

STATISTICAL ANALYTICAL METHODS FOR COMPARING THE INCIDENCE OF TUMORS TO THE HISTORICAL CONTROL DATA

Katsumi KOBAYASHI and Hiroyuki INOUE

*Biosafety Research Center, Foods, Drugs and Pesticides,
(An-Pyo Center)*

582-2 Arahama, Shioshinden, Fukude-cho, Iwata-gun, Shizuoka 437-12, Japan

(Received September 10, 1993 ; accepted October 19, 1993)

ABSTRACT— Statistical analysis for comparing the incidence of tumors in treated groups to historical control data is not generally performed. In the present study, a number of data exhibiting the lowest, low, moderate and highest incidence of tumors from long-term rodent bioassay studies has been compared with the historical control data. In studies exhibiting the lowest incidence (less than a few percent) of tumors, the Kastenbaum and Bowman test was found to be relevant since it takes into account the sample size of both the historical control data base and each treated group in the study. In studies where a wider range of tumor incidence was exhibited, a statistical method which employs a rejection limits based on the range of incidence in the historical data is recommended. When malignant tumors are evident in treated groups, no matter how low the incidence, they should be analyzed statistically and compared with the incidence in historical control data as well as those in the concurrent control group. Statistical analytical comparisons of study results to historical control data may contribute to more meaningful evaluations in carcinogenicity studies by eliminating possible false positive or false negative results.

KEY WORDS : Statistical analyses, Historical control data, Tumors.

INTRODUCTION

Statistical methods for analyzing the occurrence of tumors are generally employed to compare incidence data between control and treated groups in oncogenicity studies. Categorical analyses including the Chi-square test (Colquhoun, 1982) and the Fisher exact test (Shayne *et al.*, 1982) have been used in Japan. Currently, the National Toxicology Program (NTP, 1985 and 1989) in U. S. A. is using the adjusted occurrence rates obtained by the Kaplan-Meier test, the Cochran-Armitage trend test, Fisher exact test

and life table analysis, for this purpose. The Peto-test (Peto *et al.*, 1980) is frequently used in Europe.

Neoplastic finding showings statistically significant differences in incidence between the control and treated groups is occasionally observed. In addition, there may be no significant differences between the control and treated groups due to a low incidence of lesions in the treated groups as compared to control animals. In such cases, it would be more meaningful to compare the incidence of these tumors to that of historical control data. A trend method analysis for comparing studied control and treated groups with historical control data has been reported in the literature (Krewski *et al.*, 1988), however there are no

Correspondence : Katsumi KOBAYASHI at the above address.

standard statistical method for tumorigenesis assay (Haseman, 1984 and 1992). Main problem is the selection of appropriate analytical method arises when the incidence of malignant tumors in both the treatment group and historical control data is less than a few percent. Three statistical methods having different criteria may be employed.

- 1) A ninety-five % confidence interval using historical data (Yoshida, 1980) in which any incidence outside this limit suggests a significant difference.
- 2) Comparison with historical data based on the null hypothesis or categoric test as a second approach.
- 3) Normal ranges can be established by comparisons to rejection limits and variation in historical data. Confidence intervals of historical data are then determined from the standard deviation and t-distribution of parametric data.

The purpose of this study is to establish the most appropriate statistical method for comparing four types of incidence, a lowest (less than 1%), low (around 10%), moderate (around 30%), and high incidence (more than 80%), with corresponding historical data.

MATERIALS AND METHODS

Estimation of tumor occurrence rate based on historical control data : The historical control data on the occurrence of tumors were obtained from 11 oncogenicity studies using F344 male rats (Charles River Japan Inc.) conducted at the An-Pyo Center from 1980 to 1986 (Table 1). Analyzed tumors were : hepatocellular carcinoma which occurred in 2 of the 11 studies, large granular cell leukemia (L. G. L.) of the spleen (8% incidence), pituitary-adenoma (30% inci-

dence), and interstitial cell tumors of the testes (more than 80% incidence). The 95% confidence interval of the historical data was determined by 3 models (Yoshida, 1980) as follows :

- a. Approximation method applying normal distribution

$$p \pm \{Z (\alpha/2) \cdot \sqrt{pg/n + 1/2n}\} \dots\dots \text{Equation 1}$$

This equation is generally applied when the No. of test animals is greater than 100 in number. The 95, 99 and 99.9% confidence intervals were determined by using the areas of a normal curve table at 2.5, 0.5 and 0.05% levels.

- b. Exact estimation method using binomial and F distributions

$$pU = f_1 F (f_1, f_2 ; \alpha/2) / [f_1 F (f_1, f_2 ; \alpha/2) + F_2] \dots\dots \text{Equation 2}$$

$$pU ; F_1 (n_1) = 2 (n+1),$$

$$F_2 (n_2) = 2 (n-k)$$

$$pL = f_2 / [f_1 F (f_1, f_2 ; \alpha/2) + f_2] \dots\dots \text{Equation 3}$$

$$pL ; f_1 = 2 (n-k+1), F_2 = 2k$$

The upper and lower limits are determined from these equations.

- c. Approximation methods using poisson and Chi-squared distributions

$$pU = 1/2n \cdot X^2 (fU, \alpha/2) \dots\dots \text{Equation 4}$$

$$fU = 3 (r+1)$$

$$pL = 1/2n \cdot X^2 (fL, 1-\alpha/2) \dots\dots \text{Equation 5}$$

$$fL = 2r$$

The poisson distribution may employ a Chi-squared distribution when the incidence (p) of tumor occurrence is small or large. However, the sample number must be greater than 50. The upper and lower limits are determined from equations 4 and 5.

Using null hypothesis and categoric analysis :

The comparison of historical data in Table 1 and findings in Table 2 were analyzed for the incidence of hepatocellular carcinoma and the L. G.

Table 1. Historical control data of incidence of tumors for 11 oncogenicity feeding studies using F344 male rats for 104 weeks.

Organ	Finding	Total rates, (11 studies) (%)	Number of occurrence / 50 rats on each study											
			A	B	C	D	E	F	G	H	I	J	K	Mean
Liver	Hepatocellular carcinoma	3/550 (0.5)	0	0	0	0	0	1	2	0	0	0	0	0.3/50
Spleen	Leukemia/malignant lymphoma	41/550 (7.5)	8	2	0	0	4	5	5	3	1	7	6	4/50
Pituitary	Adenoma	160/550 (29.1)	12	13	9	25	9	23	21	10	11	10	17	15/50
Testis	Interstitial cell tumor	465/550 (84.5)	38	44	43	36	41	41	45	41	43	46	47	42/50

L. of the spleen by 5 statistical models as follows : These findings were analyzed between control and treated groups by the Fisher exact and Peto-test. Same statistically significant differences existed in the incidence of two tumor types in these analyses.

a. Normal distribution test

$$Z_c = \frac{|p-P| - 1/2n}{\sqrt{pg/n}} \quad \dots\dots \text{Equation 6}$$

Comparing the value from the normal distribution table.

b. F distribution test

The null hypothesis, $p = P$, is abandoned if value $F(f_1, f_2; \alpha)$ is higher than that of the F-table value.

$$F = f_2P/f_1(1-P) \quad \dots\dots \text{Equation 7}$$

$$f_1 = 2(\gamma + 1), f_2(n - \gamma)$$

c. Chi distribution test

The incidence (p) of occurrence in historical data is small and a poisson distribution can be applied. The value of $2nP$ is calculated from equations 4, and 5 and the resulting value is compared with the Chi-table.

$p < P$, D. F. (degree of freedom) = $2(\gamma + 1)$, level of significance $\alpha = 5\%$

d. Categorical Chi-squared test

$$X^2 = \sum [(Observed - Expected)^2 / Expected] \quad \dots\dots \text{Equation 8}$$

e. Kastenbaum and Bowman test (Kastenbaum and Bowman, 1966)

This test, reported by Kastenbaum and Bowman, is used for analyzing the micronucleus assay data as compared to historical data. The incidence of microcleated polychromatic erythrocytes in non-treated mice is very low, around 0.21%, at the An-Pyo Center.

$$\sum_{r=X_1}^{X_1+X_2} \binom{X_1+X_2}{r} p^r(1-P)^{X_1+X_2-r} \leq \alpha \quad \dots\dots \text{Equation 9}$$

Analysis using distribution of historical data :

a. Application of rejection limits

Masuyama (Shibata, 1964) recommended that the normal range would be determined using the range of incidence in historical data shown in Table 1. This model is used to establish rejection limits for outliers in parametric data.

$$\text{Mean} \pm \text{S. D.} \cdot \sqrt{n+1/n \cdot t_{(n-1)}^{0.05, 0.01 \text{ or } 0.001}} \quad \dots\dots \text{Equation 10}$$

RESULTS

The estimation of occurrence in historical data :

a. The confidence intervals were determined by an approximation method using normal distribution as shown in Table 3. Hepatocellular carcinoma appearing at the lowest incidence showed a significant difference of the 5% level (confidence interval : 1.1-0.0%) when one in 50 animals were observed, compared to historical data (3/550 = 0.5%). The L. G. L. of the spleen showed a significant difference at the 5% (confidence interval : 9.7-5.3%) and 1% (confidence interval : 10.4-4.6%) where the incidence were 5 and 6 in 50 animals, respectively, were observed compared to historical data (41/550 = 7.5%). Pituitary adenoma also showed a significant difference at the 5% (confidence interval : 32.9-25.3%) and 1% (confidence interval : 34.1-24.1%) where the L. G. L. was evident in 17 and 18 of 50 animals were observed compared to historical data (160/550 = 29.1%). Interstitial cell tumors of the testes showed a significant difference at the 5% (confidence interval : 87.5-81.5%) and 1% (confidence interval : 88.5-80.5%) where incidence of 44 and 45 in 50 animals were observed compared to historical data (465/550 = 84.5%). Since 5, 1 and 0.1% levels of the significance reveal fixed values of 1.96, 2.57 and 3.29 as μ values respectively, calcu-

Table 2. Results of categoric test of lowest incidence findings of hepatocellular carcinoma and low incidence of leukemia/malignant lymphoma in an oncogenicity study using F344 male rats.

Organ	Finding	Incidences		Fisher exact test	Peto test
		Control	Treatment		
Liver	Hepatocellular carcinoma	0/50	2/50	P > 0.05	P > 0.05
Spleen	Leukemia/malignant lymphoma	0/50	5/50	P < 0.05	P < 0.05

lating the range of confidence intervals of this model is simple. A very narrow confidence interval is required in this approximation method. Even the lowest incidence, such as the occurrence of one case of hepatocellular carcinoma can reveal significance.

b. Exact estimation methods using binomial and F distribution: It is impossible to determine the confidence intervals by this method when the sample size of the historical data is greater than 120 because f_1 and f_2 are approaching 1.

c. Approximation method using poisson and Chi-squared distributions: It is not possible to calculate confidence intervals by this method in cases of increased incidence of occurrence since D. F. is infinite in the Chi-table and since pU equal to pL .

Analysis between historical data and current test findings:

Table 4 shows the results of statistical analysis comparing historical data to bioassay data following a, b, c, d and e models. Results of statistical analysis on the incidence of hepatocellular carcinoma and the L. G. L. of the spleen from the lowest and low incidence, respectively, are presented.

a. Normal distribution test. Significant differences at the 5% level were observed in the

occurrences of hepatocellular carcinoma and the L. G. L. of the spleen in the treated groups as compared to historical data. However, this model does not include the factor of a sample number in the treated group.

b. F distribution test. No significant differences were observed in the occurrence of hepatocellular carcinoma and the L. G. L. of the spleen as compared to historical data. The factor of sample number in the treated group is not taken into account in this test model.

c. Chi distribution test. Significant differences were evident in both incidence at the 5% level.

d. Categorical test by Chi-squared. Significant differences were evident in both incidence of occurrence at the 5% level as in the Chi distribution test. This test model is relatively simple. However, it does not include the factor of sample number in the treated group, as in the normal distribution test.

e. Kastenbaum and Bowman test. Table 5 shows the results of statistical analysis using the Kastenbaum and Bowman test on occurrences of hepatocellular carcinoma and the L. G. L. of the spleen. However, no significant differences were evident between bioassay data and historical data at the 5% level in both tumor types. Further, the number of occurrences showing a

Table 3. Confidence intervals of historical data determined by the approximation method using normal distribution.

Organ	Finding	Total rates, (11 studies) (%)	Confidence intervals (range), %		
			95%	99%	99.9%
Liver	Hepatocellular carcinoma	3/550 (0.5)	1.1- 0.0 (1.1)	1.3- 0.0 (1.3)	1.5- 0.0 (1.5)
Spleen	Leukemia/malignant lymphoma	41/550 (7.5)	9.7- 5.3 (4.4)	10.4- 4.6 (5.8)	11.2- 3.8 (7.4)
Pituitary	Adenoma	160/550 (29.1)	32.9-25.3 (7.6)	34.1-24.1 (10.0)	35.5-22.7 (12.8)
Testes	Interstitial cell tumor	465/550 (84.5)	87.5-81.5 (6.0)	88.5-80.5 (8.0)	89.6-79.4 (10.2)

Table 4. Results of statistical analysis of low incidence findings, hepatocellular carcinoma and leukemia/malignant lymphoma analyzed by various models between historical data and a study finding.

Organ	Finding	Historical data	Treated vs. group	Models				
				Distribution			Categoric test by chi-squared	Kastenbaum and Bowman
				Normal	F	Chi		
Liver	Hepatocellular carcinoma	3/550	2/50	*	NS	*	*	NS
Spleen	Leukemia/malignant lymphoma	41/550	5/50	*	NS	*	*	NS

* $P < 0.05$, NS $P < 0.05$.

significant difference from historical data were 3 in the cases of hepatocellular carcinoma and 9 or more in the L. G. L. of the spleen. A characteristic of this test model is that it takes into account the number of samples in historical data and the current bioassay data as a ratio. However, pituitary-adenoma and interstitial cell tumors of the testes were not analyzable because of their occurrence in such high incidence.

Analysis by using distribution of historical data :

a. Application of rejection limits. Table 6 shows the confidence intervals with the rejection limits for historical data. The number of occurrences revealing significant differences at the 5% level from historical data was more than 3 in the case of hepatocellular carcinoma, greater than 11 in the L. G. L. of the spleen, greater than 29 in pituitary adenoma and less than 34 in interstitial cell tumors of the testes per 50 animals, respectively.

DISCUSSION

Confidence intervals from historical data by the approximation method using normal distribu-

tion are easily calculated and are not influenced by sample size. However, confidence intervals so calculated are narrow. Unlike other models which do not take account sample size, the Kastenbaum and Bowman test takes into account the sample sizes of both the bioassay and the historical data. A normal distribution test may be applicable only when the sample size is almost the same as the number in the historical data. Although difficult to calculate, because of the need to interpolate, the F-distribution test trends to be accurate. Bilateral testing requires a value for 2.5% in F-table. Consequently, if the incidence of the finding in bioassay study is the smallest, the Kastenbaum and Bowman test is recommended as the analysis of choice. The application of rejection limits may be recommended in all cases with low incidence.

When, the pertinent findings such as malignant tumors show very low incidence, statistical analysis employing historical data rather than the concurrent control group may be applicable. Historical data accumulated from a number of bioassay studies performed in the same institute may contribute to a more accurate analysis of tumor incidence data in the treated groups of

Table 5. Comparison of historical data and a study finding by using Kastenbaum and Bowman method.

Organ	Finding	Total rates, (11 studies) (%)	Found minimum No. of rats with significant difference in 50 rats		
			10% level	5% level	1% level
Liver	Hepatocellular carcinoma	3/550 (0.5)	2	3	4
Spleen	Leukemia/malignant lymphoma	41/550 (7.5)	7	9	11
Pituitary	Adenoma	160/550 (29.1)	-	-	-
Testes	Interstitial cell tumor	465/550 (84.5)	-	-	-

- : Unable to calculate or no table due to a high incidence of occurrence.

Table 6. Confidence intervals of historical data as determined by the application of rejection limits.

Organ	Finding	Mean occurrence ± S.D.	Confidence interval No. of occurrence per 50 rats		
			95%	99%	99.9%
Liver	Hepatocellular carcinoma	0.27 ± 0.65	0- 2	0- 2	0- 3
Spleen	Leukemia/malignant lymphoma	3.73 ± 2.76	0-10	0-13	0-17
Pituitary	Adenoma	14.54 ± 5.94	1-28	0-34	0-43
Testes	Interstitial cell tumor	42.27 ± 3.32	35-50	31-50	26-50

classical bioassays. Historical data accurately reflect the occurrence of naturally-occurring tumors in the animals as well as the range of incidence of tumors in the limited number of animals in the concurrent control group. Consequently, the utilization of historical data in statistical analysis are useful in detecting false positive or false negative results. The results of current statistical methods suggest that the utilization of historical data could prove useful for this purpose. However, assessment method of laboratory data for neoplastic activity would be needed to take into account the time-to tumor aspect.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Kaneo Ohori, Department of Public Health, Hamamatsu University School of Medicine and Dr. W. J. Tierney, D. A. B. T., consultant in Toxicity, New Jersey, U. S. A. for critically reviewing the manuscript and for their useful advice and suggestions.

REFERENCES

- Colquhoun, D. (1982): Lectures on Biostatistics. Oxford University Press 1971, pp. 69-71, Iyaku journal Co., Ltd., Tokyo.
- Haseman, J. K. (1992): Value of historical controls in the interpretation of rodent tumor data. *Drug Information Journal*, **26**, 191-200.
- Haseman, J. K., Huff J. and Boorman, G. A. (1984): Use of historical control data in carcinogenicity studies in rodents. *Toxicol. Pathol.*, **12**, 126-135.
- Kastenbaum, M. A. and Bowman, K. O. (1966): The minimum significant number of successes in a binomial sample. Oak Ridge National Laboratory, Oak, Tennessee, ORNL-3909.
- Krewski, D., Smythe, R.T., Dewanji, A. and Colin D. (1988): Statistical Test with Histological Controls. *In* Carcinogenicity: The design, analysis and interpretation of long-term animal studies, pp. 23-38, Springer-Verlag, New York.
- National Toxicology Program (NTP) (1985): NTP Technical Report Series, No. 269, NIH Publication No. 85-2525, May 1985.
- National Toxicology Program (NTP) (1989): NTP Technical Report Series, No. 355, NIH Publication No. 89-2810, September 1989.
- Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parish, S., Peto, J., Richards, S. and Wahrendrof, J. (1980): Long-term and short-term screening assay for carcinogens: A critical appraisal. International Agency for Research on Cancer, IARC Monographs, Supplement 2, IARC/1/80, LYON.
- Shayne, C.G. and Carrol, S.W. (1982): Statistics for Toxicologists, Principles and Methods of Toxicology. Edition by A. Wallace Hayes, pp. 291-292, Raven Press, New York.
- Shibata, K. (1964): Seibutsu tokeigaku kogi, p. 51, Tokyo University of Agriculture (in Japanese).
- Yoshida, M. (1980): Design of experiments for animal husbandry, pp. 199-206, Yokendo Co., Ltd., Tokyo (in Japanese).