HW #4
Biostat 515/518- Scott Emerson

2/3/14

1. The association between serum LDL levels and all cause mortality was examined. Due to the censored nature of the data, Kaplan-Meier curves and corresponding summary measures from the curves were used to descriptively summarize the data. To examine the linearity of the relationship between all cause mortality and serum LDL, LDL was turned into a categorical variable with three strata: subjects with LDL <100 mg/dL, subjects with LDL between 100-159 mg/dL and subjects with LDL 160 mg/dL or higher. Robust Cox proportional hazard regression was used to calculate the hazard ratio and compare the instantaneous risk of death between serum LDL levels. Corresponding 95% confidence intervals were generated, and a Wald test was used to test the null hypothesis that the instantaneous risk of death is the same between serum LDL levels.

	1. Below is the Kaplan-Meier curve for each LDL category described above. Notice that for all groups, survival was greater than 50% during the entire study period. Subjects with LDL greater or equal to 160 mg/dL seem to have higher survival compared to the other groups. For early time points there seems to be little difference between survival in all three groups. For later time points survival seems to be highest for subjects with LDL greater than 160 mg/dL and lowest for subjects with LDL less than 100 mg/dL. The table below also summarizes survival estimates by LDL group for each year.



|  |
| --- |
| Survival Probabilities by LDL Group |
| Year | < 100 mg/dL | 100 mg/dL -159 mg/dL | >159 mg/dL |
| 1 | 0.982 | 0.980 | 1.000 |
| 2 | 0.964 | 0.947 | 0.981 |
| 3 | 0.909 | 0.921 | 0.953 |
| 4 | 0.867 | 0.894 | 0.907 |
| 5 | 0.800 | 0.841 | 0.869 |

The estimated hazard ratio for two groups differing by 1 mg/dL in LDL is 0.993. That is, a subject with a 1 mg/dL increase in LDL has a 0.7% decrease in instantaneous risk of death. It would not be unusual to observe a hazard ratio between 0.987 and 0.998. That is, it would not be unusual to observe a decreased instantaneous risk of death between 1.3% and 0.2% for a 1 mg/dL increase in serum LDL. Note that it would be unusual if subjects with different LDL levels had the same instantaneous risk of death. The p-value yielded from the Wald test of the null hypothesis is 0.009. Therefore, we reject the null hypothesis and can say with some confidence there is an association between LDL levels and all cause mortality.

* 1. The centered LDL variables were calculated and a second Cox proportional hazard model was run to computer hazard ratios that are relative to a group having serum LDL of 160 mg/dL.
1. The association between log serum LDL levels and all cause mortality was examined. Due to the censored nature of the data, Kaplan-Meier curves and corresponding summary measures from the curves were used to descriptively summarize the data. To examine the linearity of the relationship between all cause mortality and serum LDL, LDL was turned into a categorical variable with three strata: subjects with LDL <100 mg/dL, subjects with LDL between 100-159 mg/dL and subjects with LDL 160 mg/dL or higher. Robust Cox proportional hazard regression was used to calculate the hazard ratio and compare the instantaneous risk of death between log serum LDL levels. Corresponding 95% confidence intervals were generated, and a Wald test was used to test the null hypothesis that the instantaneous risk of death is the same between log serum LDL levels.

	1. The descriptive statistics are reported and discussed above. The assumption of non-linearity was examined by stratifying LDL. The estimated hazard ratio comparing a 1 mg/dL difference in log serum LDL levels is 0.438. More clearly stated, for each doubling in serum LDL the risk of death is 0.564 times lower in the group with the higher serum LDL. This was calculated by exponentiating the hazard ratio yielded from regression by log(2). It would not be unusual to observe a hazard ratio between 0.297 and 0.645. Note that it would be unusual to observe a hazard ratio of 1, meaning it would be unusual for subjects with the same log serum LDL to have the same instantaneous risk of death. The p-value associated with the Wald test of the null hypothesis is less than 0.0001. Therefore, we reject the null hypothesis and can say with some confidence there is an association between all cause mortality and log LDL.
	2. The centered log LDL variables were calculated and a second Cox proportional hazard model was run to computer hazard ratios that are relative to a group having log serum LDL of 160 mg/dL.
2. The association between a serum LDL levels and all cause mortality was examined using a quadratic function of serum LDL. Due to the censored nature of the data, Kaplan-Meier curves and corresponding summary measures from the curves were used to descriptively summarize the data. To examine the linearity of the relationship between all cause mortality and serum LDL, LDL was turned into a categorical variable with three strata: subjects with LDL <100 mg/dL, subjects with LDL between 100-159 mg/dL and subjects with LDL 160 mg/dL or higher. Robust Cox proportional hazard regression was run using both serum LDL and squared serum LDL as predictors to compare instantaneous risk of death between subjects with different LDL levels. The Wald test was used to test if the squared LDL term was significant. Predicted hazard estimates were plotted against LDL values to assess the nonlinearity of the relationship between LDL and all cause mortality.

	1. The descriptive statistics are reported and discussed above. The assumption of non-linearity was examined by stratifying LDL. The hazard ratio estimates and confidence intervals are omitted, as the interpretation is difficult and not necessary for addressing our scientific question. The p-value yielded in the Wald test is 0.089, so we cannot reject the null hypothesis. That is, we cannot claim there is a non-linear relationship between LDL and all cause mortality. Below is a plot of the predicted hazard estimates from the model plotted again LDL values. Note there is a non-linear relationship here. The maximum predicted hazard estimate is much larger than most of the values, while the minimum is slightly larger. Although test of non-linearity was not found to be significant, based on this graph I would say there is non-linear association between LDL and all cause mortality.
	2. The centered LDL and centered LDL squared variables were calculated and a second Cox proportional hazard model was run to computer hazard ratios that are relative to a group having serum LDL of 160 mg/dL.
3. Below is a graph of the different hazard ratio estimates calculate in problems 1-3 (b). The general trend of the curves are similar, however the hazard ratio estimates calculated using the quadratic model is more non-linear. For very high and very low LDL estimates all three models estimated a higher hazard ratio. For mid ranged values of LDL the quadratic term estimated hazard ratios lower than the other two models. The models from 2 and 3 seem somewhat non-linear, but the model from 1 seems linear.

