Biost 517: Applied Biostatistics I
Emerson, Fall 2010

Homework #6 Key
Wednesday November 10, 2010

Written problems: To be handed in at the beginning of class on Wednesday, November 17, 2010.

Questions for Biost 514 and Biost 517:

The written problems all refer to the laboratory data for the clinical trial of methotrexate in PBC as stored on the class web pages. In this homework we will consider the serum bilirubin value five years post randomization (week 260) for those patients for whom data is available. These measurements will be compared to patient's baseline (week 1) measurement.

Some overall comments about this problem:

In lecture we covered one-sample tests, so in problems 1-3 I only asked you to analyze each dose group separately. This is NOT the way a clinical trial should really be analyzed. As noted below in some of the answers, we would really want to compare the results for the placebo group to the results for the methotrexate group to account for any natural trends over time in bilirubin measurements. That is, the placebo group might naturally have decreased or increased bilirubin measurements due to aging (they are a five years older at the end of the study), due to environmental effects (who knows, maybe pollution causes changes in bilirubin levels), due to secular trends in behavior (maybe diet is changing), due to seasonal effects in diet (although we are less worried about this when the treatment period is 5 years), or due to progression of disease.

It would not be acceptable in a proper scientific report (and this is the definition of "proper") to merely report the analyses given below in problems 1-3 and to conclude a treatment effect because significant differences were (at times) observed in the methotrexate group but not the placebo group, or vice versa. (We could have done that if there were nonoverlapping CI, but since the CI for the two groups do overlap, we would have to perform a proper test comparing the two groups directly.)

I also note that I had you do several different analyses. In real life you would have *a priori* chosen one of the analyses (mean difference, mean ratio, proportion decreasing, ratio of geometric means, median difference) as your primary analysis. Otherwise, there is a multiple comparison issue.

This exercise was meant to have you gain experience in performing and interpreting the tests, and to gain a little insight into how they behave relative to each other. In real life, I looked at geometric means.

1. Perform an analysis to assess whether the placebo group had a change in mean bilirubin level five years post randomization. Use both the difference between measurements and the ratio between measurements as a measure of comparison. Provide relevant point estimates, 95% confidence intervals, and P values. Do the same for the methotrexate group. Make clear the interpretation of your confidence interval and P values, including the scientific relevance of your results.

Ans: **One hundred thirty-three (133) subjects were randomized to receive placebo, and all of those patients had bilirubin measurements available at time of randomization. However, only 116 of those subjects had bilirubin measurements available at 5 years. Reasons for missing measurements at 5 years included study discontinuation due to prior death (*n* subjects), liver transplantation (*n* subjects), withdrawal of consent (*n* subjects), or patient loss to follow-up (*n* subjects). Patients who discontinued their use of the study drug (*n* subjects) prior to 5 years were still included in this analysis. The distribution of bilirubin, albumin, PTINR, FVC, and DLCO measurements at the time of randomization did not differ systematically between the 110 subjects for whom 5 year measurements were available and the 23 subjects missing measurements at 5 years.** *(The annotated Stata log file contains the descriptive statistics that I looked at for this latter comparison. I might also have looked at the values of the laboratory and pulmonary measurements at other times for those patients who dropped out relative to those who continued. Nothing can tell us whether the measurements we would have obtained were unusual, but had we seen trends in the baseline or intermediate outcome data, we would certainly worry more about nonignorable missing data, i.e., we would have worried that the patients for whom we were missing data might have had very different values than we observed in the other patients.)*

Analysis of the change in bilirubin levels in the placebo group is restricted to those 116 patients with data available at 5 years. At the time of randomization, the average bilirubin measurement in that group was 0.680 mg/dl (SD 0.364 mg/dl). After the 5 year treatment period, the average bilirubin measurement was 0.988 mg/dl (SD 1.41 mg/dl). *(Note my careful wording "after the 5 year treatment period" rather than "after 5 years of treatment". Fact is, there are almost always some patients who do not take the treatment the full time, but we still analyze their data with the rest of the patients. This is called an "intent to treat" analysis, and that forms the standard for reporting clinical trial results. Ideally we would have obtained measurements on the patients who dropped out for reasons other than death or perhaps liver transplant.)*

(An analysis based on the mean difference in bilirubin measurements)
The placebo group thus averaged an increase of 0.308 mg/dl (SD 1.34 mg/dl) over the 5 year treatment period. Such an observed increase was statistically significantly different from 0 (two-sided P = 0.015). A 95% confidence interval suggests that the observed results were not unusual if the true average change in bilirubin over the 5 year period in a similar population treated with placebo were an increase anywhere between 0.0609 mg/dl to 0.554 mg/dl. *(Note that I used a two-sided P value. We are hoping that methotrexate might lead to either a decrease in bilirubin or at least less of an increase. It is easiest in this case just to describe the two-sided P value and make sure that the concept of increases or decreases is clear., No matter whether you use one-sided or two-sided P values, make clear which you use. When I use one-sided P values, I will only declare it statistically significant if the $P < 0.025$, while with a two-sided P value I would call it statistically significant if $P < 0.05$. In either case, I still report a 95% confidence interval, which by its very nature is two-sided.)*

(An analysis based on the mean ratio of bilirubin measurements)
When analyzed as a proportionate change, after the 5 year treatment period the placebo group averaged a 28.7% increase in bilirubin measurements relative to each patient's measurement at baseline. Such an observed proportionate increase was statistically significantly different from 1 (two-sided P < 0.0001). A 95% confidence interval suggests that the observed results were not unusual if the true average proportionate change in bilirubin over the 5 year period in a similar population treated with placebo were an

increase anywhere between 25.2% to 32.2%. *(Note that the estimated average ratio was 1.287, with a 95% confidence interval for the ratio of 1.252 to 1.322. I chose to express these results as a percentage change, rather than as a ratio. Had the average ratio been greater than 2, I would be more likely to use wording based on the ratio. For this data, my wording would have been "When analyzed as a proportionate change, after the 5 year treatment period the placebo group averaged spermidine measurements that were 1.29 times higher than each patient's measurement at baseline. Such a measurement would strongly reject the null hypothesis of no change in spermidine measurements (two-sided P< 0.0001). A 95% confidence interval suggests that the observed results were not unusual if the true average proportionate change in individual spermidine levels were anywhere between measurements that were 1.25 to 1.32 times higher than the measurement made at randomization.")*

One hundred thirty-two (132) subjects were randomized to receive methotrexate, and all of those patients had bilirubin measurements available at time of randomization. However, only 116 of those subjects had bilirubin measurements available at 5 years. Reasons for missing measurements at 5 years included study discontinuation due to prior death (*n* subjects), liver transplantation (*n* subjects), withdrawal of consent (*n* subjects), or patient loss to follow-up (*n* subjects). Patients who discontinued their use of the study drug (*n* subjects) prior to 5 years were still included in this analysis. The distribution of bilirubin, albumin, PTINR, FVC, and DLCO measurements at the time of randomization did not differ systematically between the 110 subjects for whom 5 year measurements were available and the 23 subjects missing measurements at 5 years.

Analysis of the change in bilirubin levels in the methotrexate group is restricted to those 116 patients with data available at 5 years. At the time of randomization, the average bilirubin measurement in that group was 0.651 mg/dl (SD 0.391 mg/dl). After the 5 year treatment period, the average bilirubin measurement was 0.844 mg/dl (SD 0.874 mg/dl).

(An analysis based on the mean difference in bilirubin measurements)
The placebo group thus averaged an increase of 0.192 mg/dl (SD 0.733 mg/dl) over the 5 year treatment period. Such an observed increase was statistically significantly different from 0 (two-sided P = 0.0055). A 95% confidence interval suggests that the observed results were not unusual if the true average change in bilirubin over the 5 year period in a similar population treated with placebo were an increase anywhere between 0.0577 mg/dl to 0.327 mg/dl.

(An analysis based on the mean ratio of bilirubin measurements)
When analyzed as a proportionate change, after the 5 year treatment period the placebo group averaged a 25.1% increase in bilirubin measurements relative to each patient's measurement at baseline. Such an observed proportionate increase was statistically significantly different from 1 (two-sided P < 0.0001). A 95% confidence interval suggests that the observed results were not unusual if the true average proportionate change in bilirubin over the 5 year period in a similar population treated with placebo were an increase anywhere between 22.9% to 27.3%. *(It is not unusual for a comparison of mean ratios to not be statistically significant even when the analysis based on mean differences does suggest a decrease. Very low measurement values can create a lot of variability in the estimated ratios. Generally it is better to analyze data on a log scale if you are interested in proportionate change, but when we do that, we have shifted our focus to geometric means instead of means. All of that having been said, we actually seem to have had greater precision in this analysis than when analyzing the mean change.)*

2. Perform an analysis to assess the proportion of the placebo group that had lower bilirubin levels 5 years post randomization (week 260) than they did at the time of randomization (week 1). Provide relevant point estimates, 95% confidence intervals, and P values for the test of a relevant hypothesis. Do the same for the methotrexate group. Make clear the interpretation of your confidence interval and P values, including the scientific relevance of your results.

(An analysis based on the proportion of patients showing a decrease in spermidine measurements)

Over the 5 year treatment period, 34.4% of the patients in the placebo group were observed to have a decrease in bilirubin levels. This observation was statistically significantly different from a 50% rate of observed decreases that might be expected if there were no systematic trend toward lower or higher bilirubin levels over time (P= 0.0011 in a two-sided test looking for rates different from 50%). A 95% confidence interval suggests that the observed results were not unusual if the true percentage of patients expected to have observed decreases in a population treated with placebo were anywhere between 25.9% and 43.9%. *(Note that this approach based on testing the proportion decreasing against a null hypothesis of 50% relies on a belief that bilirubin measurements would not naturally increase or decrease with age or time. Of course, the reason we do controlled clinical trials is to be able to assess what would naturally happen in the absence of treatment.)*

Over the 5 year treatment period, 26.7% of the patients in the methotrexate group were observed to have a decrease in bilirubin levels. This observation was statistically significantly different from a 50% rate of observed decreases that might be expected if there were no systematic trend toward lower or higher bilirubin levels over time (P< 0.0001 in a two-sided test looking for rates different from 50%). A 95% confidence interval suggests that the observed results were not unusual if the true percentage of patients expected to have observed decreases in a population treated with methotrexate were anywhere between 18.9% and 35.7%.

3. Perform an analysis to assess whether the placebo group had a change in geometric mean bilirubin level 5 years post randomization (week 260). Use the ratio of geometric means as a measure of comparison. (Note: Inference on the geometric mean is easily obtained by taking the log transform of your data, and then comparing means using differences. When you exponentiate the resulting estimates, you will have inference based on the geometric means and the ratios of geometric means. There is a handout on the class web pages which deals with the interpretation of log transformed data.) Provide relevant point estimates, 95% confidence intervals, and P values. Do the same for the methotrexate group. Make clear the interpretation of your confidence interval and P values, including the scientific relevance of your results.

(An analysis based on the ratio of geometric means of bilirubin measurements, which would also be the geometric mean of ratios computed for each individual.)

Analysis of the change in bilirubin levels in the placebo group is restricted to those 116 patients with data available at 5 years. At the time of randomization, the geometric mean of bilirubin measurements in that group was 0.583 mg/dl. After the 5 year treatment period, the geometric mean of bilirubin measurements was 0.700 mg/dl. Thus, after the 5 year treatment period the geometric mean of bilirubin levels in the placebo group was 20.1% higher than the geometric mean for those patients at randomization. Such a measurement would reject the null hypothesis of no change in bilirubin measurements (two-sided P=

0.0032). A 95% confidence interval suggests that the observed results were not unusual if the true average percentage change in geometric means were anywhere between a 6.45% increase and a 35.5% increase.

Analysis of the change in bilirubin levels in the methotrexate group is restricted to those 116 patients with data available at 5 years. At the time of randomization, the geometric mean of bilirubin measurements in that group was 0.561 mg/dl. After the 5 year treatment period, the geometric mean of bilirubin measurements was 0.685 mg/dl. Thus, after the 5 year treatment period the geometric mean of bilirubin levels in the methotrexate group was 22.0% higher than the geometric mean for those patients at randomization. Such a measurement would reject the null hypothesis of no change in bilirubin measurements (two-sided P= 0.0002). A 95% confidence interval suggests that the observed results were not unusual if the true average percentage change in geometric means were anywhere between a 10.2% increase and a 35.2% increase.

4. Perform an analysis to assess whether the treatment groups (placebo and methotrexate) differ with respect to their mean bilirubin levels 5 years post randomization. Provide relevant point estimates, 95% confidence intervals, and P values. Make clear the interpretation of your confidence interval and P values, including the scientific relevance of your results.

Ans: After 5 years of treatment, serum bilirubin levels were available on 116 of the 133 subjects in the group randomized to receive placebo and 116 of the 132 subjects in the group randomized to receive methotrexate. In the placebo group, bilirubin levels at 5 years were observed to range between 0.2 and 11.2 mg/dl, and averaged 0.988 mg/dl. The 95% CI for the true mean bilirubin level in a population receiving placebo is 0.726 to 1.25 mg/dl. In the group randomized to receive methotrexate, bilirubin levels after 5 years on study ranged from 0.2 to 8.2 mg/dl and averaged 0.844 mg/dl. The 95% CI for the true mean bilirubin level in a population prescribed methotrexate is 0.683 to 1.00 mg/dl. Based on these data, we thus estimate that prescription of a methotrexate is associated with an average bilirubin level that is 0.144 mg/dl lower than what it would be in the absence of treatment with methotrexate (95% CI 0.448 mg/dl lower 0.159 mg/dl higher). These results are not statistically significant (two-sided P= 0.349), and thus these data are consistent with results that might be observed by random chance in the absence of a treatment effect. *(Note that I did provide the minima and maxima post treatment in order to be able to assess individual extremes. I also provided CI for the group means, as those could conceivably be of interest for quantifying what we might expect for bilirubin over a five year period. However, the major interest is in the estimated differences between the two treatment arms.)*

5. Perform an analysis to assess whether the treatment groups (placebo and methotrexate) differ with respect to their change in mean bilirubin levels 5 years post randomization. Provide relevant point estimates, 95% confidence intervals, and P values. Make clear the interpretation of your confidence interval and P values, including the scientific relevance of your results.

Ans: After 5 years of treatment, serum bilirubin levels were available on 116 of the 133 subjects in the group randomized to receive placebo and 116 of the 132 subjects in the group randomized to receive methotrexate. In the placebo group, the change in bilirubin levels over 5 years were observed to range between a decrease of 1.2 mg/dl and an increase of 10.0 mg/dl, and averaged an increase of 0.308 mg/dl. The 95% CI for the true mean change in bilirubin level in a population receiving placebo is an increase of 0.0609 to 0.544

mg/dl. In the group randomized to receive methotrexate, the change in bilirubin levels after 5 years on study ranged from a decrease of 1.1 mg/dl to an increase of 6.4 mg/dl and averaged an increase of 0.192 mg/dl. The 95% CI for the true mean change in bilirubin level in a population prescribed methotrexate is an increase of 0.0577 to 0.327 mg/dl. Based on these data, we thus estimate that prescription of methotrexate is associated with an average change in bilirubin level that is 0.115 mg/dl less of an increase than what it would be in the absence of treatment with methotrexate (95% CI 0.395 mg/dl less of an increase to 0.164 mg/dl more of an increase). These results are not statistically significant (two-sided P= 0.418), and thus these data are consistent with results that might be observed by random chance in the absence of a treatment effect. *(These are not wordings that can be read quickly. Presentation of such results in a table is usually advised. That way you can get across the contrasts within treatment groups over time and contrasts across randomized treatment groups.)*

6. Perform an analysis to assess whether the treatment groups (placebo and methotrexate) differ with respect to their geometric mean bilirubin levels 5 years post randomization. Provide relevant point estimates, 95% confidence intervals, and P values. Make clear the interpretation of your confidence interval and P values, including the scientific relevance of your results.

Ans: After 5 years of treatment, serum bilirubin levels were available on 116 of the 133 subjects in the group randomized to receive placebo and 116 of the 132 subjects in the group randomized to receive methotrexate. In the placebo group, bilirubin levels at 5 years were observed to range between 0.2 and 11.2 mg/dl, with geometric mean 0.700 mg/dl. The 95% CI for the true geometric mean bilirubin level in a population receiving placebo is 0.615 to 0.796 mg/dl. In the group randomized to receive methotrexate, bilirubin levels after 5 years on study ranged from 0.2 to 8.2 mg/dl with geometric mean 0.685 mg/dl. The 95% CI for the true geometric mean bilirubin level in a population prescribed methotrexate is 0.617 to 0.760 mg/dl. Based on these data, we thus estimate that prescription of methotrexate is associated with an geometric mean bilirubin level that is 2.22% lower than what it would be in the absence of treatment with methotrexate (95% CI 20.6% lower to 13.3% higher). These results are not statistically significant (two-sided P= 0.794), and thus these data are consistent with results that might be observed by random chance in the absence of a treatment effect.