time varying variable(s) to the dataset, one can proceed with modeling as usual. Graphing of the estimated log odds of the hazard function will necessarily be participant specific.

Lets consider a study by Wheaton, Rozell, and Hall (1997), who examined the link between stressful life experiences and the risk of a psychiatric disorder. Using a random sample of adults, ages 17 to 57, in metropolitan Toronto, the researchers conducted a structured interview that allowed them to determine whether, and if so at what age (in years), each individual first experienced a depressive episode.

Among the 1393 respondents, 387 (27.8%) experienced a first onset between ages 4 and 39. Using the same interview, the researchers also ascertained whether, and if so at what age, each respondent first experienced 19 traumatic events, including major hospitalization, physical abuse, and parental divorce.

Here, we focus on one of these stressors, first parental divorce (pd), experienced by one-tenth of the sample (n = 145) at risk of an initial depressive episode. We will consider the time-varying predictor [pd] indicating whether the parents of individual i divorced during, or before, time period j. In the time periods before the divorce, $pd_{ij} = 0$; in time periods coincident with, or subsequent to, the divorce $pd_{ij} = 1$ Coding pd_{ij} in this way allows one to capture both the immediate and long-term impacts of parental divorce.

Following our earlier approach to discrete time, 36 time indicator variables could be considered but this option needs some consideration of a diagnostic. Collapsing of some time intervals is required here. Some authors consider polynomials to capture the salient features of the baseline hazard. We will consider the time fixed variable 'female' and the time varying variable 'pd' for illustrative purposes.

```
use wheaton_pp, clear
logit event i.agea pd female
gen ageaa=agea
replace ageaa=6 if ageaa<6
logit event i.ageaa pd female
predict loh,xb
twoway (lowess loh agea if id==24 & agea<21)(lowess loh agea if id ==24 & agea
>20),legend(off)
drop loh
logit event age_18 age_18sq age_18cub pd female
predict loh, xb
twoway (line loh agea if id==24 & agea<21)(line loh agea if id ==24 & agea >20),legend(off)
clear
set obs 1001
range age 4 40
gen lohb=-4.58664+0.0595987*(age-18)-0.0073603*(age-18)^2+0.0001847*(age-18)^3
gen lohf=lohb+0.5454514
gen lohpd=lohb+0.4150557
gen lohpdf=lohf+0.4150557
twoway (line lohb age)(line lohf age)(line lohpd age)(line lohpdf age)
use wheaton_pp.dta,clear
stset agea event, id(id)
stvary
stcox pd female,nohr
stcurve, haz at1(pd=0 female=0) at2(pd=1 female=0) at3(pd=0 female=1) at4(pd=1 female=1)
yscale(log)
cloglog event age_18 age_18sq age_18cub pd female
```