

Homework #3 Key

1. We pooled the group 1B with Group IIA to form an adjuvant chemotherapy (treatment) group to be compared with group 1A (untreated). In this case “failure” is a relapse and we observe that there are 9 unique failure times (I=9) when all the data are pooled. A conservative approximation to the log-rank statistic is achieved by:

$$T_{con} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

where,

$$O_k - E_k = \sum_{i=1}^9 (O_{ki} - E_{ki}), \text{ sum of differences over all 9 times, } k=1,2$$

with

$$O_{ki} = d_{ki}, \text{ the number of observed failures in group } k \text{ at time } i, \text{ and}$$

$$E_{ki} = \frac{n_{ki} d_i}{n_i}.$$

I designated Group 1 as the untreated or control, 1A, and the adjuvant chemotherapy group, combined 1B and IIA, as Group 2.

From the tables on the next page, the test statistic is:

$$T_{cons} = \frac{(8 - 3.113)^2}{3.113} + \frac{(4 - 8.887)^2}{8.887} = 10.359$$

Under the null hypothesis of no difference in the survival of the two groups  $T_{con}$  has an approximate  $\chi_1^2$  distribution. The p-value associated with the value of the test statistic under this distribution is 0.00129, which is highly significant. So we can conclude that there is strong evidence to suggest that the survival experience of the two groups differs.

The exact log-rank statistic as presented on pages 5.6 and 5.7 of your notes is:

$$T_L^2 = \frac{\left[ \sum_{i=1}^9 (O_i - E_i) \right]^2}{\sum_{i=1}^9 V_i}, \text{ where } O_i = d_{ki}, E_i = \frac{n_{ki} d_i}{n_i}, \text{ where } k=1 \text{ or } 2, \text{ since only}$$

one group needs to be calculated (under  $H_0$ ). using:

$$V_i = \frac{n_{1i}n_{2i}d_i s_i}{n_i^2(n_i - 1)}$$

The variances are included in the table on the following page and the

completed calculation of the test statistic results in a value of  $T_L^2=10.78$  which has a p-value of 0.001 under an approximate  $\chi_1^2$  distribution. This test statistic is identical to that produced by STATA.

```
. sts test trt, logrank
```

```
Log-rank test for equality of survivor functions
```

| Group | Events<br>observed | expected |  |  |
|-------|--------------------|----------|--|--|
| 1     | 8                  | 3.11     |  |  |
| 2     | 4                  | 8.89     |  |  |
| Total | 12                 | 12.00    |  |  |

chi2(1) = 10.78  
Pr>chi2 = 0.0010

There are 9 2x2 tables needed for the calculations in problem #1:

| t <sub>1</sub> = 2  | Untreated | Treated | Total | t <sub>6</sub> = 16 | Untreated | Treated | Total |
|---------------------|-----------|---------|-------|---------------------|-----------|---------|-------|
| Relapsed            | 1         | 0       | 1     | Relapsed            | 1         | 1       | 2     |
| Survived            | 14        | 28      | 42    | Survived            | 6         | 25      | 31    |
| At Risk             | 15        | 28      | 43    | At Risk             | 7         | 26      | 33    |
|                     |           |         |       |                     |           |         |       |
| t <sub>2</sub> = 3  | Untreated | Treated | Total | t <sub>7</sub> = 19 | Untreated | Treated | Total |
| Relapsed            | 1         | 0       | 1     | Relapsed            | 0         | 1       | 1     |
| Survived            | 13        | 28      | 41    | Survived            | 5         | 24      | 29    |
| At Risk             | 14        | 28      | 42    | At Risk             | 5         | 25      | 30    |
|                     |           |         |       |                     |           |         |       |
| t <sub>3</sub> = 9  | Untreated | Treated | Total | t <sub>8</sub> = 30 | Untreated | Treated | Total |
| Relapsed            | 1         | 1       | 2     | Relapsed            | 1         | 0       | 1     |
| Survived            | 12        | 27      | 39    | Survived            | 3         | 16      | 19    |
| At Risk             | 13        | 28      | 41    | At Risk             | 4         | 16      | 20    |
|                     |           |         |       |                     |           |         |       |
| t <sub>4</sub> = 10 | Untreated | Treated | Total | t <sub>9</sub> = 37 | Untreated | Treated | Total |
| Relapsed            | 2         | 0       | 2     | Relapsed            | 0         | 1       | 1     |
| Survived            | 10        | 27      | 37    | Survived            | 2         | 12      | 14    |
| At Risk             | 12        | 27      | 39    | At Risk             | 2         | 13      | 15    |
|                     |           |         |       |                     |           |         |       |
| t <sub>5</sub> = 15 | Untreated | Treated | Total |                     |           |         |       |
| Relapsed            | 1         | 0       | 1     |                     |           |         |       |
| Survived            | 8         | 26      | 34    |                     |           |         |       |
| At Risk             | 9         | 26      | 35    |                     |           |         |       |

| Untreated |          | Treated  |          | Variance |
|-----------|----------|----------|----------|----------|
| Observed  | Expected | Observed | Expected |          |
| 1         | 0.349    | 0        | 0.651    | 0.227    |
| 1         | 0.333    | 0        | 0.667    | 0.222    |
| 1         | 0.634    | 1        | 1.366    | 0.422    |
| 2         | 0.615    | 0        | 1.385    | 0.415    |

|   |       |   |       |       |  |
|---|-------|---|-------|-------|--|
|   |       |   |       |       |  |
| 1 | 0.257 | 0 | 0.743 | 0.191 |  |
| 1 | 0.424 | 1 | 1.576 | 0.324 |  |
| 0 | 0.167 | 1 | 0.833 | 0.139 |  |
| 1 | 0.211 | 0 | 0.800 | 0.160 |  |
| 0 | 0.133 | 1 | 0.867 | 0.116 |  |

2 a) Using STATA, determine the value of the log rank test for the effect of gender on the risk of recurrence as a function of time since end of primary episode.

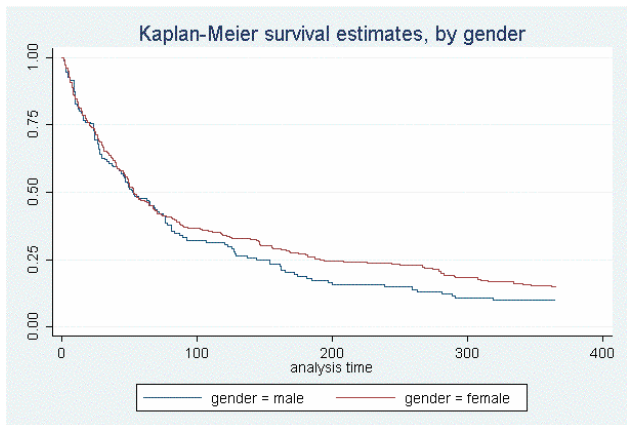
When we compare survival function by gender, Logrank test shows p-value of 0.1968, which is less than critical value of 0.005. The evidence against the null hypothesis that there is no difference between female and male groups is not enough to reject the null hypothesis. In other words, based on Logrank test, the survival functions between male and female groups is not statistically significantly different. This is not surprising when we look at the survival curves plotted by sex.

```
. sts test gender, logrank

      failure _d:  cens
      analysis time _t:  tmos

Log-rank test for equality of survivor
functions
-----
gender | Events
      | observed   expected
-----+-----
Male   |      128     116.53
Female |      244     255.47
-----+-----
Total  |      372     372.00

      chi2(1) =      1.67
      Pr>chi2 =      0.1968
```



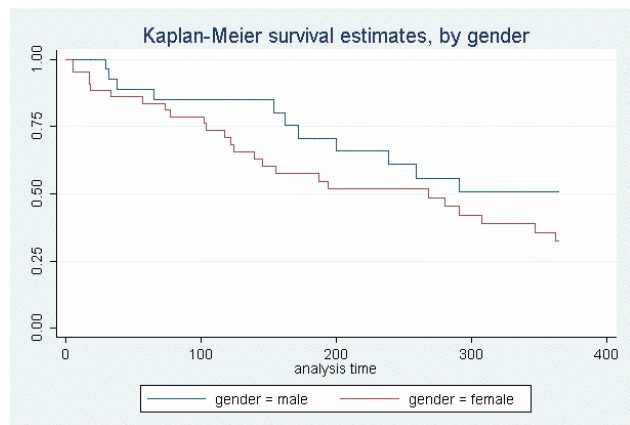
2 b) Now stratify the analysis into three strata on the basis of HSV type. Determine (i) a separate log-rank test for gender within each stratum; and (ii) an adjusted log rank test for gender effect. Explain and interpret your findings.

```
. sts test gender, strata(group)
detail

Stratified log-rank test for
equality of survivor functions
-----
-> group = 1

gender | Events
      | observed   expected
-----+-----
Male   |      11     15.05
Female |      25     20.95
-----+-----
Total  |      36     36.00

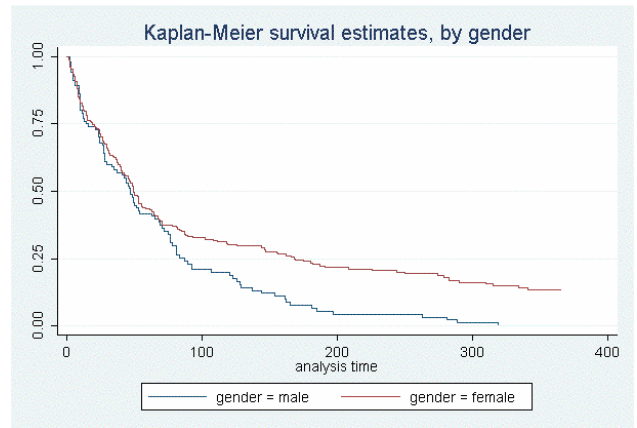
      chi2(1) =      1.88
      Pr>chi2 =      0.1704
```



```
-> group = 2
```

| gender | Events   |          |
|--------|----------|----------|
|        | observed | expected |
| Male   | 96       | 74.06    |
| Female | 185      | 206.94   |
| Total  | 281      | 281.00   |

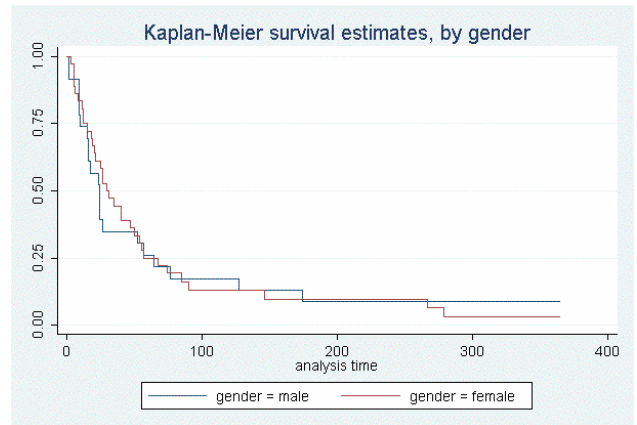
chi2(1) = 9.16  
Pr>chi2 = 0.0025



```
-> group = 3
```

| gender | Events   |          |
|--------|----------|----------|
|        | observed | expected |
| Male   | 21       | 20.72    |
| Female | 34       | 34.28    |
| Total  | 55       | 55.00    |

chi2(1) = 0.01  
Pr>chi2 = 0.9382



```
-> Total
```

| gender | Events   |             |
|--------|----------|-------------|
|        | observed | expected(*) |
| Male   | 128      | 109.83      |
| Female | 242      | 262.17      |
| Total  | 374      | 374.00      |

(\*) sum over calculations within group

chi2(1) = 4.47  
Pr>chi2 = 0.0344

When this cohort is divided based on the type of primary episode, Logrank tests evaluating whether gender is related to recurrence rates yield somewhat conflicting results. Among individuals with a primary HSV1 infection, gender is not statistically significantly associated with event times, though males appear to have consistently later times to first recurrence. Conversely, among HSV1 antibody-negative individuals with a primary HSV2 infection, males have earlier times to first recurrence, and this difference is statistically significant. Among HSV1 antibody-positive individuals with a primary HSV2 infection, the curves depicting the recurrence experience of males and females cross several times, barely distinguishable from one another. This similarity results in a highly nonsignificant log rank p-value.

In conclusion, it appears that after adjusting for HSV primary infection type, the log-rank test indicates that males do indeed have shorter times to first HSV recurrence. The comparison of event rates across genders appears to be confounded by type of primary episode.

Cross-tabulating HSV type and gender confirms that the frequencies of HSV types vary by gender. Furthermore, time to first recurrence depends on type of primary episode (log rank test for group, p-value < 0.00005).

```
. tab group gender, col
```

| group | sex           |               | Total         |
|-------|---------------|---------------|---------------|
|       | Male          | Female        |               |
| 1     | 27<br>18.00   | 43<br>14.05   | 70<br>15.35   |
| 2     | 100<br>66.67  | 227<br>74.18  | 327<br>71.71  |
| 3     | 23<br>15.33   | 36<br>11.76   | 59<br>12.94   |
| Total | 150<br>100.00 | 306<br>100.00 | 456<br>100.00 |

The HSV-type-adjusted p-value is less significant (p=0.0344) than the p-value for group 2 alone (p=0.0025), even though there are more events in the former comparison. The differences observed in group 2 are being “diluted” by combining results for it with those from the other two groups where no association was observed. Using the Logrank statistic alone, however, we cannot test whether such aggregation is appropriate, i.e., whether the gender effects are roughly the same across HSV infection types. We will be able to evaluate effect modification using Cox regression.

2 c)

a) We observe that the peto-peto test for equality of survivor functions is not significant when we compare survival by sex. In addition, p-value of Peto-peto test is much larger than Logrank test, which can be explained on survival curves in 2 a). Because Logrank test emphasizes later differences of survival and Peto-peto test emphases early differences of survival, evidence against null hypothesis using Peto-peto is weaker than that using Logrank-test.

```
. sts test gender, p
```

```
      failure _d:  censor
analysis time _t:  rectime
```

```
Peto-peto test for equality of survivor functions
```

| gender | Events<br>observed | Events<br>expected | Sum of<br>ranks |
|--------|--------------------|--------------------|-----------------|
| Male   | 128                | 116.53             | 4.321321        |
| Female | 244                | 252.47             | -4.321321       |
| Total  | 372                | 372.00             | 0               |

```
chi2(1) =      0.58
Pr>chi2 =      0.4447
```

b) Different from Logrank tests, Peto-peto tests show consistent results for all cases. Among individuals with a primary HSV1 infection, gender is not statistically significantly associated with event times, though males appear to have consistently later times to first recurrence. Among HSV1 antibody-negative individuals with a primary HSV2 infection, as opposed to the result of Logrank test, difference between female and mail difference in recurrence time is not statistically significant. Among HSV1 antibody-positive individuals

with a primary HSV2 infection, there was not statistically significant difference in survival time between male and female groups.

In conclusion, it appears that after adjusting for HSV primary infection type, the Peto-peto test indicates that first HSV recurrence time is not statistically significantly different between male and female groups.

```
. sts test gender, p strata(group) detail
```

```
Stratified peto-peto test for equality of
survivor functions
```

```
-> group = 1
```

| gender | Events<br>observed | Events<br>expected | Sum of<br>Ranks |
|--------|--------------------|--------------------|-----------------|
| Male   | 11                 | 15.05              | -2.9848611      |
| Female | 25                 | 20.95              | 2.9848611       |
| Total  | 36                 | 36.00              | 0               |

chi2(1) = 1.92  
Pr>chi2 = 0.1664

```
-> group = 2
```

| gender | Events<br>observed | Events<br>expected | Sum of<br>Ranks |
|--------|--------------------|--------------------|-----------------|
| Male   | 96                 | 74.06              | 6.935609        |
| Female | 185                | 206.94             | -6.935609       |
| Total  | 281                | 281.00             | 0               |

chi2(1) = 2.21  
Pr>chi2 = 0.1373

```
-> group = 3
```

| gender | Events<br>observed | Events<br>expected | Sum of<br>Ranks |
|--------|--------------------|--------------------|-----------------|
| Male   | 21                 | 20.72              | 1.1409097       |
| Female | 34                 | 34.28              | -1.1409097      |
| Total  | 55                 | 55.00              | 0               |

chi2(1) = 0.30  
Pr>chi2 = 0.5845

```
-> Total
```

| gender | Events<br>observed | Events<br>expected | Sum of<br>Ranks |
|--------|--------------------|--------------------|-----------------|
| Male   | 128                | 109.83             | 7.2766312       |
| Female | 244                | 262.17             | -7.2766312      |
| Total  | 372                | 372.00             | 0               |

```
(*) sum over calculations within group
```

chi2(1) = 1.71  
Pr>chi2 = 0.1913

3. There are two key components to the “at risk” concept: 1) a person must be capable of having the outcome of interest, and 2) if the event (outcome) occurs, it will be observed.

Many of you referred to an example of the herpes study from homework and class notes, where the people could not be “at risk” for a recurrence of infection until the primary infection had cleared. Starting the time interval before the end of the primary infection would also bias downward our hazard estimate by including people in the “at risk” count who are not capable of experiencing an event. While the hazard will appear to be less than it should be, the survival will appear to be greater. Many of you referred to a “flat” survival curve, constant at 1.0 until the events begin. It is also possible that the effect would be to decrease the downward slope of the survival, making it appear less steep than it otherwise would.

Also, we can consider “immortal time bias” which was discussed in the Oscar winner paper. The initial time point was the date of birth. However the Oscar winners had survival advantage during immortal time period from the date of birth to winning time, which leads bias. During this time period, the winners were not capable of having outcome of interest (death). In other words, the winners were not included in people at risk during the immortal time, as opposed to the non-winners who were at risk from the initial time of the study to the event or censoring time. The hazard estimation for the Oscar winners would be less than it should be.

Another example would be the study of leukemia deaths among persons exposed to radiation resulting from the atomic bomb explosions in Japan, 1945. Since the study did not begin until 1950, only persons still alive in 1950 could be observed. An unknown number of people could have experienced the event of interest, died of leukemia, in the 5 years prior to the start of the study. To include these years in the period of time for “at risk” would make no sense since clearly no deaths were recorded and all persons entering the study survived until at least 1950, the start of the study. The impact on the hazard estimation would be to falsely reduce the estimated hazard by inflating the number at risk while underestimating the events (since none were recorded).