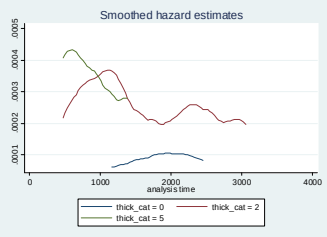
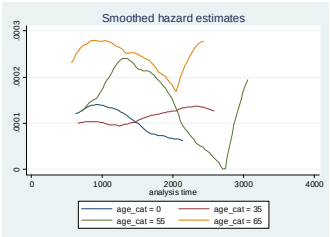
```
. use melanoma

. gen id=_n

. stset survtime , failure(status==1) id(id)

                id:  id
     failure event:  status == 1
obs. time interval:  (survtime[_n-1], survtime]
 exit on or before:  failure

------------------------------------------------------------------------------
      205  total obs.
        0  exclusions
------------------------------------------------------------------------------
      205  obs. remaining, representing
      205  subjects
       57  failures in single failure-per-subject data
   441324  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      5565

. // point is to learn about effect of the continuous variables
. egen thick_cat = cut(thick) , at(0,2,5,20)

. egen age_cat = cut(age) , at(0,35,55,65,110)
```

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

         failure _d:  status == 1
   analysis time _t:  survtime
                 id:  id
```

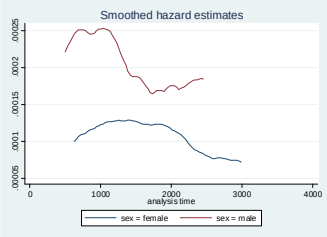



```
. // most info is binary
. replace thick_cat = thick >=2
(96 real changes made)
```

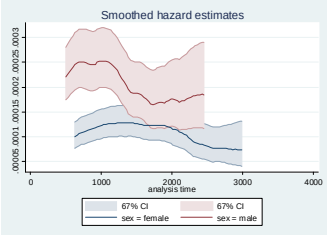
```
. sts graph, by(age_cat)  hazard

         failure _d:  status == 1
   analysis time _t:  survtime
                 id:  id
```



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

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id
```

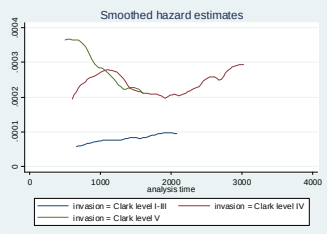



```
. sts graph, by(sex) hazard ci level(67)

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id
```



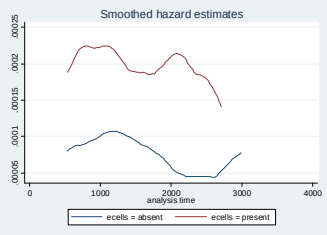

```
. stcox sex , tvc(sex) texp(log(survtime/1600)) nohr

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id

Iteration 0:   log likelihood = -283.10925
Iteration 1:   log likelihood = -279.41449
Iteration 2:   log likelihood =  -279.4055
Iteration 3:   log likelihood =  -279.4055
Refining estimates:
Iteration 0:   log likelihood =  -279.4055

Cox regression -- no ties

No. of subjects =          205                 Number of obs   =      1870
No. of failures =           57
Time at risk    =       441324
                                                LR chi2(2)      =      7.59
Log likelihood  =    -279.4055                  Prob > chi2     =    0.0225

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
main         |
         sex |   .4545677   .3203576     1.42   0.156    -.1733217    1.082457
-------------+----------------------------------------------------------------
tvc          |
         sex |  -.4666954   .3932363    -1.17   0.241    -1.231424    .3100337
------------------------------------------------------------------------------
Note: variables in tvc equation interacted with log(survtime/1600)
```

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

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id
```

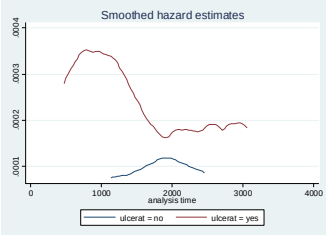



```
. // most info is binary
. gen invas_cat = invasion > 0
```

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

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id
```



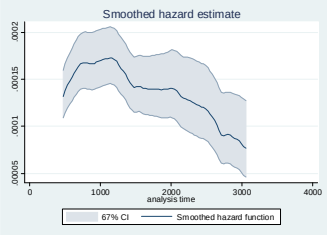

```
. sts graph, by(ulcerat)  hazard

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id
```



```
. replace age_cat = age > 35
(171 real changes made)

. // could use age as continuous, but most info is again binary
. // do not need xi here since all covars are binary 0,1
. stcox  sex invas_cat ecells ulcerat thick_cat age_cat

        failure _d:  status == 1
  analysis time _t:  survtime
               id:  id

No. of subjects =          205                     Number of obs   =       205
No. of failures =           57
Time at risk    =       441324
                                                    LR chi2(6)      =     47.92
Log likelihood  =   -259.23961                      Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   1.683799    .473142     1.85   0.064      .9707453    2.92062
   invas_cat |   1.497389   .5872226     1.03   0.303      .6942626    3.229576
      ecells |   1.688354   .5237427     1.69   0.091      .9192861    3.101087
     ulcerat |   2.370916   .7904118     2.59   0.010      1.233514    4.557097
   thick_cat |     2.0064    .934487     1.86   0.007      .8751131    5.022231
     age_cat |   1.338283   .5257613     0.74   0.458      .6196287    2.890358
------------------------------------------------------------------------------

. // plausible that invas_cat and thick_cat are carrying same info
. estimates store full_mod
```

```
. stcox sex invas_cat ecel ulce age_cat

        failure _d:  status == 1
  analysis time _t:  survtime
               id:  id

No. of subjects =          205                     Number of obs   =       205
No. of failures =           57
Time at risk    =       441324
                                                    LR chi2(5)      =     46.81
Log likelihood  =   -259.79209                      Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   1.581821   .4355338     1.67   0.096      .9221259    2.713468
      ecells |    1.64861   .5094349     1.62   0.106      .8996822    3.020971
     ulcerat |     2.3881   .7973938     2.61   0.009      1.241212    4.594876
   thick_cat |   2.778232    .989867     2.87   0.004      1.381945    5.585209
     age_cat |   1.289179   .5045376     0.65   0.516      .5986651    2.776147
------------------------------------------------------------------------------

. lrtest full_mod

Likelihood-ratio test                                 LR chi2(1)  =      1.10
(Assumption: . nested in full_mod)                    Prob > chi2 =    0.2932
```

```
. // nature of baseline hazard
. sts graph , hazard ci level(67)

        failure _d:  status == 1
  analysis time _t:  survtime
               id:  id
```



```
. // explore using Pan regr
. stsplit st_cat , every(250) after(0)
(1665 observations (episodes) created)

. // do not use collapse without understanding it better
. gen risk_time = _t - _t0

. gen log_st = log(survtime/1500)

. gen log_st_sq = log_st * log_st

. poisson _d    log_st log_st_sq sex ecel ulcer thick_cat age_cat , exp(risk_time)

Iteration 0:   log likelihood =  -295.0184
Iteration 1:   log likelihood = -295.01347
Iteration 2:   log likelihood = -295.01347

Poisson regression                              Number of obs   =       1870
                                                LR chi2(7)      =      49.73
                                                Prob > chi2     =     0.0000
Log likelihood = -295.01347                     Pseudo R2       =     0.0777

------------------------------------------------------------------------------
          _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      log_st |  -.4692238   .3976451    -1.53   0.127    -1.072001    .1335536
   log_st_sq |  -.3783639   .2279961    -1.66   0.097    -.8252321    .0683043
         sex |   .4398331   .2764831     1.56   0.119    -.1120849    .9917512
       ecells |   .5849381   .3987458     1.64   0.102    -.1982095    1.118086
      ulcerat |   .8585528   .3344559     2.57   0.010     .2030319    1.514074
   thick_cat |   1.000565   .3561958     2.81   0.005     .3023717    1.698758
     age_cat |   .2378041   .3912715     0.61   0.543    -.5290739    1.004682
       _cons |  -10.62994   .4946477   -21.82   0.000    -11.59825   -9.661623
  risk_time  |  (exposure)
------------------------------------------------------------------------------
```

```
. // compare to coxreg
. stcox sex ecel ulcer thick_cat age_cat ,nohr

        failure _d:  status == 1
  analysis time _t:  survtime
                id:  id

No. of subjects =          295                  Number of obs   =      1870
No. of failures =           57
Time at risk    =       441324

                                                LR chi2(5)      =     46.81
Log likelihood  =    -259.79289                 Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |    .458577   .2753369     1.67   0.096    -.0810735    .9982274
      ecells |   .4999523   .3898088     1.62   0.105    -.1637137    1.185578
     ulcerat |   .8785149   .3338974     2.61   0.009     .218868    1.538162
   thick_cat |   1.021815   .3562938     2.87   0.004     .3234918    1.720138
     age_cat |   .2540855   .3913935     0.65   0.516    -.5130529    1.021224
------------------------------------------------------------------------------
```

```
. // simplify baseline model
. poisson _d    log_st sex ecel ulcer thick_cat age_cat , exp(risk_time)

Iteration 0:   log likelihood = -296.52572
Iteration 1:   log likelihood = -296.52557
Iteration 2:   log likelihood = -296.52557

Poisson regression                              Number of obs   =       1870
                                                LR chi2(6)      =      48.71
                                                Prob > chi2     =     0.0000
Log likelihood = -296.52557                     Pseudo R2       =     0.0730

------------------------------------------------------------------------------
          _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      log_st |  -.0604921   .1668094    -0.36   0.717    -.3875892    .2666051
         sex |   .4421873   .2753889     1.61   0.118    -.0975695    .9819198
      ecells |   .5110669   .3899322     1.85   0.068    -.0541051    1.116239
     ulcerat |   .8920625   .3339436     2.67   0.008     .237535    1.54659
   thick_cat |   .9979985   .3639249     2.74   0.006     .2846185    1.711379
     age_cat |   .3187529   .3985461     0.82   0.414    -.4623934    1.099899
       _cons |  -10.82481   .4854094   -22.38   0.000    -11.7756    -9.872026
  risk_time  |  (exposure)
------------------------------------------------------------------------------
```

```
. poisson _d    sex ecel ulcer thick_cat age_cat , exp(risk_time)

Iteration 0:   log likelihood = -296.59095
Iteration 1:   log likelihood = -296.59083
Iteration 2:   log likelihood = -296.59083

Poisson regression                                Number of obs   =       1870
                                                  LR chi2(5)      =      46.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -296.59083                       Pseudo R2       =     0.0728

------------------------------------------------------------------------------
          _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .4477188   .2758471     1.62   0.105    -.0929316    .9883692
      ecells |    .513362   .3088862     1.66   0.097    -.0920438    1.118768
     ulcerat |   .9051367   .3316681     2.73   0.006     .2550791    1.555194
   thick_cat |   .9662888   .3533995     2.73   0.006     .2736385    1.658939
     age_cat |   .3364317   .3876971     0.87   0.386    -.4234406    1.096304
       _cons |  -10.82913   .4855467   -22.30   0.000    -11.78078   -9.877471
   risk_time | (exposure)
------------------------------------------------------------------------------

. // constant hazard gives essentially coxreg estimates
```

```
. // it was surprising that in the model before last that the log_st effect was so small
. // it is interesting in this regard that if we omit all the covariables
. // the log_t effect becomes larger --- my interpretation of this is that when we ignor
> e
. // the covariates there is some individual heterogeneity, and selection causes the
. // hazard to decrease.  adjusting for the covariates removes this, so they may be
. // carrying most of the "heterogeneity"
. poisson _d  log_st

Iteration 0:   log likelihood = -255.08779
Iteration 1:   log likelihood = -255.08779

Poisson regression                                Number of obs   =       1870
                                                  LR chi2(1)      =       1.76
                                                  Prob > chi2     =     0.1849
Log likelihood = -255.08779                       Pseudo R2       =     0.0034

------------------------------------------------------------------------------
          _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      log_st |  -.2109464   .1560591    -1.35   0.176    -.5168167    .0949239
       _cons |  -3.558442    .146062   -24.36   0.000    -3.844719   -3.272166
------------------------------------------------------------------------------
```