Homework #1

12 January 2015

1. Among those with censored observations (death=0), the minimum time to death is 1480 days. This is slightly more than 4 years, so at 4 years or 1461 days the vital status is known for every person in this dataset.

2. Methods: Descriptive statistics are presented in the table below within groups defined by serum CRP levels, as well as for all groups combined. These groups are defined as serum CRP less than 1 mg/L, between 1 and 3 mg/L inclusive, and greater than 3 mg/L. For continuous variables (age, BMI, and cholesterol), the mean, standard deviation, minimum, and maximum are shown in the table below. For binary variables (sex, smoking, prior history of cardiovascular disease [CVD], and death within 4 years), percentages are shown in the table below.

Results: Subjects missing serum CRP measurements (n=67) are excluded both from this table and from subsequent analyses. There were 428 subjects in the lowest serum CRP group (<1 mg/L, 2629 in the intermediate group (1-3 mg/L), and 1876 in the highest group (>3 mg/L). There were similar proportions of males to females in the two lowest groups, with a lower percentage of males in the highest group. Mean age and cholesterol was similar across all three groups. Mean BMI increased across increasing serum CRP groups, as did the proportion of those with a prior history of CVD. An increasing proportion of subjects died within 4 years of study enrollment across increasing serum CRP groups.

|  |  |
| --- | --- |
|  | **Serum CRP Levels** |
|  | **Below 1 mg/L (n=428)** | **1 – 3 mg/L****(n=2629)** | **Above 3 mg/L (n=1876)** | **All levels****(n=4933)** |
| **Male (%)** | 45.6% | 44.5% | 37.7% | 42.0% |
| **Age** | 73.5 (5.8; 65-94) | 72.8 (5.6; 65-100) | 72.6 (5.5; 65-93) | 72.8 (5.57; 65-100) |
| **BMI** | 23.8 (3.6; 15.6-38.6) | 26.1 (4.2; 14.7-53.2) | 28.1 (5.2; 15.3-58.8) | 26.7 (4.7; 14.7-58.8) |
| **Smoking history** | 9.6% | 10.1% | 15.7% | 12.2% |
| **Cholesterol** | 206 (40.5; 109-407) | 213 (38.6; 73-363) | 212 (39.7; 96-430) | 212 (39.2; 73-430) |
| **Prior history of CVD (%)** | 18.2% | 20.6% | 27.2% | 22.9% |
| **Death within 4 years (%)** | 4.9% | 7.7% | 13.9% | 9.8% |

 \*Continuous variables are mean (standard deviation; minimum – maximum).

3. Methods: The mean serum CRP level in the group that died within 4 years of study enrollment was compared to the mean serum CRP level in the group that survived at least 4 years, and the difference in means was tested using a t-test and 95% confidence interval allowing for unequal variances.

Results: The mean serum CRP level in the group that died within 4 years was 5.38 mg/L and the mean serum CRP level in the group that survived at least 4 years was 3.42 mg/L. The observed difference of 1.95 mg/L lower mean serum CRP level among the group that survived at least 4 years would not be judged unusual if the true difference in means were between 1.21 and 2.7 mg/L (two-sided P<0.001). We conclude that the distribution of serum CRP levels differs between those with a high risk of dying within 4 years and those with a low risk of dying within 4 years.

4. Methods: Geometric mean serum CRP levels were compared between those who died within 4 years and those who survived at least 4 years. Some measurements of CRP were reported as zeros; in order compare geometric means these measurements were set to ½ of the next lowest measurement, which was 1 (zeroes set to 0.5). Differences in the mean of log-transformed serum CRP levels were tested using a t-test that allows for unequal variances. Results of the t-test (the point estimate and the 95% confidence interval) were then exponentiated.

Results: Those who died within 4 years were observed to have a geometric mean serum CRP level of 2.97 mg/L, while those who survived at least 4 years were observed to have a geometric mean serum CRP level of 2.03 mg/L. The observed 68.3% higher geometric mean in the group that died within 4 years would not be judged unusual if the true ratio of geometric means is between 62.1% and 75.1%. This observation is statistically significant at a 0.05 level significance (two-sided P=<0.001) and thus we can reject the null hypothesis that the geometric mean serum CRP levels are not different by vital status at 4 years.

5. Methods: The proportion of subjects dying within 4 years were compared between subjects who had ‘high’ serum CRP (>3 mg/L) and those who had ‘low’ serum CRP (<3 mg/L). Pearson’s chi squared test was used to test for differences in the probability of death within 4 years.

Results: Based on a 95% confidence interval, this 8.9% lower absolute survival probability in those participants with high CRP would not be judged to be unusual if the true difference in survival probabilities were anywhere between 7.4% and 10.4% (two-sided P < 0.001) and thus we can with confidence reject the null hypothesis of no difference in the proportion of survival at 4 years across CRP groups.

6. Methods: The odds of dying within the group with high serum CRP levels (CRP > 3 mg/L) was compared to the odds of dying in the low serum CRP group. Fisher’s exact test was used to test for an odds ratio different than 1.

Results: Based on a 95% confidence interval, the observed odds ratio of 3.41 for the comparison of the high serum CRP group to the low serum CRP group would not be judged unusual if the true odds ratio is between 2.75 and 4.33 (Fisher’s exact two-sided P<0.001) and thus we reject the null hypothesis .

7. Perform a statistical analysis evaluating an association between serum CRP and all-cause mortality over the entire period of observation of these subjects by comparing the instantaneous risk of death across groups defined by whether the subjects have high serum CRP (“high” = CRP > 3 mg/L).

Methods: Kaplan-Meier estimates were used to estimate the survival distributions. The strata were defined as low (less than or equal to 3 mg/L) or high (greater than 3 mg/L) serum CRP. The logrank test was used to test the difference in the survival distributions.

Reults: The following graph shows the Kaplan-Meier estimates of survival probability for those subjects low CRP and those with high CRP. This graph shows the tendency for higher survival probabilities in the low serum CRP group. Using a logrank test to test the difference in instantaneous hazards we see a 14.3% difference in hazard (two-sided P<0.001).

8. A number of factors should be considered in determining *a priori* which test to use. Conditioning on the serum CRP levels makes sense as they were collected and measured prior to the outcome of interest. A continuous exposure variable is more precise than dichotomizing that variable. Finally, while a multiplicative model (geometric means) might be more relevant biologically due to the involvement of enzymatic reactions, a comparison of means rather than geometric means is a simpler comparison. Given these considerations and the fact that we are considering both a dichotomized exposure and outcome (even though this reduces precision), *a priori* I would chose a t-test of means, comparing probability of survival to 4 years across exposure (serum CRP) groups.