**Biostats 518 HW4**

**Question 1**

**a. Methods:** Proportional Hazards (Cox) regression with robust standard error was performed to evaluate the instantaneous risk of death (any cause) over time across groups defined by serum LDL modeled continuously. The Wald statistic was used to make statistical inference and was calculated from the regression slope parameter and its robust standard error. The Breslow method was used for ties. Two-sided p-values and 95% confidence intervals were also computed.

**Inference:** 725 subjects had baseline serum LDL measurements with a mean LDL of 126 mg/dL, (SD: 33.6 mg/dL, range 11-247 mg/dL). The instantaneous rate of death (hazard) was 0.993 times as much for each 1 mg/dL increase in serum LDL. Thus, the hazard ratio decreased by a factor of 0.007 for each 1 mg/dL increase in LDL. Based on 95% confidence intervals, this hazard ratio of 0.993 would not be unusual if the true hazard ratio were between 0.987 and 0.998. A two-sided p-value of 0.009 suggests we can reject the null hypothesis that the instantaneous risk of death is not associated with serum LDL, in favor of a lower hazard associated with higher LDL levels.

**b.** Completed for question 4

**Question 2**

**a. Methods:** Proportional Hazards (Cox) regression with robust standard error was performed to evaluate the instantaneous risk of death (any cause) over time across groups defined by the log serum LDL. The Wald statistic was used to make statistical inference and was calculated from the regression slope parameter and its robust standard error. The Breslow method was used for ties. Two-sided p-values and 95% confidence intervals were also computed.

**Inference:** 725 subjects had baseline serum LDL measurements with a mean log LDL of 4.80 (SD: 0.293, range: 2.40 to 5.51). This corresponds to a geometric mean of 121 mg/dL (SD: 1.34; range: 11 to 247). The hazard ratio for an e-fold difference in LDL was 0.438. Thus, the hazard ratio decreased by 56.2% for each e-fold increase in LDL. Based on 95% confidence intervals, this hazard ratio of 0.438 would not be unusual if the true hazard ratio were between 0.297 and 0.645 (or in other words between 35.5% and 70.3% lower hazard). A two-sided p-value less than 0.001 suggests we can reject the null hypothesis that the instantaneous risk of death is not associated with serum LDL, in favor of a lower hazard of death associated with higher LDL levels.

**b.** Completed for question 4

**Question 3**

***a.*****Methods:** Proportional Hazards (Cox) regression with robust standard error was performed to evaluate the instantaneous risk of death (any cause) over time across groups defined by the serum LDL modeled quadradically. When treating LDL as a quadratic term, we can determine whether or not the hazard of death is linearly associated with LDL, based on the overall p-value for a test of both the first order and second order terms simultaneously. However, we lose the ability to easily interpret the magnitude of any association observed.

**Inference:**725 subjects had baseline serum LDL measurements with a mean LDL of 126 mg/dL, (SD: 33.6 mg/dL, range 11-247 mg/dL). When treating LDL as a quadratic term, we can determine whether or not the hazard of death is linearly associated with LDL. Based on the overall two-side p-value of 0.0005 testing for linearity when modeling LDL quadratically, we can reject the null hypothesis that the relationship is linear, and conclude that the relationship between the risk of death and LDL is non-linear (suggesting a possible u-shaped curve). We find strong evidence supporting a relationship between the hazard of death and serum LDL.

**b.** Completed for question 4

**Question 4**

****

The above graph illustrates the models for estimating the hazard ratios when LDL is modeled continuously (A), as a log (B), and quadratically (C). In model A (blue), the trend is decreasing and linear, showing that the instantaneous risk of death decreases with increasing LDL. In model B (red), the trend decreases, though it is an exponential relationship with lower LDL levels having the greatest decreases in the relative hazard, and higher LDL levels having less impact on the relative hazard as LDL increases. In model C (green), the relative hazard has a u-shaped relationship with increasing LDL. Thus at lower levels of LDL in model C, the relative hazard decreases with increasing LDL, and at higher LDL, the relative hazard increases (though much less dramatically) with increasing LDL. Generally, all 3 models show an overall decreasing trend in the relative hazard with increasing LDL, except for model 3 that increases slightly for LDL levels greater than ~160. Additionally, the hazard for all 3 models are very similar to each other for serum LDL values between 90 mg/dL and 160 mg/dL, diverging at the more extreme LDL values. Therefore, for the majority of the range of LDL levels, the models similarly predict the hazard ratios, and only disagree at the more extreme (low and high) values of LDL.