Environmental Health Perspectives Supplements
Vol. 102 Suppl 1:49-52 1994

Statistical Analysis of Data in Mutagenicity Assays: Rodent Micronucleus Assay

by M. Hayashi,[1] S. Hashimoto,[2] Y. Sakamoto,[3] C. Hamada,[4] T. Sofuni,[1] and I. Yoshimura[5]

To evaluate chemical safety, many kinds of short-term mutagenicity assays are performed together with long-term assays in animals. Rationales and methodology for these assays have been well discussed and documented. No statistical method, however, has been singled out as the method of choice for the evaluation of mutagenicity assay data, although a number of reports on statistical methods to evaluate such data have been published. Among the mutagenicity assays, the micronucleus assay using mouse bone marrow erythropoietic cells have been widely used to assess cytogenetic activities of test chemicals. A statistical evaluation procedure for this assay is proposed herein, combining the use of historical control data and dose-response relationships. The probability of type I errors and the power of this method are compared with those of some other conventional methods by Monte Carlo simulation.

Introduction

Several published guidelines for mutagenicity studies require appropriate statistical treatment of assay data. Although there are a large number of publications on the statistical evaluation of mutagenicity test data, no one method is recommended exclusively. Many of the methods are statistically validated but toxicologically impractical. Thus, many empirical methods are still used for data evaluation. This situation may be attributable to inadequate communication between biologists and statisticians. An attempt at improvement has been made by the U.K. Environmental Mutagen Society, and an overview and recommendations have been published (*1*).

Many kinds of mutagenicity assays have been performed in the course of safety assessments of chemicals. Gene mutation, chromosomal aberration, and DNA modification are main end points. The micronucleus assay, an *in vivo* chromosomal aberration assay, is targeted here to establish a statistical evaluation procedure that is practical and should be readily accepted by biologists (*2*).

Micronucleus Assay

Male ddY mice, 8 weeks old, were used for the micronucleus assay. The experiment consisted of at least three dose groups and concurrent negative and positive control groups. Each group had six mice. The highest dose level and the sampling times, which were the most important factors in the assay, were optimized experimentally (*3*). Second and third dose levels were one-half and one-quarter of the highest dose level, respectively. After chemical treatment, mice were killed and femoral marrow cells were smeared on clean glass slides, fixed with methanol for 5 min at room temperature, and stained with Giemsa or acridine orange (*4*). The frequencies of micronucleated polychromatic erythrocytes (MNPCEs; young erythrocytes with one or more micronuclei) were scored microscopically based on an observation of 1000 polychromatic erythrocytes (PCEs) per animal.

Characteristics of the Micronucleus Assay Data

The characteristics of data obtained from the micronucleus assay are as follows: *a*) The frequency distribution of MNPCE, the target cell being analyzed, is binomial, at least in the negative control groups (Fig. 1). *b*) The historical negative and positive control data can be constructed relatively easily. Control charts of MNPCE data from negative and positive control groups accumulated at the National Institute of Hygienic Sciences, Tokyo, are shown in Figures 2 and 3, respectively. *c*) Each chemical is assayed at at least three dose levels, to assess a dose-response relationship; *d*) It is not difficult to repeat the assay if the results are marginal or problematic, unlike long-term animal experiments.

[1]Division of Genetics and Mutagenesis, National Institute of Hygienic Science, Tokyo, Japan.
[2]Department of Public Health Statistics, National Institute of Public Health, Tokyo, Japan.
[3]Biology Department, Takeda Analytical Research Laboratories, Ltd., Osaka, Japan.
[4]Research and Development Division, Takeda Chemical Industries, Ltd., Osaka, Japan.
[5]Faculty of Engineering, Science University of Tokyo, Tokyo, Japan.
Address reprint requests to M. Hayashi, Division of Genetics and Mutagenesis, National Institute of Hygienic Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158, Japan.
This paper was presented at the International Biostatistics Conference on the Study of Toxicology that was held May 13-25, 1991, in Tokyo, Japan.

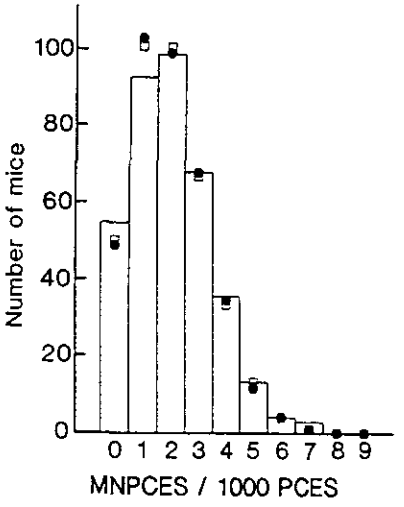

FIGURE 1. The distribution of frequency of micronucleated polychromatic erythrocytes (PCEs) based on 1000 PCEs observation in the negative historical control (bars), the fitted binomial distribution B(0.002, 1000) (□), and binomial random numbers (●) (*2*).

Strategy of the Evaluation

First, the concurrent negative and positive control data are tested to validate the assay system. Second, the treatment-induced response is evaluated for each dose group. Third, the dose-response relationship is assessed. This strategy is comparable to that of the empirical evaluation process used by biologists in toxicological studies.

Proposed Procedure

The procedure consists of the following three steps (Fig. 4) and accepts as significant an overall p-value of ≤ 0.01.

Step 1. The concurrent negative and positive controls are compared with the historical control. If the mean frequency of MNPCEs deviates from the historical mean by ≥ 3 standard deviations, the experiment is discounted and a new one is performed for the same chemical.

Step 2. After passing step 1, the observed frequencies of MNPCEs in treatment groups are compared with the frequency distribution of the negative historical control at a significance level of 0.05 divided by the number of dose groups. If no dose group shows significant difference, the test chemical is classified as negative.

Step 3. If the test chemical is not classified as negative in step 2, the monotonous dose–response relationship is tested by the Cochran-Armitage trend test using the concurrent negative control at the nominal significance level of 0.05. The test chemical is classified as positive only when a positive dose–response relationship is shown.

Evaluation of the Procedure

Type I errors and statistical power were compared by the Monte Carlo method among four statistical methods: the Cochran-Armitage trend test [significance level: 0.01 (*5,6*)]; a binomial test (significance level: 0.01/number of dose groups); conditional binomial test [significance level: 0.01 (*7*)]; and the proposed three-step method [significance level: 0.05/(number of dose groups) for step 2, 0.05 for step 3, and 0.01 overall (*2*)].

When type I error was simulated against the number of mice per group, the Cochran-Armitage trend test showed almost constant type I error near the nominal level (Fig. 5). Although type I errors of the other methods were not exactly controlled at 0.01, the probabilities were comparable among the four methods. The statistical powers of the methods were simulated with four groups. The doses were set as d_0, d_1, d_2, d_3 = 0, 1, 2, 4 mg/kg. The population proportions of MNPCEs were set as π_0, π_1, π_2, π_3 = 0.2, 0.25, 0.3, 0.4%. The number of mice per group was set as 2–10. The simulated powers of the four methods were ranked as binomial test > three-step method >> Cochran-Armitage trend test > conditional binomial test (Fig. 6).

The experiment could be rejected if the concurrent control data (e.g., step 1 of the three-step method) deviated greatly from the historical control data, depending on the current experimental conditions. When the deviation was small, however, the robustness of the method was important. When data were biased downward by 10%, i.e., when the frequencies of MNPCEs in all groups were lowered 10% from the expected values (based on historical data), the power of the three-step method was the highest (Fig. 7) in spite of the decrease of a type I error. On the other hand, when data were biased upward by 10% (Fig. 8), the binomial test showed the highest power, with the three-step

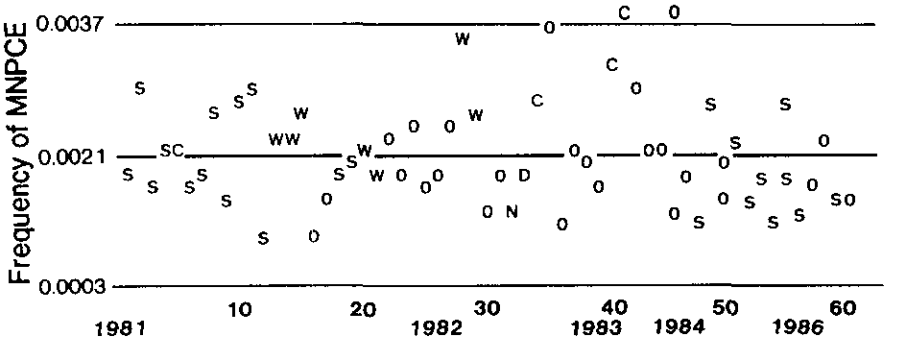

FIGURE 2. A control chart of the historical data of the proportion of micronucleated polychromatic erythrocytes in the negative control groups using six mice per group during the period from 1981 to 1986 (*2*). C, 0.5% carboxymethyl cellulose sodium salt aqueous solution; D, dimethyl sulfoxide; N, no treatment; O, olive oil; S, physiological saline; W, distilled water.

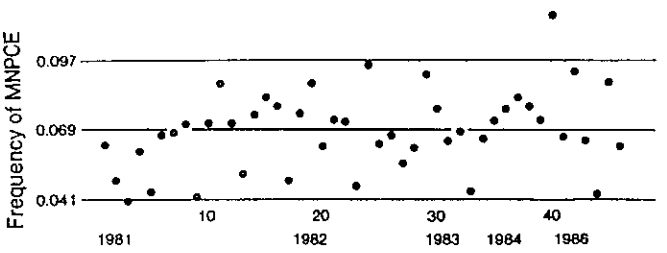

FIGURE 3. A control chart of the historical data of the proportion of micronucleated polychromatic erythrocytes in the positive control groups using six mice per group during the period from 1981 to 1986 (*2*). Mice received mitomycin C once IP at a dose of 2 mg/kg and were killed 24 or 30 hr after treatment.

method as a close second; moreover, the increase of type I error for the three-step method was not as great as the binomial test. Although the type I error for the Cochran-Armitage trend test was constant and nominal even when data were biased in both directions, the power of this method, as well as of the conditional binomial test, was lower than that of the binomial test or the three-step method.

The strategy of the proposed three-step method is practical and should be readily accepted by toxicologists. The results concur with toxicological judgment (*2*). For the overall evaluation of micronucleus test data, reproducibility is also important. If the results of the statistical evaluation disagree with the intuition of the investigator, an additional experiment is recommended to confirm the test result.

Prerequisites to apply the proposed three-step method are as follows: *a*) negative and positive historical control data must be

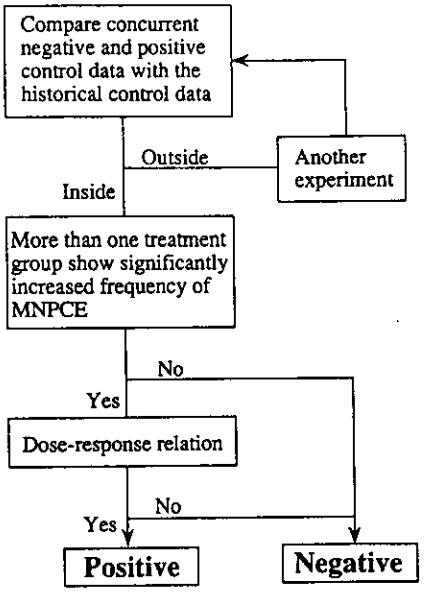

FIGURE 4. Strategy of the evaluation of the micronucleus assay data. First step: the concurrent negative and positive control data are tested to validate the assay system itself, and, if it is not acceptable, a new experiment should be performed. Second step: data from each dose group are evaluated to determine the increase of response compared with the historical control. Third step: the dose-response relationship is assessed. After testing these steps, chemicals would be declared negative or positive in the micronucleus assay.

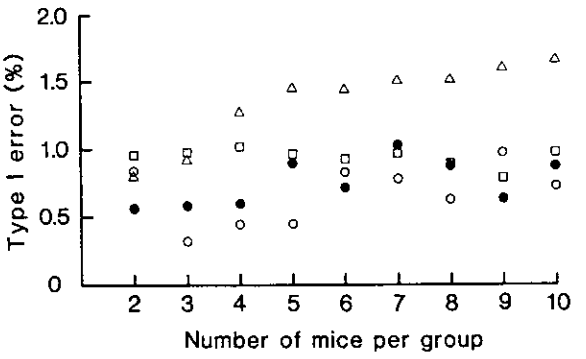

FIGURE 5. The probability of a type I error in experiments with three dose groups and a negative control group (*2*). (□) Cochran-Armitage trend test [significance level: 0.01 (*5,6*)] (○) binomial test (significance level: 0.01/no. of dose groups); (△) conditional binomial test [significance level: 0.01 (*7*)]; (●) the proposed three-step method (significance level: 0.05/no. of dose groups for step 2, 0.05 for step 3, and 0.01 overall).

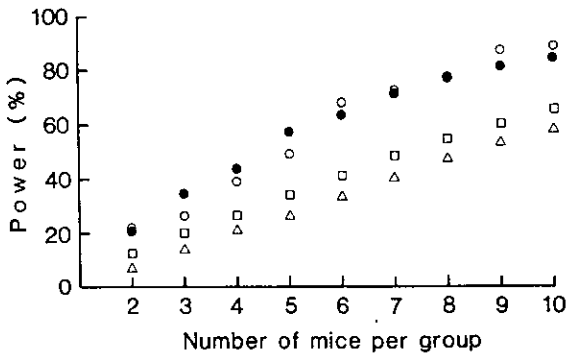

FIGURE 6. The power of the four methods in experiments with three dose groups and a negative control group (*2*). (□) Cochran-Armitage trend test [significance level: 0.01 (*5,6*)] (○) binomial test (significance level: 0.01/no. of dose groups); (△) conditional binomial test [significance level: 0.01 (*7*)]; (●) the proposed three-step method (significance level: 0.05/no. of dose groups for step 2, 0.05 for step 3, and 0.01 overall).

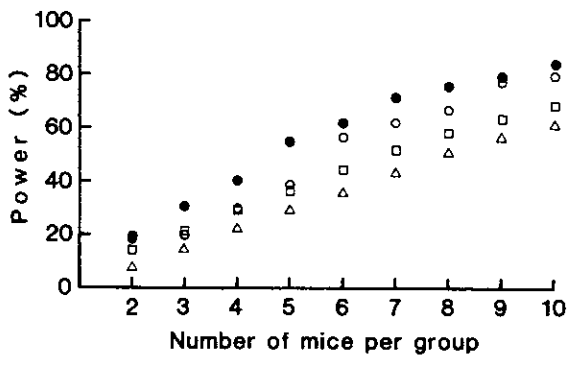

FIGURE 7. The power of the four methods in experiments with three dose groups and a negative control group biased downward by 10% (*2*). (□) Cochran-Armitage trend test [significance level: 0.01 (*t,6*)] (○) binomial test (significance level: 0.01/no. of dose groups); (△) conditional binomial test [significance level: 0.01 (*7*)]; (●) the proposed three-step method (significance level: 0.05/no. of dose groups for step 2, 0.05 for step 3, and 0.01 overall).

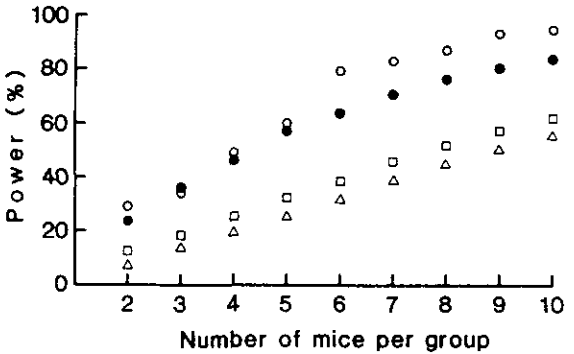

FIGURE 8. The power of the four methods in experiments with three dose groups and a negative control group biased upward by 10% (*2*). (□) Cochran-Armitage trend test [significance level: 0.01 (*5,6*)] (○) binomial test (significance level: 0.01/no. of dose groups); (△) conditional binomial test [significance level: 0.01 (*7*)]; (●) the proposed three-step method (significance level: 0.05/no. of dose groups for step 2, 0.05 for step 3, and 0.01 overall).

available for the relevant mouse strain, and the distribution of the negative control is binomial. *b*) For every new chemical, dose levels and sampling times must be optimized, possibly by a dose-and sampling-time-finding pilot experiment. *c*) Slides should be coded and examined without any knowledge about treatment, preferably by the same investigator(s). *d*) The frequency of MNPCEs should be based on the observation of at least 1000 PCEs per animal. *e*) Both negative and positive control groups must be included in an experiment.

It is most important that the test results are credible and reliable technically and biologically. After statistical evaluation of data, the results of an experiment are often interpreted as fact. But statistical methods cannot evaluate the quality of the data, and they might lead to the impression that there were no problems in the data. Therefore, it is the responsibility of experimental toxicologists to generate reliable data for statistical analyses.

REFERENCES

1. U.K. Environmental Mutagen Society Subcommittee on Guidelines for Mutagenicity Testing. Report, Part III, Statistical Evaluation of Mutagenicity Test Data (D. J. Kirkland, Ed.), Cambridge University Press, Cambridge, 1989.
2. Hayashi, M., Yoshimura, I., Sofuni, T., and Ishidata, M., Jr. A procedure for data analysis of the rodent micronucleus test involving a historical control. Environ. Mol. Mutagen. 13: 347–356 (1989).
3. Hayashi, M., Sofuni, T., and Ishidata, M., Jr. A pilot experiment of the micronucleus test. The multi-sampling at multi-dose levels method. Mutat. Res. 141: 165–169 (1984).
4. Hayashi, M., Sofuni, T., and Ishidate, M. Jr. An application of acridine orange fluorescent staining to the micronucleus test. Mutat. Res. 105: 253–256 (1983).
5. Armitage, P. Test for linear trends in proportions and frequencies. Biometrics 11: 375–386 (1955).
6. Cochran, W. G. Some methods for strengthening the common χ^2 tests. Biometrics 10: 417–451 (1954).
7. Kastenbaum, M. A., and Bowman, K. O. Tables for determining the statistical significance of mutation frequencies. Mutat. Res. 9: 527–549 (1970).