1.

It is valid to dichotomize time to death at 5 years because there are no censoring events occurring within 5 years of an individual’s entrance into the study period; all 121 observations whose follow-up ended within 5 years had deaths observed. This is shown in the table below.

|  |  |  |
| --- | --- | --- |
| **Total Follow-Up Time** | **Death Observed** | **Death Not Observed** |
| **≤ 5 Years** | 121 | 0 |
| **> 5 Years** | 12 | 602 |

2.

We focus on exploring an association between serum LDL levels and 5-year all-cause mortality. Below we present a table of descriptive statistics for our sample of 735 adults aged 65 or older in our cohort study conducted over an 11-year period. There were 10 individuals with missing values of LDL in the dataset; we exclude these 10 observations from this table and all additional analysis. We report distributions for the full sample and for subsets of the sample restricted to observations with low LDL (<160 mg/dL) or high LDL (≥160 mg/dL).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1. Descriptive Statistics** | | | | | | | |
| **Variable** | **Arithmetic  Mean** | **SD** | **Min** | **25th  %tile** | **Median** | **75th  %tile** | **Max** |
|  | ***Overall Sample (N=725)*** | | | | | | |
| Serum LDL level (mg/dL) | 125.80 | 33.60 | 11 | 102 | 125 | 147 | 247 |
| Death Within 5 Years | 0.16 | -- | -- | -- | -- | -- | -- |
| Age | 74.57 | 5.45 | 65 | 71 | 74 | 78 | 99 |
| Weight (lbs) | 159.86 | 30.77 | 74 | 138.5 | 158 | 179 | 264 |
| Male | 0.50 | -- | -- | -- | -- | -- | -- |
| History of Smoking | 0.56 | -- | -- | -- | -- | -- | -- |
| Congestive Heart Failure | 0.06 | -- | -- | -- | -- | -- | -- |
| Angina | 0.09 | -- | -- | -- | -- | -- | -- |
| Myocardial Infarction | 0.12 | -- | -- | -- | -- | -- | -- |
| Transient Ischemic Attack | 0.03 | -- | -- | -- | -- | -- | -- |
|  | ***Observations with LDL Below 160 mg/dL (N=618)*** | | | | | | |
| Serum LDL level (mg/dL) | 116.36 | 25.73 | 11 | 98.25 | 118 | 137 | 159 |
| Death Within 5 Years | 0.17 | -- | -- | -- | -- | -- | -- |
| Age | 74.51 | 5.39 | 65 | 71 | 73 | 78 | 99 |
| Weight (lbs) | 159.36 | 30.78 | 86 | 138 | 158 | 178 | 264 |
| Male | 0.51 | -- | -- | -- | -- | -- | -- |
| History of Smoking | 0.56 | -- | -- | -- | -- | -- | -- |
| Congestive Heart Failure | 0.06 | -- | -- | -- | -- | -- | -- |
| Angina | 0.09 | -- | -- | -- | -- | -- | -- |
| Myocardial Infarction | 0.12 | -- | -- | -- | -- | -- | -- |
| Transient Ischemic Attack | 0.03 | -- | -- | -- | -- | -- | -- |
|  | ***Observations with LDL of At Least 160 mg/dL (N=117)*** | | | | | | |
| Serum LDL level (mg/dL) | 180.36 | 18.26 | 160 | 166.5 | 175 | 188 | 247 |
| Death Within 5 Years | 0.13 | -- | -- | -- | -- | -- | -- |
| Age | 74.88 | 5.77 | 65 | 70 | 74 | 78 | 94 |
| Weight (lbs) | 162.74 | 30.68 | 74 | 143 | 159 | 181 | 257 |
| Male | 0.42 | -- | -- | -- | -- | -- | -- |
| History of Smoking | 0.54 | -- | -- | -- | -- | -- | -- |
| Congestive Heart Failure | 0.03 | -- | -- | -- | -- | -- | -- |
| Angina | 0.07 | -- | -- | -- | -- | -- | -- |
| Myocardial Infarction | 0.12 | -- | -- | -- | -- | -- | -- |
| Transient Ischemic Attack | 0.06 | -- | -- | -- | -- | -- | -- |

Our outcome of interest is death within 5 years (5-year all-cause mortality) and our predictor of interest is serum LDL level. Table 1 shows that observed 5-year all-cause mortality was not higher in the group with elevated LDL (13%) than in the group with low LDL (17%), in the study sample; in fact, mortality was higher in the low LDL group. If we had expected to observe a positive association between LDL level and mortality, this result may have been surprising. As judged by the descriptive statistics presented in Table 1, the suspected risk factors for heart disease (age, sex, smoking history, prior diagnosis of heart disease or stroke) do not vary considerably between observations with high and low LDL. This gives some evidence that these risk factors do not act as confounders in our analysis of the association between LDL level and 5-year all-cause mortality in our sample.

3.

**Method:** We compare the mean serum LDL values between observations who died within 5 years of study enrollment and subjects who were still alive 5 years after enrollment. To do this, we use a two-sample t-test that does not assume equal variance between mortality groups. The two-sample t-test estimates whether the mean LDL levels in our sample are consistent with a true difference in mean LDL across groups defined by 5-year mortality.

**Inference:** We are testing

vs.

,

where is the mean LDL for individuals who died within 5 years of study enrollment and is the mean LDL for individuals who were still alive 5 years after enrollment. Our alternative hypothesis for the two-sample t-test is that the true difference in population means We find a 95% confidence interval for this true difference in means is (1.44, 15.56) with an associated p-value of 0.019. Our sample is consistent with a true value of between 1.44 mg/dL and 15.56 mg/dL and, thus, is consistent with a positive association between mean LDL level and 5-year survival.

4.

**Method:** To compare the geometric mean of LDL observations between 5-year all-cause mortality groups, we first conduct a two-sample t-test (not assuming equal variances) on LDL levels that have been subjected to the natural log transformation. Next, we compute a 95% confidence interval for the ratio of geometric means between the two mortality groups by exponentiating each bound of the 95% CI generated by the two-sample t-test for the difference in means. Our point estimate of the ratio of geometric means between these two groups is found by exponentiating the difference in means used by the two-sample t-test.

**Inference:** For our sample, we find that is the 95% confidence interval for the ratio of geometric mean LDL values for 5-year survivors relative to those who died within 5 years of enrollment. We have a corresponding point estimate of 1.097 for this ratio of geometric mean LDL values between groups. Since our sample is consistent with a true ratio of geometric mean LDL values of between 1.02 and 1.18, we have evidence of a negative association between geometric mean LDL and 5-year all-cause mortality.

5.

**Method:** To test whether 5-year all-cause mortality rates differ between observations with high LDL (≥160 mg/dL) and low LDL (<160 mg/dL), we construct a 95% confidence interval of the following form:

Here, is the 5-year all-cause mortality rate (probability of death) among individuals with high LDL, is the 5-year all-cause mortality rate (probability of death) among individuals with low LDL, is the sample size in the high LDL group, and is the sample size in the low LDL group. For our data, we have and .

We also report the results of a chi-squared goodness-of-fit test based on the 2x2 contingency table shown below. We have 1 degree of freedom for this test. Our null hypothesis is that high or low LDL group is not associated with 5-year all-cause mortality.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **High LDL** | **Low LDL** | **Total** |
| **Died Within 5 Years** | 14 | 105 | **119** |
| **Alive After 5 years** | 93 | 513 | **606** |
| **Total** | **107** | **618** | **725** |

**Inference:** Using our values of and we compute a 95% confidence interval of

Since 0 is contained in this confidence interval, our observed sample would not be surprising if the 5-year all-cause mortality rate was the same in both LDL groups. In fact, our observed sample is consistent with the 5-year all-cause mortality rate being up to 10.95 percentage points higher in the low LDL group, or up to 3.14 percentage points higher in the high LDL group. As a result, we do not find evidence that there is an association between 5-year all-cause mortality and high or low LDL level.

In addition, the chi-squared goodness of fit test on our data yields a chi-squared test statistic of 1.014 and an accompanying p-value of 0.319. If were no association between high LDL or low LDL group and 5-year all-cause mortality, there is roughly a 31.9% chance of observing a sample as or more extreme than our actual data. As a result, we do not have sufficient evidence to reject our null hypothesis of high LDL/ low LDL group membership being independent of 5-year all-cause mortality.

6.

**Method:** Here, we use the odds ratio to evaluate a potential association between LDL and 5-year all-cause mortality. Again partitioning observed LDL into high (≥160 mg/dL) and low (<160 mg/dL), our sample yields the following 2x2 table:

|  |  |  |
| --- | --- | --- |
|  | **High LDL** | **Low LDL** |
| **Died Within 5 Years** | 14 | 105 |
| **Alive After 5 years** | 93 | 513 |

This table corresponds to the following odds ratio, standard error estimate, and accompanying 95% confidence interval:

**Inference:** We estimate that the risk of 5-year all-cause mortality for individuals with high LDL is 73.5% of the risk for individuals with low LDL. The accompanying 95% confidence interval for the odds ratio is [0.40, 1.34]. The sample that underlying this result would not be unusual if the true risk of 5-year all-cause mortality for observations with high LDL were between 40% and 134% of the risk for observations with low LDL. Our data are consistent an odds ratio of 1, since this value is contained in our 95% confidence interval for the odds ratio. In this framework, our sample does not provide evidence that odds of 5-year all-cause mortality varies between high LDL/low LDL groups.

7.

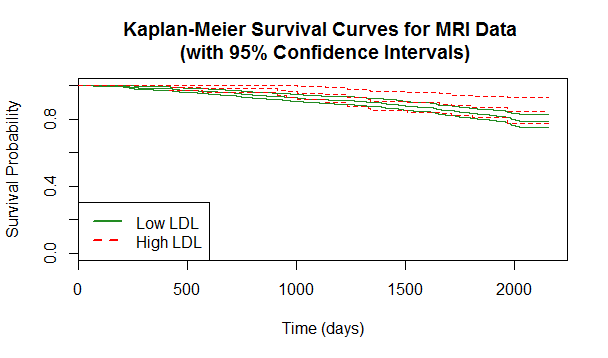
**Method:** We will evaluate whether the instantaneous rate of death varies between high and low LDL groups in two ways. First, we test the hypothesis:

where is the survival function for individuals with high LDL, is the survival function for individuals with low LDL, and varies over an individual’s entire period of observation (not just his or her first five years in the study). We will evaluate this hypothesis using the logrank test.

Second, we also compute a hazard ratio and corresponding confidence interval. To interpret the hazard ratio from our data, we assume that the survival curves for both groups are proportional; that is, there is a constant distance between the survival curves for both groups in a Kaplan-Meier plot. We present these Kaplan-Meier curves to check the validity of this assumption of proportional hazards underlying our comparison of instantaneous rate of death across high/low LDL groups.

**Inference:** The logrank test yields a p-value of 0.225. There is a 22.5% chance that, given survival functions between high and low LDL groups that are identical, we would have observed a difference in survival functions as or more extreme than what is actually present in our sample. As a result, the logrank test does not provide evidence to conclude that there is a difference in instantaneous risk of death between low LDL and high LDL groups represented in our data.

We find an estimated hazard ratio of 0.718 with a corresponding 95% confidence interval of [0.42, 1.23] and a p-value of 0.23. Under the assumption of proportional hazards, our data would be consistent with the hazard rates being equivalent between groups because 1 is contained within this confidence interval.. However, the Kaplan-Meier plot shown below suggests that the survival curves between the two groups are not equally spaced for all follow-up times (the gap noticeably widens at the end of the time interval). As a result, it does not seem reasonable to assume the presence of proportional hazards for these data, and it is not meaningful to interpret the hazard ratio we have computed.



8.

*A priori,* I would have preferred to test for an association between mortality and serum LDL by conducting a two-sample t-test to evaluate whether arithmetic mean LDL differs between 5-year all-cause mortality groups in our random sample of the elderly population. I would prefer not to dichotomize the LDL variable, which we would need to do to compare survival odds across high and low categories of LDL. By keeping LDL as it is reported in our data, we do not lose information about the distribution of LDL within our sample and we do not have to worry about finding the appropriate levels for a meaningful categorization of LDL into risk categories. I also think that determining whether arithmetic mean LDL differs between mortality groups would be helpful in answering our guiding scientific question.