Cancer survival: principles, methods and analysis
Session 16
Practical 5: Modelling net survival

Paul W. Dickman
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden
paul.dickman@ki.se

London School of Hygiene and Tropical Medicine

1 Exercises

1. **Model excess mortality using Poisson regression**

We will now model relative survival (excess mortality) in patients diagnosed with melanoma as a function of time since diagnosis (annual intervals), sex, age at diagnosis, depression, and period for the first 10 years of follow-up. We restrict to the first 10 years of follow-up since this is where most of the deaths occur and the proportional excess hazards assumption is more likely to be appropriate during a shorter follow-up interval.

The first step is to estimate relative survival, using `strs`, for each combination of covariates. You can copy the commands from the PDF file or run the commands in `practical5.do`.

```
cd "H:\Cancer Survival\Data"

use melanoma_2013, clear

stset finmdy, failure(dead==1) origin(diagmdy) scale(365.24) id(id)

strs using Lifetable_2013, br(0(1)10) mergeby(_year sex _age dep) ///
    by(dep sex agecat period_diag) save(replace) ///
    diagage(agediag) diagyear(ydiag) notables
```

Be sure to study these commands and ensure you understand them. Ask if you are unsure.

We can model excess mortality using Poisson regression using the following commands.

```
use grouped, clear
glm d i.end i.sex i.period_diag i.agecat i.dep, ///
    fam(pois) link(rs d_star) lnoff(y) eform
est store Grouped
```

The data set `grouped.dta` is output by `strs` and contains one observation for each row in each life table (with one life table for each combination of the variables specified in the `by` option in `strs`). The variable `end` contains the time at the end of each life table interval and is included in the model to allow the excess hazard to vary by time since diagnosis.

The `eform` option requests that the estimates be presented as the exponential of the estimated parameters (i.e. relative excess risks), rather than the estimated parameters.

(a) Let's now study the estimated excess hazard ratio for `sex`.
   i. Which of the two sexes has the lowest excess mortality, and by how much? You may find it helpful to use the `codebook` command to study how the variable `sex` is coded.
   ii. Is there evidence of a statistically significant difference in excess mortality between males and females?
   iii. The variable `period_diag` is coded 1 for patients diagnosed 1990-1999 and 2 for patients diagnosed 2000-2009. Based on the model we just fitted (i.e., a main effects model), what is the estimated effect of sex (i.e., the hazard ratio) for each of the two calendar periods.

(b) In what manner does excess mortality vary by time since diagnosis?

(c) In what manner does excess mortality vary by deprivation? Is the effect of deprivation statistically significant?

(d) Fit the main effects model without the `eform` option and ensure you understand why some values in the output change and some do not. Note that you can obtain these estimates simply by typing

```
. glm
```

which results in Stata displaying the estimates from the last model estimated using the `glm` command.

(e) In the main effects model, the estimated effect of each covariate is assumed to be the same for all combinations of other covariates. We will now relax this assumption, and allow the effect of sex to vary across the two calendar periods.

```
. glm d i.end i.sex##i.period_diag i.agecat i.dep, ///
      fam(pois) link(rs d_star) lnoff(y) eform
. lrtest Grouped
```

i. What is the interpretation of the estimate rate ratio for sex? This has value .5129927 and in Stata 14 is labelled 'f'. This is a hazard ratio, but what type of hazard and which two groups are being compared?

ii. What is the estimated effect of sex for each of the two periods?

iii. Is there evidence that the effect of sex differs between the two periods?

iv. When testing the effect of sex, do the Wald test and LR test give similar results?

v. Lets's now reparamaterise the model so we get the two sex effects (with CIs) directly.

```
. glm d i.end i.period_diag i.sex#i.period_diag i.agecat i.dep, ///
   fam(pois) link(rs d_star) lnoff(y) eform
```

Is there anything we don't get with this parameterisation that we did get with the previous one.

(f) Test the assumption of proportional excess hazards for sex by fitting an appropriate interaction term.

```
. glm d i.end##i.sex i.period_diag i.agecat i.dep, ///
   fam(pois) link(rs d_star) lnoff(y) eform
. lrtest Grouped
```

i. Is there evidence of statistically significant non-proportional excess hazards?

ii. What is the interpretation of the estimate rate ratio for sex? This has value .5287481 and in Stata 14 is labelled 'f'. This is a hazard ratio, but what type of hazard and which two groups are being compared?

iii. What is the estimated excess hazard ratio, for females/males, for the second year of follow-up?

iv. In what manner does the effect of sex vary by time since diagnosis?

2. **Model excess mortality using flexible parametric models**

We will now fit flexible parametric relative survival models (an extension of Royston-Parmar survival models). Before fitting the model, we need to add a variable to the dataset containing the expected hazard at the end of follow-up. We do this by generating variables for age and year at the end of follow-up and merging with the popmort file.

```
use melanoma_2013, clear
stset finmdy, failure(dead==1) origin(diagmdy) scale(365.24) id(id)
gen _age=floor(ageout)
gen _year = year(finmdy)
sort _year sex _age dep
merge m:1 _year sex _age dep using Lifetable_2013, keep(match master)
```

We will also restrict to the first 10 years of follow-up so as to obtain results comparable with the Poisson regression models.

```
stsplit foo, at(0 10) trim
```

The stpm2 command does not support the `i.` syntax for time varying coefficients so we will generate dummy variables.

```
tab agecat, gen(agecat)
tab dep, gen(dep)
replace sex=sex-1
replace period_diag=period_diag-1
```

(a) Now fit the main effects model.

```
stpm2 agecat2 agecat3 agecat4 agecat5 dep2 dep3 dep4 dep5 sex period_diag, ///
   bhazard(rate) df(5) scale(hazard) eform
```

How do the parameter estimates compare to the corresponding Poisson regression model?

(b) The estimates associated with the spline variables do not have a simple interpretation, so unlike with Poisson regression we cannot see from the parameter estimates how the excess hazard varies as a function of time since diagnosis. We can, however, plot the estimated excess hazard as a function of time since diagnosis. Here we will plot an estimate for each sex.

```
predict h2, hazard per(1000) ci
predict s2, survival ci

twoway (line h2 _t if agecat == 1 & sex == 0 & period_diag == 1 & dep == 1, sort) ///
(line h2 _t if agecat == 1 & sex == 1 & period_diag == 1 & dep == 1, sort)

twoway (line h2 _t if agecat == 1 & sex == 0 & period_diag == 1 & dep == 1, sort) ///
(line h2 _t if agecat == 1 & sex == 1 & period_diag == 1 & dep == 1, sort), ///
yscale(log)
```

(c) Now relax the assumption that hazards must be proportional by sex. That is, allow the effect of sex to vary with time. Effectively we are fitting an interaction between sex and time since diagnosis, just as we did in the Poisson regression model.

```
stpm2 agecat2 agecat3 agecat4 agecat5 dep2 dep3 dep4 dep5 sex period_diag, ///
   bhazard(rate) df(5) scale(hazard) eform tvc(sex) dftvc(3)
```

(d) Now plot the hazards for each sex.

```
predict h3, hazard per(1000) ci
predict s3, survival ci

twoway (line h3 _t if agecat == 1 & sex == 0 & period_diag == 1 & dep == 1, sort) ///
```

```
(line h3 _t if agecat == 1 & sex == 1 & period_diag == 1 & dep == 1, sort)

// Now on log scale
twoway (line h3 _t if agecat == 1 & sex == 0 & period_diag == 1 & dep == 1, sort) ///
(line h3 _t if agecat == 1 & sex == 1 & period_diag == 1 & dep == 1, sort), ///
yscale(log)
```

(e) Now plot the hazard ratio for sex as a function of time since diagnosis.

```
predict hr, hrnum(sex 1) ci

twoway (line hr _t if agecat == 1 & period_diag == 1 & dep == 1, sort)

twoway (rarea hr_lci hr_uci _t, sort pstyle(ci)) ///
(line hr _t, sort), yline(1) legend(off) ///
        xtitle("Years from Diagnosis") ///
        ytitle("Excess Mortality Rate Ratio")
```

(f) In theory, estimates from a flexible parametric model are sensitive to the choice of number of knots and location of knots. In practice we have not found this to be an issue. Here we fit the null model (no covariates) with different values of degrees of freedom and compare the estimates.

```
foreach df in 2 4 6 {
  stpm2, bhazard(rate) df(‘df’) scale(hazard)
  predict h_df‘df’, hazard ci
  replace h_df‘df’ = h_df‘df’ * 1000
  predict s_df‘df’, survival ci
  estimates store df‘df’
}
twoway (line h_df2 h_df4 h_df6 _t, sort lcolor(red blue black))
twoway (line s_df2 s_df4 s_df6 _t, sort lcolor(red blue black))
```

Compare the models using the AIC and BIC.

```
estimates stats df2 df4 df6
```

Which is the best fitting model?

About AIC and BIC

AIC (Akaike information criterion) and BIC (Bayesian information criterion) are two popular measures for comparing the relative goodness-of-fit of statistical models. The AIC and BIC are defined as:

$$AIC = -2\ln(\text{likelihood}) + 2k$$

$$BIC = -2\ln(\text{likelihood}) + \ln(N)k$$

where k = number of parameters estimated and N = number of observations.

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC/BIC value. Hence, the measures not only reward goodness of fit, but also include a penalty that is an increasing function of the number of estimated parameters. AIC uses a fixed constant, 2, in the penalty term whereas the penalty in BIC is a function of the number of observations. It is not always obvious how ‘number of observations’ should be defined for time-to-event data, particularly for grouped or split data. Volinsky and Raftery (2000) suggest using the number of events for N in the BIC penalty term for survival models. The `estimates stats` command contains an option `n(#)` for specifying N.

In many circumstances both the AIC and BIC will suggest the same model. For population-based survival data, the number of observations is large so BIC will penalize models with additional parameters more strongly than AIC.

2 Solutions to practical 5

1. (a)
```
. glm d i.end i.sex i.period_diag i.agecat i.dep, ///
>     fam(pois) link(rs d_star) lnoff(y) eform

(output omitted)
------------------------------------------------------------------------------
             |                 OIM
           d |      exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         end |
           1 |          1  (base)
           2 |   .9269327   .0216551    -3.25   0.001     .8854465    .9703626
           3 |   .8341632   .0213215    -7.09   0.000     .7934032    .8770171
           4 |    .64624   .0194685   -14.49   0.000     .6091871    .6855466
           5 |   .5001996   .0182512   -18.99   0.000     .4656771    .5372815
           6 |   .3823165   .0172909   -21.26   0.000     .3498855    .4177534
           7 |   .3650101   .0182197   -20.19   0.000     .3309914    .4025251
           8 |   .2925747   .0179102   -20.08   0.000     .2594954    .3298708
           9 |   .2096609   .0170127   -19.25   0.000     .1788331     .245803
          10 |    .159858   .0164677   -17.80   0.000     .1306316    .1956232
             |
         sex |
           m |          1  (base)
           f |   .5175755   .0090537   -37.65   0.000     .5001312    .5356282
             |
  period_diag|
           1 |          1  (base)
           2 |   .6635498   .0114696   -23.73   0.000     .6414463     .686415
             |
      agecat |
       15-44 |          1  (base)
       45-54 |   1.363393   .0375022    11.27   0.000     1.291836    1.438913
       55-64 |   1.678121   .0442196    19.65   0.000     1.593652    1.767067
       65-74 |   2.023413   .0547902    26.03   0.000     1.918826      2.1337
         75+ |   3.511707   .0946319    46.61   0.000     3.331044    3.702167
             |
         dep |
    Affluent |          1  (base)
           2 |   1.067071   .0264467     2.62   0.009     1.016475    1.120185
           3 |   1.117721   .0283828     4.38   0.000     1.063454    1.174758
           4 |   1.262463   .0332292     8.85   0.000     1.198986      1.3293
    Deprived |   1.577702   .0450678    15.96   0.000     1.491798    1.668553
             |
       _cons |   .0405636   .0012439  -104.51   0.000     .0381975    .0430764
      ln(y)  |          1  (exposure)
------------------------------------------------------------------------------
```

   i. Based on the fitted model, we estimate that females experience only 52% of the excess mortality experienced by males. That is, females have a 48% lower excess mortality.

   ii. Yes, it is statistically significant (p-value is low and CI does not contain 1).

   iii. Same as in part (i). The effect of sex is assumed constant for all combinations of covariates.

   (b) Excess mortality becomes progressively lower with increasing follow-up (excess hazard rations for `end` become smaller).

(c) Excess mortality is higher among the most deprived. Cannot use the pairwise tests to formally test statistical significance, even if they give us a good idea of what to expect from the global test; must test the 4 parameters as a group.

```
. test 2.dep 3.dep 4.dep 5.dep

 ( 1)  [d]2.dep = 0
 ( 2)  [d]3.dep = 0
 ( 3)  [d]4.dep = 0
 ( 4)  [d]5.dep = 0

           chi2(  4) =  299.56
         Prob > chi2 =    0.0000
```

The differences in excess mortality by deprivation are highly statistically significant.

(d) We now get log excess hazard ratios rather than hazard ratios. Note that the test statistics are unchanged, but we see that they are now the estimate divided by the SE which was not the case in the previous output. CI's are now symmetric around the point estimate.

(e)
```
. glm d i.end i.sex##i.period_diag i.agecat i.dep, ///
>     fam(pois) link(rs d_star) lnoff(y) eform

(output omitted)
------------------------------------------------------------------------------
             |                 OIM
           d |      exp(b)   Std. Err.      z    P>|z|     [95% CI]
-------------+----------------------------------------------------------------
         sex | (output omitted)
           m |          1  (base)
           f |   .5129927    .012295   -27.85   0.000  .4894521 .5376655
             |
  period_diag |
           1 |          1  (base)
           2 |   .6586112   .0145444   -18.91   0.000  .6307129 .6877435
             |
  sex#period_diag
        f#2 |   1.019123    .035518     0.54   0.587  .9518329 1.091169
             |
      agecat |
       15-44 |          1  (base)
       45-54 |    1.36349   .0375058    11.27   0.000  1.291927 1.439018
       55-64 |   1.678314   .0442256    19.65   0.000  1.593833 1.767272
       65-74 |   2.024273   .0548348    26.03   0.000  1.919602 2.134652
         75+ |   3.513666   .0947445    46.60   0.000  3.332792 3.704357
             |
         dep |
    Affluent |          1  (base)
           2 |   1.067103   .0264466     2.62   0.009  1.016508 1.120217
           3 |   1.117745   .0283827     4.38   0.000  1.063477 1.174781
           4 |   1.262458    .033228     8.85   0.000  1.198984 1.329293
    Deprived |   1.577954   .0450751    15.97   0.000  1.492036 1.668819
             |
       _cons |   .0407071   .0012754  -102.18   0.000  .0382827 .0432851
       ln(y) |          1  (exposure)
------------------------------------------------------------------------------
```

i. The model includes a sex by period interaction, which means the effect of sex is now estimated separately for each period. The parameter estimate that looks like the main effect of sex is the effect of sex during the first calendar period (the reference level of the other factor in the interaction).

ii. For period 1, 0.513, and for period 2, $0.513 \times 1.019 = 0.523$.

```
. lincom 2.sex + 2.sex#2.period_diag, eform

------------------------------------------------------------------------------
           d |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .5228024   .0132916   -25.51   0.000     .4973897    .5495135
------------------------------------------------------------------------------
```

iii. No, the z test for the interaction effect is not significant and the CI contains 1.

iv.
```
. glm d i.end i.sex i.period_diag i.agecat i.dep, ///
>    fam(pois) link(rs d_star) lnoff(y) eform
. est store Grouped

. glm d i.end i.sex##i.period_diag i.agecat i.dep, ///
>    fam(pois) link(rs d_star) lnoff(y) eform

. lrtest Grouped

Likelihood-ratio test                 LR chi2(1)  =      0.30
(Assumption: Grouped nested in .)     Prob > chi2 =    0.5869
```

The Wald test reported a Z statistic of 0.54. The LR test reported a χ^2_1 test statistic of 0.30. Note that Z^2 is χ^2_1 if Z is standard normal. We see that $0.54^2 = 0.29$ which is very close to 0.30.

v. We now get estimates of the effect of sex for each period, whereas all other parameter estimates are unchanged. This is the same model, but with a different parameterisation.

```
------------------------------------------------------------------------------
             |                 OIM
           d |   exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
sex#period_diag |
          f#1  | .5129927    .012295   -27.85   0.000     .4894521    .5376655
          f#2  | .5228024   .0132916   -25.51   0.000     .4973897    .5495135
------------------------------------------------------------------------------
```

(f)
```
. glm d i.end##i.sex i.period_diag i.agecat i.dep, ///
>     fam(pois) link(rs d_star) lnoff(y) eform

(output omitted)
------------------------------------------------------------------------------
             |                 OIM
           d |      exp(b)   Std. Err.      z    P>|z|     [95% CI]
-------------+----------------------------------------------------------------
         end |
           1 |          1  (base)
           2 |   .9573507    .028276    -1.48   0.140  .9035043 1.014406
           3 |   .8589457   .0279475    -4.67   0.000  .8058797 .9155061
           4 |   .6647691      .0257   -10.56   0.000  .6162591 .7170978
           5 |   .4858419   .0236259   -14.84   0.000  .4416743 .5344264
           6 |   .3812067   .0226661   -16.22   0.000  .3392729 .4283235
           7 |   .3526958   .0240374   -15.29   0.000  .3085945 .4030997
           8 |   .2514869   .0229593   -15.12   0.000  .2102837 .3007634
           9 |   .1705677   .0213173   -14.15   0.000  .1335104 .2179108
          10 |   .1477953     .02186   -12.93   0.000  .1106016 .1974966
             |
         sex |
           m |          1  (base)
           f |   .5287481   .0169307   -19.90   0.000  .4965843 .5629951
             |
     end#sex |
        2#f  |   .9183334   .0443228    -1.77   0.078  .8354447 1.009446
        3#f  |   .9274007   .0486698    -1.44   0.151  .8367516  1.02787
        4#f  |   .9312216   .0572191    -1.16   0.246  .8255641 1.050401
        5#f  |   1.065706   .0780272     0.87   0.385  .9232417 1.230154
        6#f  |   1.003787   .0913091     0.04   0.967  .8398702 1.199695
        7#f  |   1.073454   .1067562     0.71   0.476  .8833455 1.304477
        8#f  |    1.35052   .1654668     2.45   0.014  1.062212 1.717082
        9#f  |   1.506164    .245417     2.51   0.012    1.0944 2.072854
       10#f  |    1.17191   .2404668     0.77   0.439  .7838525 1.752081
             |
     period_diag
           1 |          1  (base)
           2 |   .6632693   .0114701   -23.74   0.000   .641165 .6861356
             |
      agecat |
       15-44 |          1  (base)
       45-54 |   1.364029   .0375067    11.29   0.000  1.292463 1.439558
       55-64 |   1.678921     .04424    19.66   0.000  1.594413 1.767908
       65-74 |   2.022364   .0547677    26.01   0.000   1.91782 2.132607
         75+ |   3.508631   .0946045    46.55   0.000  3.328024 3.699039
             |
         dep |
    Affluent |          1  (base)
           2 |   1.067592   .0264618     2.64   0.008  1.016968 1.120737
           3 |   1.117929   .0283942     4.39   0.000   1.06364 1.174989
           4 |   1.263187   .0332495     8.88   0.000  1.199672 1.330065
    Deprived |   1.578162   .0450867    15.97   0.000  1.492223 1.669052
             |
       _cons |   .0402423   .0013021   -99.29   0.000  .0377694 .0428772
       ln(y) |          1  (exposure)
------------------------------------------------------------------------------
```

i. The time by sex interaction is highly significant, indicating evidence of non-proportional excess hazards by sex.

```
. lrtest Grouped

Likelihood-ratio test                     LR chi2(9)  =     24.59
(Assumption: Grouped nested in .)         Prob > chi2 =    0.0035
```

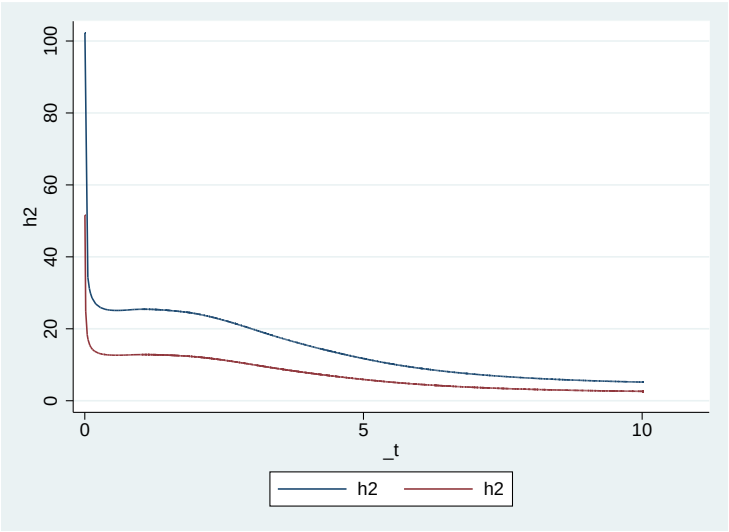
ii. The effect of sex during the first year of follow-up.

iii. $0.5287481 \times 0.9183334 = 0.49$

iv. The effect of sex (i.e., the female superiority in survival) is greater early in the follow-up compared to later.

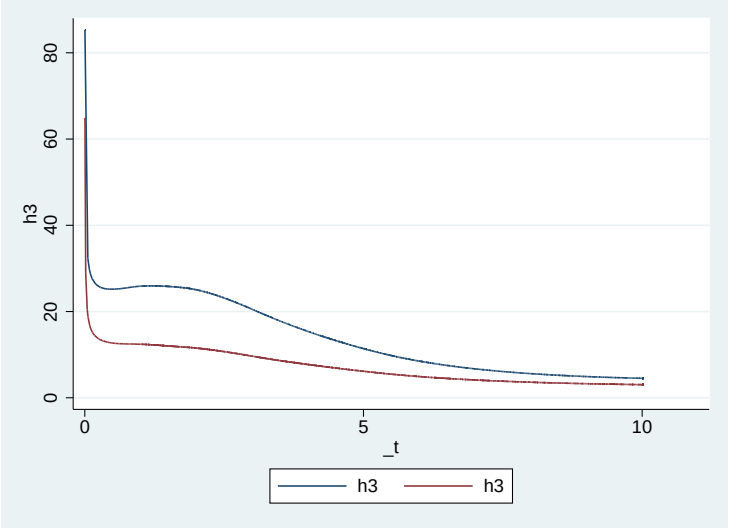
2. (a) With the exception of the effect of time, the estimated excess hazard ratios are very similar to those obtained in Q 1 (a). We are now modelling the effect of time using a spline whereas in Q1 we modelled it using a step function.

(b) Following is the plot of the two hazard functions on the hazard scale. The two hazards are constrained to be proportional. On the log-hazard scale we see that there is a constant difference between the two lines.

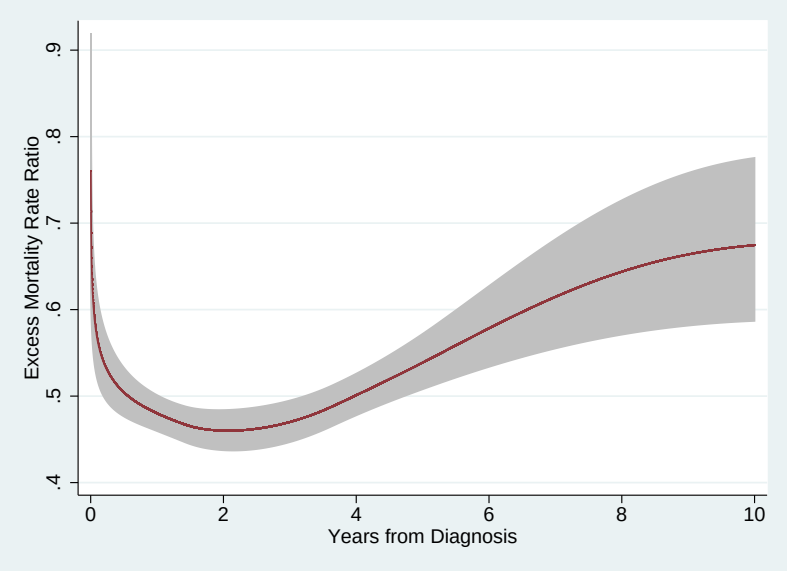


(c) Output not shown since the graphs are more interesting than the parameter estimates.

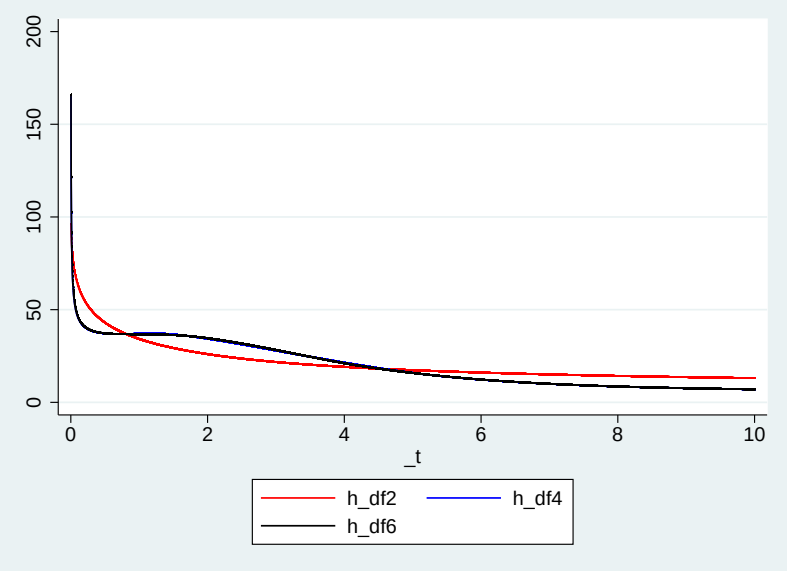
(d) The curves are no longer proportional. Compared to the proportional hazards model fitted in (b), we see that there are now slightly bigger differences between males and females during the first 5 years.



(e) As we saw with the Poisson regression model, the effect of sex is larger during the earlier follow-up years.



(f) We see that 2df is not able to capture the shape of the hazard, but there is little difference between 4 and 6 df.



The AIC and BIC both suggest the model with 4 df over the one with 6 (with 2df being the worst fit).

```
. estimates stats df2 df4 df6, n(38143)
-----------------------------------------------------------------------------
Model |    Obs    ll(null)   ll(model)     df          AIC          BIC
------+----------------------------------------------------------------------
  df2 |  38143           .   -109056.6      3     218119.2     218144.8
  df4 |  38143           .   -108668.5      5       217347     217389.8
  df6 |  38143           .   -108668.1      7     217350.1       217410
-----------------------------------------------------------------------------
        Note:  N=38143 used in calculating BIC
```