**Biost 515/518**

Homework 4

02/03/2014

1. Perform a statistical regression analysis evaluating an association between serum LDL and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum LDL modeled as a continuous variable.
	1. Include full description of your methods, appropriate descriptive statistics, and full report of your inferential statistics.

|  |  |
| --- | --- |
|  | **Serum LDL (mg/dL)** |
| **Time (mo)** | **<100** **(n=165)** | **100-129****(n=228)** | **130-159****(n=225)** | **>=160****(n=107)** |
| **12** | 0.982 | 0.983 | 0.978 | 1.000 |
| **24** | 0.964  | 0.939  | 0.956  | 0.981 |
| **36** | 0.909 | 0.912  | 0.929 | 0.953 |
| **48** | 0.867 | 0.877  | 0.911 | 0.907  |
| **60** | 0.800 | 0.811 | 0.871 | 0.869 |

**Descriptive statistics: The graph and table above depict Kaplan-Meier estimates of survival probabilities for the 725 subjects for whom serum LDL measurements are available, stratified into four groups as specified in the table. From the graph, we see that the group with the highest serum LDL (≥160 mg/dL) tends to have the highest survival probability at most time points, though the curve overlaps with the next lower group (LDL between 130 and 159) between 3.5 and 5 years of follow-up. The group with lowest serum LDL (<100 mg/dL) also has lowest survival, by nearly 10% at the end of follow-up time. This group is closely matched by the next-lowest group (LDL between 100 and 130) until about the five-year timepoint, and both are quite significantly lower than the two groups with higher LDL.**

**Methods: The hazard ratio for the death of subjects were compared between subjects who differed in serum LDL using a proportional hazards (Cox) regression model. Statistical inference on the difference in hazards as a function of serum LDL, modeled as a continuous variable, was based on the Wald-based calculation of p-values and confidence intervals. Parameter estimates are baesd on maximum partial likelihood estimation. Robust standard errors were used in the regression model.**

**Inference: Data was available on 725 subjects having mean serum LDL of 126 mg/dL (SD 33.6 mg/dL; range 11-247 mg/dL). From proportional hazards regression analysis, we estimate that for each 1 mg/dL unit difference in serum LDL, the risk of death is 0.738% lower (hazard ratio 0.9926) in the group with the higher LDL. This estimate is statistically significant (p=0.0093). A 95% confidence interval suggests that this observation is not unusual if a group that has 1 mg/dL higher serum LDL has a true instantaneous risk of death between 1.29% lower and 0.182% lower than the group with the lower serum LDL. Therefore, we reject the null hypothesis of no association between survival time and serum LDL in favor of a trend toward lower risk of death among subjects with higher serum LDL levels.**

* 1. For each population defined by serum LDL value, compute the hazard ratio relative to a group having serum LDL of 160 mg/dL. **An appropriate variable was created.**
1. Perform a statistical regression analysis evaluating an association between serum LDL and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum LDL modeled as a continuous logarithmically transformed variable.
	1. Include full description of your methods, appropriate descriptive statistics (you may refer to problem 1, if the descriptive statistics presented there are adequate for this question), and full report of your inferential statistics.

**Descriptive statistics: The table and Kaplan-Meier survival curves in Question 1 apply to this question as well. It is difficult to judge whether a logarithmic trend exists in the data, but based on prior scientific knowledge, we expect the effect of serum LDL to be multiplicative and therefore suppose that a logarithmic transformation of LDL levels is appropriate.**

**Methods: The hazard ratio for the death of subjects were compared between subjects who differed in serum LDL using a proportional hazards (Cox) regression model. Statistical inference on the difference in hazards as a function of the natural log of serum LDL, modeled as a continuous variable, was based on the Wald-based calculation of p-values and confidence intervals. Parameter estimates are baesd on maximum partial likelihood estimation. Robust standard errors were used in the regression model.**

**Inference: Data was available on 725 subjects having mean serum LDL of 126 mg/dL (SD 33.6 mg/dL; range 11-247 mg/dL). From proportional hazards regression analysis, we estimate that for each 10% difference in serum LDL, the instantaneous risk of death is 7.58% lower (hazard ratio 0.9242) in the group with the higher LDL. This estimate is highly statistically significant (p<0.0001). A 95% confidence interval suggests that this observation is not unusual if the true instantaneous risk of death was anywhere between 10.94% and 4.09% lower for each 10% higher serum LDL level. Therefore, we can with high confidence reject the null hypothesis of no association between survival time and serum LDL in favor of a trend toward lower risk of death among subjects with higher serum LDL levels.**

* 1. For each population defined by serum LDL value, compute the hazard ratio relative to a group having serum LDL of 160 mg/dL. **An appropriate variable was created.**
1. Perform a statistical regression analysis evaluating an association between serum LDL and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum LDL modeled quadratically (so include both a term for serum LDL modeled continuously and a term for the square of LDL).
	1. Include full description of your methods, appropriate descriptive statistics (you may refer to problem 1, if the descriptive statistics presented there are adequate for this question), and full report of your inferential statistics. In the inferential statistics, include your conclusion regarding the linearity of the association of serum LDL and the log hazard.

**Descriptive statistics: Please refer to the descriptive statistics included in Question 1. There is no apparent U-shaped trend; the two groups with higher serum LDL have higher survival over nearly all time points compared to the two groups with lower serum LDL. If a U-shaped trend were present, we would expect to see the middle two groups cluster together and the two extremes perhaps cluster at either higher or lower survival probabilities than the central LDL values.**

**Methods: The hazard ratio for the death of subjects were compared between subjects who differed in serum LDL using a proportional hazards (Cox) regression model. Statistical inference on the difference in hazards as a function of serum LDL and the square of serum LDL, modeled as continuous variables, was based on the Wald-based calculation of p-values and confidence intervals. Parameter estimates are baesd on maximum partial likelihood estimation. Robust standard errors were used in the regression model.**

**Inference: Data was available on 725 subjects having mean serum LDL of 126 mg/dL (SD 33.6 mg/dL; range 11-247 mg/dL). From proportional hazards regression analysis, the p-value for the ldlsqr term is 0.055, indicating that we do not have sufficient evidence to conclude that there is a nonlinear association between survival and LDL levels. However, an overall p-value of 0.0005 still suggests that there is an association between survival and LDL levels, since every term in the model is related to LDL.**

* 1. For each population defined by serum LDL value, compute the hazard ratio relative to a group having serum LDL of 160 mg/dL. **An appropriate variable was created.**
1. Display a graph with the fitted hazard ratios from problems 1 – 3. Comment on any similarities or differences of the fitted values from the three models.

**The graph below shows fitted hazard ratios for populations defined by serum LDL values, relative to a group having serum LDL of 160 mg/dL. Points labeled A (blue) are fitted hazard ratios based on serum LDL, B (red) are ratios using log serum LDL, and C (green) are ratios using a model fit with terms for serum LDL and the square of serum LDL. The hazard ratios are plotted on the log scale.**

**We note that all three lines intersect around 80 and 160 mg/dL. The intersection point at 160 mg/dL was fixed by our method of determining hazard ratios: since we used 160 mg/dL as our reference, all three curves should be exactly 1 at 160 mg/dL. At LDL levels less than 80 mg/dL or greater than 160 mg/dL, the fitted ratios using model C are highest, followed by model B and then model A. The reverse is true between 80 and 160 mg/dL.**

**For the hazard ratios from the model that was quadratic in LDL (C), the apparent minimum hazard ratio is 1 – that is, the lowest probability of death occurs at serum LDL values around 160 mg/dL. In contrast, the other two curves (A and B) suggest that hazard ratios are monotonically decreasing, so that as serum LDL increases, hazard ratios decrease.**

