POWER CALCULATIONS for the Traditional (Radioactive) LLNA

Introduction

During their review of the LLNA: BrdU-ELISA test method, some members of the ICCVAM LLNA expert peer review panel requested information on statistical power vs. number of animals used for this assay. They wanted know how many animals would be adequate for detecting the associated threshold stimulation index for a positive response (e.g., SI > 3). Dr. Joe Haseman (a consultant biostatistician for ILS, Inc., the NICEATM support contractor) subsequently conducted power calculations for each nonradiolabeled method as well as the traditional (i.e., radiolabeled) LLNA to determine the number of animals needed to demonstrate statistical significance between control and treatment groups. The results of this analysis (http://iccvam.niehs.nih.gov/restrict/llnapanel/docs/HasemanReport14Feb08FD.pdf) provided the adequate number of animals to achieve an SI ≥ 3 (and a number of alternative SI thresholds) for each method.

An assumption made in the original analysis was that any “treatment effect” produced would have essentially the same SD (on a log-transformed scale) as the control data (i.e. that the treatment will change only the mean response and not the variability). Therefore, a follow-up analysis (described herein) was conducted to test this assumption. Dr. Haseman was provided vehicle control and corresponding treatment group data (disintegrations per minute) from five to seven different traditional LLNA studies from three different laboratories in order to establish the variability among control and treated animals. The power calculations were based on a one-sided $p<0.05$ Student’s t test applied to the log-transformed data (just as in the previous power calculations). Dr. Haseman's analysis follows:

27 **Table 1** **Data for Summary Statistics**

Experiment	Test Substance	Original Scale		Log Scale		Laboratory
		Mean	SD	Mean	SD	
1A	Veh	443.4	233.86	5.976	0.5531	1
	Oxy	2145.0	621.21	7.625	0.3671	1
1B	Veh	410.2	100.30	5.994	0.2421	1
	Dino-4	8073.6	3155.55	8.942	0.3603	1
	Dino-10	14235.5	3725.77	9.538	0.2649	1
1C	Veh	462.2	172.26	6.078	0.3874	1
	Q-C-33	2949.8	551.49	7.974	0.1998	1
	C-C1100	6160.8	1242.31	8.709	0.2104	1
1D	Veh	397.8	92.64	5.968	0.2092	1
	Tri-33	3181.0	697.81	8.040	0.2616	1
	Tri-100	7941.0	2912.59	8.933	0.3335	1
1E	Veh	466.8	154.26	6.104	0.3262	1
	HCA	1906.6	376.37	7.538	0.1938	1
1F	Veh	352.6	118.53	5.826	0.3211	1
	Form	1589.2	878.33	7.223	0.6381	1
1G	Veh	333.0	167.74	5.702	0.5336	1
	KCr2	1278.8	779.84	6.969	0.7357	1
2A	Veh	487.8	164.01	6.142	0.3649	2
	KCr2-0.1	2194.4	1583.04	7.461	0.7852	2
	KCr2-0.5	6188.2	5764.00	8.658	0.8541	2
2B	Veh	729.2	314.07	6.496	0.5214	2
	Form-5	2739.9	902.44	7.866	0.3636	2
	Form-10	7724.2	1226.35	8.941	0.1719	2
2C	Veh	586.6	279.96	6.296	0.4252	2
	HCA-10	2720.8	523.28	7.894	0.1939	2

Experiment	**Test Substance**	**Original Scale**		**Log Scale**		**Laboratory**
		Mean	**SD**	**Mean**	**SD**	
	HCA-30	10548.9	4527.28	9.193	0.4408	2
2D	Veh	618.4	103.27	6.416	0.1644	2
	Oxy	3033.3	438.87	8.010	0.1379	2
2E	Veh	487.4	80.26	6.178	0.1585	2
	Dino-4	12558.2	2403.34	9.421	0.2161	2
	Dino-21	7007.8	(n=1)	8.855	(n=1)	2
2F	Veh	304.1	208.62	5.402	0.9937	2
	Tri-7	1811.5	915.09	7.328	0.7646	2
	Tri-33	9136.3	3307.04	9.054	0.4322	2
	Tri-100	22884.4	3115.72	10.031	0.1372	2
2G	Veh	309.4	110.19	5.686	0.3512	2
	Q-C-33	3298.6	1079.53	8.054	0.3521	2
	Q-C-100	6275.1	2180.15	8.688	0.3924	2
3A	Veh	330.5	145.26	5.706	0.5184	3
	HCA	3394.9	851.67	8.104	0.1147	3
3B	Veh	288.5	229.15	5.338	1.0113	3
	Form	1449.4	947.98	7.086	0.7575	3
	HCA	1017.2	673.94	6.752	0.6697	3
3C	Veh	152.5	31.78	5.008	0.2275	3
	Form	1932.5	16334.1	7.205	1.0508	3
	HCA	1103.1	451.06	6.939	0.4102	3
3D	Veh	296.2	126.07	5.604	0.4820	3
	HCA	2219.9	914.30	7.635	0.4271	3
3E	Veh	215.3	149.44	5.104	0.9148	3
	Form	2519.7	224.81	7.829	0.0899	3
	HCA	2055.6	1031.49	7.384	0.9743	3

Abbreviations: SD = Standard deviation; Veh = Vehicle

Table 2 Summary Statistics

Laboratory	Test Substance	Original Scale Averages		Log-transformed Scale Averages	
		Mean	SD	Mean	SD
1	Veh	409.4	148.51	5.950	0.3675
	Rx	4946.1	1494.10	8.149	0.3565
2	Veh	503.3	180.05	6.088	0.4256
	Rx	7008.7	2151.24	8.532	0.4032
3	Veh	256.6	136.34	5.352	0.6308
	Rx	1961.5	828.67	7.367	0.5618
All 3	Veh	403.8	156.93	5.843	0.4582
	Rx	5102.3	1597.95	8.121	0.4291

Abbreviations: Rx = Treatment groups; SD = Standard deviation; Veh = Vehicle

Discussion and Conclusion

Overall, these are not subtle effects being detected. Although there are some ratios close to 3, the average ratio seen in these data is approximately 12 to 13, with some as high as 75. All of these effects detected by the ratio rule would also have been detected as significant ($p<0.05$) by a simple Student's t test applied to the logged data.

For each of the three labs (and overall), the SD's for the treated groups (on a log scale) track very closely those for the vehicle controls, confirming the assumption that we made in the power calculations that although the chemical may increase the mean response, it does not increase the (logged) variability. Frankly, I was surprised to see how closely the average (logged) SD's agreed for the control and treated groups (see table above). Lab 3 was more variable than Labs 1 and 2 for the treated groups, just as they were for the controls.

One potential problem with the Ratio Rule is that it ignores the underlying variability in the data, and I had speculated that the responses for certain chemicals may be estimated more accurately than for others. To check this out, I did a simple one-way ANOVA on the logged SD's. The overall difference among chemicals was marginally significant ($p=0.064$). Looking at the data, it does indeed appear that some chemicals (e.g., KCr2 with SD's of 0.7357, 0.7852, and 0.8541) have more variability associated with the estimation process than do other chemicals (e.g., oxyfluorfer [0.3671, 0.1379], dinocap [0.3603, 0.2649, 0.2161, and Q-C [0.1998, 0.2104, 0.3521, 0.3924]). This difference in variability among chemicals would have been statistically significant, were it not for the highly variable results observed for FORM [ranging from 0.0899 to 1.0508], which itself may be of interest.

These latest calculations support the earlier power calculations by validating one of the key assumptions made. It also provides further evidence that a formal statistical test may be superior to the “Ratio Rule” for evaluating the data, at least from the standpoint of power.

Joe Haseman
2-23-08