**Biost 518 / 515: Applied Biostatistics II**

**Homework #4**

This homework investigates associations between death from any cause and age, sex, and serum bilirubin in a population of patients with primary biliary cirrhosis who were enrolled in a randomized clinical trial (RCT) of D-penicillamine.

1. Provide suitable descriptive statistics pertinent to the scientific questions addressed in this homework.

**Methods**: Descriptive statistics are displayed by observation type (censored/death) and for the entire sample of 418 subjects. Within each group, we presented the mean, standard deviation, minimum, and maximum for bilirubin and age, as there were no missing cases here. For sex, percentage female and number of missing cases was displayed.

Descriptive statistics were also presented for the censoring distribution for all the subjects in this trial. Here, the minimum and maximum censored times, the Kaplan-Meier estimates of the 25th and 50th percentile, and the Kaplan-Meier estimate of the mean time to follow up are displayed.

Because subjects in this study have relatively high levels of bilirubin (since they have primary biliary cirrhosis (PBC)) and since it seems that bilirubin acts on a multiplicative scale, the following categories of bilirubin are displayed in the table: <1 mg/dl, 1-2 mg/dl, 2-4 mg/dl, and 4-28 mg/dl. These values were also chosen since there was a fairly even distribution of sample sizes among them.

The number of subjects and deaths is presented within each category and overall, as are the survival probabilities at 2, 6, and 12 years, and the restricted mean survival when all categories had some subjects at risk except for the 4-28 mg/dl category (11.45 years). A graph with Kaplan-Meier survival curves for each group is also displayed.

**Inference**: There were 418 subjects in this randomized clinical trial, but 106 of them were missing data on sex (including 70 who were censored). Those individuals were excluded from the analysis looking at the association between mortality and bilirubin when adjusting for sex, though we cannot be sure as to how those dropped subjects affect the generalizability of our results.

Of the 418 subjects, the mean bilirubin level was 3.22 mg/dl, though the subjects who died had much higher levels of bilirubin (5.54 mg/dl) than those who were censored (1.77 mg/dl). The mean age overall was 50.7 years old, though those who died were older than the censored subjects (53.9 vs 48.7 years old). Among the 312 subjects with available information, censored individuals were also more likely to be female (92.5%) than those who died (82.4%).

Table 2 displays Kaplan-Meier based estimates of the survival distribution within 4 categories of bilirubin and among the entire group of subjects. The 418 subjects were followed for an average of 6.88 years (median=6.48 years; min=1.46 years; max=13.13 years). During this time, 161 deaths were observed. The progressively large difference in survival probabilities can be seen at 6 and 12 years between the <1 mg/dl category and >4 mg/dl group. In fact, everyone with bilirubin greater than 4 mg/dl died by 12 years of enrollment. Additionally, during the first 11.45 years, subjects with <1 mg/dl of bilirubin lived 10.16 years on average while those with >4 mg/dl of bilirubin lived 3.90 years on average.

Next, a graph with Kaplan-Meier curves is displayed for the 4 categories of bilirubin. Again, there is a trend for those with lower bilirubin to survive for a longer period of time. At the end of the 13.13 years of observation, more than 70% of subjects with <1 mg/dl of bilirubin survived, about 45% of those with 1-2 mg/dl of bilirubin survived, about 5% of those with 2-4 mg/dl of bilirubin survived, and there were no survivors for those individuals with bilirubin levels greater than 4 mg/dl.

 Table 1: Bilirubin, age, and sex, by type of observation and across all subjects

|  |  |  |
| --- | --- | --- |
|   | **Observation** | **All subjects (n=418)** |
|   | **Censored (n=257)** | **Death (n=161)** |
| Bilirubin (md/dl) | 1.77 (2.20; 0.3-18) | 5.54 (5.84; 0.3-28) | 3.22 (4.41; 0.3-28) |
| Age (years) | 48.7 (10.4; 26.3-78.4) | 53.9 (9.81; 30.9-76.7) | 50.7 (10.45; 26.3-78.4) |
| Sex (% female) | 92.5% Missing = 70 | 82.4% Missing = 36 | 88.5% Missing = 106 |

Note: Descriptive statistics presented for continuous variables are the mean (standard deviation; minimum – maximum). For binary variables, percentages and number of missing cases are displayed.

|  |  |
| --- | --- |
| Table 2: Kaplan-Meier based estimates of time to death for all subjects |  |
|   | **Bilirubin at Study Enrollment** | **All subjects** |
|   | **< 1 mg/dL** | **1-2 mg/dL** | **2-4 mg/dL** | **4-28 mg/dL** |
| # Subjects | 142 | 107 | 78 | 91 | 418 |
| # Deaths | 20 | 30 | 44 | 67 | 161 |
| 2 year Survival Probability | 99.3% | 97.2% | 91.0% | 79.1% | 92.8% |
| 6 year Survival Probability | 90.0% | 79.2% | 50.0% | 24.8% | 66.4% |
| 12 year Survival Probability | 70.9% | 44.9% | 4.8% | 0.0% | 35.3% |
| 25th percentile of Survival | 10.55 y | 7.11 y  | 2.74 y | 1.34 y | 4.00 y |
| Median Survival | -- | 11.47 y | 5.70 y | 2.95 y | 9.30 y |
| 11.45 Years Restricted Mean of Survival 1 | 10.16 y | 8.85 y | 6.03 y | 3.90 y 2 | 7.76 y |
| -- = The median was not able to be estimated with the available data. |
| 1 Average number of years alive during the first 11.45 years. |
| 2 The last observation was a death, so this is an actual mean, not restricted mean. |



1. In prior homeworks using the Cardiovascular Health Study datasets, we were able to use logistic regression to investigate associations between mortality and various covariates. Why might such an approach not seem advisable with these data? (Consider the extent to which such analyses might be confounded and/or lack precision.)

**Answer**: First, right-censoring occurs in this dataset starting at 1.46 years. Therefore, we should not use the other forms of regression that do not take into account censoring when looking at the entire time of observation (or anything past 1.4 years). Next, logistic regression does not improve when using robust standard error estimates, but we would like to allow our analyses to allow for heteroscedasticity. It also appears that sex or age might confound our analysis and/or be a precision variable, so these need to be addressed before running a logistic regression (or any regression).

1. Perform a statistical regression analysis evaluating an association between serum bilirubin and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum bilirubin modeled as a continuous variable.
	1. Include a full report of your inference about the association.

**Methods:** Using proportional hazards regression for all 418 subjects (where there were 161 observed deaths over 13.1 years), we estimate the hazard ratio over the entire period of observation for groups of subjects varying by 1 mg/dl of bilirubin, using robust standard error estimates. A hazard ratio of 1 was tested, and 95% confidence intervals and p-values were Wald-based.

**Inference:** Using proportional hazards regression, we estimate that for each difference of 1 mg/dl of bilirubin, the risk of (all-cause) mortality is 15.2% higher for those subjects with greater bilirubin. This observation is consistent with a group that has a 1 mg/dl higher bilirubin having a risk of mortality between 12.1% and 18.5% higher than the lower bilirubin group, based on a 95% confidence interval. This estimate is statistically significant (p < 0.001), so we can with high confidence reject the null hypothesis that there is no association between bilirubin and all-cause mortality.

* 1. For each population defined by serum bilirubin value, compute the hazard ratio relative to a group having serum bilirubin of 1 mg/dL. (This will be used in problem 6).
1. Perform a statistical regression analysis evaluating an association between serum bilirubin and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum bilirubin modeled as a continuous logarithmically transformed variable.
	1. Why might this analysis be preferred *a priori?*

**Answer**: Before performing the analysis, log transformed bilirubin might be preferred since bilirubin acts on a multiplicative scale. A constant difference in bilirubin is not expected to have the same increase in risk, whereas a doubling would be more likely. Log transforming bilirubin would allow us to report association on a scientifically meaningful scale and add precision to our analysis.

* 1. Include a full report of your inference about the association.

**Methods:** Using proportional hazards regression for all 418 subjects (where there were 161 observed deaths over 13.1 years), we estimate the hazard ratio over the entire period of observation for groups of subjects varying by a 2-fold difference in bilirubin, using robust standard error estimates. Thus, analysis was performed on log-transformed bilirubin and reported on a 2-fold scale. A hazard ratio of 1 was tested, and 95% confidence intervals and p-values were Wald-based.

**Inference:** Using proportional hazards regression, we estimate that for each doubling of bilirubin, the risk of (all-cause) mortality is 98.4% higher for those subjects with greater bilirubin. This observation is consistent with a group that has twice as much bilirubin as another having a risk of mortality between 1.78 to 2.21 times higher than the lower bilirubin group, based on a 95% confidence interval. This estimate is statistically significant (p < 0.001), so we can with high confidence reject the null hypothesis that there is no association between bilirubin and all-cause mortality.

* 1. For each population defined by serum bilirubin value, compute the hazard ratio relative to a group having serum LDL of 1 mg/dL. (This will be used in problem 6).
1. One approach to testing to see whether an association between the response and the predictor of interest is adequately modeled by an untransformed continuous variable is to add some other transformation to the model and see if that added covariate provides statistically significant improved “fit” of the data. In this case, we could test for “linearity” of the bilirubin association with the log hazard ratio by including both the untransformed and log transformed bilirubin. (Other alternatives might have been bilirubin and bilirubin squared, but in this case our *a priori* interest in the log bilirubin might drive us to the specified analysis.)
	1. Provide full inference related to the question of whether the association is linear.

**Methods:** We used a proportional hazards regression model that included linear and log-transformed bilirubin terms for all 418 subjects (where there were 161 observed deaths over 13.1 years). First, the association between bilirubin and mortality was assessed by testing whether both the terms were simultaneously zero. The p-value for this 2 df test was based on Wald estimates that used robust standard error estimates. Next, the linearity of this association was tested using a Wald test that the slope of the log-transformed term was zero, using robust standard error estimates again.

**Inference:** Using a proportional hazards regression that included linear and log-transformed bilirubin terms, we see that there is an association between bilirubin and death that is statistically significant (p<0.001). Testing for nonlinearity by looking at the log bilirubin term in the model, we see that the value is highly statistically significant (p<0.001). Therefore, we can say with high confidence that the association between bilirubin and mortality is not adequately modeled by a log hazard function that uses untransformed (linear) bilirubin.

* 1. Again, save the fitted values from this model by obtaining the estimated HRs relative to a group with bilirubin of 1 mg/dl. (This will be used in problem 6.)
1. Display a graph with the fitted hazard ratios from problems 3 - 5. Comment on any similarities or differences of the fitted values from the three models.



**Answer**: The graph above presents fitted hazard ratios from the three models in problems 3-5. All three models predict that individuals with higher bilirubin will have a higher relative hazard of mortality. Both Models 2 and 3 are similar in their approximations until about 15 mg/dl of bilirubin, whereas Model 1 estimates a lower relative hazard than the other models until about 20 mg/dl of bilirubin, when it estimates a higher relative hazard than the others. Model 1 appears to approximate an upward curvilinear trend in relative hazards, whereas Model 3 approximates a downward curvilinear trend. Model 2 seems to predict a linear trend in the relative hazards.

1. We are interested in considering analyses of the association between all cause mortality and serum bilirubin after adjustment for age and sex.
	1. What evidence is present in the data that would make you think that either sex or age might have confounded the association between death and bilirubin?

**Answer**: When adding sex and age into the model, we see that the effect of bilirubin on death increases from the unadjusted model. With a doubling of bilirubin, the hazard ratio changes from 1.98 times higher (unadjusted) to 2.11 times higher (adjusted). We could also see from Table 1 that age and sex appear to be associated with mortality in the sample. Furthermore, sex appears to be associated with bilirubin (table not presented).

* 1. What evidence is present in the data that would make you think that either sex or age might have added precision to the analysis of the association between death and bilirubin? (In real life, we would ideally decide whether to adjust in our pre-specified SAP).

**Answer**: Here, we see that including sex would lower the sample size available and thus lower precision, and additionally we see that the unadjusted standard error is smaller than the adjusted standard error. However, we know that PBC affects women more than it does men, so stratifying by sex may result in a more precise analysis that does not mask the effect of sex on the association between bilirubin and mortality. Also, we know that older individuals who develop liver disease are likely to die within 5-10 years of the initial PBC diagnosis (which typically occurs between 35-60 years of age), so stratifying on age would enable the association to be more precise between bilirubin and mortality.

* 1. Provide full inference regarding an association between death and bilirubin after adjustment for sex and age.

**Methods:** After adjusting for sex and age, we used proportional hazards regression for the 312 subjects with available data for sex (106 missing cases), age, bilirubin, and mortality, where there were 125 observed deaths over 13.1 years. We estimated the hazard ratio over the entire period of observation for groups of subjects varying by a 2-fold difference in bilirubin with the same sex and age, using robust standard error estimates. Thus, analysis was performed on log-transformed bilirubin and reported on a 2-fold scale. A hazard ratio of 1 was tested, and 95% confidence intervals and p-values were Wald-based.

**Inference:** Using proportional hazards regression, we estimate that for each doubling of bilirubin, the risk of (all-cause) mortality is 2.11 times higher for those subjects with greater bilirubin of the same sex and age. This observation is consistent with a group that has twice as much bilirubin as another having a risk of mortality between 1.84 to 2.42 times higher than the lower bilirubin group with the same sex and age, based on a 95% confidence interval. This estimate is statistically significant (p < 0.001), so we can with high confidence reject the null hypothesis that there is no association between bilirubin and all-cause mortality after adjusting for sex and age.

1. Note that in the above analyses, we completely ignored the intervention in the RCT? What impact could this have had on our results?

**Answer**: Because this is a randomized control trial, we generally suppose that no confounding is present in our data. However, we are limited to PBC patients, which typically are women. Therefore we need to clearly specify the population that we can generalize our results to. Now, we found a statistically significant association after our results, but if we were unlucky in our sampling for the RCT, this result might be a Type I error. Again, before we conduct our analysis, we need to lay out which variables we will stratify by and why, as well as consider the study design.