Biost 517: Applied Biostatistics I
Emerson, Fall 2011

Homework #1 Key
September 28, 2011

Written problems: To be handed in at the beginning of class on Wednesday, October 5, 2011 (See the end of this handout for the Data Analysis problem to be discussed in Discussion Section October 5, 7, 10, 11.)

> *On this (as all homeworks) unedited Stata 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.)*

Questions for Biost 514 and Biost 517:

The class web pages contains a description of a dataset regarding a clinical trial of DFMO in the prevention of colon polyps (dfmo.doc and dfmowide.txt). For this homework, we are only interested in the 7 variables measuring patient **age**, sex (**female**), dose of DFMO **(dose),** and serum spermidine at baseline (randomization) (**spd0**), after 6 months of treatment with study drug (**spd6**), after 12 months of treatment with study drug (**spd12**), and 3 months after discontinuation of treatment with study drug (15 months after randomization) (**spd15**). Where relevant, provide descriptive statistics for each of these variables in the entire sample, as well as within groups defined by dose. The descriptive statistics should provide information on the number of valid observations, the number of missing observations, the mean, the standard deviation, the minimum, 25^{th} percentile, median, 50^{th} percentile, 75^{th} percentile, and the maximum, where such statistics are of scientific interest.

Comment on what the results of you analysis might say about the differences between dose groups in the sample with respect to their pre-randomization characteristics (age, sex, and baseline spermidine levels), the propensity for data to be missing at later times, and the values of the measured spermidine levels while on study drug and after discontinuation of study drug. (Just look at the numbers and describe trends in the typical values as well as in the variability of the measurements—there is no need to do any formal statistical hypothesis testing). Also compare the patients' pre-randomization characteristics across groups defined by whether later measurements of spermidine are missing or not. Can you speculate on possible reasons for the trends you note?

One hundred fourteen (114) predominantly male subjects (85.1% male, 14.9% female) between the ages of 45 and 81 were randomized to daily DFMO doses of 0 g/m^2 (n= 32), 0.075 g/m^2 (n=29), 0.2 g/m^2 (n=25), and 0.4 g/m^2 (n=28). Table 1 contains descriptive statistics for each of the variables within dose strata. The dose groups had similar mean age (one subject was missing age information in the 0.075 g/m^2 group) and mean spermidine levels at baseline, though the 0.2 g/m^2 group had no females.

Serum spermidine levels that were to be made 6 months post randomization were missing for 8 subjects, all but one of whom were also missing measurements at 12 months. In addition, 12 subjects who had spermidine levels at 6 months were missing measurements at 12 months. A total of 22 subjects were missing the spermidine measurements that were to be made 3 months after the one year period of treatment.

Among the subjects with available measurements, the groups receiving DFMO tended toward a lower average than the placebo group at 6 months, with a clearer tendency toward lower average (and median) with increasing dose after 12 months of treatment. Similar trends are observed for the minma and 25th percentiles, and to a lesser extent the 75th percentiles and the maxima. The standard deviations do not show a clear consistent tendency to be lower for the groups with lower means, though there is some tendency in that regard.

Table 1: Descriptive statistics by dose and for the entire sample for selected variables. Statistics presented include the number of subjects randomized to each dose group (N), the number of subjects missing data for the respective measurements (msng), the mean, standard deviation (SD), the median (Mdn), the 25th and 75th percentiles (interquartile range or IQR), minimum (Min), and maximum (MAX)

		N (msng)	Mean (SD)	Mdn (IQR)	(Min, Max)
Female (%)	Dose 0	32 (0)	18.8%		
	Dose 0.075	29 (0)	17.2%		
	Dose 0.200	25 (0)	0.0%		
	Dose 0.400	28 (0)	21.4%		
	TOTAL	114 (0)	14.9%		
Age (years)	Dose 0	32 (0)	65.9 (8.51)	66.4 (61.1, 73.5)	(45.5, 77.2)
	Dose 0.075	28 (1)	61.3 (7.69)	61.4 (56.5, 65.8)	(47.8, 76.9)
	Dose 0.200	25 (0)	62.8 (8.28)	63.7 (59.2, 68.3)	(45.4, 77.6)
	Dose 0.400	28 (0)	63.9 (7.81)	65.0 (60.1, 69.4)	(48.5, 81.0)
	TOTAL	113 (1)	63.6 (8.16)	64.3 (59.2, 69.6)	(45.4, 81.0)
Baseline Spermidine	Dose 0	32 (0)	3.26 (1.45)	2.93 (2.08, 4.17)	(1.40, 7.05)
	Dose 0.075	29 (0)	3.47 (1.55)	2.91 (2.29, 4.62)	(1.51, 7.02)
	Dose 0.200	25 (0)	3.35 (1.33)	2.92 (2.42, 4.01)	(1.70, 6.22)
	Dose 0.400	28 (0)	3.56 (1.88)	3.08 (2.15, 4.70)	(0.66, 7.60)
	TOTAL	114 (0)	3.41 (1.55)	2.96 (2.27, 4.29)	(0.66, 7.60)
Month 6 Spermidine	Dose 0	30 (2)	3.37 (1.53)	3.02 (2.20, 4.28)	(1.51, 6.91)
	Dose 0.075	28 (1)	2.64 (0.89)	2.46 (1.94, 3.05)	(1.39, 5.12)
	Dose 0.200	23 (2)	2.58 (1.64)	1.85 (1.49, 3.90)	(1.07, 7.84)
	Dose 0.400	25 (3)	2.68 (1.43)	2.07 (1.67, 2.90)	(1.06, 6.34)
	TOTAL	106 (8)	2.84 (1.41)	2.52 (1.80, 3.47)	(1.06, 7.84)
Month 12 Spermidine	Dose 0	28 (4)	3.26 (1.31)	2.82 (2.26, 4.27)	(1.01, 5.91)
	Dose 0.075	26 (3)	2.92 (0.99)	2.86 (2.13, 3.64)	(1.35, 4.92)
	Dose 0.200	21 (4)	2.71 (1.40)	2.51 (1.76, 3.78)	(0.29, 6.45)
	Dose 0.400	20 (8)	1.95 (0.80)	1.93 (1.48, 2.46)	(0.00, 3.42)
	TOTAL	95 (19)	2.77 (1.23)	2.55 (1.99, 3.59)	(0.00, 6.45)
Month 15 Spermidine	Dose 0	27 (5)	2.69 (0.93)	2.45 (2.04, 3.37)	(1.25, 4.62)
	Dose 0.075	26 (3)	2.95 (0.99)	2.98 (2.40, 3.63)	(0.00, 4.83)
	Dose 0.200	21 (4)	2.98 (0.90)	2.81 (2.20, 3.71)	(1.81, 4.81)
	Dose 0.400	18 (10)	2.70 (0.87)	2.69 (2.36, 3.02)	(1.29, 4.47)
	TOTAL	92 (22)	2.83 (0.92)	2.72 (2.20, 3.44)	(0.00, 4.83)

The above trends toward lower values of spermidine in the higher dose groups were necessarily based on the subjects for whom measurements were available. To the extent that these missing measurements reflect subjects dropping out of the trial, we might be concerned that the missing data reflect adverse experiences, which might in turn be caused

by the same mechanism that would cause lower spermidine values. If this is the case, then the above trends in spermidine on treatment might be an understatement of what would be observed if all subjects had completed the trial. Some evidence that might make us worry about such a mechanism of "informative missingness" is apparent in the higher rates of missingness with increasing dose.

Tables 2 and 3 further explore the patterns of missing data according to the available measurements. Table 2 compares those subjects in each dose group who do and do not have spermidine data available at 6 months, while Table 3 makes similar comparisons for the missingness at 12 months.

Table 2: Descriptive statistics comparing subjects who are and who are not missing spermidine measurements after 6 months of treatment. Statistics presented include the number of subjects with available data for the respective measurements among those randomized to each dose group (n), the mean, standard deviation (SD), minimum (Min), and maximum (MAX)

	Not Missing 6 Month Spd	Missing 6 Month Spd
	Mean (SD; Min – Max; n)	Mean (SD; Min – Max; n)
Dose 0		
Female (%)	20.0%(n=30)	0.0%(n=2)
Age (years)	65.8 (8.64; 45.5 - 77.2; n=30)	66.3 (8.90; 60.1 - 72.6; n=2)
Baseline Spermidine	3.28 (1.50; 1.40 - 7.05; n=30)	3.00 (0.53; 2.62 - 3.37; n=2)
Month 6 Spermidine	3.37 (1.53; 1.51 - 6.91; n=30)	NA (NA; NA - NA; n=0)
Month 12 Spermidine	3.20 (1.30; 1.01 - 5.91; n=27)	4.81 (NA; 4.81 - 4.81; n=1)
Month 15 Spermidine	2.66 (0.94; 1.25 - 4.62; n=26)	3.30 (NA; 3.30 - 3.30; n=1)
Dose 0.075		
Female (%)	17.9%(n=28)	0.0%(n=1)
Age (years)	61.2 (7.82; 47.8 - 76.9; n=27)	64.3 (NA; 64.3 - 64.3; n=1)
Baseline Spermidine	3.36 (1.45; 1.51 - 7.02; n=28)	6.62 (NA; 6.62 - 6.62; n=1)
Month 6 Spermidine	2.64 (0.89; 1.39 - 5.12; n=28)	NA (NA; NA - NA; n=0)
Month 12 Spermidine	2.92 (0.99; 1.35 - 4.92; n=26)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	2.95 (0.99; 0.00 - 4.83; n=26)	NA (NA; NA - NA; n=0)
Dose 0.2		
Female (%)	0.0%(n=23)	0.0%(n=2)
Age (years)	62.5 (8.56; 45.4 - 77.6; n=23)	66.4 (2.72; 64.5 - 68.3; n=2)
Baseline Spermidine	3.35 (1.38; 1.70 - 6.22; n=23)	3.35 (0.61; 2.92 - 3.78; n=2)
Month 6 Spermidine	2.58 (1.64; 1.07 - 7.84; n=23)	NA (NA; NA - NA; n=0)
Month 12 Spermidine	2.71 (1.40; 0.29 - 6.45; n=21)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	3.01 (0.91; 1.81 - 4.81; n=20)	2.34 (NA; 2.34 - 2.34; n=1)
Dose 0.4		
Female (%)	20.0%(n=25)	33.3%(n=3)
Age (years)	64.5 (8.06; 48.5 - 81.0; n=25)	59.0 (2.00; 56.8 - 60.8; n=3)
Baseline Spermidine	3.70 (1.95; 0.66 - 7.60; n=25)	2.42 (0.42; 2.09 - 2.90; n=3)
Month 6 Spermidine	2.68 (1.43; 1.06 - 6.34; n=25)	NA (NA; NA - NA; n=0)
Month 12 Spermidine	1.95 (0.80; 0.00 - 3.42; n=20)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	2.70 (0.87; 1.29 - 4.47; n=18)	NA (NA; NA - NA; n=0)

In Table 2 there is no clear trend toward particular patterns of missingness according to age, sex or the baseline spermidine levels, though we are hampered by low sample size among the subjects with missing 6 month measurements.

In Table 3, however, there is a bit more data. However, even in this table there is no consistent trend toward lower baseline or 6 month spermidine measurements among those who dropped out. Of course, whether or not such trends were evident, we are actually more interested in what the trends would have been evident among the 12 month measurements in those people who dropped out of the study early, not all of whom may have dropped out for the same reason. But there is nothing in our data that will tell us about any such trends. Hence, we must report the existence of the missing data and any trends we have noted that might predict who would be missing data. Then we just discuss possible impact and perhaps perform some sensitivity analyses.

Table 3: Descriptive statistics comparing subjects who are and who are not missing spermidine measurements after 6 months of treatment. Statistics presented include the number of subjects with available data for the respective measurements among those randomized to each dose group (n), the mean, standard deviation (SD), minimum (Min), and maximum (MAX)

	Not Missing 6 Month Spd	Missing 6 Month Spd
	Mean (SD; Min – Max; n)	Mean (SD; Min – Max; n)
Dose 0		
Female (%)	21.4%(n=28)	0.0%(n=4)
Age (years)	65.1 (8.45; 45.5 - 77.2; n=28)	71.4 (7.61; 60.1 - 76.0; n=4)
Baseline Spermidine	3.30 (1.46; 1.40 - 7.05; n=28)	3.02 (1.57; 1.68 - 5.29; n=4)
Month 6 Spermidine	3.26 (1.45; 1.51 - 6.63; n=27)	4.36 (2.28; 2.53 - 6.91; n=3)
Month 12 Spermidine	3.26 (1.31; 1.01 - 5.91; n=28)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	2.65 (0.94; 1.25 - 4.62; n=26)	3.56 (.; 3.56 - 3.56; n=1)
Dose 0.075		
Female (%)	19.2%(n=26)	0.0%(n=3)
Age (years)	61.5 (7.72; 47.8 - 76.9; n=25)	60.1 (8.98; 49.8 - 66.1; n=3)
Baseline Spermidine	3.33 (1.46; 1.51 - 7.02; n=26)	4.71 (2.12; 2.43 - 6.62; n=3)
Month 6 Spermidine	2.68 (0.91; 1.39 - 5.12; n=26)	2.16 (0.52; 1.79 - 2.53; n=2)
Month 12 Spermidine	2.92 (0.99; 1.35 - 4.92; n=26)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	2.95 (0.99; 0.00 - 4.83; n=26)	NA (NA; NA - NA; n=0)
Dose 0.2		
Female (%)	0.0%(n=21)	0.0%(n=4)
Age (years)	63.2 (8.38; 45.4 - 77.6; n=21)	61.1 (8.68; 48.5 - 68.3; n=4)
Baseline Spermidine	3.28 (1.31; 1.70 - 6.22; n=21)	3.73 (1.55; 2.33 - 5.87; n=4)
Month 6 Spermidine	2.35 (1.23; 1.07 - 5.26; n=21)	5.00 (4.02; 2.16 - 7.84; n=2)
Month 12 Spermidine	2.71 (1.40; 0.29 - 6.45; n=21)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	2.97 (0.92; 1.81 - 4.81; n=19)	3.02 (0.97; 2.34 - 3.71; n=2)
Dose 0.4		
Female (%)	20.0%(n=20)	25.0%(n=8)
Age (years)	64.5 (7.90; 48.9 - 81.0; n=20)	62.2 (7.84; 48.5 - 73.1; n=8)
Baseline Spermidine	3.71 (1.89; 0.66 - 7.21; n=20)	3.21 (1.95; 1.66 - 7.60; n=8)
Month 6 Spermidine	2.78 (1.57; 1.06 - 6.34; n=20)	2.26 (0.42; 1.89 - 2.81; n=5)
Month 12 Spermidine	1.95 (0.80; 0.00 - 3.42; n=20)	NA (NA; NA - NA; n=0)
Month 15 Spermidine	2.60 (0.77; 1.29 - 4.43; n=17)	4.47 (.; 4.47 - 4.47; n=1)

(Note that in Table 1, I grouped the rows by variable first in order to facilitate the comparisons across dose groups down the columns. But in Tables 2 and 3, I was going to make comparisons across the rows, and I thought it would be more useful to look at all variables for a given dose group: The placebo group would not be expected to have drop out due to toxicity.)