Homework 1

1/7/2015

1. Censoring and time to death

Time to death is usually right-censored data because some people drop out of the study and/ or survive until the end of the study. The minimum follow up time for participants in this study was 1480 days or 4.05 years. This means that the vital statistics (i.e. alive vs dead) were known for every person at 4 years of follow up. It is therefore valid to dichotomize death into those who died before 4 years and after 4 years. In looking at the subgroup of people who died before 4 yrs, the survival data is no longer right censored.

2. Descriptive statistics for associated variables:

Methods: Descriptive statistics are presented in groups by serum c-reactive protein (CRP) levels. Subgroups of CRP were defined according to existing recommendations for risk of cardiovascular disease by CRP level: below 1mg/L (low risk of heart disease), 1-3 mg/L (average risk of heart disease), greater than 3mg/L (high risk of heart disease). Within each group defined by CRP level, we present continuous variables (age, BMI, cholesterol) according to mean, standard deviation, minimum and maximum. For binary variables (gender, current tobacco use, prior history of cardiovascular disease), we present percentages.

Results: CRP levels are available for 4933 participants. Of note, 68 participants did not have CRP levels measured and their information has been removed from the analysis. We cannot assess whether these omissions impact our data analysis. Several of the other variables also had missing data: 13 participants did not have BMI measured, 3 participants did not have recorded cholesterols and 6 participants did not have smoking status recorded.

Of the remaining participants who had CRP levels measured, the minimum CRP level was 0mg/l (or less than assayable levels) and the maximum was 108mg/L. 1969 participants had CRP levels less than 1mg/L, 1088 had levels between 1-3 mg/l, and 1876 had levels greater than 3mg/L. The table below presents the descriptive statistics of several additional variables of interest. While the majority of participants in the study were female, there were relatively more females in the higher CRP subgroup then the other two subgroups. Participants with higher CRP also were more likely to be smokers and more likely to have had a prior cardiovascular disease.

Descriptive statistics by CRP levels (mg/L)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CRP mg/L (number participants): | < 1 (n= 1969) | 1 to 3 (n = 1088) | > 3 (n = 1876) | any level (n = 4933) |
| Age (yrs)1 | 73.09 (5.71, 65-98) | 72.55 (5.39, 65-100) | 72.63 (5.50, 65- 93) | 72.80 (5.56, 65-100) |
| Male (%) | 44.80% | 44.40% | 37.70% | 42% |
| BMI1,2 | 25.14 (3.84, 14.7-43.2) | 26.90 (4.50, 15.1-53.2) | 28.11 (5.18, 15.3-58.8) | 26.66 (4.72, 14.7-58.8) |
| Current smoker (%)3 | 9.30% | 11.30% | 15.70% | 12.20% |
| Cholesterol (mg/dl)1,4 | 210.31 (38.03, 73-407) | 213.89 (40.44, 78-354) | 211.84, (39.71, 96-430) | 211.68 (39.22, 73-430) |
| prior CVD (%) | 18.90% | 22.70% | 27.20% | 22.90% |

1 Descriptive statistics presented are the mean (standard deviation, minimum - maximum).

2 Data missing from 13 participants

3 Data missing from 6 participants

4 Data missing from 3 participants

3. 4 year all- cause mortality

Methods: Mean CRP levels were compared between participants who died within 4 years of study enrollment and participants who survived at least 4 years from study enrollment. The difference in mean CRP level was calculated using a t-test which did not assume equal variances between the two groups. The 95% confidence interval was also obtained using the same method.

Results: The mean CRP level in the 484 participants who died within 4 years was 5.38mg/L and the mean CRP level in the 4449 participants who survived for at least 4 years from study enrollment was 3.42mg/L resulting in an observed mean difference in serum crp level of 1.95mg/L. With 95% confidence, we would not be surprised by our results if the true difference in mean CRP level between those who died within 4 years and those who survived at least 4 years was between 1.21 and 2.70mg/L. The t-test gives a 2 sided p-value of < 0.001; so with high certainty, we can reject the null hypothesis that mean CRP levels are not different in the participants who died before 4 years versus participants who survived 4 years.

4. 4 year all- cause mortality and geometric means

Methods: To obtain geometric mean CRP levels, the recorded serum CRP was log transformed and used in the ~~below~~ statistical tests below. Of note, to avoid the error of generating inaccurately missing data when calculating geometric means, participants whose serum CRP levels were recorded as 0 were replaced by 0.5mg/L. This was based on the assumption that participants did not actually have a CRP level of 0mg/L, but in reality a smaller value then was clinically assayable by laboratory techniques. A level of 0.05mg/L was chosen arbitrarily because the lowest, non-zero recorded CRP level was 1mg/L (in reality, most labs can assay down to a level of 0.04mg/L).

Geometric mean serum crp levels were compared between participants who died within 4 years of study enrollment and study participants who survived for at least 4 years using a t-test that allowed for unequal variances. 95% confidence intervals for the population were obtained using similar methods. Mean estimates and CI were then exponentiated to gain inference on geometric means.

Results: The geometric mean crp level in the 484 participants who died within 4 years of study enrollment was 2.97 mg/L and the geometric mean crp level in the 4449 participants who survived for at least 4 years from study enrollment was 2.03mg/L, resulting in an observed 38% higher geometric mean crp in participants who died within 4 years. Based on a 95% confidence interval, it would not be unusual to get these results if the true population had anywhere between 28.6% to 47.6% higher geometric mean serum crp levels in those who died within 4 years compared to those who survived to 4 years. The t-test gives a 2 sided p-value of < 0.001; so with high certainty, we can reject the null hypothesis that geometric mean CRP levels are not different in the participants who died before 4 years versus participants who survived 4 years.

5. All cause 4 year mortality by low and high crp

Methods: An indicator variable was created to divide serum crp levels into low (less then or equal to 3mg/L) and high values (greater than 3mg/L). The proportion of participants dying within 4 years of study enrollment was compared between participants who had low serum crp levels and participants who had high crp levels. Differences in probability of death were computed using the chi-squared test for independence and 95% confidence intervals were obtained using Wald statistics.

Results: Based on 4 year vital statistics, 8.0% of the 3758 participants with low serum crp levels died compared to 15.6% of the 1243 participants with high serum crp levels. With 95% confidence, this 7.6% lower survival in the participants with high crp levels would not be unusual if the true difference in population survival was between 5.4% and 9.8%. Using a chi-squared test of independence, this difference in survival is statistically significant at a level of 0.05 (2 sided p- value < 0.0001).

6. Odds Ratio for all cause 4 –year mortality

Methods: The odds of participants dying within 4 years of study enrollment were compared among participants with low serum crp levels (less than or equal to 3mg/L) and participants with high serum crp levels (greater than 3mg/L). An odds ratio different than 1 was tested and 95% confidence intervals were computed using the Fischer Exact test.

Results: In the 3758 participants with low serum crp levels, the odds of dying within 4 years was 0.087. In the 1243 participants with high serum crp levels, the odds of dying within 4 years was 0.18. Based on 95% confidence intervals, the observed odds ratio of 2.12 for the comparison of the high crp to low crp group would not be unusual if the true odds ratio were anywhere between 1.74 and 2.69. The Fischer’s exact test gives a p-value of < 0.001, suggesting that with high confidence we can reject the null hypothesis that the odds ratio comparing 4 year mortality in participants with high and low serum crp levels is 1.

7. All-cause mortality and serum CRP

Methods: Survival analysis was conducted using Kaplan Meier methods. Survival curves for participants with low (less than or equal to 3mg/L) vs high (greater than 3mg/L) crp levels were compared using the log rank test. The difference in hazard ratio for participants with low serum crp levels was compared to participants with high serum crp levels using Cox proportional hazard regression. 95% confidence intervals were computed using the Huber- White Sandwich estimator.

Results: Using the cox proportional hazard regression, the observed hazard ratio of 1.66 compares the instantaneous risk of death for participants with high serum crp levels to the instantaneous risk of death for participants with low serum crp levels. With 95% confidence, we would not be surprised to get this hazard ratio if the true hazard ratio for the population was between 1.47 and 1.89. The logrank test gives a p-value of <0.0001, so we can reject with high certainty the null hypothesis that the survival time of participants with low and high serum crp levels are equal.

8. The decision on which statistical tests to use for describing the association between crp levels and survival time should occur a priori to any analysis of the data (and determination of statistical significance). Basically I am looking for a test which I know how to perform with my limited knowledge, is easy and intuitive for others to understand, uses all my data (i.e. try to avoid dichotomization of data) and makes biologic sense (i.e. conditions based on predictor of interest/ crp level and summarizes survival). Based on all these principals, I would probably use the t-test to compare mean crp levels across groups defined by vital status at 4 years. Although comparing geometric mean crp levels might make more sense from a biologic standpoint because levels are multiplicative, it is less intuitive for most readers to understand.