Biost 517: Applied Biostatistics I
Emerson, Fall 2010

Homework #7 Key
November 17, 2010

Written problems: To be handed in at the beginning of class on Wednesday, November 24, 2010.

Questions for Biost 514 and Biost 517:

The written problems all refer to the data on the usefulness of post-treatment PSA levels for prognosis in hormonally treated prostate cancer (the file psa.txt on the class web pages). We consider several alternative strategies to assess whether there is an association between time in remission and nadir PSA level. In all problems, provide relevant descriptive statistics and as complete statistical inference as possible (i.e., provide point estimates, confidence intervals, and p values where possible, along with a statement of your scientific/statistical conclusions).

1. Base your analysis on a comparison of mean nadir PSA across groups defined by whether the patient relapsed within the specified timeframe or not.

    a. Consider relapse within 12 months. Can the requested analysis be performed with this data? If so, provide the analysis. If not, explain why not.

Ans: Because no subjects were censored prior to 12 months, we know for each subject whether he relapsed within 12 months. We are thus free to divide this cross-sectional sample into groups defined by whether the patient relapses within 12 months.

The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. Relapse within 12 months was observed in 13 patients, and they were observed to have an average nadir PSA of 46.4 ng/ml (SD 64.4 ng/ml, range 0.5 – 183 ng/ml). The remaining 37 patients stayed in remission for at least 12 months, and their average nadir PSA was observed to be 5.79 ng/ml (SD 16.5 ng/ml, range 0.1 – 92 ng/ml). A comparison of the two groups finds that nadir PSA for patients relapsing within 12 months is estimated to average 40.6 ng/ml higher than that for patients who remain in remission for at least 12 months. Such an observation is statistically different from 0 (two-sided P= 0.0432 by a t test that allows for the possibility of unequal variances), with a 95% confidence interval suggesting that the observed result is not unusual if the true average nadir PSA were anywhere from 1.46 ng/ml higher to 79.8 ng/ml higher for the group relapsing within 12 months than for the group remaining in remission at least 12 months.

(We could have considered the t test that presumes equal variances. Had I done so, the above paragraph might have (erroneously) read:

> *The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. Relapse within 12 months was observed in 13 patients, and they were observed to have an average nadir PSA of 46.4 ng/ml (SD 64.4 ng/ml, range 0.5 – 183 ng/ml). The remaining 37 patients stayed in remission for at least 12 months, and their average nadir PSA was observed to be 5.79 ng/ml (SD 16.5 ng/ml, range 0.1 – 92 ng/ml). A comparison of the two groups finds that nadir PSA for patients relapsing within 12 months is estimated to average 40.6 ng/ml higher than that for patients who remain in*

remission for at least 12 months. Such an observation is statistically different from 0 (two-sided P= 0.0008 by a t test that presumes equality of variances across groups), with a 95% confidence interval suggesting that the observed result is not unusual if the true average nadir PSA were anywhere from 17.8 ng/ml higher to 63.5 ng/ml higher for the group relapsing within 12 months than for the group remaining in remission at least 24 months.

The only thing that changed was the p value and the CI. Note however that the standard deviations are strikingly different between the two groups. Note further that the group with the higher standard deviation has the smaller sample size. If the variances are truly different between the groups in such a setting (i.e., a setting in which the group with the larger variance has the smaller sample size), the inference about the mean tends to be anti-conservative: the p value is spuriously low and the CI is too narrow. The inference based on the p value is valid about a difference in the distributions, but we cannot be confident that it is a difference in the means—it could be a difference in the variances that leads to an observed spurious difference in sample means. I believe that the inference based on the t test that allows for the possibility of unequal variances to be vastly better for this problem.

I also note that when variances are truly unequal and the larger sample size is in the group with the larger variance, the t test that presumes equal variances leads to conservative inference about the mean—the p values are too large and the CI too wide.)

b. Consider relapse within 24 months. Can the requested analysis be performed with this data? If so, provide the analysis. If not, explain why not.

Ans: Because no subjects were censored prior to 24 months, we know for each subject whether he relapsed within 24 months. We are thus free to divide this cross-sectional sample into groups defined by whether the patient relapses within 24 months.

The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. Relapse within 24 months was observed in 22 patients, and they were observed to have an average nadir PSA of 31.9 ng/ml (SD 52.5 ng/ml, range 0.5 – 183 ng/ml). The remaining 28 patients stayed in remission for at least 24 months, and their average nadir PSA was observed to be 4.12 ng/ml (SD 17.3 ng/ml, range 0.1 – 92 ng/ml). A comparison of the two groups finds that nadir PSA for patients relapsing within 24 months is estimated to average 27.8 ng/ml higher than that for patients who remain in remission for at least 24 months. Such an observation is statistically different from 0 (two-sided P= 0.0250 by a t test that allows for the possibility of unequal variances), with a 95% confidence interval suggesting that the observed result is not unusual if the true average nadir PSA were anywhere from 3.79 ng/ml higher to 51.9 ng/ml higher for the group relapsing within 24 months than for the group remaining in remission at least 24 months.

(The comments made in part a about the t test that presumes equal variances apply here as well.

In the Stata file, note that the CI for the mean nadir PAS for the group remaining in remission for 24 months includes impossible negative values. This certainly suggests that the sample size of 28 is not very large relative to the highly skewed distribution of data. I would infer the same about the sample size of 22 in the other group.

I also note that we had a more statistically significant difference in part b than part a. Pretending for a moment that we absolutely know that the differences are real, we might have predicted

having the greater precision in part b, even though the actual difference between the groups is less: 40.6 vs 27.8. The optimal sample size ratio in order to gain the greatest precision is to have sample sizes proportional to standard deviations. In part a, the ratio of sample standard deviations for the relapsing group relative to the non-relapsing group was 3.90, while the ratio of sample sizes was 0.35. In part b, the ratio of sample standard deviations for the relapsing group relative to the non-relapsing group was 3.04, while the ratio of sample sizes was 0.786—closer agreement.)

c. Consider relapse within 36 months. Can the requested analysis be performed with this data? If so, provide the analysis. If not, explain why not.

Ans: One subject was censored as early as 24 months, thus we cannot know for each subject whether he relapsed within 36 months. We are thus not free to divide this cross-sectional sample into groups defined by whether the patient relapses within 36 months, and we would have to use alternative analysis strategies.

2. Base your analysis on a comparison of geometric mean nadir PSA across groups defined by whether the patient relapsed within 24 months or not.

Ans: Because no subjects were censored prior to 24 months, we know for each subject whether he relapsed within 24 months. We are thus free to divide this cross-sectional sample into groups defined by whether the patient relapses within 24 months.

The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. Relapse within 24 months was observed in 22 patients, and they were observed to have an geometric mean nadir PSA of 8.31 ng/ml (range 0.5 – 183 ng/ml). The remaining 28 patients stayed in remission for at least 24 months, and their geometric mean nadir PSA was observed to be 0.520 ng/ml (range 0.1 – 92 ng/ml). A comparison of the two groups finds that geometric mean nadir PSA for patients relapsing within 24 months is estimated to be 16.0-fold that for patients who remain in remission for at least 24 months. Such an observation is statistically different from 0 (two-sided P < .0001 by a t test that allows for the possibility of unequal variances), with a 95% confidence interval suggesting that the observed result is not unusual if the true geometric nadir PSA for the group relapsing within 12 months were anywhere from 5.98-fold to 42.8-fold that of the group remaining in remission at least 24 months.

(Wording in multiplicative models is always difficult, and you must be very careful to make sure you are as clear as possible when writing the results and that you study the results when you read someone else's analysis—statistical analyses never make for good prose.

Suppose that the estimated ratio of group A to group B is 1.24. Then we might say the GM for group A is 24% higher or 1.24-fold that of group B.

If the estimated ratio of group A to group B is 0.13, then we might say the GM for group A is 87% lower or 0.13 times as high as that of group B.

We can always invert ratios if we think it is less prone to confusion. I did so for this analysis. I found it easier to stress that the GM for relapsers was 16-fold that of the non-relapsers, rather than saying that the GM for the non-relapsers was only 0.06 times as high that of the relapsers. But there are always common misconceptions in the non-statistical readership: If Group A has a GM of 90 and Group B has a GM of 60, then the GM of Group A is 50% higher than (1.5-fold)

that of Group B. On the other hand, Group B as a GM that is 33.3% lower than (0.667-fold) that of Group A.)

3. Base your analysis on a comparison of the probability of remaining in remission for 24 months as a function of nadir PSA levels above the specified thresholds.

    a. Consider groups defined by nadir PSA greater than 1 ng/ml.

Ans: Because no subjects were censored prior to 24 months, we know for each subject whether he relapsed within 24 months. We are thus free to use the binary variable indicating whether the patient relapses within 24 months as a response variable in this cross-sectional sample.

The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. The nadir PSA was observed to be greater than 1 ng/ml in 24 subjects, and 18 (75.0%) of them were observed to relapse within 24 months. The nadir PSA was observed to be less than or equal to 1 ng/ml in 26 subjects, and 4 (15.4%) of them were observed to relapse within 24 months. The comparison of the two groups defined by nadir PSA thus finds that the group with higher nadir PSA had a probability of relapse within 24 months that was 59.6% higher (absolute difference) than that in the low nadir PSA group. Such an observation is statistically different from 0 (two-sided P < 0.0001 by a chi square test for independence), with a 95% confidence interval suggesting that the observed result is not unusual if the true absolute difference in relapse probabilities at 24 months were anywhere from 37.4% higher to 81.8% higher for the group with nadir PSA greater than 1 ng/ml than for the group with the lower nadir PSA.

(I could have used Fisher's exact test, but with the average event rate of 44% and roughly equal sample sizes, the expected counts in each cell of the 2 x 2 contingency table is greater than 5, and standard criteria would say that the asymptotic distribution was okay to use.

Note the need to be careful when describing differences in proportions. The "absolute difference" (or "attributable risk" or "risk difference") is probably used most often, but the "relative difference" (or "relative risk reduction/increase") is sometimes used because with rare events it is thought to be less subject to imprecision in eligibility criteria (ask me about this if you are interested in why and when this obtains). So if Group A had an event probability of 30% and Group B had an event probability of 35%, then there was an absolute difference of 5%, but Group A had a relative decrease of 0.05 / 0.35 = 14.3%.)

    b. Consider groups defined by nadir PSA greater than 2 ng/ml.

Ans: Because no subjects were censored prior to 24 months, we know for each subject whether he relapsed within 24 months. We are thus free to use the binary variable indicating whether the patient relapses within 24 months as a response variable in this cross-sectional sample.

The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. The nadir PSA was observed to be greater than 2 ng/ml in 19 subjects, and 15 (78.9%) of them were observed to relapse within 24 months. The nadir PSA was observed to be less than or equal to 2 ng/ml in 31 subjects, and 7 (22.6%) of them were observed to relapse within 24 months. The comparison of the two groups defined by nadir PSA thus finds that the group with higher nadir PSA had a probability of relapse

within 24 months that was 56.4% higher (absolute difference) than that in the low nadir PSA group. Such an observation is statistically different from 0 (two-sided P = 0.0001 by a chi square test for independence), with a 95% confidence interval suggesting that the observed result is not unusual if the true absolute difference in relapse probabilities at 24 months were anywhere from 32.9% higher to 79.9% higher for the group with nadir PSA greater than 2 ng/ml than for the group with the lower nadir PSA.

4. Base your analysis on a comparison of the odds of remaining in remission for 24 months as a function of nadir PSA levels above the specified thresholds.

    a. Consider groups defined by nadir PSA greater than 1 ng/ml.

Ans: Because no subjects were censored prior to 24 months, we know for each subject whether he relapsed within 24 months. We are thus free to use the binary variable indicating whether the patient relapses within 24 months as a response variable in this cross-sectional sample.

The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. The nadir PSA was observed to be greater than 1 ng/ml in 24 subjects, and 18 (75.0%) of them were observed to relapse within 24 months (odds of relapse 3.00). The nadir PSA was observed to be less than or equal to 1 ng/ml in 26 subjects, and 4 (15.4%) of them were observed to relapse within 24 months (odds of relapse 0.182). The comparison of the two groups defined by nadir PSA thus finds that the odds of relapse in the group with higher nadir PSA is 16.5-fold that in the low nadir PSA group. Such an observation is statistically different from 0 (two-sided P < 0.0001 by a chi square test for independence), with a 95% confidence interval suggesting that the observed result is not unusual if the true odds of relapse at 24 months for the group with nadir PSA greater than 1 ng/ml were anywhere from 3.42-fold to 88.2-fold that of the group with the lower nadir PSA.

    b. Consider groups defined by nadir PSA greater than 2 ng/ml.

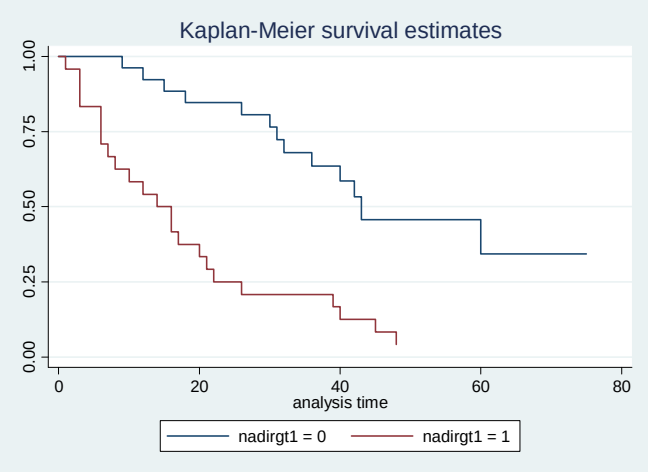
Ans: The study data consist of measurements made on 50 consecutive hormonally treated prostate cancer patients. The nadir PSA was observed to be greater than 2 ng/ml in 19 subjects, and 15 (78.9%) of them were observed to relapse within 24 months (odds of relapse 3.75). The nadir PSA was observed to be less than or equal to 2 ng/ml in 31 subjects, and 7 (22.6%) of them were observed to relapse within 24 months (odds of relapse 0.292). The comparison of the two groups defined by nadir PSA thus finds that the odds of relapse in the group with higher nadir PSA is 12.9-fold that in the low nadir PSA group. Such an observation is statistically different from 0 (two-sided P = 0.0001 by a chi square test for independence), with a 95% confidence interval suggesting that the observed result is not unusual if the true odds of relapse at 24 months for the group with nadir PSA greater than 2 ng/ml were anywhere from 2.74-fold to 67.5-fold that of the group with the lower nadir PSA.

5. Base your analysis on a comparison of the instantaneous risk of relapse as a function of nadir PSA levels above the specified thresholds.

    a. Consider groups defined by nadir PSA greater than 1 ng/ml.

Ans: **Fifty men with hormonally treated prostate cancer were followed for cancer relapse for an average of 52.1 months (range 24 – 75 months). During the period of follow-up, relapse was observed in 36 patients.** *(Note that I took these estimates from the censoring distribution, not from descriptive statistics based solely on the variable obstime. This is how it should be done. The statistical information in survival analyses is more proportional to the number of events than to the number of observations, so it is sometimes useful to provide the number of observed events. Sometimes I might also have reported the number of events observed in each group in order to give a feel for the statistical information available to estimate the survival probabilities and quantiles. The precision of the logrank statistic, however, relates to the number of subjects at risk in each group at the time of each observed event and the total number of events in both groups combined, rather than the number of events in each group. To that end, I could have also reported the censoring distribution in each group.)*

The nadir PSA was observed to be greater than 1 ng/ml in 24 subjects, and to be less than or equal to 1 ng/ml in 26 subjects. The following figure displays Kaplan-Meier estimates of the relapse-free survival of men within those two strata. The estimated probability of relapse-free survival at 1, 2, 3, and 4 years was 54.2%, 25.0%, 20.8%, and 4.17%, respectively, in the group having higher nadir PSA, and was 92.3%, 84.6%, 63.5%, and 45.7% at the corresponding times in the group having lower nadir PSA. The 25th percentile and median of the relapse-free survival time distribution was 6 months and 14 months, respectively in the group having higher nadir PSA, and the corresponding quantiles were 31 months and 43 months in the group having lower nadir PSA. *(I probably would have given either the estimated survival probabilities at fixed points in time or the estimated quantiles, but not both in real life. Note that I chose not to report the 75th percentile, because it was not estimable in the low nadir group. Sometimes, I would have reported that percentile in the high nadir group and then reported that it was inestimable in the low nadir group.)*

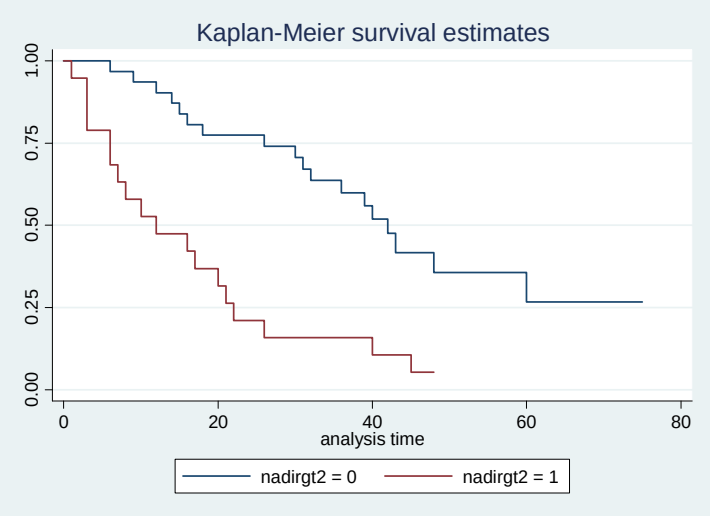


The comparison of the two groups defined by nadir PSA in a proportional hazards analysis estimates that the instantaneous risk of the earlier of relapse of death in the group having nadir PSA greater than 1 ng/ml is 4.19-fold that in the low nadir PSA group. Such an observation is statistically different from 1 (two-sided P < 0.0001 by a logrank test), with a 95% confidence interval suggesting that the observed result is not unusual if the true instantaneous risk for the group with nadir PSA greater than 1 ng/ml were anywhere from 2.06-fold to 8.53-fold that of the group with the lower nadir PSA.

b. Consider groups defined by nadir PSA greater than 2 ng/ml.

Ans: Fifty men with hormonally treated prostate cancer were followed for cancer relapse for an average of 52.1 months (range 24 – 75 months). During the period of follow-up, relapse was observed in 36 patients.

The nadir PSA was observed to be greater than 2 ng/ml in 19 subjects, and to be less than or equal to 2 ng/ml in 31 subjects. The following figure displays Kaplan-Meier estimates of the relapse-free survival of men within those two strata. The estimated probability of relapse-free survival at 1, 2, 3, and 4 years was 47.4%, 21.1%, 15.8%, and 5.26%, respectively, in the group having higher nadir PSA, and was 90.3%, 77.4%, 59.9%, and 35.7% at the corresponding times in the group having lower nadir PSA. The 25th percentile and median of the relapse-free survival time distribution was 6 months and 12 months, respectively in the group having higher nadir PSA, and the corresponding quantiles were 26 months and 42 months in the group having lower nadir PSA.



The comparison of the two groups defined by nadir PSA in a proportional hazards analysis estimates that the instantaneous risk of the earlier of relapse of death in the group having nadir PSA greater than 1 ng/ml is 3.64-fold that in the low nadir PSA group. Such an observation is statistically different from 1 (two-sided P = 0.0001 by a logrank test), with a 95% confidence interval suggesting that the observed result is not unusual if the true instantaneous risk for the group with nadir PSA greater than 2 ng/ml were anywhere from 1.85-fold to 7.18-fold that of the group with the lower nadir PSA.

6. How similar are the decisions you make about associations in problems 1- 5? Which analyses would you have preferred *a priori*?

Ans: In each of the analyses, we concluded with 95% confidence that there is an association between nadir PSA and the probability of remaining in remission. Though it need not be the case in every instance, the tendency is towards higher nadir PSAs being associated with shorter times in remission, no matter which of the various summary measures were being considered.

There is of course a multiple comparison problem if we adopt a strategy of performing each one of these analyses and then choosing the one with the lowest P value. Hence it is extremely important that a single primary analysis be identified. In this case, I would

typically choose an analysis based on the hazard ratio, because it tends to consider a general tendency over all follow-up times, and thus would generally provide greater precision.

Then as secondary analyses, I might have considered comparisons of relapse free survival probabilities at selected clinically important times (which clinically important times would not necessarily be restricted to those times that I had no censoring—see HW # 8).

I note that the comparison of distributions of mean or geometric mean nadir PSA across groups defined by relapse status are not as desirable, because they are conditioning on the future and looking at the past. When given the choice, I prefer to condition on the past (i.e., nadir PSA) and look to the future (i.e., time in remission). But if I did not know how to do anything else, I could still answer my scientific question about associations using the approach in either problem 1 or 2. Of those two, I would prefer the geometric means based on my knowledge that PSA measurements can become quite large (heavily skewed distributions) in disease. (Such an observation is not limited to PSA, by any means.) In such settings, it is often the case that geometric means are estimated more precisely than are means.