BIOSTAT III: Survival Analysis

Examination

December 15, 2010

Time: 12:30 -14:30

Code (please do not write your name):

- Time allowed is 2 hours.
- Please try and write your answers on the exam sheet. You may use separate paper if absolutely necessary. Your working and motivation for your answer, not just the final answer, will be assessed when grading the examination.
- The exam contains 6 question with several parts. The marks available for each part are indicated.
- A score of 13 marks or more out of a possible 25 will be required to obtain a passing grade.
- The questions may be answered in English or Swedish (or a combination thereof).
- A non-programmable scientific calculator (i.e., with ln() and exp() functions) will most probably be useful. You may not use a mobile phone or other communication device as a calculator or for any other purpose.
- The exam is not 'open book' but each student will be allowed to bring one A4 sheet of paper into the exam room which may contain, for example, hand-written notes or photocopies from textbooks/lecture notes etc. Both sides of the page may be used.
- The exam supervisors have been advised not to answer any questions you may have regarding the content of the exam. If you believe a question contains an error or is ambiguous then please write a note with your answer indicating how you have interpreted the question.
- Tables of critical values of the χ^2 distribution are provided on the last page.

A study was performed to quantify the benefit of a new inflatable device to protect elderly persons from hip fractures resulting from falls. The device is worn around the hips at all times. It is hypothesized that the device will reduce the incidence of hip fractures in this population. Forty-eight women over the age of 60, with no previous history of hip trauma, were recruited for this study. Of these 48 women 28 were randomly given the device and instructed how to wear it. The remaining 20 women were not provided with the device. In addition all women, whether of not they were assigned the protective device, were given a new experimental bone fortifying drug at enrollment of the study. The initial dosage of the drug came in two levels (50 mg or 100 mg) and was assigned randomly to the women. At inclusion in the study blood calcium levels were assessed for all women. All 48 women entered the study the 1 September 1995 and were followed up for hip fracture, with attained age as the underlying time-scale. Women who had not experienced the event at the closing date of the study were censored. Moreover, it was decided at study onset that, if a woman was ever hospitalised during follow up, she would also be censored at the date of hospitalisation (as she would no longer be considered at risk of falling and fracturing her hip). The time to hip fracture or censoring was recorded in days.

```
. /** stset the data using attained age as the timescale **/
. stset exitdate, failure(fracture == 1) enter(entrydate) origin(birthdate)
                  id(id) scale(365.24)

                id:  id
     failure event:  fracture == 1
obs. time interval:  (exitdate[_n-1], exitdate]
 enter on or after:  time entrydate
 exit on or before:  failure
    t for analysis:  (time-origin)/365.24
            origin:  time birthdate

--------------------------------------------------------------------------------
       48  total obs.
        0  exclusions
--------------------------------------------------------------------------------
       48  obs. remaining, representing
       48  subjects
       31  failures in single failure-per-subject data
  61.1105  total analysis time at risk, at risk from t =         0
                              earliest observed entry t =        62
                                   last observed exit t =   83.3142



/*Distribution of calcium levels among the study participants*/
. sum cal

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     calcium |        48    9.940208    1.394182       7.25      12.32
```

```
. tab protect

      wears |
     device |      Freq.     Percent        Cum.
------------+-----------------------------------
     0 = No |         20       41.67       41.67
    1 = Yes |         28       58.33      100.00
------------+-----------------------------------
      Total |         48      100.00
```

```
. tab dose (initial dose of the bone-fortifying drug)

       dose |      Freq.     Percent        Cum.
------------+-----------------------------------
         50 |         26       54.17       54.17
        100 |         22       45.83      100.00
------------+-----------------------------------
      Total |         48      100.00
```

```
. /** split the data according to attained age **/
. stsplit ageband, at(60 70 80 90)
(1 observation (episode) created)

. tab ageband

    ageband  |       Freq.     Percent        Cum.
-------------+-----------------------------------
60 = 60 - 69 |         19       38.78       38.78
70 = 70 - 79 |         25       51.02       89.80
80 = 80 - 89 |          5       10.20      100.00
-------------+-----------------------------------
      Total  |         49      100.00
```

1. We now estimate a Cox regression model.

```
/*Model A*/
. stcox i.protect i.dose calcium

Cox regression -- no ties

No. of subjects =           48                     Number of obs   =        58
No. of failures =           31
Time at risk    =  61.11050268
                                                   LR chi2(3)      =     22.15
Log likelihood  =   -27.278096                     Prob > chi2     =    0.0001

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.protect |   .1114077   .0636005    -3.84   0.000     .0363899    .3410746
    100.dose |   .5174499     .29034    -1.17   0.240     .1722904    1.554088
     calcium |   1.522428   .7548618     0.85   0.397      .576086    4.023333
------------------------------------------------------------------------------
```

(a) (1 mark) Is the effect of calcium adjusted for age? Motivate your answer.

(b) (1 mark) Interpret the estimated effect of calcium. You do not need to comment on statistical significance.

(c) (1 mark) What is the effect of receiving the higher dose of the bone fortifying drug among individuals with a calcium level of 10?

2. We now fit another Cox model in which we include an interaction term between the variables calcium and initial dosage of the new drug.

```
. /*Model B*/
. stcox  i.protect i.dose*calcium

Cox regression -- no ties

No. of subjects =           48                     Number of obs   =        58
No. of failures =           31
Time at risk    =  61.11050268
                                                   LR chi2(4)      =     23.56
Log likelihood  =   -26.572158                     Prob > chi2     =    0.0001

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.protect |    .093167   .0576652    -3.83   0.000     .0276959     .313407
   100.dose  |   .0016046    .007995    -1.29   0.197     9.21e-08    27.95682
     calcium |   1.081991   .6367542     0.13   0.893     .3414203    3.428923
100.dose*cal |   1.786063   .8826155     ****   *****     ********    ********
------------------------------------------------------------------------------
```

(a) (1 mark) Interpret the incidence rate ratio for the variable labelled 100.dose.

(b) (2 marks) What is the effect of initial dosage of the new bone-fortifying drug for an individual with calcium level 8?

(c) (2 marks) Is the effect of initial dosage of the new bone-fortifying drug modified by calcium level? Assess this formally. Remember to state the null hypothesis, alternative hypothesis, value of the test statistic, assumed distribution of the test statistic under the null hypothesis, and a comment on statistical significance.

3. For model A, the proportional hazards assumption was tested using Schoenfeld residuals. The output from the formal test and a plot of the residuals for the calcium variable are displayed on the next page.

   (a) (1 mark) In the plot, explain what characteristics you would look for when assessing the proportional hazards assumption.

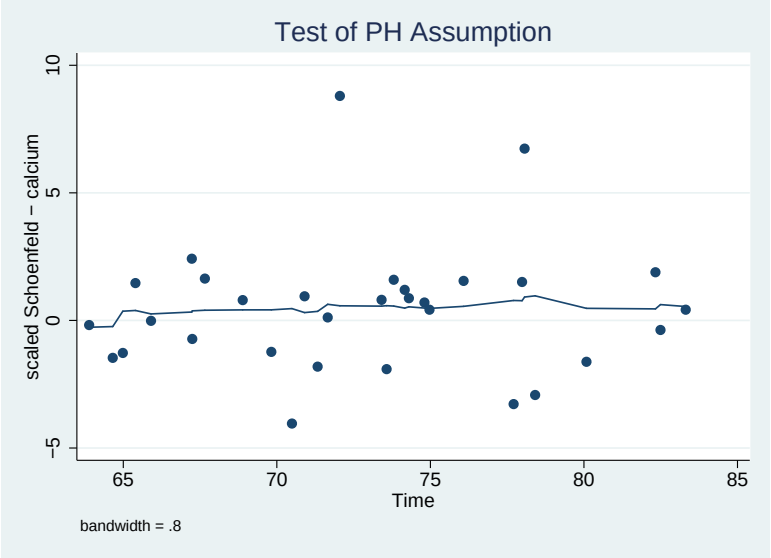
   (b) (2 marks) State the formal null hypothesis and the alternative hypothesis for the test labelled `1.protect` in the table.

   (c) (1 mark) What is your conclusion about the assumption of proportional hazards for the effect of calcium? Motivate your answer.

```
. estat phtest, detail

      Test of proportional-hazards assumption

      Time:  Time
      ------------------------------------------------------------------------
                  |       rho             chi2       df       Prob>chi2
      ------------+-----------------------------------------------------------
      0b.protect  |            .             .        1             .
      1.protect   |      0.06818          0.11        1          0.7405
      50b.dose    |            .             .        1             .
      100.dose    |     -0.22310          1.67        1          0.1966
      calcium     |      0.06460          0.11        1          0.7438
      ------------------------------------------------------------------------
```



4. We now use the same data and instead fit a Poisson regression model. The output from the fitted model is provided below.

```
# MODEL C
. streg i.ageband i.protect i.dose calcium, dist(exp) nohr

Exponential regression -- log relative-hazard form

No. of subjects =           48                     Number of obs   =        49
No. of failures =           31
Time at risk    =  61.11050268
                                                   LR chi2(5)      =     29.93
Log likelihood  =    95.771057                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ageband |
         70  |   .1426858   .5605953     0.25   0.799    -.9560609    1.241432
         80  |   .5348323   .9683479     0.55   0.581    -1.363095    2.432759
             |
   1.protect |  -1.798042   .3779128    -4.76   0.000    -2.538738   -1.057347
    100.dose |  -.6854703   .4085292    -1.68   0.093    -1.486173    .1152322
     calcium |  -.2755033   .2279115    -1.21   0.227    -.7222017    .1711951
       _cons |   3.298738   2.609902     1.26   0.206    -1.816576    8.414052
------------------------------------------------------------------------------
```

(a) (2 marks) What is the hazard ratio comparing an individual with calcium level 12 to an individual with calcium level 8? Is this hazard ratio statistically significant? Motivate your answer.

(b) (2 marks) What is the estimated incidence rate for an individual with the protective device, low dose of the fortifying drug, a calcium level of 10 and aged 72?

5. The following questions are not based on the models above.

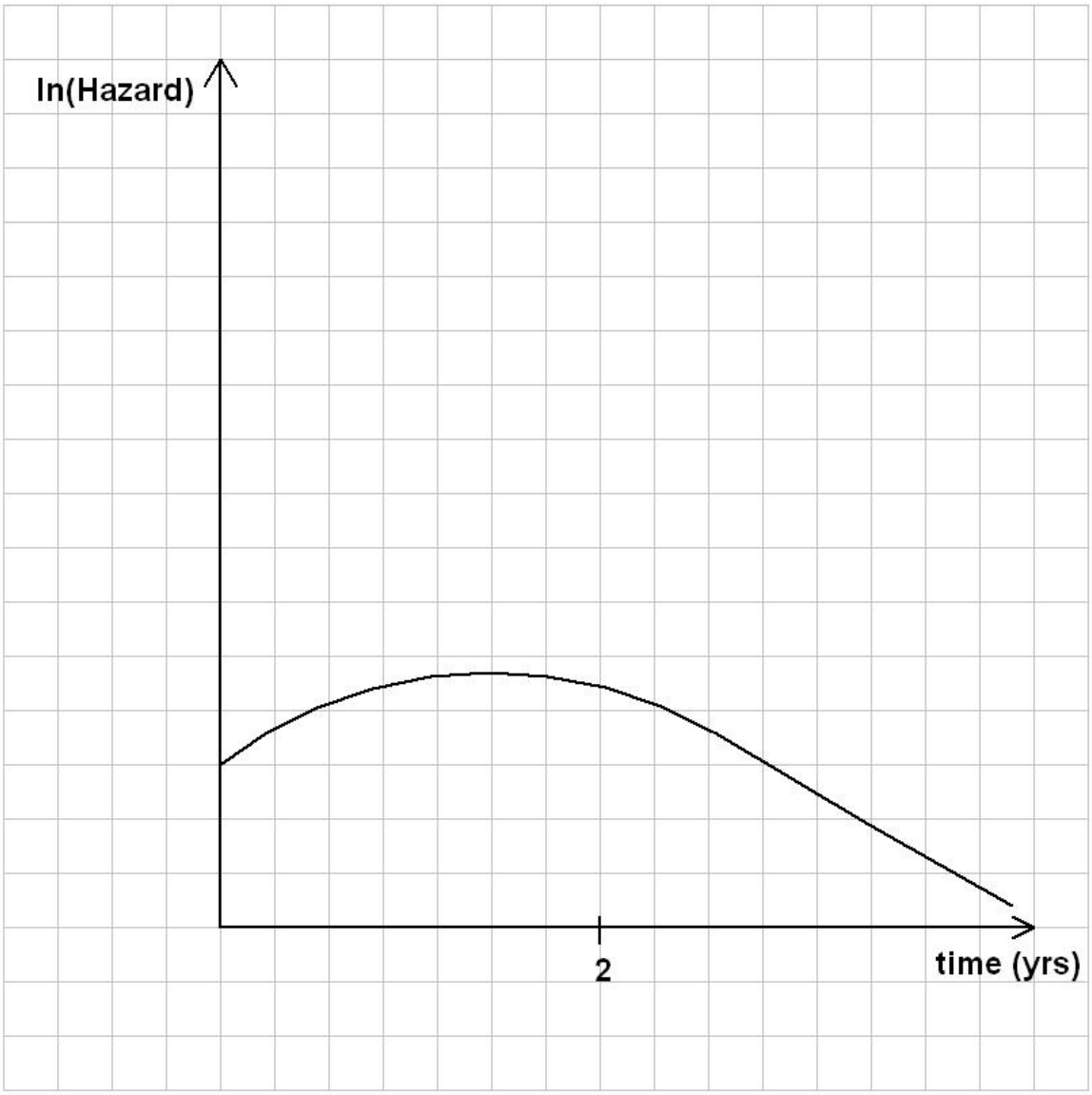
   (a) (1 mark) Provide the definition of a competing risk.

   (b) (1 mark) Describe a scenario when a nested case-control study is preferable to a cohort study.

   (c) (1 mark) Describe how controls are selected in a nested case-control study.

6. The mortality rate after a diagnosis of cancer depends on time since diagnosis as well as the presence of metastasis at diagnosis. Assume that a cohort study is conducted to determine whether the mortality rate depends on a binary exposure of interest. A Cox regression model is fitted to the data with time since diagnosis as the timescale. In the fitted model, the estimated hazard ratio measuring the effect of metastases at diagnosis (compared to no metastasis) was 3.

   (a) (1 mark) Imagine that, for the patients with no metastasis (i.e. the reference group), the logarithm of the mortality rate has the form shown in the figure below. In the figure below draw the line representing the logarithm of the mortality rates for patients with metastasis. You should indicate how the estimated hazard ratio is represented on the graph.



   (b) (2 marks) Imagine that we now instead fit a Poisson model where time since diagnosis is split into 4 annual intervals. Please add to the previous graph, how you would expect the log hazard rates from the Poisson model to look for the two groups respectively.

(c) (1 mark) Describe under what situation the two models provide the same estimate for the effect of metastasis at diagnosis.

(d) (1 mark) Give one example of a situation of when a Poisson model would be preferable to a Cox model.

(e) (1 mark) Explain how the proportional hazards assumption can be assessed in the Poisson model.