Statistical evidence favoring Poisson variation in *Rb9 trans* and *Rb9 cis*

Michael A. Newton

Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53792

The possibility of Poisson variation of tumor multiplicities in each of the four groups [*Rb9*-free (+/+), *Rb9 trans*, *Rb9 cis*, *Rb9/Rb9*] is assessed in two steps. First, the null hypothesis of Poisson variation is tested by exact conditional inference[A]. Conditionally on the total tumor count in each line, the numbers of tumors per animal has a multinomial distribution. The null distribution of the sample coefficient of dispersion (CD) test statistic is obtained by Monte Carlo simulation, and a small group-specific p-value indicates evidence against Poisson variation (Table A).

Table A • p-values based on 10^4-1 simulated data sets			
+/+	*Rb9 trans*	*Rb9 cis*	*Rb9/Rb9*
0.0001	0.37	0.38	0.15

These results indicate that the *Rb9*-free line is not consistent with Poisson variation but the *Rb9*-carrying lines are consistent with such variation. Calculations were done using R (www.r-project.org). As extra-Poisson variation is often observed with tumor multiplicity data, these results do not in themselves provide positive evidence favoring Poisson variation. Such evidence is assessed in a second calculation.

The two-parameter Negative Binomial distribution describes variation in tumor multiplicity data and includes the Poisson distribution as a special case[B]. Specifically, the probability mass assigned to x tumors is

$$p(x|\lambda,\kappa) = c\, \lambda^x\, \Gamma(x+1/\kappa)/\, \Gamma(x+1)$$

where c is a normalizing constant, λ is the expected tumor count, $\Gamma(\ )$ is the Gamma function, and $\kappa \geq 0$ characterizes over-dispersion relative to the Poisson in that the variance is $\lambda\,(1+\kappa\,\lambda)$. The Poisson distribution arises in the limit when $\kappa = 0$.

To evaluate the evidence supporting $\kappa = 0$ in the *Rb9* groups, we note that each of the four groups may have its own value of λ and its own value of κ. Different hypotheses may be posed regarding the equality of the λ's among some of the groups, equality of the κ 's among some groups, and whether or not $\kappa = 0$ for any groups. In total, there are 780 distinct hypotheses (15 patterns among the λ's and, for each, 52 patterns among the κ 's). We estimated by the method of maximum likelihood the free parameters under each hypothesis, thus obtaining a maximized log-likelihood (L) and a number of parameters (p).   The best hypothesis maximizes L subject to a complexity penalty. Using either the Akaike information criterion (AIC)[C] or the Bayes information criterion (BIC)[D], the best model entails Poisson variation for both *Rb9 trans* and *Rb9 cis*, equality of means for these groups, different means but common shape between the +/+ case and the *Rb9/Rb9* case. Table B indicates the position of several hypotheses. A traditional hypothesis testing approach would not evaluate the evidence in favor of Poisson variation, which is supported by the global analysis based on BIC and AIC.

Table B • Statistical evidence favoring Poisson variation in *Rb9 trans* and *Rb9 cis*

Pattern of means				Pattern of κ							
+/+	*Rb9 trans*	*Rb9 cis*	*Rb9/Rb9*	+/+	*Rb9 trans*	*Rb9 cis*	*Rb9/Rb9*	L	p	BIC(r)	AIC(r)
1	2	2	3	1	0	0	1	-211.62	4	1	1
1	2	2	3	1	0	0	0	-211.68	4	2	2
1	2	2	3	1	0	1	1	-212.43	4	3	6
1	2	3	4	1	0	0	1	-210.71	5	8	3
1	2	2	3	1	1	0	1	-212.46	4	4	7
1	2	3	4	1	0	0	0	-210.76	5	9	4
1	2	3	4	1	1	1	1	-212.55	5	40	52
1	2	3	4	1	2	3	4	-210.50	8	160	91

In columns 1-4 equality of the label corresponds to equality of the means. In columns 5-8 equality of the label corresponds to equality of κ, and, further, the label 0 indicates Poisson variation in that group. L is the maximized log-likelihood; p is the number of free parameters; BIC(r) and AIC(r) are, respectively, the rank (out of 780) of the hypothesis in terms of the BIC and the AIC. Models were fit using Splus 3.4 (Mathsoft).

References

A. Agresti, A. Categorical data analysis. Wiley, New York, 1990.

B. Drinkwater, N.R., Klotz, J.H. Statistical methods for the analysis of tumor multiplicity data. *Cancer Research* **41**, 113-119 (1981).

C. Akaike, H. Information measures and model selection. *Bulletin of the International Statistical Institute* **50**, 277-290 (1983).

D. Schwartz, G. Estimating the dimension of a model. *The Annals of Statistics* **6**, 461-464 (1978).