BIOST 515/518 Winter 2014

Homework 4

ID 4086

1. (a) Methods: An association between serum LDL and all-cause mortality was investigated by comparing the instantaneous risk (hazard) of death across groups defined by serum LDL modeled linearly as a continuous variable using a proportional hazards regression model with robust standard errors. Descriptive statistics based on Kaplan-Meier survival curves and survival probabilities from 1 – 5 years were calculated, based on a categorical variable of serum LDL dichotomized into 3 groups (<129mg/dL, 130-159mg/dL, >160mg/dL). Statistical inference on the hazard ratio and its standard error estimates was based on the (Wald statistic). Two-sided p-values and 95% confidence intervals were computed using the same estimates for the standard error.

Descriptive statistics:



**Table of survival probabilities among groups categorized by serum LDL levels**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time (years) | < 120mg/dL | 130 - 159mg/dL | > 160mg/dL | Total |
| 1 | 0.9822 | 0.9778 | 1.0000 | 0.9810 |
| 2 | 0.9491 | 0.9556 | 0.9813 | 0.9537 |
| 3 | 0.9109 | 0.9289 | 0.9533 | 0.9211 |
| 4 | 0.8728 | 0.9111 | 0.9065 | 0.8884 |
| 5 | 0.8066 | 0.8711 | 0.8692 | 0.8354 |

From the above graph and table, it appears that survival is the lowest among the group with serum LDL <120mg/dL overall. At years 1-3, subjects with serum LDL >160mg/dL have the highest survival probabilities. At years 4-5, subjects with serum LDL between 130 and 159mg/dL have the highest survival probabilities. There is some crossover in the survival curves of the three groups.

Results: There are 725 available observations (11 missing, 131 deaths). When comparing two groups with different cholesterol levels, the instantaneous risk of death is estimated to be0.738% lower (hazard ratio 0.9926) for each 1mg/dL difference in serum LDL levels, with the group having the higher serum LDL level tending towards a lower instantaneous risk of desk. This observed difference is statistically different from a hazard ratio of 1 (p=0.0093), with a 95% confidence interval suggesting that the observed hazard ratio is what might be typically observed if the true instantaneous risk of dying was anywhere between0.182% and 1.29% lower for each 1mg/dL higher in serum LDL levels (hazard ratio between 0.9871 and 0.9982). We thus reject the null hypothesis of no association between serum LDL and all-cause mortality in favor of a trend towards lower risk of death among subjects with higher serum LDL levels.

(b) STATA code not included. See Q4 for results.

2. (a) Methods: An association between serum LDL and all-cause mortality was investigated by comparing the instantaneous risk (hazard) of death across groups defined by serum LDL modeled as a continuous logarithmically transformed variable using a proportional hazards regression model with robust standard errors. Here serum ldl was log-transformed before the cox regression model was run in STATA. Descriptive statistics based on Kaplan-Meier survival curves and survival probabilities from 1 – 5 years were calculated, based on a categorical variable of serum LDL dichotomized into 3 groups (<129mg/dL, 130-159mg/dL, >160mg/dL). Statistical inference on the hazard ratio and its standard error estimates was based on the (Wald statistic). Two-sided p-values and 95% confidence intervals were computed using the same estimates for the standard error.

Descriptive statistics: See question 1 (a) for table and graph. Since logldl is a one-to-one transformation of ldl, the descriptive statistics are the same. It is difficult to eyeball whether there is a logarithmic relationship (our eyes are bad for that) but it does seem like there is some sort of relationship.

Results: There are 725 available observations (11 missing, 131 deaths). When comparing two groups with different cholesterol levels, the instantaneous risk of death is estimated to be7.58% lower (hazard ratio 0.9242) for each 10% difference in serum LDL levels, with the group having the higher serum LDL level tending towards a lower instantaneous risk of death. This observed difference is statistically different from a hazard ratio of 1 (p < 0.0005), with a 95% confidence interval suggesting that the observed hazard ratio is what might be typically observed if the true instantaneous risk of death was anywhere between 4.09% and 10.94% lower for each 10% higher in serum LDL levels (hazard ratio between 0.8906 and 0.9591). We thus reject the null hypothesis of no association between serum LDL and all-cause mortality in favor of a trend towards lower risk of death among subjects with higher serum LDL levels.

(b) STATA code not included. See Q4 for results.

3. (a)  Methods: An association between serum LDL and all-cause mortality was investigated by comparing the instantaneous risk (hazard) of death across groups defined by serum LDL modeled quadratically using a proportional hazards regression model with robust standard errors. Here both a term for serum LDL modeled continuous and a term for the square of LDL was included in the cox regression model in STATA. Descriptive statistics based on Kaplan-Meier survival curves and survival probabilities from 1 – 5 years were calculated, based on a categorical variable of serum LDL dichotomized into 3 groups (<129mg/dL, 130-159mg/dL, >160mg/dL). Statistical inference on the hazard ratio and its standard error estimates was based on the (Wald statistic). The p-value for the individual estimates

Descriptive statistics: See question 1 (a) for table and graph. Since ldl-squared is a one-to-one transformation of ldl, the descriptive statistics would be similar for the groupings. From the data it does not appear that there is a U-shaped trend, although again, this is difficult to eyeball.

Results: There are 725 available observations (11 missing, 131 deaths). The p-value for the squared ldl term, holding the linear term constant, is 0.055. This is not statistically significant at the 0.05 level, thus we cannot with confidence reject the null hypothesis that there is not a quadratic association between ldl and mortality. However, this does not mean there is no association or even no nonlinear associations – there could exist other nonlinear associations. The overall F-test (of both the linear and quadratic terms) gives a statistically significant p-value (p = 0.0005). Thus we can reject the null hypothesis of no association between serum LDL and all-cause mortality. However we cannot know from this model which direction the association is.

(b) STATA code not included. See Q4 for results.

4. Methods: The fitted hazard ratios are as described in problems 1-3 (b). That is, the hazard ratio relative to a group with serum LDL of 160mg/dL was computed. These estimated hazard ratios (fithrA for the linear model, fithrB for the log transformation model, fithrC for the quadratic transformation model) were plotted against LDL, with the hazard ratios on the logarithmic scale.

Results:



Compared to a group with serum LDL of 160mg/dL, the fitted hazard ratios of all models were above 1 at serum LDL levels up until 160mg/dL. At levels below 75mg/dL, the quadratic fit has the highest hazard ratio, then the hazard ratios become about the same. At serum LDL levels above 160mg/dL, the quadratic fit gives a hazard ratio above 1 while both the linear and logarithmic fits give hazard ratios below 1. Overall the linear and logarithmic fits look similar, while the quadratic fit gives a higher hazard ratio at both lower and higher serum LDL levels.