Biost 518: Applied Biostatistics II
Emerson, Winter 2007

Homework #4
February 7, 2007

1. Problem 1 relates to the data set from the clinical trial of beta carotene supplementation. For each of the following models, provide inference (P values, and where appropriate, 95% confidence intervals with scientific interpretation of the parameters) regarding the effect of beta carotene supplementation on the plasma beta carotene levels after 9 months of treatment. Also provide a table of predicted values for each of these models.

    a. Provide descriptive statistics for plasma beta carotene levels after 9 months of treatment by dose group.

Ans: I chose to compare the distribution of plasma beta carotene levels across the dose groups using the means.

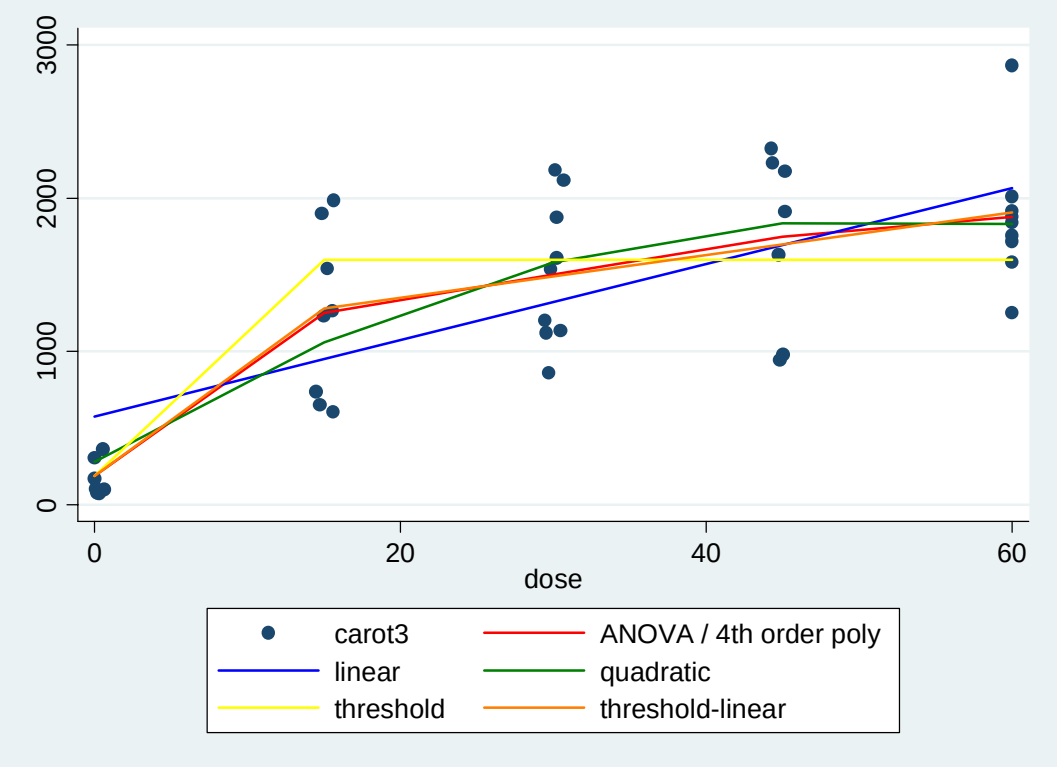
Table 1: Descriptive statistics for plasma beta carotene levels after 9 months of treatment by dose.

Dose	N	Mean	SD	Min	25th%	Mdn	75th%	Max
0	7	186	88	85	126	149	286	323
15	8	1254	570	577	695	1250	1771	2019
30	9	1505	479	849	1157	1499	1840	2249
45	7	1749	579	950	993	1848	2248	2310
60	9	1878	430	1233	1725	1865	1918	2855

The following table contains the fitted values from each of the six models. I note that the fitted means from the dummy variables (Model B) and the quartic (fourth order) polynomial (Model G) each correspond exactly to the sample means for each dose group. This correspondence between the quartic polynomial and the dummy variables is due to the fact that there were only five levels of dose sampled (and four is one less than five). The estimates from Model C would lie exactly on a straight line. The estimates from Model D would not lie on a straight line, because the sample means from each group are not very linear. In Model E, the estimated means are the same for all dose groups above 0, and the dose 0 group estimate corresponds exactly to the sample mean for that group. In Model F, the estimated means for dose groups higher than 0 lie exactly on a straight line, and the dose 0 group estimate corresponds exactly to the sample mean for that group. These relationships are displayed graphically in the figure.

Table 2: Fitted means from the six models.

Dose	Sample Means	Model B	Model C	Model D	Model E	Model F	Model G
0	186.321	186.321	577.247	279.551	186.321	186.321	186.321
15	1253.583	1253.583	949.859	1061.776	1597.345	1279.650	1253.583
30	1504.611	1504.611	1322.470	1581.024	1597.345	1489.329	1504.611
45	1749.081	1749.081	1695.082	1837.295	1597.345	1699.008	1749.081
60	1877.630	1877.630	2067.694	1830.589	1597.345	1908.687	1877.630



b. Model dose as dummy variables.

Ans: An analysis of variance of the plasma beta carotene levels after 9 months of supplementation finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect (P < 0.0001). The placebo group is estimated to have a mean plasma beta carotene level of 186 µg / dl (95% confidence interval unadjusted for multiple comparisons: 2.58 to 3.51 µg / dl). The dose 15 group is estimated to have a mean plasma beta carotene 1067 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 652 to 1482 µg / dl higher), the dose 30 group is estimated to have a mean plasma beta carotene 1318 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 985 to 1652 µg / dl higher), the dose 45 group is estimated to have a mean plasma beta carotene 1563 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 1118 to 2008 µg / dl higher), and the dose 60 group is estimated to have a mean plasma beta carotene 1691 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 1391 to 1992 µg / dl higher). *(Note that the P value comes from a overall F test in this setting of dose modeled with four covariates.)*

c. Model dose continuously as a linear predictor.

Ans: An analysis performed by regressing the plasma beta carotene levels on a linear dose variable estimates that the mean tends to increase 24.8 µg / dl (95% CI:

increases 17.7 to 32.0 µg / dl) for each 1 mg/day difference in dose. Such a difference is beyond that which might be reasonably expected to be observed when there is no true effect of beta carotene supplementation on plasma beta carotene levels (P < .0001). (*From that regression model, the placebo group is estimated to have a mean plasma beta carotene level of 577 µg / dl (95% confidence interval unadjusted for multiple comparisons: 296 to 859 µg / dl), however nonlinearities evident in the dose-response curve would lessen our confidence in the accuracy of that estimate. There is a multiple comparison issue as I might have considered providing estimates for other dose groups.)*

d. Model dose as two variables: a continuous linear predictor along with a quadratic term (so an additional predictor equal to the square of dose).

Ans: An analysis performed by regressing the plasma beta carotene levels on a quadratic polynomial in dose finds that the observed differences between the dose groups means is greater than what might reasonably be expected when DFMO had no true effect (P < .0001). A statistically significant second order term in dose (P = 0.001) suggests strong evidence against a linear dose response. *(Note that the P value comes from the overall F test in this setting of dose modeled with two covariates. Interpreting the slope parameters is difficult here, though we can interpret the intercept: The placebo group is estimated to have a mean plasma beta carotene level of 280 µg / dl (95% confidence interval unadjusted for multiple comparisons: 111 to 448 µg / dl). We could, of course, obtained predicted means and 95% CI for the other dose groups. If I did so, I would likely comment on the degree to which I thought a quadratic curve described the data, because every dose group's estimate is influenced by the "borrowing of information" across all dose groups. Visually, this model does a reasonable job, though I am a little bothered by the fitted values for the highest dose group being lower than those for the 45 mg / day group.)*

e. Model dose as a binary variable indicating whether dose was greater than 0.

Ans: An analysis comparing the placebo group to the combined groups receiving some dose of beta carotene supplementation finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect on plasma beta carotene (P < .0001). The placebo group is estimated to have a mean plasma beta carotene level of 186 µg / dl (95% confidence interval unadjusted for multiple comparisons: 123 to 250 µg / dl). *(Note that I had no qualms providing the predicted value for the placebo group, because it is estimated without borrowing information from any other group. There is still a bit of a multiple comparison issue when providing predicted means for a single dose group, even though I do not try to provide it for other groups. Interpreting the slope parameters is difficult here, because there is no good scientific reason to estimate the effect of beta carotene supplementation across combined dose groups. Had we obtained predicted means and 95% CI for the dose groups, they would have been the same for all doses higher than 0. I would have commented on the fact that I do not think this model is a particularly good fit.)*

f. Model dose as two variables: a binary variable indicating whether dose was greater than 0 and a continuous linear term.

Ans: Analysis finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect on plasma beta carotene levels (P < .0001). From the modeling of dose as a linear continuous predictor and a threshold effect at dose 0, we have sufficient evidence to reject the hypothesis of a linear relationship among means across all dose levels (P < .0005). The estimated linear trend across dose groups 15 – 60 mg / day suggests the mean tends to increase 14.0 µg / dl (95% CI unadjusted for multiple comparisons: increases 3.74 to 24.3 µg / dl) for each 1 mg / day difference in dose when comparing doses above 15 mg / day and above. The placebo group is estimated to have a mean plasma beta carotene level of 186 µg / dl (95% confidence interval unadjusted for multiple comparisons: 122 to 251 µg / dl). *(Again, I present estimates for the placebo group because that estimate is not influenced by borrowing information from the other dose groups. It is interesting to note that the CI is wider in this model compared to the simple threshold model in part e. This is due to the extra covariate in the model: The standard error involves a term n-p, where n is the number of observations and p is the number of estimated parameters—intercept and slopes.)*

g. Model dose as four variables: a continuous linear predictor, a quadratic term, a cubic term, and a fourth order term (i.e., dose raised to the fourth power).

Ans: An analysis performed by regressing the plasma beta carotene levels on a quartic (fourth order) polynomial in dose finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect (P < .0001). *(Note that the P value comes from the overall F test in this setting of dose modeled with four covariates. Interpreting the slope parameters is difficult here, though we can interpret the intercept: The placebo group is estimated to have a mean plasma beta carotene level of 186 µg / dl (95% confidence interval unadjusted for multiple comparisons: 120 to 253 µg / dl). We could, of course, obtained predicted means and 95% CI for the other dose groups, but that would be easier just using the ANOVA model, which is of course equivalent to this quartic polynomial: There are only five dose groups, so a fourth order polynomial fits the dose groups' sample means perfectly—there is no "borrowing of information".)*

2. Repeat the analyses in problem 1 adjusting for the baseline plasma beta carotene levels. Note that the Stata functions "test" and "testparm" can be used to perform Wald tests of multiple parameters adjusted for other covariates. You do not need to consider the descriptive statistics or the fitted values for this problem.

a. Model dose as dummy variables.

Ans: An analysis of covariance of the plasma beta carotene levels after 9 months of supplementation adjusted for baseline values of plasma beta carotene finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect (P < 0.0001). Compared to a placebo group having similar baseline values, the dose 15

group is estimated to have a mean plasma beta carotene 1224 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 790 to 1658 µg / dl higher), the dose 30 group is estimated to have a mean plasma beta carotene 1440 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 1123 to 1756 µg / dl higher), the dose 45 group is estimated to have a mean plasma beta carotene 1679 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 1339 to 2018 µg / dl higher), and the dose 60 group is estimated to have a mean plasma beta carotene 1791 µg / dl higher than the placebo group (95% CI unadjusted for multiple comparisons: 1480 to 2102 µg / dl higher). *(Note that the P value comes from the multiple partial F test in this setting of dose modeled with four covariates in the presence of another covariate. We do have greater precision in general with this model, but this does not always translate into a narrower CI, because there is some slight correlation between dose and baseline values. Also note that the coefficient for the baseline value is not 1, as might at first be expected. The coefficient will tend to be some sort of average within group correlation between the follow-up and baseline values times the ratio of the follow-up and baseline standard deviations. As the standard deviation of the plasma beta carotene levels increased markedly with supplementation, this coefficient is greater than 1 Note also that I did not provide estimates of the mean plasma beta carotene in any of the dose groups, because to do such would require conditioning upon some particular value for the baseline value. This is sometimes done using the mean baseline value, but I don't know that this would add very much.)*

b. Model dose continuously as a linear predictor.

<u>Ans</u>: A baseline-adjusted analysis performed by regressing the plasma beta carotene levels on a linear dose variable estimates that the mean tends to increase 25.5 µg / dl (95% CI: increases 18.1 to 32.8 µg / dl) for each 1 mg/day difference in dose when baseline values are similar. Such a difference is beyond that which might be reasonably expected to be observed when there is no true effect of beta carotene supplementation on plasma beta carotene levels (P < .0001). *(Because dose is modeled with a single covariate, we can obtain the P value from the partial t test. The comments made above in part a regarding precision, the coefficient of the baseline term, and the value of estimating group specific fitted values hold here as well.)*

c. Model dose as two variables: a continuous linear predictor along with a quadratic term (so an additional predictor equal to the square of dose).

<u>Ans</u>: A baseline-adjusted analysis performed by regressing the plasma beta carotene levels on a quadratic polynomial in dose finds that the observed differences between the dose groups' means is greater than what might reasonably be expected when DFMO had no true effect (P < .0001). A statistically significant second order term in dose (P = 0.001) suggests strong evidence against a linear dose response. *(Because dose is modeled with two covariates in the presence of a non-dose related covariate, we must obtain the P value from the multiple partial F test. The comments made above in part a regarding precision, the coefficient of the baseline term, and the value of estimating group specific fitted values hold here as well.)*

d. Model dose as a binary variable indicating whether dose was greater than 0.

Ans: A baseline-adjusted analysis comparing the placebo group to the combined groups receiving some dose of beta carotene supplementation finds that the observed differences between the dose groups' mean is greater than what might reasonably be expected when beta carotene supplementation had no true effect on plasma beta carotene (P < .0001). *(Because dose is modeled with a single covariate, we can obtain the P value from the partial t test. The comments made above in part a regarding precision, the coefficient of the baseline term, and the value of estimating group specific fitted values hold here as well. As with problem 1, I find little value in describing estimated differences between the combined dose groups and the placebo group.)*

e. Model dose as two variables: a binary variable indicating whether dose was greater than 0 and a continuous linear term.

Ans: A baseline-adjusted analysis finds that the observed differences between the dose groups' means is greater than what might reasonably be expected when beta carotene supplementation had no true effect on plasma beta carotene levels (P < .0001). From the modeling of dose as a linear continuous predictor and a threshold effect at dose 0, we have sufficient evidence to reject the hypothesis of a linear relationship among means across all dose levels (P < .0005). The estimated linear trend across dose groups 15 – 60 mg / day suggests the mean tends to increase 12.8 µg / dl (95% CI unadjusted for multiple comparisons: increases 3.09 to 22.5µg / dl) for each 1 mg / day difference in dose when comparing doses above 15 mg / day and above in patients with similar baseline values. *(Because dose is modeled with two covariates in the presence of a non-dose related covariate, we must obtain the P value from themultiple partial F test. The comments made above in part a regarding precision, the coefficient of the baseline term, and the value of estimating group specific fitted values hold here as well.*

f. Model dose as four variables: a continuous linear predictor, a quadratic term, a cubic term, and a fourth order term (i.e., dose raised to the fourth power).

Ans: A baseline-adjusted analysis performed by regressing the plasma beta carotene levels on a quartic (fourth order) polynomial in dose finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect (P < .0001). *(Note that the P value comes from a multiple partial F test in this setting of dose modeled with four covariates in the presence of a non-dose related covariate. This analysis is equivalent to that using dummy variables for dose (the ANCOVA model).)*

3. Now consider the effect of beta carotene supplementation on plasma vitamin E levels. Repeat the analyses in problem 2 (i.e.adjusting for the baseline plasma vitamin E levels). Note that the Stata functions "test" and "testparm" can be used to perform Wald tests of multiple parameters adjusted for other covariates. You do not need to consider the descriptive statistics or the fitted values for this problem.

   a. Model dose as dummy variables.

Ans: An analysis of covariance of the plasma vitamin E levels after 9 months of supplementation adjusted for baseline values of plasma vitamin E finds that the observed differences between the dose groups' means is greater than what might reasonably be expected when beta carotene supplementation had no true effect (P = 0.0053). Compared to a placebo group having similar baseline values, the dose 15 group is estimated to have a mean plasma vitamin E 1.24 mg / dl lower than the placebo group (95% CI unadjusted for multiple comparisons: 0.620 to 1.85 mg / dl lower), the dose 30 group is estimated to have a mean plasma vitamin E 0.837 mg / dl lower than the placebo group (95% CI unadjusted for multiple comparisons: 1.75 mg / dl lower to 0.816 mg / dl higher), the dose 45 group is estimated to have a mean plasma vitamin E 1.20 mg / dl lower than the placebo group (95% CI unadjusted for multiple comparisons: 0.427 to 1.97 mg / dl lower), and the dose 60 group is estimated to have a mean plasma vitamin E 1.02 mg / dl lower than the placebo group (95% CI unadjusted for multiple comparisons: 0.199 to 1.84 mg / dl lower). *(Note that the P value comes from the multiple partial F test in this setting of dose modeled with four covariates in the presence of another covariate. Note also that I did not provide estimates of the mean plasma vitamin E in any of the dose groups, because to do such would require conditioning upon some particular value for the baseline value. This is sometimes done using the mean baseline value, but I don't know that this would add very much.)*

b. Model dose continuously as a linear predictor.

Ans: A baseline-adjusted analysis performed by regressing the plasma vitamin E levels on a linear dose variable estimates that the mean tends to decrease 0.012 mg / dl (95% CI: decreases 0.0252 mg / dl to increases 0.000399 mg / dl) for each 1 mg/day difference in dose when baseline values are similar. Such a difference is not beyond that which might be reasonably expected to be observed when there is no true effect of beta carotene supplementation on plasma vitamin E levels (P = 0.057). *(Because dose is modeled with a single covariate, we can obtain the P value from the partial t test. The comments made above in part a regarding precision, the coefficient of the baseline term, and the value of estimating group specific fitted values hold here as well.)*

c. Model dose as two variables: a continuous linear predictor along with a quadratic term (so an additional predictor equal to the square of dose).

Ans: A baseline-adjusted analysis performed by regressing the plasma vitamin E levels on a quadratic polynomial in dose finds that the observed differences between the dose groups' means is greater than what might reasonably be expected when DFMO had no true effect (P = 0.0223). A statistically nonsignificant second order term in dose (P = 0.108) suggests no strong evidence against a linear dose response. *(Because dose is modeled with two covariates in the presence of a non-dose related covariate, we must obtain the P value from the multiple partial F test.)*

d. Model dose as a binary variable indicating whether dose was greater than 0.

Ans: A baseline-adjusted analysis comparing the placebo group to the combined groups receiving some dose of beta carotene supplementation finds that the

observed differences between the dose groups' mean is greater than what might reasonably be expected when beta carotene supplementation had no true effect on plasma vitamin E (P = .001). *(Because dose is modeled with a single covariate, we can obtain the P value from the partial t test. As with problem 1, I find little value in describing estimated differences between the combined dose groups and the placebo group.)*

e. Model dose as two variables: a binary variable indicating whether dose was greater than 0 and a continuous linear term.

Ans: A baseline-adjusted analysis finds that the observed differences between the dose groups' means is greater than what might reasonably be expected when beta carotene supplementation had no true effect on plasma vitamin E levels (P = 0.0033). From the modeling of dose as a linear continuous predictor and a threshold effect at dose 0, we have sufficient evidence to reject the hypothesis of a linear relationship among means across all dose levels (P = 0.006). The estimated linear trend across dose groups 15 – 60 mg / day suggests the mean tends to increase 0.00196 mg / dl (95% CI unadjusted for multiple comparisons: decreases 0.0130 mg / dl to increases 0.0169 ng / dl) for each 1 mg / day difference in dose when comparing doses above 15 mg / day and above in patients with similar baseline values. The lack of statistical significance suggests that we do not have strong evidence for an additional effect of beta carotene supplementation for doses above 15 mg / day. *(Because dose is modeled with two covariates in the presence of a non-dose related covariate, we must obtain the P value from the multiple partial F test.)*

f. Model dose as four variables: a continuous linear predictor, a quadratic term, a cubic term, and a fourth order term (i.e., dose raised to the fourth power).

Ans: A baseline-adjusted analysis performed by regressing the plasma vitamin E levels on a quartic (fourth order) polynomial in dose finds that the observed differences between the dose groups means is greater than what might reasonably be expected when beta carotene supplementation had no true effect (P = 0.0053). *(Note that the P value comes from a multiple partial F test in this setting of dose modeled with four covariates in the presence of a non-dose related covariate. This analysis is equivalent to that using dummy variables for dose (the ANCOVA model).)*

4. For each of the following models, provide inference (P values, and where appropriate, 95% confidence intervals with scientific interpretation of the parameters) regarding the effect of beta carotene supplementation on the odds of a decreased plasma vitamin E levels of at least 1.5 mg/dl after 9 months of treatment (i.e., the difference between plasma vitamin E level at 9 months and plasma vitamin E level at baseline was -1.5 mg/dl or less). Also provide a table of predicted values for the odds of decreased plasma vitamin E as well as the predicted values for the probability of decreased plasma vitamin E for each of these models.

   a. Provide descriptive statistics for the probability and odds of decreased plasma vitamin E levels after 9 months of treatment by dose group.

Ans: The following tables contain the fitted values (both fitted proportions and fitted odds) from each of the six models. I note that the fitted proportions (odds) from the dummy

variables (Model B) and the quartic (fourth order) polynomial (Model G) each correspond exactly to the sample proportions (odds) for each dose group. This correspondence between the quartic polynomial and the dummy variables is due to the fact that there were only five levels of dose sampled (and four is one less than five). The log odds estimates from Model C would lie exactly on a straight line. The estimates from Model D would not lie on a straight line, because the sample means from each group are not very linear, instead fitting the threshold model better. In Model E, the estimated means are the same for all dose groups above 0, and the dose 0 group estimate corresponds exactly to the sample mean for that group. In Model F, the estimated log odds for dose groups higher than 0 lie exactly on a straight line, and the dose 0 group estimate corresponds exactly to the sample proportion for that group.

Table 3: Fitted proportions from the six models.

Dose	Sample Proportions	Model B	Model C	Model D	Model E	Model F	Model G
0	0.286	0.286	0.408	0.351	0.286	0.286	0.286
15	0.625	0.625	0.452	0.473	0.545	0.555	0.625
30	0.444	0.444	0.497	0.547	0.545	0.549	0.444
45	0.571	0.571	0.541	0.569	0.545	0.542	0.571
60	0.556	0.556	0.585	0.539	0.545	0.536	0.556

Table 4: Fitted odds from the six models.

Dose	Model B	Model C	Model D	Model E	Model F	Model G
0	0.400	0.689	0.540	0.400	0.400	0.400
15	1.667	0.825	0.897	1.200	1.248	1.667
30	0.800	0.987	1.209	1.200	1.216	0.800
45	1.333	1.180	1.321	1.200	1.185	1.333
60	1.250	1.412	1.171	1.200	1.155	1.250

b. Model dose as dummy variables.

Ans: An analysis of the odds of markedly decreased vitamin E levels (i.e., a decrease of 1.5 mg / dl or more) after 9 months of supplementation when modeling each dose group as a dummy variable finds that the observed differences between the dose groups means is not greater than what might reasonably be expected when beta carotene supplementation had no true effect (P =0.736). The dose 15 group is estimated to have odds of markedly decreased vitamin E 4.17 times higher than that for the placebo group (95% CI unadjusted for multiple comparisons: odds ratio of 0.460 to 37.8), the dose 30 group is estimated to have odds of markedly decreased vitamin E 2.00 times higher than that for the placebo group (95% CI unadjusted for multiple comparisons: odds ratio of 0.238 to 16.8), the dose 45 group is estimated to have odds of markedly decreased vitamin E 3.33 times higher than that for the placebo group (95% CI unadjusted for multiple comparisons: odds ratio of 0.352 to 31.6), and the dose 60 group is estimated to have odds of markedly decreased vitamin E 4.17 times higher than that for the placebo group (95% CI unadjusted for

multiple comparisons: odds ratio of 0.372 to 26.3). *(Note that the P value comes from a overall F test in this setting of dose modeled with four covariates. For greater clarity, I could have included an estimate of the odds of markedly decreased vitamin E levels for the placebo group. That would have only taken using the Stata "logit" command and exponentiating the output.)*

c. Model dose continuously as a linear predictor.

Ans: An analysis of the odds of markedly decreased vitamin E levels (i.e., a decrease of 1.5 mg / dl or more) after 9 months of supplementation when modeling a linear dose variable estimates that the odds of a markedly decreased vitamin E tends to be 1.20% higher (95% CI: increases 1.78% lower to 4.28% higher) for each 1 mg/day difference in dose. Such a difference is not beyond that which might be reasonably expected to be observed when there is no true effect of beta carotene supplementation on plasma beta carotene levels (P = 0.434).

d. Model dose as two variables: a continuous linear predictor along with a quadratic term (so an additional predictor equal to the square of dose).

Ans: An analysis of the odds of markedly decreased vitamin E levels (i.e., a decrease of 1.5 mg / dl or more) after 9 months of supplementation when modeling a quadratic polynomial in dose finds that the observed differences between the dose groups odds is not greater than what might reasonably be expected when beta carotene supplementation had no true effect (P = 0.630). *(Note that the P value comes from the overall F test in this setting of dose modeled with two covariates. Having found no effect of dose whatsoever, it seems silly to me to belabor whether there might be a nonlinear effect: A flat line is linear.)*

e. Model dose as a binary variable indicating whether dose was greater than 0.

Ans: An analysis of the odds of markedly decreased vitamin E levels (i.e., a decrease of 1.5 mg / dl or more) after 9 months of supplementation when comparing the placebo group to the combined groups receiving some dose of beta carotene supplementation finds that the observed differences between the dose groups is not greater than what might reasonably be expected when beta carotene supplementation had no true effect on the odds of markedly decreased vitamin E levels (P = 0.232).

f. Model dose as two variables: a binary variable indicating whether dose was greater than 0 and a continuous linear term.

Ans: An analysis of the odds of markedly decreased vitamin E levels (i.e., a decrease of 1.5 mg / dl or more) after 9 months of supplementation based on a threshold at 0 mg / day and a linear effect on the log odds above that dose finds that the differences between dose groups is not greater than what might reasonably be expected when beta carotene supplementation had no true effect levels (P = 0.487). *(Note that the P value comes from the overall F test in this setting of dose modeled with two covariates. Having found no effect of dose whatsoever, it seems silly to me to belabor whether there might be a nonlinear effect: A flat line is linear.)*

g. Model dose as four variables: a continuous linear predictor, a quadratic term, a cubic term, and a fourth order term (i.e., dose raised to the fourth power).

Ans: An analysis of the odds of markedly decreased vitamin E levels (i.e., a decrease of 1.5 mg / dl or more) after 9 months of supplementation based on a quartic (fourth order) polynomial in dose finds that the observed differences between the dose groups means is not greater than what might reasonably be expected when beta carotene supplementation had no true effect (P = 0.736). *(Note that the P value comes from the overall F test in this setting of dose modeled with four covariates.)*

5. Repeat problem 4, but consider the odds of increased plasma beta carotene levels (i.e., a higher plasma beta carotene level at 9 months than at baseline) as a function of dose.

Ans: Unfortunately, sample sizes in this study were small enough (and the effect of supplementation on plasma beta carotene high enough) that there was little variability in response at the highest dose groups. That is, all of the subjects receiving beta carotene supplementation were observed to have higher plasma beta carotene levels after 9 months of treatment. In such a setting, it is impossible to fit a logistic regression model: The estimated odds ratio is infinite. Other techniques would have to be used in this setting.

6. Which of the above analyses would you prefer *a priori* to test for an effect of beta-carotene supplementation on plasma levels of beta-carotene? Which of the above analyses would you prefer *a priori* to test for an effect of beta-carotene supplementation on plasma levels of vitamin E? Justify your answer.

Ans: First, I would tend to always adjust for baseline in a randomized clinical trial. Then, I would choose an appropriate contrast of effects across dose groups. In order to protect the validity of statistical inference, an appropriate model would have to be chosen prior to looking at the data. We must therefore make judgements based on our beliefs about the type of trends that might be present in the data. Lacking any prior knowledge, I might use the linear continuous model in order to look for a first order trend. This would likely be my first choice for vitamin E, because I had no good reason to suspect any relationship at all. In the case of plasma beta carotene, it does not come as a big surprise that supplementation would have an effect, so the threshold-linear model would probably be my choice for its ability to parsimoniously look for an effect, while also allowing some insight into the dose response relationship. Indeed, I might even choose this for vitamin E, as it would allow greater flexibility while not completely losing power in the presence of a linear trend. In any case, all of the above models would have their adherents. The key point is to choose the model beforehand.