**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #4**

February 2, 2015

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

**Method:** The predictor of interest defined as baseline serum bilirubin measured in mg/dL was categorized into six groups (0, 1, 2-4, 5-8, 9-16, 17-28) based on the expected multiplicative biological increments of bilirubin levels in pathological states (all the participants have primary biliary cirrhosis). For baseline serum bilirubin total and subcategories, the sample size, number of missing cases, mean, standard deviation, minimum and maximum were used to describe continuous variables like age (years), while the sample size, number of missing cases and proportions were used for categorical variables like sex (female) and those who received D-Penicillamine treatment which was specifically included from the data set because of its potential confounding effect in our research question. The outcome “death” was described as count within each category and total serum bilirubin levels. Kaplan-Meier estimates were used to describe the time to censoring and time to death distribution, 25th, 50th and 75th percentiles of survival distribution for both time to death and censoring was reported and survival probabilities from enrolment to death at 2, 4, 6, 8 and 12 years were reported for total and each category of baseline serum bilirubin levels.

**Inference:** A total of 418 study participants were followed up with 106 and 108 missing cases for sex and treatment variables. The mean number of years from enrollment to death was 8.35[[1]](#footnote-1) years and a median of 9.295 (0.112 to 11.474 years). A total of 161 deaths occurred during this period of follow up. In contrast, the time from enrolment to censoring was lower; a mean of 6.879 years and a median of 6.475 (1.459 to 13.128) –see Figure 2. There is a decreasing trend in survival probability as the levels of baseline serum bilirubin increase and the trend is more dramatic as the number of years increase from 2 to 12 years. This is seen more clearly in Figure 1 that shows categories >9mg/dL having the least survival probability estimates, followed by categories 2 – 8mg/dL and then categories <1mg/dL.

The mean baseline serum bilirubin level was 3.221 mg/dL with a standard deviation of 4.408mg/dL indicating a right skewed distribution. The minimum serum bilirubin levels ranged from 0.3 to 28 mg/dL. The study participants had a mean age of 50.74 years (SD 10.45; 26.28 – 78.44) and were mostly females (88.46%). The mean age within each bilirubin category was similar in general with less age variability within the lowest (0mg/dL) and higher bilirubin categories (>9mg/dL). Female sex distribution within categories was higher ranging from 93.27% to 100% within the lowest (0mg/dL) and higher bilirubin categories (>9mg/dL) compared to 80.0 to 88.31% in other categories.

On average, 50.65% of the total study population received treatment, however those with the highest baseline serum bilirubin level had lowest proportion (25%) assigned to treatment compared to other categories that were more or less about half the population within each category.

Figure1: Kaplan-Meier estimates for time distribution from study enrolment to death (any cause) among 418 study participants with different baseline levels of serum bilirubin



Figure2: Kaplan-Meier estimates for time distribution from study enrolment to censoring among 418 study participants



Table 1: Descriptive statistics of study population’s demographic characteristics with Kaplan-Meier based estimates of distribution of time from study enrolment to death from any cause and censoring for total study population and within baseline serum bilirubin categories

|  |  |  |
| --- | --- | --- |
|  | Serum Bilirubin Level at Study Enrollment  |  |
|  | 0 mg/dL | 1 mg/dL | 2 – 4 mg/dL | 5 – 8 mg/dL | 9 – 16mg/dL | 17 – 28 mg/dL | Total |
| Number of Subjects | 142 | 107 | 91 | 43 | 22 | 13 | 418 |
| Age (years) | 51.19; 9.8 (30.57 – 75) | 50.56; 10.79 (26.28 – 76.71) | 50.15; 11.43 (29.56 – 71.89) | 50.92; 11.12(30.86 – 78.44) | 50.55; 9.98 (33.18 –- 70.56) | 51.13; 6.57 (43.52 – 65.88) | 50.74; 10.45 (26.28 –78.44) |
| Female N3, % | 10493.27% | 7788.31% | 7281.94% | 3580% | 12100% | 12100% | 312188.46% |
| Received D-Penicillamine N3, % | 10348.54% | 7655.26% | 7252.78% | 3551.43% | 1250% | 1225% | 310150.65% |
| Deaths N3 | 20 | 30 | 53 | 26 | 20 | 12 | 161 |
| Death Time Distribution  | Min – Max (0.112 – 11.474 years) |
| Survival Probability2 |  |  |  |  |  |  |  |
| At 2 years  | 96.47% | 95.32% | 83.52% | 79.07% | 54.55% | 53.85% | 88.02% |
| At 4 years  | 95.70% | 87.11% | 65.90% | 43.32% | 18.18% | 7.69% | 75.16% |
| At 6 years  | 89.95% | 79.15% | 49.24% | 29.70% | 13.64% | 7.69% | 66.43% |
| At 8 years  | 83.34% | 63.57% | 41.48% | 18.56% | NA2 | NA2 | 56.89% |
| At 10 years  | 76.33% | 53.85% | 11.67% | NA2 | NA2 | NA2 | 44.22% |
| At 12 years  | 70.88% | 44.88% | 3.89% | NA2 | NA2 | NA2 | 35.34% |
| 25th percentile2  | 10.549 | 7.113 | 2.560 | 2.546 |  0 .605 | 0.591 | 4.003 |
| 50th percentile2 | NA2 | 11.474 | 5.697 | 3.907 | 2.007 | 2.335 | 9.295 |
| 75th percentile2 | NA2 | NA2 | 9.295 | 6.177 | 2.839 | 3.203 | NA2 |
| Censor Time Distribution | Min – Max (1.459 – 13.128 years) |
| 25th percentile2  | 5.013 | 3.984 | 4.118 | 3.614 | 4.526 | 6.207 | 4.359 |
| 50th percentile2 | 6.938 | 6.401 | 6.379 | 4.832 | 5.153 | 6.207 | 6.475 |
| 75th percentile2 | 9.027 | 10.467 | 12.192 | 8.643 | NA2 | 6.207 | 9.133 |

* 1106 missing values for sex variable, 108 missing values for treatment variable
* 2 Kaplan-Meier estimates of distribution of time from study enrolment to death from any cause within baseline serum bilirubin level total and subcategories. NA indicates non estimable survival probability with available data within stratum
* 3 Sample size within each category

Finally, there is some evidence of effect modification in outcome, the survival curve between males and females is different – see figure 3.

Figure 3: Survival estimates stratified by sex



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.)

The least censored time in this data set is 1.46 years with 36/382 (9.42%) deaths of any cause occurring during this time period. If we created a binomial variable of dead within 1.46 years and its alternative will comprise of dead and individuals with censored information/unknown status therefore reducing precision in determining associations with deaths.

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.

**Method:** We used proportional hazards regression to model distributions of time to death of any cause among groups of baseline serum bilirubin levels. The serum bilirubin level measurements used were in continuous untransformed format. Hazard ratios from this regression, had 95% confidence intervals and two-sided p values generated using Wald statistics based on Huber-White sandwich estimator. The HR estimates and 95% CI generated were exponentiated by a factor of 10 to make inference based on a 10mg/dL difference in baseline serum bilirubin levels.

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

**Inference:** The 418 study participants had a mean baseline serum bilirubin level of 3.22 mg/dL with a standard deviation of 4.41mg/dL and a range from 0.3 to 28mg/dL. The proportional hazards regression analysis showed that with a 10 mg/dl unit difference in baseline serum bilirubin, the risk of instantaneous death is 4.12 fold higher in the group with the higher bilirubin. The 95% CI indicate that this estimate will not be unusual if the true risk is anywhere between 3.13 and 5.45 fold higher in the group with higher baseline serum bilirubin levels. This estimate is statistically significant with a two-sided p value (P<0.0001) thus rejecting the null hypothesis that states that the risk of death from any cause is not associated with baseline serum bilirubin levels.

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.

**Method:** Baseline serum bilirubin was log transformed to the base of two. We used proportional hazards regression to model distributions of time to death of any cause among groups of baseline serum log transformed bilirubin levels. Hazard ratios from this regression, had 95% confidence intervals and two-sided p values generated using Wald statistics based on Huber-White sandwich estimator.

* 1. Why might this analysis be preferred *a priori?*

The serum bilirubin levels increase in a multiplicative scale in pathological states per empirical data. The study population all have liver disease and will therefore expect multiplicative differences between groups that could be compared using constant ratios of log transformed bilirubin.

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

**Inference:** The 418 study participants had a mean baseline serum bilirubin level of 3.22 mg/dL with a standard deviation of 4.41mg/dL and a range from 0.3 to 28mg/dL. The proportional hazards regression analysis showed that with each doubling of baseline serum bilirubin levels, the risk of instantaneous death is 1.98 fold higher in the group with the higher bilirubin. The 95% CI indicate that this estimate will not be unusual if the true risk of instantaneous death is anywhere between 1.78 to 2.21 fold higher in the group with higher baseline serum bilirubin levels. This estimate is statistically significant with a two-sided p value (P<0.0001) thus rejecting the null hypothesis that states that the risk of death from any cause is not associated with baseline serum bilirubin levels.

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.

**Inference:** The 418 study participants had a mean baseline serum bilirubin level of 3.22 mg/dL with a standard deviation of 4.41mg/dL and a range from 0.3 to 28mg/dL. The test for linearity of the bilirubin association with the log hazard ratio indicated a significant linearity with a two-sided p <0.0001. Thus the addition of log transformed serum bilirubin (base 2) adds a better fit of data and thus rejects the null hypothesis.

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.

Figure 4: Fitted hazard ratios relative to a group with serum bilirubin of 1 mg/dL estimated from proportional hazard regression models with baseline serum bilirubin levels as linear continuous variable (black dots) alone, as log transformed variable (base 2) – (blue dots) alone and model with both linear and log transformed terms (green dots).



All the models in Figure 4 indicate an increasing trend in hazard ratios with increasing baseline serum bilirubin levels. Both the log transformed and linear with log transformed serum bilirubin have similar fits till about 15mg/dL of serum bilirubin levels. After that point, the models have different fits and this could be attributed to low data points.

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? (In real life, we would ideally decide whether to adjust for potential confounding in our pre-specified statistical analysis plan (SAP)).

A logistic regression model for a binary death outcome defined with deadin1.49years (cutoff for censored data) with bilirubin levels generates an OR of 1.175 (95% CI 1.105 - 1.25), when we adjust for sex and age in this model the OR changes to 1.181 (95% CI 1.094 - 1.276) suggesting presence of confounding.

Alternatively, using descriptive statistics, there is some evidence suggestive of sex being a confounder based on Figure 3 that shows females are less likely to die (outcome) compared to men in the Kaplan-Meier plot while women distribution is highest within the lowest bilirubin category. The highest baseline serum bilirubin categories also had higher proportion of women but these categories had sparse data (see table below).

Figure 3: Survival estimates stratified by sex



|  |  |  |
| --- | --- | --- |
|  | Serum Bilirubin Level at Study Enrollment  |  |
|  | 0 mg/dL | 1 mg/dL | 2 – 4 mg/dL | 5 – 8 mg/dL | 9 – 16mg/dL | 17 – 28 mg/dL | Total |
| Number of Subjects | 142 | 107 | 91 | 43 | 22 | 13 | 418 |
| Female N3, % | 10493.27% | 7788.31% | 7281.94% | 3580% | 12100% | 12100% | 312188.46% |

* 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).

The hazard ratio of proportional hazards regression model of distributions of time to death of any cause among groups of baseline serum log transformed (base 2) bilirubin levels is 1.98 (95% CI 1.781 - 2.212).

In contrast, the hazard ratio of proportional hazards regression model of distributions of time to death of any cause among groups of baseline serum log transformed (base 2) bilirubin levels adjusting for age and sex is 2.108 (95% CI 1.840 - 2.416).

Because of the non-collapsibility of proportional hazard regression model to confounders, the change in hazard ratios is suggestive of a precision effect of sex or age.

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

The hazard ratio in the unadjusted model indicates a 1.98 fold higher risk in instantaneous death for every two-fold increase in baseline serum bilirubin level. After adjusting for age and sex, the proportional hazards regression increased to 2.11 fold higher in the group with the higher bilirubin. The 95% CI indicate that this estimate will not be unusual if the true risk of instantaneous death is anywhere between 1.84 to 2.42 fold higher in the group with higher baseline serum bilirubin levels. This estimate is statistically significant with a two-sided p value (P<0.0001) thus rejecting the null hypothesis that states that the risk of death from any cause is not associated with baseline serum bilirubin levels.

The difference in hazard ratios between the non-adjusted and adjusted model is suggestive of a precision effect of age or sex.

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

The intervention could be associated with exposure bilirubin level (predictor) – more likely and with outcome (death) if toxic – which is less likely as this drug would have been tested and approved before use and testing in human subjects. It therefore could be potentially a precision variable in the model and would not change the general trend of the increasing HR with increased baseline serum bilirubin levels. Even if were a confounder, because of the non-collapsable nature of proportional hazard model to confounders, the HR estimates will not have been affected.

1. Largest observed analysis time is censored, mean is underestimated [↑](#footnote-ref-1)