Solutions, Biostat III exam, December 4, 2009

1. (a) (4 marks) Based on models 1 and/or 2, is there evidence of an association between histological type and age group? If so, describe how the distribution of histological type varies by age.

   Answer:From model 1 we see that patients with papillary carcinoma experience 50% lower mortality than patients with follicular carcinoma. From model 2 we see that age is strongly associated with mortality (higher age implies higher mortality). We also see that age confounds the association between histology and mortality (HR changes from 0.5 to 0.7 on adjusting for age). This implies age must be associated with histology. Some of the apparent 'benefit' of having pappilary carcinoma (compared to follicular) is explained by age. That is, those with the less fatal histology (papillary) are more likely to have a favourable age distribution (young age). This means papillary carcinoma must be relatively more common among the young ages (and follicular carcinoma relatively more common among the older ages).

```
. tab agegrp papillary, row

Age group  | Follicula  Papillary |     Total
-----------+----------------------+----------
      0-39 |       342      1,077 |     1,419
           |     24.10      75.90 |    100.00
-----------+----------------------+----------
     40-49 |       293        667 |       960
           |     30.52      69.48 |    100.00
-----------+----------------------+----------
     50-59 |       385        659 |     1,044
           |     36.88      63.12 |    100.00
-----------+----------------------+----------
     60-69 |       490        620 |     1,110
           |     44.14      55.86 |    100.00
-----------+----------------------+----------
       70+ |       456        565 |     1,021
           |     44.66      55.34 |    100.00
-----------+----------------------+----------
     Total |     1,966      3,588 |     5,554
           |     35.40      64.60 |    100.00
```

   (b) (2 marks) Based on model 2, complete the 5 missing cells in the table below with the hazard ratio for each of the 5 categories compared to individuals diagnosed with follicular carcinoma in 1958–67. That is, the joint reference category is follicular carcinoma diagnosed in 1958–67. The hazard ratios you provide should be applicable for males aged 0–39.

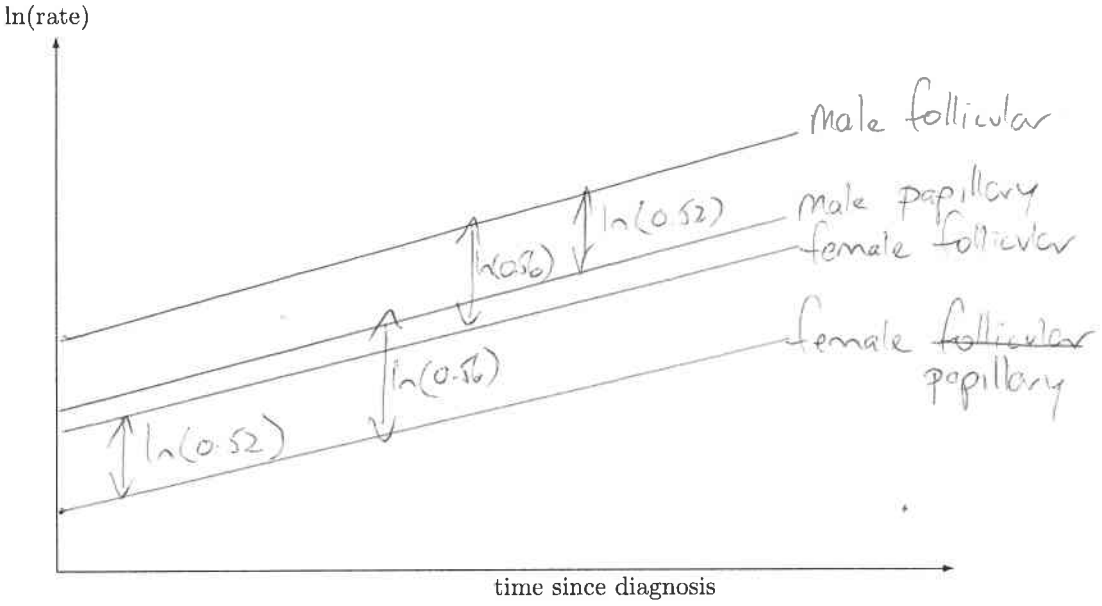
   Answer:

	follicular	papillary
1958–67	1.00	0.71
1968–77	0.70	$0.71 \times 0.70 = 0.50$
1978–87	0.40	$0.71 \times 0.40 = 0.28$

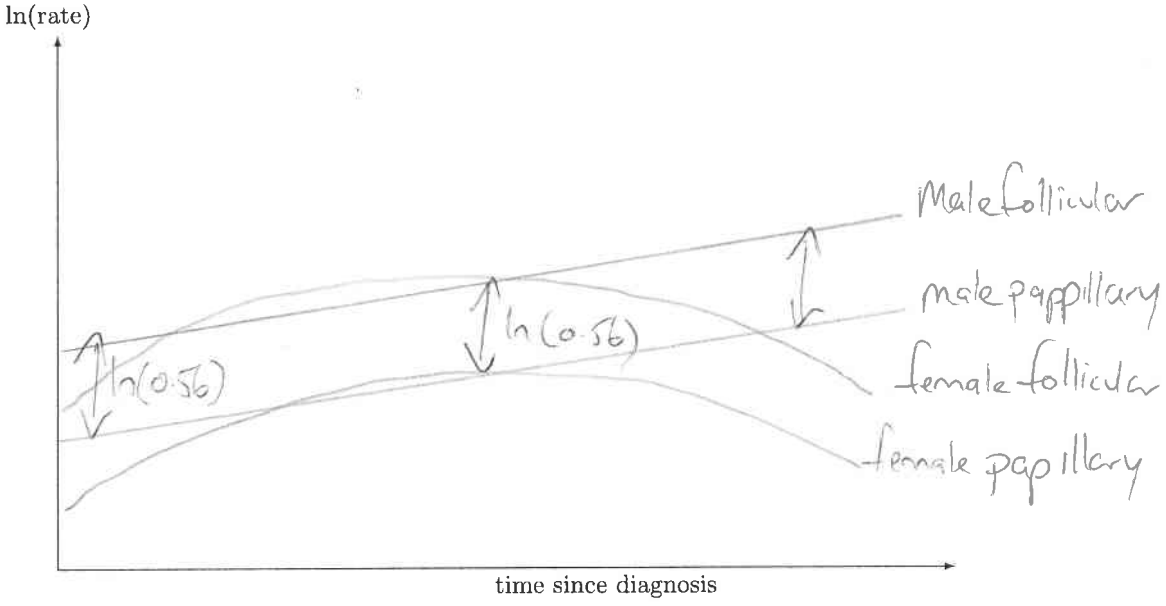
   (c) (1 mark) How would the numbers in the table in the previous question change if you instead constructed the table for females aged 0–39?

   Answer: The estimates would remain unchanged since the model does not include interaction terms.

(d) (2 marks) The log hazards do not necessarily have to be linear (as shown below), but there should be a constant difference between them and that difference is equal to the logarithm of the relevant hazard ratio.



(e) (2 marks) The stratified Cox model, stratified by sex, allows the baseline hazards to differ between males and females. The effect of histology is, however, assumed to be proportional for each gender with the same hazard ratio. That is, there is a constant difference between the log hazards.



(f) (3 marks) Based on model 3, what is the estimated hazard ratio (mortality rate ratio) and 95% confidence interval comparing papillary to follicular carcinoma?

Answer: Point estimate exp(-0.6992) = 0.497

A 95 % CI for the log hazard ratio is $-0.6992 \pm 0.0648 = (-0.826, -0.572)$.

A 95 % CI for the hazard ratio is therefore $(\exp(-0.826), \exp(-0.572)) = (0.438, 0.564)$.

(g) (3 marks) Based on model 3, perform a formal hypothesis test of the effect of sex. You should 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.

Answer: We'll perform a Wald test since we do not have all relevant information to perform a likelihood ratio test.

The null hypothesis to be tested is:

$$H_0 : \beta_{sex} = 0$$

against

$$H_A : \beta_{sex} \neq 0.$$

The test statistic is defined as

$$z^2 = (\frac{\hat{\beta}}{s.e(\hat{\beta})})^2 = 70.4$$

and has a χ^2 distribution with 1 df under H_0.

The critical value of a χ^2 with 1 degree of freedom is 3.841 at the 5 % significance level. Since our test statistic is 70.4 the Wald test is highly significant and we reject the null hypothesis that there is no effect of sex.

Alternatively, the test statistic could be defined as

$$z = (\frac{\hat{\beta}}{s.e(\hat{\beta})}) = -8.38$$

which has a standard normal distribution under H_0. The critical value at the 5% level is -1.96 (there were no tables provided but this should be known) so we reject the null hypothesis.

This type of hypothesis test (parameter estimate divided by its standard error) is provided in the output of essentially every statistical model (e.g., linear, logistic, Poisson, Cox) in every standard statistical package. See, for example, the output on the next page. Note that the testing is always performed on the original scale so the test statistics and p-values are identical for both the original scale and exponentiated scale.

(h) (3 marks) Based on model 3, what is the estimated mortality rate (deaths due to DTC per 1000 person-years) during the third year of follow-up for females diagnosed with papillary carcinoma.

Answer: The model is ln(rate) = linear predictor. The linear predictor for females diagnosed with papillary carcinoma during the third year of follow-up is

$$-2.094 - 1.0728 - 0.5704 - 0.6992 = -4.4364$$

That is, ln(rate)=-4.4364. The mortality rate per person year is given by $\exp(-4.4364) = 0.0118$. Thus, the mortality rate per 1000 person years is $0.0118 \times 1000 = 11.8$.

2. (3 marks)

   Answer: Calendar period appears to be a potential confounder, since it is associated with both the exposure (treatment) and the outcome (mortality). However, matching on calendar year of diagnosis would be overmatching and therefore not a sensible strategy. That is, by matching on calendar period the treatments would be identical for many (possibly all) of the matched sets so we would not be able to efficiently estimate the treatment effect.

Following is the full output for model 3, on both the original scale and exponentiated scale.

```
. xi: streg i.fu i.sex papillary, dist(exp) nohr
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Ifu_1 |  -.7809395   .1060801    -7.36   0.000    -.9888526   -.5730264
      _Ifu_2 |  -1.072776   .1198709    -8.95   0.000    -1.307719   -.8378337
      _Ifu_3 |  -1.018338   .1193797    -8.53   0.000    -1.252318   -.7843578
      _Ifu_4 |  -1.208827    .130702    -9.25   0.000    -1.464998   -.9526557
      _Ifu_5 |  -1.753111   .1666616   -10.52   0.000    -2.079762   -1.426461
      _Ifu_6 |  -1.649412   .1616002   -10.21   0.000    -1.966142   -1.332681
      _Ifu_7 |  -1.759166   .1723537   -10.21   0.000    -2.096973   -1.421359
      _Ifu_8 |   -1.83878   .1811221   -10.15   0.000    -2.193772   -1.483787
      _Ifu_9 |  -2.079463   .2045917   -10.16   0.000    -2.480456   -1.678471
     _Ifu_10 |  -2.053329   .2045997   -10.04   0.000    -2.454337   -1.652321
     _Ifu_11 |  -2.288301   .2310889    -9.90   0.000    -2.741227   -1.835375
     _Ifu_12 |  -2.149023   .2210386    -9.72   0.000    -2.582251   -1.715795
     _Ifu_13 |   -2.61796   .2834165    -9.24   0.000    -3.173446   -2.062473
     _Ifu_14 |  -2.716724    .307103    -8.85   0.000    -3.318634   -2.114813
     _Isex_2 |  -.5704134   .0680352    -8.38   0.000    -.7037599   -.4370669
   papillary |  -.6991869    .064821   -10.79   0.000    -.8262336   -.5721401
       _cons |  -2.094152   .0781802   -26.79   0.000    -2.247382   -1.940921
------------------------------------------------------------------------------

. streg
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Ifu_1 |   .4579756   .0485821    -7.36   0.000     .3720033    .5638165
      _Ifu_2 |   .3420575   .0410027    -8.95   0.000     .2704362    .4326467
      _Ifu_3 |   .3611949   .0431193    -8.53   0.000     .2858416    .4564127
      _Ifu_4 |   .2985473   .0390207    -9.25   0.000     .2310784    .3857153
      _Ifu_5 |   .1732341   .0288715   -10.52   0.000     .1249599    .2401574
      _Ifu_6 |   .1921629   .0310536   -10.21   0.000     .1399959    .2637691
      _Ifu_7 |   .1721885   .0296773   -10.21   0.000     .1228277    .2413858
      _Ifu_8 |   .1590114   .0288005   -10.15   0.000     .1114954    .2267773
      _Ifu_9 |   .1249973   .0255734   -10.16   0.000     .0837051    .1866592
     _Ifu_10 |   .1283071   .0262516   -10.04   0.000     .0859201    .1916047
     _Ifu_11 |   .1014386   .0234413    -9.90   0.000     .0644911    .1595536
     _Ifu_12 |    .116598   .0257727    -9.72   0.000     .0756037    .1798207
     _Ifu_13 |   .0729516   .0206757    -9.24   0.000     .0418591    .1271391
     _Ifu_14 |   .0660909   .0202967    -8.85   0.000     .0362022    .1206559
     _Isex_2 |   .5652917   .0384597    -8.38   0.000     .4947217    .6459282
   papillary |   .4969893   .0322153   -10.79   0.000     .4376947    .5643164
------------------------------------------------------------------------------
```