**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #1**

January 5, 2015

***ID: 1983***

1. The first participant censored not due to death was at day 1,480 days after enrollment (less than one month over 4 years). It is therefore valid to divide the dataset at time point 4 years after enrollment; the data within 4 years contains known life/death data on all participants, while data after 4 years contains participants censored for reasons other than death meaning the vital statistics are no longer known for all individuals. 5/5
2. Methods: CRP levels were divided according to recommendations for risk of cardiovascular disease: below 1 mg/L (low); 1-3 mg/L (average); above 3 mg/L (high) and a summary measure of all levels. Descriptive statistics were performed according to these CRP levels on age, sex, BMI, smoker, serum cholesterol, a previous history of cardiovascular disease, and death. Previous history of cardiovascular disease was assessed by any previous angina, MI, TIA or stroke. Binary variables are summarized as frequencies and continuous variables are summarized with mean, standard deviation, and range.

Inference: There are 67 missing CRP levels on the 5,000 participants enrolled in the study. Since CRP is the predictor of interest it will not be possible to draw relevant results on the missing 1.3% and therefore these subjects are omitted from all analysis. There is missing data on other variables, the most significant is 12 BMI records for the average age group, noted below the table. Distribution of sex and age across CRPs groups is very similar. Smoking increases across CRP group 9.3% in the low risk group, 11.0% in the average, and 16.4 in the high risk group. Within the first 4 years of the study, the high risk CRP group saw the highest frequency of deaths, 15.6%, compared to the low risk group who experienced 4.9%. Previous cardiovascular condition also seemed to increase across CRP levels. There does not appear to be a trend in serum cholesterol from these descriptive statistics. The average CRP group had the highest average cholesterol, 212.5 mg/dl). Also, higher BMI and lower percent male with higher CRP.

Figure

4/4 for table layout

3/3 for choice of descriptive statistic

2/3 for discussion of finding

Total = 9/10

1. Methods: A two sample t test (α = 0.05) was used to compare mean CRP levels of participants who died within 4 years of enrollment and those who lived at least to 4 years after enrollment. The one-sided alternative hypothesis that the mean CRP level of the survivors is less than the mean CRP levels of those who died will be used. Equal variances were not assumed across the two groups.

Inference: Of the 4933 subjects, 9.8% died within 4 years of enrollment. Mean CRP levels for those who died is 5.61 mg/L (SD: 8.09) compared to a mean of 3.42 mg/L (SD: 5.87) for the survivor group. Those who lived had a mean 1.95 mg/L lower CRP level. With 95% confidence interval we would expect the true population difference to be between 1.21 mg/L and 2.70 mg/L lower in those who survived than those who did not. With a P < 0.0001 we can reject the null hypothesis, with significance level of 0.05, that the two means are equal in favor of the alternative that surviving to 4 years is associated with a lower mean CRP level.

3/5 for performing an appropriate analysis and describing the methods appropriately (say what you did with missing data, how did you calculate 95% CIs?)

4/5 for reporting the association appropriately (say whether 1 sided or two sided p-value in conclusion)

Total = 7/10

1. Methods: A two sample t test (α = 0.05) was used to compare geometric mean of log transformed CRP levels of participants who died within 4 years of enrollment and those who lived at least to 4 years after enrollment. The one-sided alternative hypothesis that the mean CRP level of the survivors is less than the mean CRP levels of those who died will be used. Equal variances were not assumed across the two groups. In order to effectively perform a logarithm transformation on the CRP level we had to reevaluate the CRP levels that were reported as zero. It is unlikely that subject’s actual CRP level was zero, instead, we assume the measurement was very low. Since all CRP levels are given as positive integers, we will assign all zero CRP levels to the next lowest CRP value of 1 for this geometric analysis.

Inference: The 428 participants with CRP levels of zero were inputted as a CRP level of 1 mg/L for this analysis. The two sample t test was performed and the 484 that died within 4 years of enrollment had a geometric mean CRP level of 3.06 mg/L while the 4449 survivors had 2.16 mg/L geometric mean CRP. The survivors had a 41% larger geometric mean CRP than those that died. With 95% confidence the true population survivor mean is between 54% greater and 29% greater than the geometric mean CRP of those who died. With a P < 0.0001 we can reject the null hypothesis, with significance level of 0.05, that the two means are equal in favor of the alternative that surviving to 4 years is associated with a lower mean CRP level.

3/5 for performing an appropriate analysis and describing the methods appropriately (say what you did with missing data), how did you calculate 95% CIs?)

4/5 for reporting the association appropriately (say whether 1 sided or two sided p-value in conclusion)

Total = 7/10

1. Methods: To determine if the probability of death is independent of high and non-high CRP levels within the first 4 years of the study, the data was divided into CRP greater than 3 mg/L and CRP less than or equal to 3 mg/L.A chi squared test was performed for independence using two sided test and significance level of 0.05.

Inference: 1,175 subjects had high CRP levels of greater than 3 mg/L. 15.5% of the high CRP level group died within the first 4 years of study enrollment. Of the 3,758 non-high CRP group, 8.0% died within the first four years of their enrollment in the study. The average survival of the high CRP level group was 7.56% lower than subjects with a CRP of 3mg/L or lower. With 95% confidence the true population difference is 5.3% to 9.8% lower for the non-high CRP group. Using a significant level of 0.05, we reject (P<0.0001) the null hypothesis that the probability of death is independent of CRP levels.

3/5 for performing an appropriate analysis and describing the methods appropriately (say what you did with missing data, how did you calculate 95% CIs?)

4/5 for reporting the association appropriately (say whether 1 sided or two sided p-value in conclusion)

Total = 7/10

1. Methods: To determine if the odds of death is independent of high and non-high CRP levels within the first 4 years of the study, the data was divided into two groups: CRP greater than 3 mg/L and CRP less than or equal to 3 mg/L. A chi squared test was performed for independence using two sided test and significance level of 0.05.

Inference: The odds of dying in the first 4 years of enrollment for participants (n = 1,175) with high CRP was 0.184 while those with non-high CRP (n = 3,758) had an odds of 0.087. The odds ratio of high CRP to non-high CRP is 2.119. Using Wald based confidence interval, with 95% confidence the true population odds ratio is between 1.740 and 2.580. The two sided chi squared test with 0.05 significance level results in a p value of less than 0.0001 and we can reject the null hypothesis that the odds of death is independent of CRP levels.

3/5 for performing an appropriate analysis and describing the methods appropriately (say what you did with missing data, how did you calculate 95% CIs?)

5/5 for reporting the association appropriately

Total = 8/10

1. Methods: The survival distributions of the two groups (high CRP with levels above 3 mg/L and non-high CRP with levels of 3 mg/L or lower) were compared using the logrank test. Kaplan Meier statistics were used to descriptively assess the high and non-high CRP groups.

Inference: Survival probabilities for the two groups defined by non-high and high CRP levels are listed in the table below by years. The non-high CRP group is shown to have higher survival throughout the eight years of the study. Within 8 years of enrollment 76.9% of participants with CRP less than or equal to 3 mg/L survive while 64.7% of subjects with CRP over 3 mg/L survive at the same time point. The instantaneous risk of death is 68.71% higher for the high CRP group than for the non-high CRP group. With 95% confidence the true population instantaneous risk of death is between 48.5% and 91.7% higher for the high CRP group than for the non-high CRP group. Using a significant level of 0.05, we reject (P < 0.0001) the null hypothesis that survival is equal in the two groups.



|  |  |
| --- | --- |
|   | Survival Probabilities (Kaplan Meier) |
| Year | non-high CRP ≤ 3 mg/L [n= 3758] | high CRP > 3 mg/L [n=1175] |
| 1 Year | 0.988 | 0.967 |
| 2 Years | 0.971 | 0.926 |
| 3 Years | 0.948 | 0.881 |
| 4 Years | 0.920 | 0.844 |
| 5 Years | 0.884 | 0.800 |
| 6 Years | 0.853 | 0.755 |
| 7 Years | 0.810 | 0.710 |
| 8 Years | 0.769 | 0.647 |

3/5 for performing an appropriate analysis and describing the methods appropriately (say what you did with missing data, how did you calculate 95% CIs?)

4/5 for reporting the association appropriately (say whether 1 sided or two sided p-value in conclusion)

Total = 7/10

1. To best answer the question about associations between mortality and serum CRP, I would probably choose to provide a two sample t test across two groups comparing the means for the simplicity and ease in understanding results. There are some challenges with this data that make it so I would eliminate certain analysis types. For one, the presence of a zero result for CRP level that is likely not actually zero means that the geometric mean will need to account for the zeros in a fashion that is probably not incredibly accurate. During the analysis I chose to have the zeros equal the minimum non-zero output in the dataset, but that is less than ideal. Additionally, since CRP was recorded as an integer and the common understandings around CRP are with ranges less than 1 mg/L (low), 1-3 mg/L (average), and over 3 mg/L (high) it makes defining the appropriate CRP groups difficult. It is probably most prudent to dichotomize the groups by non-high (less than or equal to 3 mg/L) and high (over 3 mg/L) to account for the not ideal accuracy in measuring CRP. Comparing the survival distributions for two groups of CRP and looking at the hazard ratio would seem appropriate for the censorship of this data too. Which is also a fairly straightforward analysis that is easily interpretable.

8/10 - you mentioned using means (the 4th bullet point, 2 points), performing valid analyses that you know how to do (5th bullet point, 2 points), and made a valid conclusion (4 points).