Biost 517
Applied Biostatistics I

Final Examination
Tuesday, December 17, 2002

Instructions: Please provide concise answers to all questions. Rambling answers touching on topics not directly relevant to the question will tend to count against you. Nearly telegraphic writing style is permissible.

The examination is closed book and closed notes. If you come to a problem that you believe cannot be answered without making additional assumptions, clearly state the reasonable assumptions that you make, and proceed.

Problems 1 - 5 refer to a clinical trial of methotrexate in the treatment of Primary Biliary Cirrhosis, a progressive disease of the liver which often leads to liver transplantation or death. Patients were accrued to the study over a nine year period from March 1, 1989 to March 1, 1998. They were randomized in a double blind fashion to receive either methotrexate or placebo, and then followed until they received a liver transplant, they died, or until the study ended on November 1, 2002. The variables available in this data set include the following. All variables except Obstime and Failure are measured at the time of randomization.

- **Age** = the patient's age in years
- **Male** = an indicator of the patient's sex (0 = female, 1= male)
- **Race** = a code indicating the patient's race/ethnicity (1= Caucasian, 2= Black, 3= Native American, 4= Hispanic, 5= Oriental/Pacific, 6= Mideast/Arabian, 7= Indian subcontinent, 8= Other)
- **Weight** = the patient's weight in kilograms
- **QoL** = a code indicating the patient's self-reported quality of life (1= normal health, 2= regular activity but not completely well, 3= not able to carry out regular activity, 4= confined to bed most the time, 5= in hospital most the time)
- **Bili** = patient's bilirubin in mg/dl (tends to be high in liver disease)
- **Albumin =** patient's albumin in mg/dl (tends to be low in liver disease)
- **Hepmeg** = an indicator of an enlarged liver (0= no, 1= yes)
- **Tx =** an indicator of treatment received by the patient (0= placebo, 1= methotrexate)
- **Obstime** = time of follow-up in years from start of study until death, liver transplant, or the time of data analysis, whichever comes first
- **Failure** = type of failure observed (0= none, 1= liver transplant, 2= death)

	n	msng	mean	std dev	min	25%ile	median	75%ile	maximum
Age	511	1	51.918	9.459	23.000	46.000	52.000	59.000	79.000
Male	511	0	0.061	0.239	0.000	0.000	0.000	0.000	1.000
Race	511	0	1.360	0.988	1.000	1.000	1.000	1.000	8.000
Weight	511	33	70.724	15.707	42.600	59.450	67.600	79.400	150.000
QoL	511	15	1.692	0.690	1.000	1.000	2.000	2.000	4.000
Bili	511	0	1.119	2.082	0.100	0.500	0.700	1.100	35.200
Albumin	511	7	3.965	0.442	1.800	3.800	4.000	4.300	5.200
Hepmeg	511	24	0.283	0.451	0.000	0.000	0.000	1.000	1.000
Tx	511	0	0.509	0.500	0.000	0.000	1.000	1.000	1.000
Obstime	511	0	4.923	2.945	0.009	2.006	5.396	7.504	13.530
Failure	511	0	0.775	0.690	0.000	0.000	1.000	1.000	2.000

1. The above table provides the sample size, number of missing observations, sample mean, median, standard deviation, minimum, maximum, 25th and 75th percentiles for the above data. For each of the above variables briefly indicate which of the above descriptive statistics are not of any scientific relevance (use of "all OK" or "none OK" or similar statements are encouraged where appropriate). Also indicate any other descriptive statistics you might wish to look at instead.
    a. Age

Answer: All are OK. *(Age is a continuous variable not subject to censoring.)*

b. Male

Answer: All are OK, but only the sample mean is really of interest. *(Male is a binary variable so the sample mean is the proportion of subjects who are male.)*

c. Race

Answer: None are OK except for sample size and number of missing. I would prefer to have a frequency table. *(Race is a nominal categorical variable. The numbers are just labels.)*

d. Weight

Answer: All are OK. *(Weight is a continuous variable not subject to censoring.)*

e. QoL

Answer: All are OK except the sample mean and standard deviation. With so few categories, I would actually prefer to have a frequency table. *(QoL is an ordered categorical variable. The quantiles make sense, but because the numbers are not really quantitative, the mean and SD do not make sense by themselves—I might be willing to compare across groups using the means, however.)*

f. Bili

Answer: All are OK. *(Bili is a continuous variable not subject to censoring.)*

g. Albumin

Answer: All are OK. *(Albumin is a continuous variable not subject to censoring.)*

h. Hepmeg

Answer: All are OK, but only the sample mean is really of interest. *(Hepmeg is a binary variable so the sample mean is the proportion of subjects who have enlarged livers.)*

i. Tx

Answer: All are OK, but only the sample mean is really of interest. *(Tx is a binary variable so the sample mean is the proportion of subjects who were assigned to receive methotrexate.)*

j. Obstime

Answer: None are OK except for sample size and number of missing. We would want Kaplan-Meier estimates of probability of failing at particular times. *(Obstime is a censored continuous variable. Pertinent descriptive statistics that would be of use scientifically would have to be based on Kaplan-Meier estimates, taking into account the indicator of events Failure.)*

k. Failure

Answer: None are OK except for sample size and number of missing. We would want Kaplan-Meier estimates of probability of failing at particular times. *(Failure is measured over different timeframes for each subject. Pertinent descriptive statistics that would be of use scientifically would have to be based on Kaplan-Meier estimates, taking into account the time of observation as recorded in Obstime.)*

2. Is there evidence that any of the above variables are particularly prone to outlying values? If so, which ones? Briefly explain your reasoning.

Answer: Weight appears to have a very large maximum: 150 kg is over 4 SD from the mean—a slightly large distance for a sample size of 511. The maximum is also clearly further from the mean than the minimum (that is, the mean and median are not the midpoint of the range).

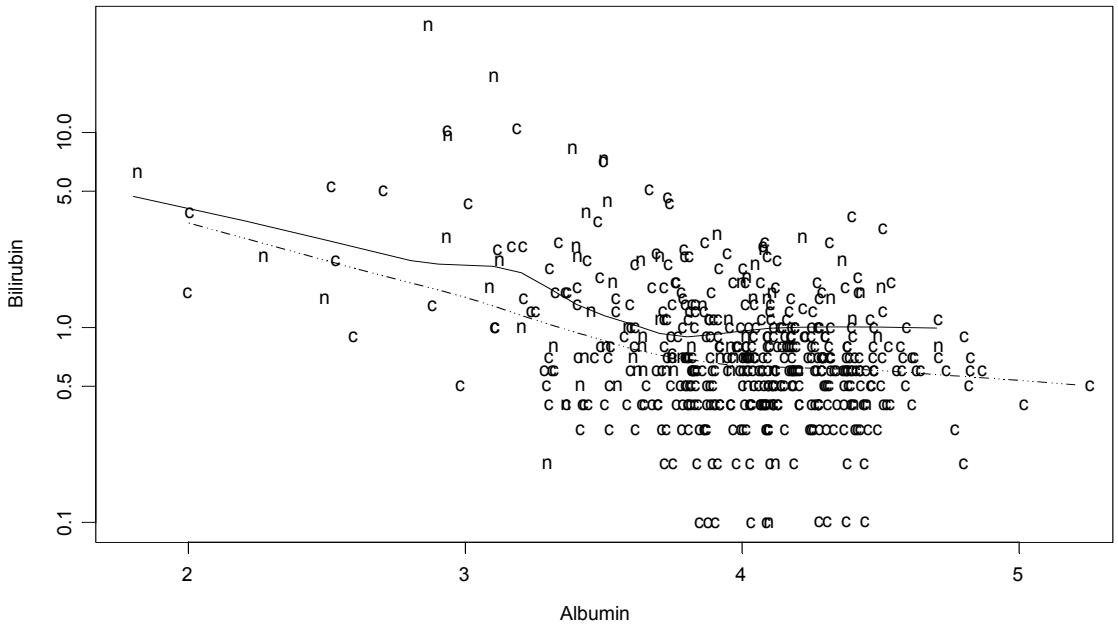
Bili is very skewed as evidenced by a SD that is much greater than the mean for this positive variable. The maximum is over 16 SD from the mean. Clearly we have some outliers here.

Albumin appears to have a minimum value that is over 4 SD from the mean. This might represent a low outlying value.

(It is not of much interest to consider the nonquantitative or censored variables.)

3. Below is a scatterplot of serum bilirubin measurements and albumin levels. Points are labeled according to race/ethnicity categorized as c= Caucasian and n= Noncaucasian, and lowess

curves for each race stratum are superimposed on the graph (broken line is Caucasian, solid line is Noncaucasian).



a. Is there evidence of an association between albumin and bilirubin levels? Explain your reasoning.

Answer: **Yes, because there appears to be a trend toward higher bilirubin for the lowest values of albumin.** *(Note that this graph plots bilirubin on a logartithmic scale. Thus we are looking at trends in the geometric mean. Note that for albumin levels in the "normal range" of about 3 – 5, there does not appear to be much of a trend in the geometric means of bilirubin. However, when the liver disease has progressed enough to have very low albumin, the bilirubin also tends to be higher.)*

b. Is there evidence of an association between race/ethnicity and bilirubin levels after adjusting for albumin levels? Explain your reasoning.

Answer: **Yes, there is vertical separation between the lowess smooths for Caucasians and Noncaucasians.**

c. Is there evidence that race modifies the association between albumin and bilirubin? Explain your reasoning.

Answer: **No, the two lowess curves look fairly parallel to me.** *(I would give you credit if you said "yes", providing you invoke nonparallel curves. I actually think that these curves are consistent with parallel curves with some random sampling variation—we don't really have all that large of a sample in the Noncaucasian group.)*

4. Suppose we are interested in studying whether treatment with methotrexate prolongs time to failure. The following table provides descriptive statistics of the variable Obstime stratified by the variable Failure.

	n	msng	mean	std dev	min	25%ile	median	75%ile	maximum
all	511	0	4.92	2.95	0.01	2.01	5.40	7.50	13.53
0	192	0	7.66	1.35	4.77	6.53	7.87	8.76	13.53
1	242	0	3.22	2.33	0.01	1.20	2.86	4.98	9.02
2	77	0	3.45	2.45	0.05	1.66	2.68	5.64	10.14

For each of the following analyses, briefly explain why the proposed analysis is problematic.

a. We create a new variable AnyFailure that indicates whether Failure is nonzero, and compare the proportion of patients having AnyFailure=1 across treatment groups.

Answer: The time of observation varies across individuals, and thus it is not immediately clear the timeframe in which the failures occurred. In a randomized trial in which subjects are randomized roughly equally as they are accrued, this is merely inefficient and difficult to communicate the results. Had this been an observational study, misleading results can be obtained from such an analysis if the time of follow-up varied across the groups being compared. Even in a randomized study, a tendency for treatments to be associated with study dropout unrelated to failure would cause problems, as well, though in this case, we would certainly worry that any differential dropout might be informative about failure.

b. We analyze the 10 year survival by considering the number of patients observed to fail within 10 years divided by the number of patients who were observed at least 10 years. (That is, we discard all patients who have been observed for less than 10 years and have not yet been observed to fail.)

Answer: This is clearly bad. We are discarding patients who are prone to have better survival. That is, patients who were accrued 6 years ago and failed are counted, while patients who were accrued 6 years ago and who did not fail are discarded.

c. We are interested in how methotrexate affects time to liver transplant, so we treat deaths as censored variables and use Kaplan-Meier estimates of the proportion of patients in each group who have liver transplant within 10 years.

Answer: This suffers from the problem of "competing risks". The subjects who died may have been more prone to get a liver transplant than the subjects who did not die. Thus we cannot regard that we are getting a good estimate of the risk of liver transplantation.

5. Suppose we calculate Kaplan-Meier estimates of 10 year survival without liver transplant (so we are counting either liver transplant or death as an event). We obtain the following results.

		10 yr Surv	Std Err	95% CI lo	95% CI hi
All	Plc	0.288	0.034	0.229	0.363
	Mtx	0.333	0.041	0.261	0.425
Stratified by Race					
Caucasian	Plc	0.268	0.035	0.207	0.347
	Mtx	0.338	0.044	0.261	0.436
Noncaucasian	Plc	0.463	0.099	0.305	0.703
	Mtx	0.206	0.152	0.049	0.872

a. A 95% confidence interval for the difference in 10 year survival probabilities comparing placebo (Plc) to methotrexate (Mtx) is –0.06 to 0.15. Provide an interpretation for this confidence interval.

Answer: The observed 4.5% difference in 10 year survival favoring the methotrexate group is not atypical of results that we would expect if the true difference in 10 year survival were between a 6% difference favoring placebo or a 15% difference favoring methotrexate. Because the 95% CI includes 0, we cannot reject the null hypothesis that methtrexate has no effect on ten year survival.

b. The two-sided P value comparing placebo (Plc) to methotrexate (Mtx) is 0.321. What do you conclude about the effect of treatment on 10 year survival?

Answer: When there is no true effect of methotrexate on 10 year survival, the probability of observing a difference as great as 4.5% between the groups is 0.321. Thus we cannot reject the null hypothesis that methotrexate is not associated with 10 year survival.

c. Is there evidence that the effect of treatment is modified by race?

Answer: For Noncaucasians, the difference in 10 year survival probabilities across treatment groups is .257 in favor of placebo. For Caucasians, the difference in 10 year survival probabilitis across treatment groups is .070 in favor of methotrexate. This difference in estimates of association is certainly suggestive of effect modification, though it is not clear whether we have enough precision to be sure that this would be true in the population.

d. Is there evidence that the effect of treatment is confounded by race?

Answer: Because there is some suggestion of effect modification, it is moot to consider confounding: There is not a common effect of methotrexate treatment in the race strata.

6. Bonus: Explain how you would use the above inferential statistics stratified by race to test the presence of effect modification by race.

Answer: We can use the standard errors for each estimate to compute the standard error for the difference: The standard error for a difference is the square root of the sum of the squared standard errors for the two statistics. The estimate for the measure of effect modification is the difference of the differences.

Thus: In the Caucasian stratum,

$$se(Mtx - Plc\ Diff) = \sqrt{se^2(Mtx) + se^2(Plc)} = \sqrt{.044^2 + .035^2} = .056$$

(We can therefore get a 95% CI for the effect of methotrexate in Caucasians as

$$.070 \pm 1.96 \times .056 = (-.040, .180)$$

In the Noncaucasian stratum,

$$se(Mtx - Plc\ Diff) = \sqrt{se^2(Mtx) + se^2(Plc)} = \sqrt{.152^2 + .099^2} = .184$$

(We can therefore get a 95% CI for the effect of methotrexate in Caucasians as

$$-.257 \pm 1.96 \times .184 = (-.618, .103)$$

The difference in the effects across strata is .070 – (-.257) = .327. Using these same methods, we find that the standard error of this measure of interaction is .192, so the 95% CI for the measure of effect modification is -.049 to .703. We therefore cannot reject the null hypothesis that there is no effect modification by race/ethnicity on the association between methotrexate and 10 year survival.