Biost 518: Applied Biostatistics II
Biost 515: Biostatistics II

Emerson, Winter 2015

Homework #4 Key

February 2, 2015

Written problems: To be submitted as a MS-Word compatible file to the class Catalyst dropbox by 9:30 am on Monday, February 9, 2014. See the instructions for peer grading of the homework that are posted on the web pages.

> *On this (as all homeworks) Stata / R code and unedited Stata / R output is **TOTALLY** unacceptable. Instead, prepare a table of statistics gleaned from the Stata output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*
>
> ***Unless explicitly told otherwise in the statement of the problem, in all problems requesting "statistical analyses" (either descriptive or inferential), you should present both***
>
> - ***Methods:** A brief sentence or paragraph describing the statistical methods you used. This should be using wording suitable for a scientific journal, though it might be a little more detailed. A reader should be able to reproduce your analysis. DO NOT PROVIDE Stata OR R CODE.*
> - ***Inference:** A paragraph providing full statistical inference in answer to the question. Please see the supplementary document relating to "Reporting Associations" for details.*

This homework investigates associations between death from any cause and age, sex, and serum bilirubin in a population of patients with primary biliary cirrhosis who were enrolled in a randomized clinical trial (RCT) of D-penicillamine. The data can be found on the class web page (follow the link to Datasets) in the file labeled liver.txt. Documentation is in the file liver.doc.

1. Provide suitable descriptive statistics pertinent to the scientific questions addressed in this homework.

***Instructions for grading:** This problem is worth 10 points. There must be suitable descriptive statistics and a discussion of what they would mean scientifically. Points to consider in the grading include:*

- ***We are missing a lot of data for sex. This should have been noted.***
- ***Descriptive statistics should be presented within bilirubin categories. As with previous homeworks, a minimum of three such categories should be used.***
  - ***In this diseased population, it does not make sense to base the categories solely on the "normal range" of bilirubin. There should be some categories corresponding to "high" and "very high".***
  - ***As we have repeatedly discussed in class, in a diseased population laboratory measurements are often better log transformed, because a multiplicative mechanism obtains. Hence, categorization of bilirubin on the basis of doubling is probably most informative, though if only three categories are presented that is a moot point.***
- ***Descriptive statistics should also be presented for the survival probabilities. As noted below, statistical information is roughly proportional to the number of observed events, so that should also be presented. I presented descriptive statistics of the time of follow-up as computed using Kaplan-Meier estimates. While I was not really expecting that anyone would do this (see the key to***

HW #4 from Biost 517 in Fall 2011 for an example on how to do this), I do expect you to understand this principle from the key and be able to answer questions about this on exams.

<u>Ans:</u> *Methods:* Descriptive statistics for the time of follow-up were computed using Kaplan-Meier estimates in which observed deaths were treated as censored observations of the time of follow-up. Mean and standard deviation of time of follow-up were then calculated using the correspondence between means and the area under a survival curve. Summary statistics (mean, standard deviation, minimum, and maximum for age and bilirubin and percentages for male sex) are presented for the overall sample, as well as within strata defined by serum bilirubin level. Kaplan-Meier curves were used to compute summary statistics for the time to death and were similarly presented for the entire population and within the bilirubin strata. Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis.

***Results:* Age and bilirubin data is available on 418 patients with primary biliary cirrhosis, however, only 312 of those subjects have sex data available. Table 1 presents summary statistics for the study subjects on these variables. Patients ranged in age from 26-78 years (mean 50.7, SD 10.4) . Among the subjects with sex information recorded, 11.5% were male, in keeping with the previously known epidemiology of this disease. Serum bilirubin ranged from 0.3 – 28 mg/dL (mean 3.22, SD 4.41). Age distributions were similar across the strata defined by serum bilirubin level. Males were observed to be more prevalent in the middle ranges of the bilirubin strata, although there were extremely small numbers of patients in the highest stratum to be able to make any definitive statements.**

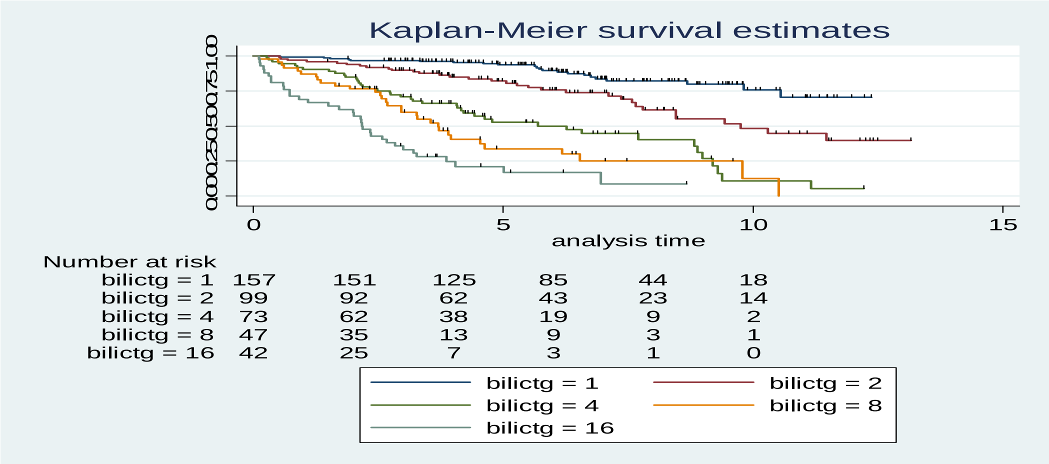
Subjects were followed for all cause mortality for an average of 6.88 years (SD 2.95) years, with deaths observed for 161 of the 418 subjects (125 of the 312 subjects with recorded sex). *(It is important to give the reader some idea of the amount of "statistical information" we have for our analysis. With complete (uncensored) data, the sample size tells us this. With censored data, however, the statistical information is roughly proportional to the number of events. Furthermore, in the presence of censored data, it is important to give some idea of the length of time subjects would have been followed for survival. It is not appropriate to just describe the distribution of the observation times for the subjects who are still alive. Instead, you should use Kaplan-Meier estimates of the censoring distribution. I will not be surprised if few (if any) students knew to summarize the censoring distribution in this way. But I do expect them to have made clear how many events were observed.)*

Table 1 also presents Kaplan-Meier estimates of patient survival probabilities at 2, 4, 6, and 8 years, with a depiction of the full curves by bilirubin strata in Figure 1. Evident from these results is a clear trend toward lessened probability of survival among groups with higher bilirubin levels at the time of study enrollment.

Table 1: Descriptive statistics for selected variables measured at time of study enrollment and probability of survival within strata defined by bilirubin. Statistics presented for age and bilirubin are the sample mean (sample SD; minimum – maximum; number of observations N with available measurements). Sex distribution among the 318 subjects with available data is characterized by the percent male, with the actual counts (n) and subjects with available data (N). Survival probabilities are computed using Kaplan-Meier estimates.

			Serum Bilirubin			
	Bili ≤ 1mg/dL	**1mg/dL < Bili ≤ 2mg/dL**	**2mg/dL < Bili ≤ 4mg/dL**	**4 mg/dL < Bili ≤ 8mg/dL**	**8 mg/dL < Bili**	**All Subjects**
Male (%, n)	6.0% (n= 7 / N= 116)	14.1% (n= 10 / N= 71)	20.3% (n= 12 / N= 59)	16.2% (n= 6 / N= 37)	3.4% (n= 1 / N= 29)	11.5% (n= 36 / N= 31
Age (y)	50.8 (10.0; 26 - 75; N=157)	51.4 (10.8; 31 - 77; N=99)	50.0 (11.2; 30 - 70; N=73)	50.9 (11.8; 31 - 78; N=47)	50.2 (8.5; 33 - 71; N=42)	50.7 (10.4; 26 - 78; N=4
Bilirubin (mg/dL)	0.69 (0.18; 0.3 - 1.0; N=157)	1.43 (0.29; 1.1 - 2.0; N=99)	2.91 (0.57; 2.1 - 4.0; N=73)	5.81 (1.03; 4.2 - 8.0; N=47)	14.53 (4.93; 8.1 - 28.0; N=42)	3.22 (4.41; 0.3 - 28.0; N=
Prob 2 yr survival	96.8%	93.9%	84.9%	76.5%	59.5%	88.0%
Prob 4 yr survival	96.1%	85.0%	66.2%	40.5%	24.6%	75.2%
Prob 6 yr survival	89.7%	75.7%	50.0%	33.7%	16.8%	66.4%
Prob 8 yr survival	82.2%	61.5%	40.1%	25.0%	8.4%	56.9%

Figure 1: Kaplan-Meier estimates of survival probabilities within strata defined by serum bilirubin level at time of study enrollment.



2. In prior homeworks using the Cardiovascular Health Study datasets, we were able to use logistic regression to investigate associations between mortality and various covariates. Why might such an approach not seem advisable with these data? (Consider the extent to which such analyses might be confounded and/or lack precision.)

***Instructions for grading*: This problem is worth 10 points. The student should have discussed the loss of information when only considering 18 months of follow-up, as well as the possibility that ignoring the censoring would have been potentially confounded. (They do not have to be as complete as I was below, but you should understand all these points in case I ever ask you the question again.)**

Ans: We are interested in survival probabilities at various points in time, but not all subjects were followed for the same period of time. Because probability of survival is very strongly influenced by the length of time you are considering, the potential observation time (the censoring time) is possibly an important effect modifier, confounder, or precision variable.

- **In the context of a survival analysis, effect modification by the censoring distribution would mean that we might get different answers if we followed subject for a different distribution of censoring times. Whether this obtains will ultimately depend on how we defined our summary measure. We generally specify our summary measure (e.g., hazard ratio or 5-year survival probability) in advance. When we use something like 5-year survival probability as our summary of association, a different censoring distribution might greatly affect the *precision* of our analysis, but it will not tend to lead to different answers. But when using the (time averaged) hazard ratio, we will likely obtain different "truths" if the survival curves do not have proportional hazards. We generally do not worry too much about effect modification even in this case for hypothesis testing, however, because under the strong null hypothesis (i.e., no difference in survival curves whatsoever), the curves do adhere to proportional hazards.**
- **Were we to just ignore the censoring, we might be concerned that the different groups were followed for different amounts of time. Since we have already noted that length of observation is strongly associated with the probability of survival, then we might worry that our analysis was confounded by differences in length of follow-up. This is the major reason that we always should use methods appropriate for censored data when we do not have complete observations. I will, however, note that in a RCT, the censoring distribution should be equal for the two arms *PROVIDED* the only reason a subject is censored is because he/she was still alive at time of data analysis. If, however, the subjects refuse to complete the study for any reason (e.g., toxic effects of treatment), then we are still worried about confounding.**
- **Even if the censoring distribution is equal across all groups, adjusting for the important precision variable of "length of follow-up" will provide much greater precision: We will not have to dichotomize our data as alive/dead and can instead consider the varying survival probabilities over time.**

We could of course have considered just dichotomizing our data at the earliest time that any subject was censored. However, as noted in the descriptive statistics, there were 161 deaths observed during the total period of follow-up. The earliest censored observation occurred at just under 18 months, by which time only 36 deaths had been observed. Hence, restricting attention to the data during the first 18 months will be extremely imprecise: we will have lost almost all our statistical information.

3. Perform a statistical regression analysis evaluating an association between serum bilirubin and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum bilirubin modeled as a continuous variable.

a. Include a full report of your inference about the association.

<u>Instructions for grading</u>*: *This problem is worth 10 points.

<u>Ans:</u> ***Methods:*** **Association between time to death and serum bilirubin at study entry was summarized by the hazard ratio estimated from a proportional hazards regression of the potentially censored time of death on serum bilirubin modeled as an untransformed continuous random variable. 95% confidence intervals and two-sided p values were computed using Wald statistics based on the Huber-White sandwich estimator for the standard error.** *(It is not crucial to use the robust SE in this case, though an argument can be made that it will relax the impact that non-proportional hazards might have on the estimation of the standard errors under an alternative hypothesis.)*

Results: **One hundred sixty-one (161) death were observed among the 418 patients enrolled in the study. Based on a proportional hazards regression analysis, when comparing populations differing in their serum bilirubin level, the instantaneous risk of death was estimated to be 15.2% higher for every 1 mg/dL difference in bilirubin levels (HR = 1.152). Based on a 95% confidence interval, such an observation is not unusual when the true hazard ratio is anywhere between 1.127 and 1.179. Thus this observation allows us to reject the null hypothesis of no association between survival time and serum bilirubin (two sided P < 0.0001).**

b. For each population defined by serum bilirubin value, compute the hazard ratio relative to a group having serum bilirubin of 1 mg/dL. (This will be used in problem 6). If *HR* is the hazard ratio (use the actual hazard ratio estimate) obtained from your regression model, this can be effected by the Stata code

```
gen fithrA = HR ^ (bili – 1)
```

It could also be computed by creating a centered bilirubin variable, and then using the Stata `predict` command

```
gen cbili = bili – 1
stcox cbili
predict fithrA
```

4. Perform a statistical regression analysis evaluating an association between serum bilirubin and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum bilirubin modeled as a continuous logarithmically transformed variable.

a. Why might this analysis be preferred *a priori?*

<u>Instructions for grading</u>*: *This problem is worth 5 points.

<u>Ans:</u> The population being studied all have liver disease that affects bilirubin metabolism and excretion. Hence, we expect most of them to have abnormal measurements, and it is not unusual that stages of disease might act multiplicatively on the bilirubin level. Hence we might expect more homogeneity of effect by considering doublings of bilirubin. Using log transformations does such a comparison.

b. Include a full report of your inference about the association.

<u>Instructions for grading</u>*: *This problem is worth 10 points.

<u>Ans:</u> ***Methods:*** **Association between time to death and serum bilirubin at study entry was summarized by the hazard ratio estimated from a proportional hazards regression of the potentially censored time of death on serum bilirubin modeled as an log transformed continuous random variable. 95% confidence intervals and two-sided p values were computed using Wald statistics based on the Huber-White sandwich estimator for the standard error.** *(It is not crucial to use the robust SE in this case,* *though an argument can be made that it will relax the impact that non-proportional hazards might have on the estimation of the standard errors under an alternative hypothesis.)*

Results: **One hundred sixty-one (161) death were observed among the 418 patients enrolled in the study. Based on a proportional hazards regression analysis, when comparing populations differing in their serum bilirubin level, the instantaneous risk of death was estimated to be 1.98-fold higher for every doubling of bilirubin levels (HR = 1.984). Based on a 95% confidence interval, such an observation is not unusual when the true hazard ratio is anywhere between 1.781 and 2.212. Thus this observation allows us to reject the null hypothesis of no association between survival time and serum bilirubin (two sided P < 0.0001).**

c. For each population defined by serum bilirubin value, compute the hazard ratio relative to a group having serum LDL of 1 mg/dL. (This will be used in problem 6). If *HR* is the hazard ratio (use the actual hazard ratio estimate) obtained from your regression model, this can be effected by the Stata code

```
gen logbili = log(bili)
stcox logbili
fithrB = HR ^ (logbili)
```

(Note that the log(1) = 0 when using any base, so there is no need to rescale by the bilirubin values. Note also that you might want to use a different base in your logarithmic transformation in order to facilitate more natural reporting of effects.)

5. One approach to testing to see whether an association between the response and the predictor of interest is adequately modeled by an untransformed continuous variable is to add some other transformation to the model and see if that added covariate provides statistically significant improved “fit” of the data. In this case, we could test for “linearity” of the bilirubin association with the log hazard ratio by including both the untransformed and log transformed bilirubin. (Other alternatives might have been bilirubin and bilirubin squared, but in this case our *a priori* interest in the log bilirubin might drive us to the specified analysis.)

a. Provide full inference related to the question of whether the association is linear.

Instructions for grading: This problem is worth 10 points.

Ans: *Methods:* **The linearity of the association between time to death and serum bilirubin at study entry was investigated using a proportional hazards regression of the potentially censored time of death on serum bilirubin modeled both as untransformed and log transformed continuous random variables included in the model simultaneously. We used the regression parameter estimate for the log transformed bilirubin to test the null hypothesis that a linear association between the log hazard ratio and untransformed bilirubin was adequate to describe the data. Evidence that the log bilirubin coefficient was different from zero was interpreted as evidence of a nonlinear association between the log hazard and bilirubin. 95% confidence intervals and two-sided p values were computed using Wald statistics based on the Huber-White sandwich estimator for the standard error.** *(It is not crucial to use the robust SE in this case, though an argument can be made that it will relax the impact that non-proportional hazards might have on the estimation of the standard errors under an alternative hypothesis.)*

Results: **Based on a proportional hazards regression analysis that included both a linear and a log transformed bilirubin measurement, the regression coefficient for the log transformed term was found to be significantly different from 0 (two-sided P < 0.0005). We thus reject the null hypothesis that the association between survival and bilirubin is well-described by a log hazard ratio that is linear in bilirubin.** *(I note that had I asked a pre-specified question of whether a multiplicative association existed, we could have used this same model to detect departures from a model in which the log hazard was linear in log bilirubin. Based on the P value of 0.148 for the untransformed bilirubin term, I would not have been able* *to reject a null hypothesis that a model with log bilirubin was adequate. That does not, however, prove that the log bilirubin model is "the truth": Absence of evidence is not evidence of absence.)*

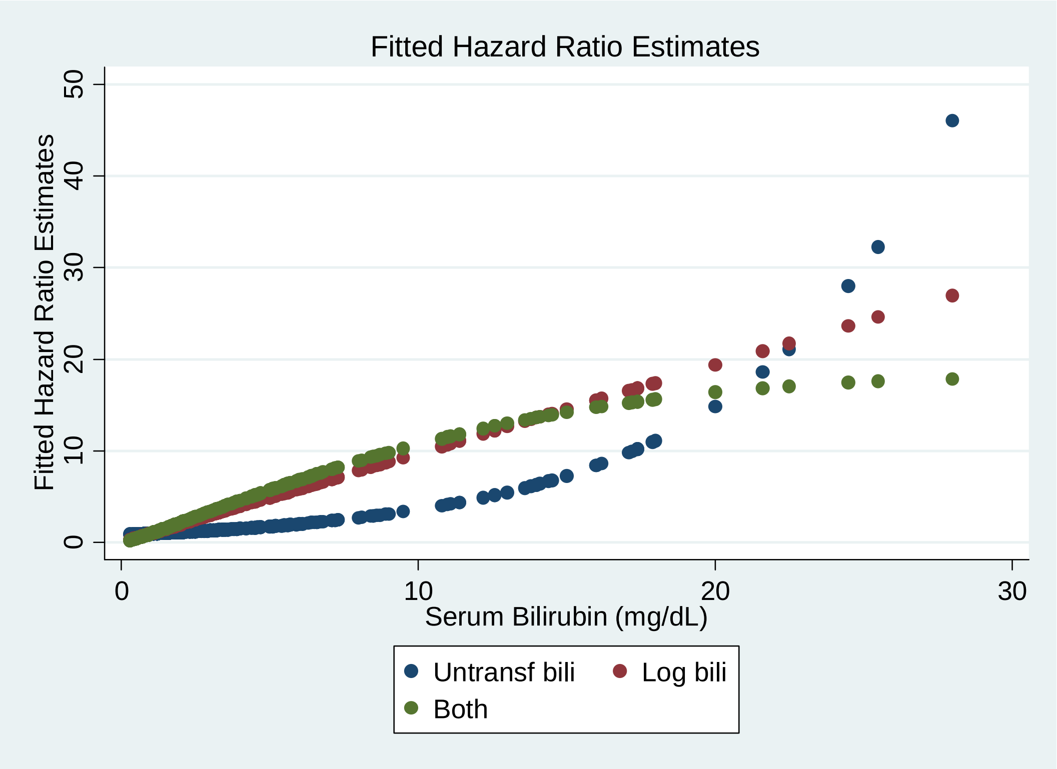
b. Again, save the fitted values from this model by obtaining the estimated HRs relative to a group with bilirubin of 1 mg/dl. (This will be used in problem 6.)

6. Display a graph with the fitted hazard ratios from problems 3 - 5. Comment on any similarities or differences of the fitted values from the three models.

***Instructions for grading*: This problem is worth 10 points.**

Ans: Figure 2 displays the fitted values from each of the models. We see that there are marked differences between the fitted values from the model with untransformed bilirubin and the model with log transformed bilirubin. The model fit in problem 5 had greater flexibility to "choose" either of the first two models or something in between. It can be seen that the model with both terms most closely agrees with the log bilirubin model, though it does depart from it somewhat at higher levels of bilirubin. Hence, the speculation that the log transformation of bilirubin might be a better "fit" is descriptively supported by these few comparisons.

Figure 2: Fitted hazard ratio estimates for the three models in problems 3 – 5.



7. We are interested in considering analyses of the association between all cause mortality and serum bilirubin after adjustment for age and sex.

a. What evidence is present in the data that would make you think that either sex or age might have confounded the association between death and bilirubin? (In real life, we would ideally decide whether to adjust for potential confounding in our pre-specified statistical analysis plan (SAP)).

***Instructions for grading*: This problem is worth 10 points. The student must address both the association between the potential confounder and the POI (bilirubin) and any causal association between the potential confounder and the response (survival time).**

<u>Ans:</u> Both age and sex are well-known to be strongly associated with probability of survival, and those associations are generally regarded as causal. Based on that knowledge, we would certainly be entertaining that idea that there might be confounding. (Evidence based on PH regression analyses is presented in the answer to part b, below.) From the descriptive statistics presented in Table 1, we did not see any particular trend toward meaningful differences in the age distribution by bilirubin, but there was a tendency for men to have higher bilirubin levels than women. Hence, there is not strong evidence that age would confound the association between survival and bilirubin, but sex might.

b. What evidence is present in the data that would make you think that either sex or age might have added precision to the analysis of the association between death and bilirubin? (In real life, we would ideally decide whether to adjust in our pre-specified SAP).

<u>Instructions for grading</u>: This problem is worth 10 points. Note that I used logarithmically transformed bilirubin in my analyses. I think that best. But a student who used untransformed bilirubin can get full credit if they answered correctly for their model. In performing analyses, students will need to have considered that a reduced sample is used when sex is included in the model, because many observations are missing sex.

<u>Ans:</u> Both age and sex are well-known to be strongly associated with probability of survival in the general population, but it is not as well-known whether such associations exist within PBC patients. To explore evidence that these variables might add precision to our analysis, we would want to consider the hazard ratio estimates from a model that includes age, sex, and log bilirubin. This is supported by a proportional hazards regression of survival time on bilirubin, sex, and age:

- **After adjusting for log transformed bilirubin and sex, we estimate a 1.039 fold increased risk of mortality for each year difference in age, Over the 50 year range of ages, such a HR would likely confer some added precision, though it is not as substantial an effect as seen with doubling of bilirubin. (I further note that the observed association between survival and age is statistically significant with P < 0.0005.)**
- **After adjusting for log transformed bilirubin and age, we estimate a 1.071 fold increased risk of mortality for males compared to females, Such a HR does not seem very substantial compared to the effects of a doubling of bilirubin or a ten year difference in age. (I further note that the observed association between survival and sex is not statistically significant with P= 0.799.)**

c. Provide full inference regarding an association between death and bilirubin after adjustment for sex and age.

<u>Instructions for grading</u>: This problem is worth 10 points.

<u>Ans:</u> ***Methods:*** **Association between time to death and serum bilirubin at study entry was summarized by the hazard ratio estimated from a proportional hazards regression of the potentially censored time of death on serum bilirubin modeled as a log transformed continuous random variable with adjustment for continuous age and sex. 95% confidence intervals and two-sided p values were computed using Wald statistics based on the Huber-White sandwich estimator for the standard error.** *(It is not crucial to use the robust SE in this case, though an argument can be made that it will relax the impact that non-proportional hazards might have on the estimation of the standard errors under an alternative hypothesis.)*

Results: **One hundred twenty-five (125) deaths were observed among the 312 patients enrolled in the study and having sex recorded. Based on a proportional hazards regression analysis, when comparing populations differing in their serum bilirubin level but having similar age and sex, the instantaneous risk of death was estimated to be 2.108-fold higher for every doubling of bilirubin levels (HR = 2.108). Based on a 95% confidence interval, such an observation is not unusual when the true adjusted hazard**

ratio is anywhere between 1.840 and 2.416. Thus this observation allows us to reject the null hypothesis of no association between survival time and serum bilirubin (two sided P < 0.0001).

(We can compare the adjusted HR for log bilirubin to the unadjusted HR in the same population (i.e., the 312 subjects who had sex recorded). In such an unadjusted analysis, the estimated HR for a doubling of bilirubin is 2.121.

- *Thus, after adjustment for age and sex, we obtained a HR estimate that was essentially unchanged. This argues that the precision added by age is not all that great compared to the high degree of association between survival and bilirubin. We can see that in the magnitude of the HR for a doubling of bilirubin and that for a difference in age: $1.039^{19.5} = 2.108$, suggesting that one doubling of bilirubin has the same effect on survival as 19.5 years in age. In the range of our data, we have more than 6 doublings of bilirubin.*
- *To the extent that we believed that there was a real difference (rather than just statistical noise) between 2.108 and 2.121, we see that the adjusted estimate is closer to the null. This is suggestive of some slight confounding, as we would have expected a pure precision variable to "deattenuate" the collapsed HR.)*

8. Note that in the above analyses, we completely ignored the intervention in the RCT? What impact could this have had on our results?

***Instructions for grading*: This problem is worth 10 points. The student either pointed out that the RCT in some sense precluded confounding, or they did not. Hence, the student either got the 10 points, or they did not.**

Ans: This was a randomized clinical trial, and hence there is no association between the treatment and the bilirubin value on average. Hence, treatment is at most a precision variable and does not present any confounding. To the extent that the treatment does not confer a great survival advantage, failure to adjust for the treatment variable will have little effect on the analysis.