Biost 518: Applied Biostatistics II
Biost 515: Biostatistics II
Emerson, Winter 2015

Homework #5 Key
February 20, 2015

Written problems: To be submitted as a MS-Word compatible file to the class Catalyst dropbox by 9:30 am on Friday, February 27, 2014. See the instructions for peer grading of the homework that are posted on the web pages.

> *On this (as all homeworks) Stata / R code and unedited Stata / R output is **TOTALLY** unacceptable. Instead, prepare a table of statistics gleaned from the Stata output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*
>
> ***Unless explicitly told otherwise in the statement of the problem, in all problems requesting "statistical analyses" (either descriptive or inferential), you should present both***
>
> 1. ***Methods:** A brief sentence or paragraph describing the statistical methods you used. This should be using wording suitable for a scientific journal, though it might be a little more detailed. A reader should be able to reproduce your analysis. DO NOT PROVIDE Stata OR R CODE.*
> 2. ***Inference:** A paragraph providing full statistical inference in answer to the question. Please see the supplementary document relating to "Reporting Associations" for details.*

All problems of the homework relate to the clinical trial of DFMO and suppression of polyamines. In this homework I ask you to sometimes use dummy variables to analyze the data. There are two approaches to performing this analysis:

- Manually create indicator variables `dose0, dose075, dose200, dose400` that are 1 if the subject received the corresponding dose, and 0 otherwise. (You will only need to include three of these variables in any particular model, because the intercept will model the remaining group.)
- Use Stata's facility to automatically create such dummy variables in a regression analysis by prefixing a variable name with `"i."`:
  - To use this feature, you must have a variable with only integer values. So you might want to create a variable `g doseint= 1000 * dose`
  - Then, to perform an analysis of mean spermidine at 12 months across dose groups using dose 0 group as reference: `regress spd12 i.doseint, robust`
  - To perform an analysis of mean spermidine at 12 months across dose groups using dose 0.075 group as reference: `regress spd12 ib75.doseint, robust`
  - (In R you can use the `dummy( )` function to do these same things.)

1. Provide suitable descriptive statistics.

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

- ***Descriptive statistics should be presented within dose group. It is useful to think about the baseline data distinct from the follow-up data. I thus separated the descriptive statistics into two tables, though I did not necessarily expect students to do the same.***
- ***We are missing a lot of data for some treatment groups. This should definitely be noted, as should be the fact that it seems to occur more at the higher dose groups.***

- ***In Table 1, I am primarily focusing on the patient characteristics (and potential for imbalances across groups), so I present the more typical measurements of mean, standard deviation, minimum and maximum. I also chose to present the median in this table, mostly just to demonstrate the extent to which the geometric means for the polyamine data might be similar to the median in cases where positive data is skewed.***
    - ***One possible explanation for such a correspondence is that if the log transformed data is symmetric, then the median of the log transformed data is the same as the mean of the log transformed data. When we then exponentiate the mean of the log transformed data, we get the geometric mean. And because exponentiation is an "order preserving transformation", exponentiation of the median of the log transformed data leads to the median of the untransformed data.***
    - ***I did not present the geometric means in Table 1, because I was going to present the geometric means for the baseline data in the "Results" descriptive statistics in Table 2. But if this had been some other potential confounder that would be analyzed using logarithmic transformations, I may well have presented the geometric mean.***
- ***In Table 2, I am primarily focusing on obtaining preliminary estimates of the treatment effect. Several points are worth special note.***
    - ***I am interested descriptively in all time points, even though the primary analysis was just the 12 month measurement.***
    - ***Here I choose to present geometric means, rather than means, because that was the basis for my primary analyses. (I do also present an alternative table based on means just for use in grading.)***
    - ***It is extremely important to present minima and maxima, as there may have been individual patients who attained "toxic" levels. We may not know exactly what such a toxic level might be, but it is always of interest to have such data presented in order to view such individual characteristics in addition to the aggregate data represented by means, geometric means, or medians.***
    - ***I want to also present a suitable measure of spread, much as I would when analyzing the means. This material is presented for future reference. I was not really expecting many of you to have had experience with this issue.***
    - ***One choice for describing the spread would be the "geometric standard deviation", which is the exponentiated standard deviation of the log transformed data. This is of value for the exact same reasons that the SD is used with the mean:***
        - ***Technical reliance on equal variances will pertain to the equality of the standard deviations of the log transformed data. Exponentiation of those values will be just as easily used to assess such equality.***
        - ***The general results about the proportion of data within a certain number of SD of the mean will now apply to the proportion of data within a proportion of the geometric standard deviation away from the geometric mean. (This property is called "Chebyshev's inequality")***

        Chebyshev's Inequality for untransformed data:

        $$Y \sim \left(\mu, \sigma^2\right) \;\Rightarrow\; \Pr\left(\mu - k\sigma \le Y \le \mu + k\sigma\right) = \Pr\left(|Y - \mu| \le k\sigma\right) \ge 1 - \frac{1}{k^2}$$

        - ***Note that Chebyshev's inequality is not an exact bound on the probability of being with an additive distance from the mean. For instance, Chebyshev's tells us that***

more than 88.89% of the data will be within 3 SD of the mean for <u>any</u> distribution that has a variance. But we know much more for the normal (Gaussian) distribution: approximately 95% of the distribution is within 1.96 SD, and 99.73% of the data is within 3 SD of the mean for the normal distribution.

- ***When we apply Chebyshev's to the geometric mean, we now talk about the proportionate distance from the geometric mean.***

$$\textit{Chebyshev's Inequality for } \log \textit{ transformed data } W = \log(Y): \quad GM(Y) = e^{\nu}; \quad GSD(Y) = e^{\tau}$$

$$W \sim (\nu, \tau^2) \;\Rightarrow\; \Pr(\nu - k\tau \le W \le \nu + k\tau) = \Pr(|W - \nu| \le k\tau) \ge 1 - \frac{1}{k^2}$$

$$\Rightarrow\; \Pr\left(e^{\nu}e^{-k\tau} \le Y \le e^{\nu}e^{k\tau}\right) = \Pr\left(\left|\frac{Y}{e^{\nu}}\right| \le e^{k\tau} = (e^{\tau})^k\right) \ge 1 - \frac{1}{k^2}$$

- ***Another alternative would be the arithmetic coefficient of variation (CV), which is defined as the ratio of the standard deviation divided by the mean. This is a very natural parameter to report when log transformed data might be approximately distributed as a normal (Gaussian) random variable.***

$$\textit{Lognormal data}: \log \textit{ transformed data } W = \log(Y) \sim N(\nu, \tau^2)$$

$$GM(Y) = e^{\nu}; \quad GSD(Y) = e^{\tau} \qquad E(Y) = e^{\nu}e^{\tau^2/2} \qquad Var(Y) = \left(e^{\tau^2} - 1\right)e^{2\nu}e^{\tau^2}$$

$$\Rightarrow\; CV(Y) = \frac{SD(Y)}{E(Y)} = \sqrt{\left(e^{\tau^2} - 1\right)} = \sqrt{\left(GSD^2(Y) - 1\right)}$$

- ***When reporting the coefficient of variation, we typically compute this using the distribution free estimate: the sample SD divided by the sample mean. This will not agree closely with the relationship with the GSD unless the data are log normally distributed <u>and</u> we have a sufficiently large sample size to get a precise estimate of the variance of Y (which might need an extremely large sample size if the log normal distribution is highly skewed).***
- ***While it is probably more common to report the CV with geometric means, I think it actually better to report the geometric standard deviation, and so that is what I did here. (I note that it is not that unusual to only use the geometric means inferentially, in which case we might report the confidence intervals instead of the GSD).***
- ***(And we can of course always consider the interquartile range as another description of the spread, but that does not help us with the technical assumptions used in statistical analysis.)***

<u>Ans:</u> ***Methods:*** **Descriptive statistics are presented within dose groups for the baseline (pre-randomization) variables and included sample proportions for patient sex and mean, standard deviation, median, minimum, and maximum for age and baseline mucosal polyamines (putrescine, spermidine, spermine, and the spermidine:spermine ratio). Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis.**

When presenting descriptive statistics for the mucosal polyamine measurements over the course of the study, descriptive statistics are presented for each dose group at each measurement time and include the geometric mean, geometric standard deviation (the exponentiated standard deviation of the log transformed measurements), arithmetic coefficient of variation, minimum, and maximum. In computing statistics for the geometric means, mucosal polyamine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement for the corresponding polyamine. Hence, 26 observed values of 0 µmol / mg protein for putrescine were replaced with 0.001, 2 observed values of 0 µmol / mg protein for spermidine were replaced with 0.1415, and 2 observed values of 0 µmol / mg protein for spermine were replaced with 0.7275. Two subjects having both spermidine and spermine values below the limit of detection were omitted from the descriptive statistics for the ratio.

Results: **Table 1 presents summary statistics for the 114 subjects (17 female, 97 male) randomized to the clinical trial. Ages ranged from 45 to 81 years (mean 63.6, SD 8.2). Dose groups were well balanced with respect to age and baseline mucosal polyamine levels, though the relatively small numbers of women randomized to the trial resulted in no women randomized to the 200 mg/m^2/day group.**

Table 2 presents summary statistics for the mucosal polyamine measurements over the course of the study. Evident in the table is the increasing numbers of subjects who dropped out of the study as time from randomization increased, with the largest numbers of study dropout occurring in the highest dose group. In that 400 mg/m^2/day group, 8 of the originally randomized 28 subjects (29%) do not have measurements available at the end of the 12 month DFMO treatment period. That dropout rate is approximately double that of the other dose groups. Such missing data might be due to toxicity, and to the extent that toxicity might be associated with polyamine suppression, the results could be biased estimates of what would be observed had all subjects completed the study.

The dose groups are well-balanced with respect to the geometric means of the polyamine levels at the baseline (pre-randomization). However at 6 months of treatment, there are trends toward lower putrescine, spermidine, and spermidine:spermine ratios as dose of DFMO increases,. No striking trends are evident for spermine during DFMO treatment. Three months after treatment stopped (15 months post randomization), the polyamine measurements are relatively similar across dose groups.

(Table 2a presents analogous descriptive statistics for mean polyamine levels, rather than geometric mean levels. While I preferred to use the geometric means owing to the biologic mechanisms (DFMO blocks the activity of the enzyme ornithine decarboxylase, and the polyamine measurements are measuring the cumulative effects of changes in the rate of enzymatic reactions), it is not incorrect to analyse the mean levels.)

Table 1: Descriptive statistics for baseline variables. Statistics presented for continuous variables are the mean, standard deviation (sd), median (mdn), and minimum-maximum (range).

	Dose of DFMO				
	Placebo N= 32	**75 mg/m^2/day N= 29**	**200 mg/m^2/day N= 25**	**400 mg/m^2/day N= 28**	**All Subjects N= 114**
Female (n (%))	6 / 32 (18.8%)	5 / 29 (17.2%)	0 / 25 (0.0%)	6 / 28 (21.4%)	17 / 114 (14.9%)
Age (y)[1]	65.9 (sd= 8.5) (mdn= 66; range: 45- 77)	61.3 (sd= 7.7) (mdn= 61; range: 48- 77)	62.8 (sd= 8.3) (mdn= 64; range: 45- 78)	63.9 (sd= 7.8) (mdn= 65; range: 49- 81)	63.6 (sd= 8.2) (mdn= 64; range: 45- 81)
Putrescine (umol/mg prot)	0.66 (sd= 0.44) (mdn= 0.57; range: 0.06- 1.98)	0.65 (sd= 0.52) (mdn= 0.54; range: 0.01- 2.59)	0.61 (sd= 0.42) (mdn= 0.60; range: 0.00- 1.96)	0.65 (sd= 0.57) (mdn= 0.60; range: 0.00- 2.30)	0.65 (sd= 0.49) (mdn= 0.57; range: 0.00- 2.59)
Spermidine (umol/mg prot)	3.26 (sd= 1.45) (mdn= 2.93; range: 1.40- 7.05)	3.47 (sd= 1.55) (mdn= 2.91; range: 1.51- 7.02)	3.35 (sd= 1.33) (mdn= 2.92; range: 1.70- 6.22)	3.56 (sd= 1.88) (mdn= 3.08; range: 0.66- 7.60)	3.41 (sd= 1.55) (mdn= 2.96; range: 0.66- 7.60)
Spermine (umol/mg prot)	8.22 (sd= 5.54) (mdn= 7.52; range: 1.46- 35.5)	8.43 (sd= 5.86) (mdn= 7.32; range: 4.13- 37.7)	9.03 (sd= 7.04) (mdn= 7.53; range: 2.54- 41.7)	8.08 (sd= 5.50) (mdn= 6.81; range: 2.28- 34.0)	8.42 (sd= 5.90) (mdn= 7.48; range: 1.46- 41.7)
Spd:Spm ratio	0.46 (sd= 0.23) (mdn= 0.37; range: 0.12- 1.16)	0.46 (sd= 0.22) (mdn= 0.43; range: 0.13- 1.11)	0.46 (sd= 0.31) (mdn= 0.34; range: 0.12- 1.73)	0.48 (sd= 0.24) (mdn= 0.41; range: 0.20- 1.16)	0.46 (sd= 0.25) (mdn= 0.40; range: 0.12- 1.73)

[1] One subject in the 75 mg/m^2/day group was missing data for age.

Table 2: Descriptive statistics for mucosal polyamine measurements at baseline (pre-treatment), over the 12 month course of treatment, and at 15 months following randomization (3 months after cessation of treatment). Statistics presented for continuous variables are the geometric mean, geometric standard deviation (gsd), number of subjects missing data at the corresponding time (msng), median (mdn), and minimum-maximum (range). For the computation of statistics in this table, subjects having mucosal polyamine measurements recorded as 0 had a value equal to one-half the lowest observed positive measurement for the corresponding polyamine substituted. Two subjects who had zero measurements for both spermidine and spermine are listed as missing data for the spermidine:spermine (Spd:Spm) ratio.

	Dose of DFMO			
	Placebo	**75 mg/m^2/day**	**200 mg/m^2/day**	**400 mg/m^2/day**
	N= 32	N= 29	N= 25	N= 28
Putrescine (umol/mg prot) Baseline (pre)	0.52 (gsd= 2.19; msng= 0) (mdn= 0.57; range: 0.06 - 1.98)	0.46 (gsd= 2.86; msng= 0) (mdn= 0.54; range: 0.01 - 2.59)	0.35 (gsd= 5.77; msng= 0) (mdn= 0.60; range: 0.00 - 1.96)	0.27 (gsd= 9.06; msng= 0) (mdn= 0.60; range: 0.00 - 2.30)
6 months (tx)	0.71 (gsd= 2.27; msng= 2) (mdn= 0.72; range: 0.05 - 9.13)	0.25 (gsd= 7.30; msng= 1) (mdn= 0.49; range: 0.00 - 1.05)	0.23 (gsd= 4.99; msng= 2) (mdn= 0.27; range: 0.00 - 2.43)	0.07 (gsd= 14.36; msng= 3) (mdn= 0.19; range: 0.00 - 1.73)
12 months (tx)	0.61 (gsd= 6.91; msng= 4) (mdn= 0.90; range: 0.00 - 3.18)	0.71 (gsd= 2.78; msng= 3) (mdn= 0.69; range: 0.04 - 4.29)	0.42 (gsd= 5.30; msng= 4) (mdn= 0.51; range: 0.00 - 3.21)	0.07 (gsd= 29.04; msng= 8) (mdn= 0.39; range: 0.00 - 5.48)
15 months (off)	0.98 (gsd= 2.45; msng= 5) (mdn= 0.80; range: 0.30 - 6.51)	0.78 (gsd= 4.45; msng= 3) (mdn= 0.91; range: 0.00 - 3.53)	0.63 (gsd= 4.96; msng= 4) (mdn= 0.87; range: 0.00 - 2.58)	0.96 (gsd= 1.94; msng= 10) (mdn= 0.85; range: 0.23 - 3.20)
Spermidine (umol/mg prot) Baseline (pre)	2.97 (gsd= 1.55; msng= 0) (mdn= 2.93; range: 1.40 - 7.05)	3.17 (gsd= 1.55; msng= 0) (mdn= 2.91; range: 1.51 - 7.02)	3.13 (gsd= 1.45; msng= 0) (mdn= 2.92; range: 1.70 - 6.22)	3.11 (gsd= 1.74; msng= 0) (mdn= 3.08; range: 0.66 - 7.60)
6 months (tx)	3.07 (gsd= 1.54; msng= 2) (mdn= 3.02; range: 1.51 - 6.91)	2.51 (gsd= 1.38; msng= 1) (mdn= 2.45; range: 1.39 - 5.12)	2.23 (gsd= 1.69; msng= 2) (mdn= 1.85; range: 1.07 - 7.84)	2.39 (gsd= 1.60; msng= 3) (mdn= 2.07; range: 1.06 - 6.34)
12 months (tx)	3.01 (gsd= 1.51; msng= 4) (mdn= 2.82; range: 1.01 - 5.91)	2.75 (gsd= 1.44; msng= 3) (mdn= 2.86; range: 1.35 - 4.92)	2.32 (gsd= 1.90; msng= 4) (mdn= 2.51; range: 0.29 - 6.45)	1.71 (gsd= 1.94; msng= 8) (mdn= 1.93; range: 0.15 - 3.42)
15 months (off)	2.54 (gsd= 1.42; msng= 5) (mdn= 2.44; range: 1.25 - 4.62)	2.64 (gsd= 1.91; msng= 3) (mdn= 2.98; range: 0.15 - 4.83)	2.85 (gsd= 1.35; msng= 4) (mdn= 2.81; range: 1.81 - 4.80)	2.57 (gsd= 1.40; msng= 10) (mdn= 2.69; range: 1.29 - 4.46)
Spermine (umol/mg prot) Baseline (pre)	7.17 (gsd= 1.69; msng= 0) (mdn= 7.52; range: 1.45 - 35.5)	7.61 (gsd= 1.47; msng= 0) (mdn= 7.32; range: 4.13 - 37.7)	7.94 (gsd= 1.56; msng= 0) (mdn= 7.53; range: 2.54 - 41.7)	7.19 (gsd= 1.56; msng= 0) (mdn= 6.81; range: 2.28 - 34.0)
6 months (tx)	6.88 (gsd= 1.45; msng= 2) (mdn= 6.50; range: 3.31 - 14.4)	7.75 (gsd= 1.34; msng= 1) (mdn= 8.01; range: 4.60 - 15.7)	6.71 (gsd= 1.49; msng= 2) (mdn= 7.37; range: 2.34 - 12.1)	7.11 (gsd= 1.69; msng= 3) (mdn= 8.01; range: 2.76 - 17.2)
12 months (tx)	5.67 (gsd= 1.73; msng= 4) (mdn= 5.23; range: 2.32 - 14.6)	7.12 (gsd= 1.54; msng= 3) (mdn= 8.35; range: 3.15 - 14.1)	6.52 (gsd= 1.56; msng= 4) (mdn= 6.63; range: 2.96 - 13.8)	5.27 (gsd= 1.81; msng= 8) (mdn= 5.97; range: 0.73 - 10.7)
15 months (off)	5.95 (gsd= 1.48; msng= 5) (mdn= 5.78; range: 2.83 - 12.1)	5.79 (gsd= 1.87; msng= 3) (mdn= 6.29; range: 0.73 - 12.4)	5.11 (gsd= 1.85; msng= 4) (mdn= 4.61; range: 1.93 - 12.4)	5.96 (gsd= 1.50; msng= 10) (mdn= 6.00; range: 2.53 - 11.5)
Spd:Spm ratio Baseline (pre)	0.41 (gsd= 1.62; msng= 0) (mdn= 0.37; range: 0.12 - 1.16)	0.42 (gsd= 1.57; msng= 0) (mdn= 0.43; range: 0.13 - 1.11)	0.39 (gsd= 1.67; msng= 0) (mdn= 0.34; range: 0.12 - 1.73)	0.43 (gsd= 1.55; msng= 0) (mdn= 0.41; range: 0.20 - 1.16)
6 months (tx)	0.45 (gsd= 1.65; msng= 2) (mdn= 0.41; range: 0.14 - 1.10)	0.32 (gsd= 1.40; msng= 1) (mdn= 0.31; range: 0.20 - 0.64)	0.33 (gsd= 1.74; msng= 2) (mdn= 0.28; range: 0.17 - 1.72)	0.34 (gsd= 2.02; msng= 3) (mdn= 0.32; range: 0.16 - 2.20)
12 months (tx)	0.53 (gsd= 1.69; msng= 4) (mdn= 0.49; range: 0.21 - 2.07)	0.39 (gsd= 1.43; msng= 3) (mdn= 0.39; range: 0.19 - 0.73)	0.36 (gsd= 2.22; msng= 4) (mdn= 0.37; range: 0.03 - 1.03)	0.33 (gsd= 1.46; msng= 9) (mdn= 0.35; range: 0.18 - 0.76)
15 months (off)	0.43 (gsd= 1.41; msng= 5) (mdn= 0.46; range: 0.24 - 0.97)	0.47 (gsd= 1.61; msng= 4) (mdn= 0.38; range: 0.27 - 1.21)	0.56 (gsd= 1.82; msng= 4) (mdn= 0.54; range: 0.25 - 2.18)	0.43 (gsd= 1.41; msng= 10) (mdn= 0.41; range: 0.22 - 0.75)

Table 2a: Descriptive statistics for mucosal polyamine measurements at baseline (pre-treatment), over the 12 month course of treatment, and at 15 months following randomization (3 months after cessation of treatment). Statistics presented for continuous variables are the mean, standard deviation (sd), number of subjects missing data at the corresponding time (msng), median (mdn), and minimum-maximum (range). Two subjects who had zero measurements for both spermidine and spermine are listed as missing data for the spermidine:spermine (Spd:Spm) ratio.

	Dose of DFMO			
	Placebo	**75 mg/m^2/day**	**200 mg/m^2/day**	**400 mg/m^2/day**
	N= 32	N= 29	N= 25	N= 28
Putrescine (umol/mg prot) Baseline (pre)	0.66 (sd= 0.44; msng= 0) (mdn= 0.57; range: 0.06 - 1.98)	0.65 (sd= 0.52; msng= 0) (mdn= 0.54; range: 0.01 - 2.59)	0.61 (sd= 0.42; msng= 0) (mdn= 0.60; range: 0.00 - 1.96)	0.65 (sd= 0.57; msng= 0) (mdn= 0.60; range: 0.00 - 2.30)
6 months (tx)	1.06 (sd= 1.59; msng= 2) (mdn= 0.72; range: 0.05 - 9.14)	0.47 (sd= 0.27; msng= 1) (mdn= 0.49; range: 0.00 - 1.05)	0.45 (sd= 0.54; msng= 2) (mdn= 0.27; range: 0.00 - 2.43)	0.33 (sd= 0.43; msng= 3) (mdn= 0.19; range: 0.00 - 1.73)
12 months (tx)	1.16 (sd= 0.83; msng= 4) (mdn= 0.90; range: 0.00 - 3.18)	1.08 (sd= 1.03; msng= 3) (mdn= 0.69; range: 0.04 - 4.29)	0.80 (sd= 0.79; msng= 4) (mdn= 0.51; range: 0.00 - 3.21)	0.88 (sd= 1.42; msng= 8) (mdn= 0.40; range: 0.00 - 5.48)
15 months (off)	1.54 (sd= 1.77; msng= 5) (mdn= 0.80; range: 0.30 - 6.51)	1.19 (sd= 0.89; msng= 3) (mdn= 0.91; range: 0.00 - 3.53)	1.00 (sd= 0.71; msng= 4) (mdn= 0.87; range: 0.00 - 2.59)	1.18 (sd= 0.83; msng= 10) (mdn= 0.85; range: 0.23 - 3.20)
Spermidine (umol/mg prot) Baseline (pre)	3.26 (sd= 1.45; msng= 0) (mdn= 2.93; range: 1.40 - 7.05)	3.47 (sd= 1.55; msng= 0) (mdn= 2.91; range: 1.51 - 7.02)	3.35 (sd= 1.33; msng= 0) (mdn= 2.92; range: 1.70 - 6.22)	3.56 (sd= 1.88; msng= 0) (mdn= 3.08; range: 0.66 - 7.60)
6 months (tx)	3.37 (sd= 1.53; msng= 2) (mdn= 3.02; range: 1.51 - 6.91)	2.64 (sd= 0.89; msng= 1) (mdn= 2.46; range: 1.39 - 5.12)	2.58 (sd= 1.64; msng= 2) (mdn= 1.85; range: 1.07 - 7.84)	2.68 (sd= 1.43; msng= 3) (mdn= 2.07; range: 1.06 - 6.34)
12 months (tx)	3.26 (sd= 1.31; msng= 4) (mdn= 2.82; range: 1.01 - 5.91)	2.92 (sd= 0.99; msng= 3) (mdn= 2.86; range: 1.35 - 4.92)	2.71 (sd= 1.40; msng= 4) (mdn= 2.51; range: 0.29 - 6.45)	1.95 (sd= 0.80; msng= 8) (mdn= 1.93; range: 0.00 - 3.42)
15 months (off)	2.69 (sd= 0.93; msng= 5) (mdn= 2.45; range: 1.25 - 4.62)	2.95 (sd= 0.99; msng= 3) (mdn= 2.98; range: 0.00 - 4.83)	2.98 (sd= 0.90; msng= 4) (mdn= 2.81; range: 1.81 - 4.81)	2.70 (sd= 0.87; msng= 10) (mdn= 2.69; range: 1.29 - 4.47)
Spermine (umol/mg prot) Baseline (pre)	8.22 (sd= 5.54; msng= 0) (mdn= 7.52; range: 1.46 - 35.5)	8.43 (sd= 5.86; msng= 0) (mdn= 7.32; range: 4.13 - 37.7)	9.03 (sd= 7.04; msng= 0) (mdn= 7.53; range: 2.54 - 41.7)	8.08 (sd= 5.50; msng= 0) (mdn= 6.81; range: 2.28 - 34.0)
6 months (tx)	7.34 (sd= 2.71; msng= 2) (mdn= 6.51; range: 3.31 - 14.4)	8.07 (sd= 2.49; msng= 1) (mdn= 8.01; range: 4.60 - 15.7)	7.18 (sd= 2.48; msng= 2) (mdn= 7.37; range: 2.35 - 12.1)	8.04 (sd= 3.95; msng= 3) (mdn= 8.01; range: 2.76 - 17.2)
12 months (tx)	6.55 (sd= 3.59; msng= 4) (mdn= 5.24; range: 2.32 - 14.6)	7.75 (sd= 3.12; msng= 3) (mdn= 8.35; range: 3.15 - 14.1)	7.16 (sd= 3.15; msng= 4) (mdn= 6.63; range: 2.96 - 13.8)	5.93 (sd= 2.58; msng= 8) (mdn= 5.97; range: 0.00 - 10.7)
15 months (off)	6.39 (sd= 2.45; msng= 5) (mdn= 5.79; range: 2.83 - 12.1)	6.69 (sd= 3.32; msng= 3) (mdn= 6.29; range: 0.00 - 12.4)	6.08 (sd= 3.56; msng= 4) (mdn= 4.61; range: 1.93 - 12.4)	6.42 (sd= 2.53; msng= 10) (mdn= 6.00; range: 2.53 - 11.5)
Spd:Spm ratio Baseline (pre)	0.46 (sd= 0.23; msng= 0) (mdn= 0.37; range: 0.12 - 1.16)	0.46 (sd= 0.22; msng= 0) (mdn= 0.43; range: 0.13 - 1.11)	0.46 (sd= 0.31; msng= 0) (mdn= 0.34; range: 0.12 - 1.73)	0.48 (sd= 0.24; msng= 0) (mdn= 0.41; range: 0.20 - 1.16)
6 months (tx)	0.51 (sd= 0.27; msng= 2) (mdn= 0.41; range: 0.14 - 1.10)	0.34 (sd= 0.12; msng= 1) (mdn= 0.31; range: 0.20 - 0.64)	0.40 (sd= 0.33; msng= 2) (mdn= 0.28; range: 0.17 - 1.72)	0.45 (sd= 0.44; msng= 3) (mdn= 0.32; range: 0.16 - 2.20)
12 months (tx)	0.62 (sd= 0.43; msng= 4) (mdn= 0.49; range: 0.21 - 2.07)	0.41 (sd= 0.14; msng= 3) (mdn= 0.39; range: 0.19 - 0.73)	0.45 (sd= 0.28; msng= 4) (mdn= 0.37; range: 0.03 - 1.03)	0.36 (sd= 0.14; msng= 9) (mdn= 0.35; range: 0.18 - 0.76)
15 months (off)	0.45 (sd= 0.16; msng= 5) (mdn= 0.46; range: 0.24 - 0.97)	0.53 (sd= 0.27; msng= 4) (mdn= 0.38; range: 0.27 - 1.21)	0.68 (sd= 0.50; msng= 4) (mdn= 0.54; range: 0.25 - 2.18)	0.46 (sd= 0.15; msng= 10) (mdn= 0.41; range: 0.22 - 0.75)

2. 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 DFMO on the mucosal spermidine levels after 12 months of treatment. ***(Recall that when multiple modeled covariates are derived from the same scientific factor, you need to test all those covariates simultaneously. When no other covariates are in the model, the "overall F" or "overall chi squared" test can do this for us.)*** Note that part h asks you to provide a table of predicted values for each of these models.

(Comments about the answers to this problem: As usual, the general format for answering questions about associations is similar for all parts. I have highlighted in blue font those parts of the answers that are markedly different from each other.)

a. Model dose as dummy variables using the dose 0 group as the reference group.

Instructions for grading: This problem is worth 5 points. Although I included inference for the specific parameter estimates from multiple covariates used to model the dose effect, I did not require that students do so on this problem. I do highly recommend that you understand these interpretations, however. Points to consider in the grading include:

- ***I analyzed the treatment effect as ratios of geometric means. A priori this would likely be anticipated to be the most precise way to detect a treatment effect.***
- ***I have also included the results that would have been obtained when using the difference in means as the measure of association. There is nothing wrong with such an analysis. (Interestingly, we obtain lower P values in these analyses, but it would of course be wrong to do both analyses and choose the one with lower P values. Thus having chosen the geometric means a priori, I would have to live with that choice.)***
- ***(Note that I did not specify what summary measure should be used as a measure of association. The obvious choices would be the mean or the geometric mean, though arguably the approach used in problem 4 could also apply to this wording.)***

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in an analysis of variance implemented using dummy variables for dose group in a linear regression model.** *(Note that by using the term "analysis of variance" it would be immediately understood that the overall F test would be being used to test for an association. However, you could include a sentence to the effect of: Tests of a treatment effect were based on a Wald test simultaneously testing all of the dummy variables modeling dose.)* **Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors.** *(Note that by using the term "analysis of variance" it would be immediately understood that the overall F test would be being used to test for an association.)* **The CI and p values for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. An analysis of variance finds that observed differences between the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0082).** *(Analysis based on the difference in means has P= .0001.)* **We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses.** *(Students did not have to report the following. In real life I would likely have just given it in a table. It is often useful to present the pairwise comparisons to placebo. I do not adjust for multiple comparisons, figuring that anybody can do a Bonferroni correction in their heads. But in the statement of the problem I asked for 95% CI, and, strictly speaking, we cannot get exact 95% CI in the presence of multiple comparisons. Instead we typically use conservative adjustments.)* **Comparisons of each of the three groups receiving DFMO to the group receiving placebo reveals:**

- **the 75 mg/m^2/day dose group is estimated to have a geometric mean only 0.914 times as large as the placebo group (95% CI unadjusted for multiple comparisons: 0.741 to 1.13 times as large; unadjusted two-sided P= .397),** *(if analyzing the means, the difference in means is -.336 with 95% CI -.965 to .292, two-sided P= .291)*
- **the 200 mg/m^2/day dose group is estimated to have a geometric mean only 0.773 times as large as the placebo group (95% CI unadjusted for multiple comparisons: 0.563 to 1.06 times as large; unadjusted two-sided P= .110),** *(if analyzing the means, the difference in means is -.544 with 95% CI -1.32 to .236, two-sided P= .169)* **and**
- **the 400 mg/m^2/day dose group is estimated to have a geometric mean only 0.569 times as large as the placebo group (95% CI unadjusted for multiple comparisons: 0.408 to 0.793 times as large; unadjusted two-sided P= .001)** *(if analyzing the means, the difference in means is -1.31 with 95% CI -1.91 to -0.698, two-sided P< .0001)***.**
- *(The intercept can be interpreted relative to the placebo group: the geometric mean spermidine 12 months after randomization is estimated to be 3.01 µmol / mg protein, with a 95% CI 2.58 to 3.51 µmol / mg protein. I would generally <u>not</u> report this, instead using the descriptive statistics. (If analyzing the means, the mean spermidine 12 months after randomization is estimated as 3.26 µmol / mg protein, with 95% CI 2.76 to 3.75 µmol / mg protein.))*

b. Model dose as dummy variables using the dose 0.075 group as the reference group. You do not have to provide a formal description of the methods or inference for this part. Instead comment on how the regression parameters from this model relate to those obtained in part a. Suppose we were to completely ignore the major multiple comparison issues and to instead trust the individual p values listed in the coefficient table, what conclusions would we reach about differences among the dose groups in part a vs in part b?

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

<u>Answer</u>: **This is just a reparameterization of the model in problem 2a. Hence, fitted values would be the same, as would inference about a treatment effect. Furthermore, if we estimate the same contrasts (e.g., comparisons between two dose groups), we will always also get the same inference for those. The only difference in the models is the ease with which we can make comparisons between pairs of groups. If we overinterpret "nonsignificant" P values for the pairwise comparisons, we may sometimes find**

that we can discriminate between , say, dose 0 and dose 400, but not between dose 0 and dose 200 or between dose 200 and dose 400. In such a setting, dropping the dummy variables with the "nonsignificant" P values would lead to very different conclusions based on the arbitrary way you fit the model. Such is the result if you did use the 200 dose group as the reference. But no such problem was evident when using the dose 75 group vs dose 0 group as reference: in each case the unadjusted pairwise P values could discriminate the reference group from the dose 400 group but not from the other groups. *(This answer would be the same no matter what summary measure—means, geometric means, odds, medians, hazards-- is used to compare the groups.)*

c. Model dose continuously as a linear predictor.

***Instructions for grading*: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. Here I do require that students provide interpretation of the slope term.**

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose as an untransformed continuous predictor. A hypothesis test for the presence of a DFMO effect on mucosal spermidine was based on the regression slope parameter. Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. ~~The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A linear regression analysis finds that observed trend across the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0006). We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses. We observed a trend toward a geometric mean spermidine level being 1.40% lower for every 100 mg/m2/day increase in DFMO dose (ratio of geometric means 0.986). The 95% confidence interval suggests that these results are consistent with a true effect in which the spermidine levels are anywhere between 0.62% lower to 2.18% lower for each 100 mg/m2/day difference in DFMO dose.** *(If making comparisons across1,000 mg/m^2/day difference in dose (the original units for dose), the geometric mean ratio would be estimated to be 0.244 with 95% CI of 0.110 to 0.537. Of course the maximum difference in dose that we investigated was 400 mg/m^2/day. If the difference in mean spermidine was used as the measure of association, the estimated treatment effect per 100 mg/m^2/day would be a difference of -.0313 in the mean spermidine level, with a 95% CI of -.0175 to -.0450 and P < 0.0001.)*

- *(The intercept can be interpreted relative to the placebo group: the geometric mean spermidine 12 months after randomization is estimated to be 3.04 µmol / mg protein, with a 95% CI 2.69 to 3.43 µmol / mg protein. I would generally not report this, instead using the descriptive statistics. This*

inference would differ from that obtained from the sample geometric mean, because this uses the linear trend across all dose groups to estimate the geometric mean for the placebo group.)

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).

***Instructions for grading*: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable.**

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose in a quadratic model including both an untransformed continuous predictor and a continuous predictor equal to dose squared. Hypothesis tests for an effect of DFMO on mucosal spermidine was based on testing that both regression coefficients were simultaneously 0 when using a Wald test with the Huber-White sandwich estimator of standard errors.** ~~Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A test for treatment effect using a quadratic model for dose finds that observed differences between the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0027). We thus with high confidence reject the null hypothesis of no treatment effect.** *(Analyses based on differences in means found P< .0001.)*

- *(We could, of course, obtained predicted geometric means and 95% CI for the dose groups, but interpretation of regression parameters is extremely difficult, as is even talking about the direction of the association.)*

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

***Instructions for grading*: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. Here I do require that students provide interpretation of the slope term.**

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415.**

Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose as a binary indicator that DFMO was administered at some level above 0. A hypothesis test for the presence of a DFMO effect on mucosal spermidine was based on the regression slope parameter. Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. ~~The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A linear regression comparing subjects taking any dose of DFMO to those taking placebo finds that observed differences between the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .008). We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with DFMO treatment. We observed a trend toward a geometric mean spermidine level being 24.7% lower for subjects taking some dose of DFMO (ratio of geometric means 0.753). The 95% confidence interval suggests that these results are consistent with a true effect in which the geometric mean spermidine levels for patients on the combined DFMO arms are anywhere between 7.25% lower to 39.9% lower than those on the placebo arm.** *(If the difference in mean spermidine was used as the measure of association, the estimated treatment effect would be a difference of -.6911 in the mean spermidine level, with a 95% CI of -1.255 to -.0.1275 and P = 0.017.)*

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

<u>Instructions for grading</u>: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. I provide estimates of the parameters, though students did not have to provide them for this homework. You should understand these interpretations, however.

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose as both a binary indicator that DFMO was administered at some level above 0 and an untransformed continuous predictor. Hypothesis tests for an effect of DFMO on mucosal spermidine was based on testing that both regression coefficients were simultaneously 0 when using a Wald test with the Huber-White sandwich estimator of standard errors. Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR),

however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A test for treatment effect using a model for a threshold effect with any DFMO exposure and a linear trend above dose 0 mg/m2/day finds that observed differences between the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0027). We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses.** *(Students did not have to report the following. I do not adjust for multiple comparisons, figuring that anybody can do a Bonferroni correction in their heads.)*
Comparisons of each of the three groups receiving DFMO to the group receiving placebo reveals:

- **When comparing two groups each receiving some dose of DFMO, we observed a trend toward a geometric mean spermidine level being 1.45% lower for every 100 mg/m^2/day increase in DFMO dose (ratio of geometric means 0.986, two-sided P= .005 unadjusted for multiple comparisons). The 95% confidence interval suggests that these results are consistent with a true effect between two groups each receiving some DFMO is one in which the spermidine levels are anywhere between 0.45% lower to 2.44% lower for each 100 mg/m^2/day difference in DFMO dose.** *(If making comparisons across 1,000 mg/m^2/day difference in dose (the original units for dose), the geometric mean ratio would be estimated to be 0.232 with 95% CI of 0.0842 to 0.638 (the P value would be unchanged). Of course the maximum difference in dose that we investigated was 400 mg/m^2/day. If the difference in mean spermidine was used as the measure of association, the corresponding estimated treatment effect per 100 mg/m^2/day difference in dose for two groups who were each taking some dose of DFMO would be a difference of -.0302 in the mean spermidine level, with a 95% CI of -.014 to -.0460 and P < 0.0005.)*
- **When comparing the geometric mean spermidine level for the placebo group to what would be predicted by the linear trend among the groups receiving some dose of DFMO, the placebo group is estimated to have a geometric mean 1.025 times what would be predicted by the linear trend (inference unadjusted for multiple comparisons yields 95% CI of 0.794 to 1.323 for the geometric mean ratio, P = 0.848). Hence, there is no strong evidence for nonlinearity between the placeo group and the active dose groups.** *(If the difference in mean spermidine was used as the measure of association, the corresponding estimates are a placebo group that has mean spermidine 0.0538 µmol / mg protein lower than would be predicted by the linear trend, with a 95% CI of -.730 to0.663 and P = .875.)*
- *(Note that interpretation of the intercept is complicated in this model, because the threshold effect must be considered as well.)*

g. Model dose as three variables: a continuous linear predictor, a quadratic term, and a cubic term (a term equal to dose raised to the third power).

<u>Instructions for grading</u>: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. There are no easy interpretation of the slope parameters, but this model (which is just a reparameterization of the ANOVA (dummy variables) model) does allow easier testing that the association might be nonlinear.

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415.**

Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose in a cubic model including all three of an untransformed continuous predictor, a continuous predictor equal to dose squared, and a continuous predictor equal to dose cubed. Hypothesis tests for an effect of DFMO on mucosal spermidine was based on testing that all three of the dose related regression coefficients were simultaneously 0 when using a Wald test with the Huber-White sandwich estimator of standard errors. ~~Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~ *(Students were not required to perform the following test. I include it to demonstrate one possible advantage of the cubic model fit to four groups relative to the ANOVA model. This is not much of an advantage, given the difficulty interpreting the regression parameters in any other way.)* **In a secondary analysis that is only relevant in the presence of a statistically significant association, a hypothesis test that a statistically significant association among the log geometric means is nonlinear across dose groups was based on testing that the quadratic and cubic terms were both 0.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. The cubic linear regression model finds that observed differences between the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0082).** *(If the difference in means was being analyzed, the P value is .0001.)* **We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses.** *(Students did not have to report the following.)* **A test for nonlinearity of the dose association with the log geometric mean spermidine levels is not statistically significant (P = 0.98), thus we do not have strong evidence that the effect of DFMO on spermidine levels is nonlinear.** *(Using analyses based on means, we would say that the test for nonlinearity of the dose association with the mean spermidine levels is not statistically significant (P=0.86))*

h. Provide a table of the fitted values for each dose group from the above models. Comment on the similarities / differences between those fitted values (and the descriptive statistics).

Instructions for grading: This problem is worth 5 points.

Answer: **The following table contains the fitted values from each of the six models. I note that the fitted geometric means from the dummy variables (Models 2a and 2b) and the cubic polynomial (Model 2g) each correspond exactly to the sample geometric means for each dose group. This correspondence between the cubic polynomial and the dummy variables is due to the fact that there were only four levels of dose sampled and in each of those models we fit four parameters (a saturated model with an intercept and three slopes). The log estimates from Model 2c would lie exactly on a straight line. The log estimates from Model 2d would lie on a U-shaped function that is very nearly a straight line, because the sample geometric means from each group are very nearly linear. In Model 2e, the estimated geometric means are the same for all dose groups above 0, and the dose 0 group estimate corresponds exactly to the sample geometric mean for that group. In Model 2f, the estimated log geometric means for dose groups higher than 0 lie exactly on a straight line, and the dose 0 group estimate corresponds exactly to the sample geometric mean for that group.** *(If you fit models for the mean, the above comments about linearity all obtain with the substitution of "mean" for "log geometric mean".)*

I included excessive significant digits to provide better evidence of exact equality of the highlighted fitted values in each row. I note that this agreement of the fitted values is due to the structure of the linear predictor. It is independent of the type of regression. Hence, those same columns would be highlighted if I considered difference of means in linear regression, odds ratios in logistic regression, rate ratios in Poisson regression, hazard ratios in proportional hazards regression, or quantile ratios in accelerated failure time regressions.

Table 3: Fitted geometric means for spermidine (µmol / mg protein) from the seven models.

DFMO Dose $mg/m^2/day$	Model 2a Dummy (ref=0)	Model 2b Dummy (ref=75)	Model 2c Linear	Model 2d Quadratic	Model 2e Threshold (0)	Model 2f Threshold (0), Linear	Model 2g Cubic
0	3.0083207	3.0083207	3.0386794	3.0092383	3.0083207	3.0083207	3.0083207
75	2.7497528	2.7497528	2.7332408	2.7479687	2.2644135	2.7633287	2.7497528
200	2.3247000	2.3247000	2.2908676	2.3258380	2.2644135	2.3017778	2.3247000
400	1.7113345	1.7113345	1.7270905	1.7111653	2.2644135	1.7181967	1.7113345

3. Repeat the analyses in problem 1 adjusting for the baseline mucosal spermidine 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.

(Comments about the answers to this problem: As usual, the general format for answering questions about associations is similar for all parts. I have highlighted in blue font those parts of the answers that are markedly different from the corresponding answer in problem 2.)

a. Model dose as dummy variables using the dose 0 group as the reference group.

Instructions for grading: This problem is worth 5 points. Although I included inference for the specific parameter estimates from multiple covariates used to model the dose effect, I did not require that students do so on this problem. I do highly recommend that you understand these interpretations, however. Points to consider in the grading include:

- ***I analyzed the treatment effect as ratios of geometric means. A priori this would likely be anticipated to be the most precise way to detect a treatment effect.***
- ***I have also included the results that would have been obtained when using the difference in means as the measure of association. There is nothing wrong with such an analysis. (Interestingly, we obtain lower P values in these analyses, but it would of course be wrong to do both analyses and choose the one with lower P values. Thus having chosen the geometric means a priori, I would have to live with that choice.)***
- ***(Note that I did not specify what summary measure should be used as a measure of association. The obvious choices would be the mean or the geometric mean, though arguably the approach used in problem 4 could also apply to this wording.)***

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups adjusted for baseline spermidine values. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the**

400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in an analysis of covariance implemented using dummy variables for dose group in a linear regression model that is also adjusted for logarithmically transformed baseline spermidine values. *(Note that by using the term "analysis of covariance" it would be immediately understood that the multiple partial F test would be being used to test for an association. However, you could include a sentence to the effect of: Tests of a treatment effect were based on a Wald test simultaneously testing all of the dummy variables modeling dose.)* **Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors.** *(Note that by using the term "analysis of variance" it would be immediately understood that the overall F test would be being used to test for an association.)* **The CI and p values for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. An analysis of covariance finds that observed differences between the dose groups' geometric means adjusted for baseline spermidine values is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0073).** *(Analysis based on the difference in means has P= .0001.)* **We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses.** *(Students did not have to report the following. In real life I would likely have just given it in a table. It is often useful to present the pairwise comparisons to placebo. I do not adjust for multiple comparisons, figuring that anybody can do a Bonferroni correction in their heads. But in the statement of the problem I asked for 95% CI, and, strictly speaking, we cannot get exact 95% CI in the presence of multiple comparisons. Instead we typically use conservative adjustments.)* **Comparisons of each of the three groups receiving DFMO to the group receiving placebo reveals:**

- **the 75 mg/m^2/day dose group is estimated to have a geometric mean only 0.911 times as large as the placebo group having similar baseline spermidine values (95% CI unadjusted for multiple comparisons:** *0.749* **to** *1.11* **times as large; unadjusted two-sided P= .349),** *(if analyzing the means, the difference in means is -.347 with 95% CI -.934 to .240, two-sided P= .244)*
- **the 200 mg/m^2/day dose group is estimated to have a geometric mean only 0.770 times as large as the placebo group having similar baseline spermidine values (95% CI unadjusted for multiple comparisons: 0.564 to 1.05 times as large; unadjusted two-sided P= .099),** *(if analyzing the means, the difference in means is -.556 with 95% CI -1.31 to .198, two-sided P= .146)* **and**
- **the 400 mg/m^2/day dose group is estimated to have a geometric mean only 0.56 times as large as the placebo group having similar baseline spermidine values (95% CI unadjusted for multiple comparisons: 0.401 to 0.787 times as large; unadjusted two-sided P= .001)** *(if analyzing the means, the difference in means is -1.35 with 95% CI -1.98 to -0.721, two-sided P= .0003)***.**
- *(The intercept can***not** *be interpreted relative to the placebo group* ***as a whole, because we adjusted for the logarithmically transformed baseline values. It would not correspond to a group having baseline spermidine value of 1—a baseline log spermidine value of 0.)***

b. Model dose continuously as a linear predictor.

***Instructions for grading*: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. Here I do require that students provide interpretation of the slope term.**

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups adjusted for baseline spermidine values. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose as an untransformed continuous predictor adjusted for logarithmically transformed baseline spermidine values. A hypothesis test for the presence of a DFMO effect on mucosal spermidine was based on the dose related regression slope parameter. Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. ~~The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A linear regression analysis finds that observed trend across the dose groups' geometric means adjusted for baseline spermidine values is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .001). We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses. We observed a trend toward a geometric mean spermidine level being 1.40% lower for every 100 mg/m2/day increase in DFMO dose in the presence of similar baseline spermidine values (ratio of geometric means 0.986). The 95% confidence interval suggests that these results are consistent with a true effect in which the spermidine levels are anywhere between 0.64% lower to 2.22% lower for each 100 mg/m2/day difference in DFMO dose.** *(If making comparisons across1,000 mg/m^2/day difference in dose (the original units for dose), the geometric mean ratio would be estimated to be* **0.237** *with 95% CI of* **0.106** *to* **0.528***. Of course the maximum difference in dose that we investigated was 400 mg/m^2/day. If the difference in mean spermidine was used as the measure of association, the estimated treatment effect per 100 mg/m^2/day would be a difference of* **-.0322** *in the mean spermidine level, with a 95% CI of* **-.0179** *to* **-.0466** *and P* **< 0.0005***.)*

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>Instructions for grading</u>: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable.

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups adjusted for baseline spermidine values. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose in a quadratic model including both an untransformed continuous predictor and a continuous predictor equal to dose squared with additional adjustment for logarithmically transformed baseline spermidine values. Hypothesis tests for an effect of DFMO on mucosal spermidine was based on testing that both dose related regression**

coefficients were simultaneously 0 when using a Wald test with the Huber-White sandwich estimator of standard errors. ~~Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A test for treatment effect using a quadratic model for dose adjusted for baseline spermidine values finds that observed differences between the dose groups' geometric means is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0024). We thus with high confidence reject the null hypothesis of no treatment effect.** *(Analyses based on differences in means found P=* **.0001***.)*

- *(We could, of course, obtained predicted geometric means and 95% CI for the dose groups, but interpretation of regression parameters is extremely difficult, as is even talking about the direction of the association.)*

- *(I could also have commented on whether the association appears linear, as I did in some other parts.)*

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

Instructions for grading: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. Here I do require that students provide interpretation of the slope term.

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups adjusted for baseline spermidine values. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose as a binary indicator that DFMO was administered at some level above 0, with adjustment for logarithmically transformed baseline spermidine values. A hypothesis test for the presence of a DFMO effect on mucosal spermidine was based on the dose related regression slope parameter. Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors.** ~~The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25**

patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A linear regression comparing subjects taking any dose of DFMO to those taking placebo finds that observed differences between the dose groups' geometric means adjusted for baseline spermidine values is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .005). We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with DFMO treatment. We observed a trend toward a geometric mean spermidine level being 25.1% lower for subjects taking some dose of DFMO but having similar baseline spermidine values (ratio of geometric means 0.749). The 95% confidence interval suggests that these results are consistent with a true effect in which the geometric mean spermidine levels for patients on the combined DFMO arms are anywhere between 8.38% lower to 38.8% lower than those on the placebo arm. *(If the difference in mean spermidine was used as the measure of association, the estimated treatment effect would be a difference of -.710 in the mean spermidine level, with a 95% CI of* **-1.249** *to* **-.0.172** *and P =* **0.010***.)*

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

***Instructions for grading*: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. I provide estimates of the parameters, though students did not have to provide them for this homework. You should understand these interpretations, however.**

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups adjusted for baseline spermidine values. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose as both a binary indicator that DFMO was administered at some level above 0 and an untransformed continuous predictor, with additional adjustment for logarithmically transformed baseline spermidine values. Hypothesis tests for an effect of DFMO on mucosal spermidine was based on testing that both dose related regression coefficients were simultaneously 0 when using a Wald test with the Huber-White sandwich estimator of standard errors. Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. The CI for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. A test for treatment effect using a model for a threshold effect with any DFMO exposure and a linear trend above dose 0 mg/m2/day finds that observed differences between the dose groups' geometric means adjusted for baseline spermidine values is greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0024). We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses.** *(Students did not have to report the following. I do not adjust for multiple comparisons, figuring that anybody can do a*

Bonferroni correction in their heads.) **Comparisons of each of the three groups receiving DFMO to the group receiving placebo reveals:**

- **When comparing two groups each receiving some dose of DFMO, we observed a trend toward a geometric mean spermidine level being 1.48% lower for every 100 mg/m^2/day increase in DFMO dose (ratio of geometric means 0.985, two-sided P= .005 unadjusted for multiple comparisons). The 95% confidence interval suggests that these results are consistent with a true effect between two groups each receiving some DFMO is one in which the spermidine levels are anywhere between 0.46% lower to 2.49% lower for each 100 mg/m^2/day difference in DFMO dose, but with both groups having similar baseline spermidine values.** *(If making comparisons across 1,000 mg/m^2/day difference in dose (the original units for dose), the geometric mean ratio would be estimated to be* **0.225** *with 95% CI of* **0.0804** *to* **0.631** *(the P value would be unchanged). Of course the maximum difference in dose that we investigated was 400 mg/m^2/day. If the difference in mean spermidine was used as the measure of association, the corresponding estimated treatment effect per 100 mg/m^2/day difference in dose for two groups who were each taking some dose of DFMO would be a difference of* **-.0312** *in the mean spermidine level, with a 95% CI of* **-.014** *to* **-.0480** *and P =* **0.0001***.)*
- **When comparing the geometric mean spermidine level for the placebo group to what would be predicted by the linear trend among the groups receiving some dose of DFMO, conditional on having similar baseline spermidine values, the placebo group is estimated to have a geometric mean 1.025 times what would be predicted by the linear trend (inference unadjusted for multiple comparisons yields 95% CI of 0.802 to 1.31 for the geometric mean ratio, P = 0.848). Hence, there is no strong evidence for nonlinearity between the placebo group and the active dose groups.** *(If the difference in mean spermidine was used as the measure of association, the corresponding estimates are a placebo group that has mean spermidine* **0.0532** *µmol / mg protein lower than would be predicted by the linear trend, with a 95% CI of* **-.694** *to* **0.587** *and P =* **.869***.)*
- *(Note that interpretation of the intercept is complicated in this model, because the threshold effect must be considered as well.)*

f. Model dose as three variables: a continuous linear predictor, a quadratic term, and a cubic term (a term equal to dose raised to the third power).

***Instructions for grading*: This problem is worth 5 points. As in part a, I used the geometric means, but analyses based on means are equally acceptable. There are no easy interpretation of the slope parameters, but this model (which is just a reparameterization of the ANOVA (dummy variables) model) does allow easier testing that the association might be nonlinear.**

Methods: **The effect of DFMO on mucosal spermidine measurements after 12 months of treatment was assessed by considering the ratio of geometric means across treatment groups adjusted for baseline spermidine values. In computing statistics for the geometric means, mucosal spermidine measurements recorded as 0 were merely assumed to be below the detectable limit. Such measurements were replaced with a value equal to one-half the lowest positive measurement observed in all patients during the study. Hence, 1 observed value of 0 µmol / mg protein for spermidine in the 400 mg/m^2/day dose group was replaced with 0.1415. Logarithmically transformed spermidine measurements were then analyzed in a linear regression modeling dose in a cubic model including all three of an untransformed continuous predictor, a continuous predictor equal to dose squared, and a continuous predictor equal to dose cubed, with additional adjustment for logarithmically transformed baseline spermidine values. Hypothesis tests for an effect of DFMO on mucosal spermidine was based on testing that all three of the dose related regression coefficients were simultaneously 0 when using a Wald test with the Huber-White sandwich estimator of standard errors.** ~~**Two-sided p values and 95% confidence intervals (CI) for the geometric mean ratios were computed using Wald statistics and the Huber White sandwich estimator for standard errors. The CI for the individual regression parameters**~~

~~were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.~~ *(Students were not required to perform the following test. I include it to demonstrate one possible advantage of the cubic model fit to four groups relative to the ANOVA model. This is not much of an advantage, given the difficulty interpreting the regression parameters in any other way.)* **In a secondary analysis that is only relevant in the presence of a statistically significant association, a hypothesis test that a statistically significant association among the log geometric means is nonlinear across dose groups was based on testing that the quadratic and cubic terms were both 0.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

Results: **Measurements of mucosal spermidine levels after 12 months of DFMO treatment are available for 28 of 30 patients randomized to placebo, 26 of 29 patients randomized to 75 mg/m^2/day, 21 of 25 patients randomized to 200 mg/m^2/day, and 20 of 28 patients randomized to 400 mg/m^2/day. The cubic linear regression model finds that observed differences between the dose groups' geometric means adjusted for baseline spermidine values are greater than what might reasonably be expected when DFMO has no true effect (two sided P= .0073).** *(If the difference in means was being analyzed, the P value is .0002.)* **We thus with high confidence reject the null hypothesis of no treatment effect, in favor of the observed trend toward lower levels with higher doses.** *(Students did not have to report the following.)* **A test for nonlinearity of the dose association with the log geometric mean spermidine levels is not statistically significant (P = 0.98), thus we do not have strong evidence that the effect of DFMO on spermidine levels is nonlinear.** *(Using analyses based on means, we would say that the test for nonlinearity of the dose association with the mean spermidine levels is not statistically significant (P=0.82))*

g. Provide a table of the fitted values for each dose group from the above models. Comment on the similarities / differences between those fitted values (and the descriptive statistics).

Instructions for grading: I specifically said you did not have to answer this question. However, I want to point out that the correspondence among the models would be the same as in problem 2h. It is just that every subject will likely have different fitted values, because there are many different baseline values. If I held the baseline value constant at some particular level, then I would get fitted values in the same general pattern as seen in problem 2.

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 DFMO on the odds of decreased spermidine levels after 12 months of treatment (i.e., a lower spermidine level at 12 months than at baseline). Note that in part g you are asked to provide a table of predicted values for the odds of decreased spermidine as well as the probability of decreased spermidine for each of these models.

    a. Model dose as dummy variables.

Instructions for grading: This problem is worth 5 points. For this problem, I gave CI in a more abbreviated form that might be more acceptable for publications (where space is considered a premium)

Methods: **The effect of DFMO on the odds of decreased spermidine measurements after 12 months of treatment was analyzed in a logistic regression model of dose as dummy variables. A p value for the test of association was based on the Wald test that the regression coefficients for all dose derived variables were simultaneously equal to zero. The CI and p values for the individual regression**

parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

***Results*: The observed differences between the dose groups with respect to the odds of decreased spermidine levels is not greater than what might reasonably be expected when DFMO had no true effect (P = .1594). The dose 75 group is estimated to have odds of decreased spermidine 1.85 times as large as the placebo group (95% CI unadjusted for multiple comparisons: 0.62 to 5.49 times as large), the dose 200 group is estimated to have odds of decreased spermidine 1.88 times as large as the placebo group (95% CI unadjusted for multiple comparisons: 0.59 to 5.97 times as large), and the dose 400 group is estimated to have odds of decreased spermidine 4.62 times as large as the placebo group (95% CI unadjusted for multiple comparisons: 1.22 to 17.5 times as large).**

b. Model dose continuously as a linear predictor.

***Instructions for grading*: This problem is worth 5 points. For this problem, I gave CI in a more abbreviated form that might be more acceptable for publications (where space is considered a premium)**

***Methods:* The effect of DFMO on the odds of decreased spermidine measurements after 12 months of treatment was analyzed in a logistic regression model of dose as dummy variables. A p value for the test of association was based on the Wald test that the regression coefficients for all dose derived variables were simultaneously equal to zero. The CI and p values for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

***Results*: An logistic regression analysis using a linear dose variable estimates that the odds of decreased spermidine tends to be 30.9 times higher (95% CI: 1.40 to 682.0 times higher) for each 1,000 mg/m^2/day difference in dose (or $30.9^{0.1}$= 1.41 times higher (95% CI: $1.40^{0.1}$=1.03 to $682^{0.1}$=1.92 times higher) for each 100 mg/m^2/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 DFMO on mucosal spermidine levels (P= .0299).** (*Note the much lower P value obtained when the analysis borrows strength across the ordered dose groups compared to the dummy variable model.)*

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).

***Instructions for grading*: This problem is worth 5 points. For this problem, I gave CI in a more abbreviated form that might be more acceptable for publications (where space is considered a premium)**

***Methods:* The effect of DFMO on the odds of decreased spermidine measurements after 12 months of treatment was analyzed in a logistic regression model of dose as a quadratic function including both a linear term and a term equal to dose squared. A p value for the test of association was based on the Wald test that the regression coefficients for all dose derived variables were simultaneously equal to**

zero. The CI and p values for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

***Results*: A logistic regression analysis performed using a quadratic polynomial in dose finds that the observed differences between the doses with respect to the odds of decreased spermidine is not greater than what might reasonably be expected when DFMO had no true effect (P = .0931). (Interpreting the slope parameters is difficult here.)** *(We could, of course, obtain predicted proportions/odds and 95% CI for the dose groups. Note that the fitted values for this model and for the model in part c are nearly identical, but that we do not have statistical significance here. This is because we are having to test two parameters here without any particular gain in the statistical precision. This leads to a loss of precision, thereby illustrating the advantages of "parsimony": using as few predictors as possible to model the true relationship. But we do not, of course, know the true relationship, so we have to make tradeoffs when we fere there might be nonlinearities.)*

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

***Instructions for grading*: This problem is worth 5 points. For this problem, I gave CI in a more abbreviated form that might be more acceptable for publications (where space is considered a premium)**

***Methods:* The effect of DFMO on the odds of decreased spermidine measurements after 12 months of treatment was analyzed in a logistic regression model of dose asbinary indicator that dose was greater than 0. A p value for the test of association was based on the Wald test that the regression coefficientsfor all the dose derived variable was equal to zero. The CI and p values for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.**

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

***Results*: An analysis comparing the placebo group to the combined groups receiving some dose of DFMO finds that the observed differences between the dose groups odds of decreased spermidine is not greater than what might reasonably be expected when DFMO had no true effect (P = .0631).** *(Interpreting the slope parameters is difficult here, because there is no good scientific reason to estimate the effect of DFMO across combined dose groups.We could of course have interpreted parameters as was done in problems 2 and 3. Had we obtained predicted odds ratios and 95% CI for the dose groups, they would have been the same for all doses higher than 0, with an estimated odds ratio of 2.36.)*

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

***Instructions for grading*: This problem is worth 5 points. For this problem, I gave CI in a more abbreviated form that might be more acceptable for publications (where space is considered a premium)**

***Methods:* The effect of DFMO on the odds of decreased spermidine measurements after 12 months of treatment was analyzed in a logistic regression model of dose as both as an indicator of dose above 0 and an untransformed linear term. A p value for the test of association was based on the Wald test that**

the regression coefficients for all dose derived variables were simultaneously equal to zero. The CI and p values for the individual regression parameters were not adjusted for the multiple comparisons inherent in such analyses that model dose using multiple covariates.

Subjects missing data were omitted only if they were missing data for a variable needed for the specific analysis. Such an analysis is most appropriate for data that is missing completely at random (MCAR), however the observed trend toward higher rates of missing data in higher dose groups would suggest that MCAR is not likely to be true.

***Results*: A logistic regression analysis finds that the observed differences between the dose groups' odds of decreased spermidine is not greater than what might reasonably be expected when DFMO had no true effect (P = .0765). From the modeling of dose as a linear continuous predictor and a threshold effect at dose 0, we do not have sufficient evidence to reject the hypothesis of a linear relationship among log odds of decreased spermidine across all dose levels (P= .623). The estimated linear trend across dose groups suggests the odds of decreased spermidine tends to be 15.9 times higher (95% CI unadjusted for multiple comparisons: 0.301 times as high to 833 times higher) for each 1,000 mg/m^2/day increase in dose when comparing doses above 0 (or 1.32 times higher (95% CI: 0.887 times as high to 1.96 times higher) for each 100 mg/m^2/day increase in dose fwhen comparing doses above 0).** *(Note that the fitted values for this model differed more from those in part c than did the quadratic model, even though both used two predictors to model dose. This model borrowed data less distantly to estimate the linear trend (the linear trend did not use the dose 0 group in its estimate), hence we would expect less power.)*

f. Model dose as three variables: a continuous linear predictor, a quadratic term, and a cubic term (a term equal to dose raised to the third power).

Instructions for grading: This problem is worth 5 points.

Ans: As noted in the answers to problem 2, this is of course the exact same model as the dummy variable model in part a, so the inference will be the exact same. This parameterization is much more difficult to interpret, as noted in problems 2 and 3

g. Provide a table of the fitted values for each dose group from the above models. Comment on the similarities / differences between those fitted values (and the descriptive statistics).

Instructions for grading: This problem is worth 5 points.

***Ans:* The following tables contain the fitted values for probabilities (Table 3), odds (Table 4), and log odds (Table 5) from each of the six models, as well as from sample descriptive statistics (I used Excel to convert the probabilities to odds and log odds). I note that the fitted proportions/odds from the dummy variables (Model A) and the cubic polynomial (Model F) each correspond exactly to the sample proportions/odds for each dose group. This correspondence between the cubic polynomial and the dummy variables is due to the fact that there were only four levels of dose sampled (and three is one less than four). The estimates of the log odds from Model B would lie exactly on a straight line. The estimates from Model C would very nearly lie on a straight line, because the sample log odds departures from a straight line are not particularly better fit by a quadratic. In Model D, the estimated proportions/odds are the same for all dose groups above 0, and the dose 0 group estimate corresponds exactly to the sample proportions/odds for that group. In Model E, 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 proportions/odds for that group.**

Table 4: Fitted probabilities from the six models.

Dose	Sample Probs	Model A	Model B	Model C	Model D	Model E	Model F
0	0.464	0.464	0.493	0.491	0.464	0.464	0.464
75	0.615	0.615	0.557	0.559	0.672	0.590	0.615
200	0.619	0.619	0.659	0.662	0.672	0.670	0.619
400	0.800	0.800	0.793	0.791	0.672	0.779	0.800

Table 5: Fitted odds from the six models.

Dose	Sample Probs	Model A	Model B	Model C	Model D	Model E	Model F
Dose	0.317	0.317	0.330	0.329	0.317	0.317	0.317
0	0.381	0.381	0.358	0.358	0.402	0.371	0.381
75	0.382	0.382	0.397	0.398	0.402	0.401	0.382
200	0.444	0.444	0.442	0.442	0.402	0.438	0.444

Table 6: Fitted log odds from the six models.

Dose	Sample Probs	Model A	Model B	Model C	Model D	Model E	Model F
Dose	-0.499	-0.499	-0.481	-0.482	-0.499	-0.499	-0.499
0	-0.419	-0.419	-0.446	-0.446	-0.396	-0.431	-0.419
75	-0.418	-0.418	-0.401	-0.400	-0.396	-0.397	-0.418
200	-0.352	-0.352	-0.354	-0.355	-0.396	-0.359	-0.352

5. Which of the above analyses would you prefer *a priori* to test for an effect of DFMO on mucosal levels of polyamines?

***Instructions for grading*: This problem is worth 5 points.**

Ans: I would generally prefer using the continuous spermidine levels, as there is no compelling scientific threshold, and the continuous measurements provide greater statistical power than would dichotomized data. I would also prefer adjusting for baseline, as that will tend to provide greater precision. An advantage to the model in part g of problem 2, is that it allows some flexibility in fitting dose response while maintaining some interpretability of parameters: In addition to testing for an effect by DFMO, I can assess evidence against linear relationships as well as whether there is any advantage in giving a dose above the lowest positive dose tested. In this analysis, I would conclude that the data are relatively well fit by a straight line relationship, and thus higher doses tend to provide greater suppression of polyamines. But the same effect could be obtained by using the linear continuous model as the primary analysis, and specifying a secondary analysis to look for nonlinear associations in the event a statistically significant association was found in the primary analysis.