表 1　第 26 回 SIAM で審議された化学物質と合意結果

CAS	化学物質名・物質カテゴリー名	スポンサー	勧告	
			HH	ENV
67-68-5	Dimethyl sulfoxide	BIAC/ICCA	LP	LP
79-21-0	Peroxyacetic acid	NL/ICCA	LP	FW
98-01-1	2-Furaldehyde	NL:eu	FW	LP
99-94-5	Benzoic acid, 4-methyl-	JP	LP	LP
105-08-8	1,4-Cyclohexanedimethanol	KO	LP	LP
541-05-9	Cyclotrisiloxane, hexamethyl-	US/ICCA	LP	LP
919-31-3	Propionitrile, 3-(triethoxysilyl)-	US/ICCA	LP	LP
1341-49-7	Ammonium hydrogendifluoride ((NH4)(HF2))	NL/ICCA	-	LP
3033-77-0	2,3-Epoxypropyl trimethyl ammonium chloride	FI:eu	FW	FW
3327-22-8	3-Chloro-2-hydroxypropyl trimethylammonium chloride	FI:eu	FW	FW
7697-37-2	Nitric acid	US/ICCA	LP	LP
7757-83-7	Sodium sulfite	JP/ICCA	FW	LP
16961-83-4	Hexafluorosilicic acid	NL/ICCA	-	LP
52829-07-9	Decanedioic acid, bis(2,2,6,6-tetra methyl-4-piperidinyl) ester	CH/ICCA	LP	LP
物質カテゴリー	C5 Aliphatic Hydrocarbon Solvents Category	US/ICCA	(LP)	(LP/FW) *1
78-78-4	Butane, 2-methyl-			
109-66-0	Pentane			
287-92-3	Cyclopentane			
物質カテゴリー	Formic acid and formats Category	US/ICCA	(LP/FW) *2	(LP)
64-18-6	Formic acid			
107-31-3	Methyl formate			
141-53-7	Sodium formate			
540-69-2	Ammonium formate			
544-17-2	Calcium diformate			
590-29-4	Potassium formate			
20642-05-1	Potassium hydrogen diformate			

化学生物総合管理　第 4 巻第 2 号 (2008.12)　237-245 頁
連絡先：〒158-8501 世田谷区上用賀 1-18-1　E-mail: hirose@nihs.go.jp
受付日：2008 年 6 月 16 日　　受理日：2008 年 8 月 21 日

	(KHF)			

FW = The substance is a candidate for further work.（追加の調査研究作業が必要）
LP = The substance is currently of low priority for further work.（現状では追加作業の必要なし）
括弧内の勧告は SIAM 後の CDG 上の審議により合意が得られたものを示す。
ICCA は国際化学工業協会協議会による原案提出を示す。
eu は欧州連合でのリスク評価文書を基にしたことを意味する。
略号は、BIAC：経済産業諮問委員会、CH：スイス、FI：フィンランド、JP：日本、KO：韓国、NL：オランダ、US：米国である。
*1：Cyclopentane（CAS：287-92-3）のみ FW
*2：Methyl formate（CAS：107-31-3）のみ FW

化学生物総合管理　第 4 巻第 2 号 (2008.12)　237-245 頁
連絡先：〒158-8501 世田谷区上用賀 1-18-1　E-mail: hirose@nihs.go.jp
受付日：2008 年 6 月 16 日　　受理日：2008 年 8 月 21 日

The Journal of Toxicological Sciences (J. Toxicol. Sci.)
Vol.33, No.1, 97-104, 2008

Original Article

Evaluation of statistical tools used in short-term repeated dose administration toxicity studies with rodents

Katsumi Kobayashi[1], K. Sadasivan Pillai[2], Yuki Sakuratani[1], Takemaru Abe[1], Eiichi Kamata[3] and Makoto Hayashi[3]

[1]Chemical Management Center, National Institute of Technology Evaluation, 49-10 Nishihara-Nichome, Shibuya, Tokyo 151-0066, Japan
[2]Orchid Research Laboratories Ltd., R&D Centre, Plot No. 476/14, Old Mahabalipuram Road, Sholinganallur, Chennai 600119, India
[3]National Institute of Health Sciences 18-1 Kamiyoga-Ichome, Setagaya-ku, Tokyo 158-8501, Japan

(Received November 20, 2007; Accepted November 27, 2007)

ABSTRACT — In order to know the different statistical tools used to analyze the data obtained from twenty-eight-day repeated dose oral toxicity studies with rodents and the impact of these statistical tools on interpretation of data obtained from the studies, study reports of 122 numbers of twenty-eight-day repeated dose oral toxicity studies conducted in rats were examined. It was found that both complex and easy routes of decision trees were followed for the analysis of the quantitative data. These tools include Scheffe's test, non-parametric type Dunnett's and Scheffe's tests with very low power. Few studies used the non-parametric Dunnett type test and Mann-Whitney's *U* test. Though Chi-square and Fisher's tests are widely used for analysis of qualitative data, their sensitivity to detect a treatment-related effect is questionable. Mann-Whitney's *U* test has better sensitivity to analyze qualitative data than the chi-square and Fisher's tests. We propose Dunnett's test for analysis of quantitative data obtained from twenty-eight-day repeated dose oral toxicity tests and for qualitative data, Mann-Whitney's *U* test. For both tests, one-sided test with p=0.05 may be applied.

Key words: Statistics; 28-day repeated toxicity study; Rodents; Dunnett's test; Mann-Whitney's *U* test

INTRODUCTION

Short-term repeated oral toxicity study conducted for 14 or 28 days is aimed to (1) predict appropriate doses of test substance for future subchronic or chronic toxicity studies, (2) determine NOELs for some toxicology endpoints and (3) to allow future studies in rodents to be designed with special emphasis on identified target organs (USFDA, 2000). This study also provides information on the possible health hazards likely to arise from repeated exposure over a relatively limited period of time (USEPA, 2000; OECD, 1995). Though these guidelines provide all the information required for the conduct of the study, no information is provided on the appropriate statistical tools to be used to analyze the data obtained from the study. Use of right statistical tool to analyze the data obtained from theses studies is very crucial as the interpretation of the data is mostly based on the results of the statistical analysis.

The statistical tools used to analyze the data obtained from 122 numbers of twenty-eight-day repeated dose oral toxicity tests in rats were examined in the present study. The objective of the study was to know the different statistical tools that are used in these studies and the possible impact of these statistical tools on interpretation of the data. A brief discussion on the use and the property of the different statistical tools used in the studies are also given. The purpose of this article wished for the standardization of statistics and the analysis methods. Finally, the authors made an attempt to suggest statistical techniques that may best suit twenty-eight-day repeated dose oral toxicity studies in rodents.

Correspondence: Katsumi Kobayashi (E-mail: kobayashi-katsumi@nite.go.jp)

MATERIALS AND METHODS

Studies examined

A total number of 122 studies conducted in various test facilities in Japan were examined (MHLW, 2006). The chemical of these examinations was executed with existing chemical substances by the guideline of the Chemical Substance Control Law (1986). The number of studies conducted in each test facility is given in parenthesis: Food and Drug Safety Center, Kanagawa (22), An-Pyo Center, Shizuoka (22), Mitsubishi Chemical Safety Institute Ltd., Ibaraki (18), Safety Research Institute for Chemical Compounds Co., LTD, Hokkaido (15), Bozo Research Center Inc., Shizuoka (12), Research Institute for Animal Science in Biochemistry & Toxicology, Kanagawa (11), Panapharm Laboratories, Kumamoto (10), Nihon Bioresearch Inc., Gifu (9) and National Institutes of Health, Tokyo (3).

Quantitative and qualitative items

Several quantitative and qualitative items are evaluated in twenty-eight-day repeated dose oral toxicity tests in rats, as per the regulatory guidelines. The quantitative items that require statistical analysis are body weight, food consumption, water consumption, leucocytes, erythrocytes, hemoglobin, hematocrit, platelets, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, differential leucocyte counts, prothrombin time, activated partial thromboplastin time, total protein, albumin, albumin/globulin ratio, total bilirubin, alanine aminotransferase, aspartate aminotransferase, γ-glutamic transaminase, alkaline phosphatase, acetylcholinesterase, total cholesterol, tryglycerides, phospholipids, glucose, blood urea nitrogen, creatinine, inorganic phosphorous, calcium, sodium, potassium, chlorides, urine volume, specific gravity of urine, absolute organ weights and relative organ weights. Qualitative items that require statistical analysis are mortality, functional observation battery, clinical signs, urinalysis (color, pH, protein, glucose, ketone bodies, bilirubin, occult blood, urobilinogen, epithelial cells, erythrocytes, leucocytes, casts and crystals) and pathological findings (macroscopic and microscopic). But the regulatory guidelines do not indicate the specific statistical techniques to be used to analyze these data.

Which test to be used - One-sided or two-sided?

When the *t*-test and Dunnett's multiple comparison test (Dunnett's test) are used, the significant difference detection rate of a two-sided test is about 85% as compared with a one-sided test (Kobayashi, 1997a). In toxicological studies, usually a dosed group is compared with the control group. For this comparison, one-sided test is ideal, hence Yoshimura and Ohashi (1992) recommend the one-sided test for comparing a dosed group with the control group.

Is analysis of variance (ANOVA) necessary?

It is a common practice to subject the data, if they are from more than two groups, ANOVA. If ANOVA shows a significant difference among the groups, multiple comparison tests are used to find the significant difference between any two groups. In recent years, several authors suggested that the error of the second kind can be prevented by carrying out direct multiple comparison tests, without subjecting the data to ANOVA (Hamada *et al.*, 1998; Kobayashi *et al.*, 2000a; Sakaki *et al.*, 2000). It may be worth mention in this context that Dunnett (1964) did not recommend ANOVA prior to multiple comparison tests.

Is Bartlett's homogeneity test necessary?

Generally Bartlett's test is used to examine the homogeneity of variance if the number of animals in a group is 10 or more. Therefore, this test is not used in the toxicity studies with dogs, where the number of animals in the group is less. According to Kobayashi *et al.* (1998), Bartlett's test is not required to examine the homogeneity of variance, when the number of animals in a group is less.

Non-parametric type Dunnett's test

The non-parametric Dunnett's multiple comparison test has two techniques - 'joint type' and 'separate type' or Steel's test. When the Steel's test shows the highest dosage correlation, the number of animals required in the dosage groups to detect a significant difference in the low dosage group is four (Inaba, 1994; Kobayashi *et al.*, 1995). On the contrary, 'joint type' needs 15 animals in each group.

Transformation of data

If the data show heterogeneity of variance as per Bartlett's test, sometimes the data are transformed, for example to logarithmic values and then they are subjected to non-parametric tests. According to Finney (1995), "when a scientist measures a quantity such as concentration of a chemical compound in body fluid, his interest usually lies in the scale, perhaps mg/ml, that he has used; he is less likely to be interested in a summary of results relating to a transformed quantity such as the logarithm of blood concentration. If he analyzes in terms of logarithms, encouraged perhaps by an elementary but uncritical statistical textbook or by a convenient software package, he may find significant differences but to express his conclusions in meaningful numbers may be impossible. I do not assert

that a scientist should never transform data before analysis; I urge that data should be transformed only after careful consideration of all consequence". Therefore, transformation should be done cautiously.

Power of Scheffe's test

Use of Scheffé's test is discouraged in recent years because this test may not show a significant difference in the dosage groups even if the dosage groups show a difference of 60-53% compared to control group (Kobayashi *et al.*, 1997b).

Power of non-parametric tests using ranked data

In four groups setting with the highest dosage correlation, the minimum numbers of animals required in the low-dose group to detect a significant difference, compared to control, using the statistical tools of Scheffé's type, Dunn's test, Tukey type, Dunnett type, Williams-Wilcoxon test, Steel test and Mann-Whitney's *U* test are 22, 19, 18, 15, 8, 4 and 3, respectively. Therefore, in the twenty-eight-day repeated dose oral toxicity tests in rats, where the number of animals is 5/sex/group, except Steel and Mann-Whitney's *U* tests, other tests are not used. Inaba (1994) made a similar observation on the power of the above tests.

Power of Chi-square and Fisher's tests

When a finding in the animals of a control group is 0, in order to find a significant difference of the finding between the control group (n=5) and dosage group (n=5) by chi-square test, all the 5 animals in the dosage group (n=5) should show the finding, whereas by Fisher's test 4 animals should show the finding. When 1 animal in the control group shows a finding, even if the finding is seen in all the animals in the dosage group, a significant difference is not detected by chi-square test, but it is detected by Fisher's test. In the light of the above it may be stated that power of one-sided Fisher's test is better than the Chi-square test.

Dunnett's test is the expanded version of *t*-tests

Dunnett's test becomes *t*-test when two groups are analyzed (Kobayashi *et al.*, 1997c). Therefore, when comparing the recovery groups in the twenty-eight-day repeated dose oral toxicity tests in rats, where number of the groups is 2, it does not make any difference, whether the analysis is carried out by Dunnett test or *t*-test.

Power of Mann-Whitney's *U* test

This test is generally used for the analysis of pathology data (Kobayashi *et al.*, 1997d). A significant difference by a one-sided test is detected if the calculated *U* value is four or less. Since one-side is expected in studies like twenty-eight-day repeated dose oral toxicity tests in rats, a one-sided Mann-Whitney's *U* test is used to analyze pathology data obtained from these studies.

RESULTS

Quantitative data

Out of 122 studies examined, 79 studies used statistical tools that follow a complicated course (tool numbers; 2, 3, 4, 5, 8, 9, 10, 12, 15, 16 and17) and 43 studies used statistical tools that follow simple course (tool numbers; 1, 6, 7, 11, 13 and14) (Table 1; Fig. 1). The statistical tools describing the method of analyzes, in the case of three or more groups and two groups were mentioned in 6 studies, whereas this description was not found in 11 studies. Only eight studies used trend test (Jonckheere, 1954). In the tool number 10, the significance level of ANOVA and Kruskal-Walis's *H* test were set at p=0.10. For comparing with the control, this tool set the significance level of p=0.05. Tool numbers 13 and 14 did not perform Bartlett's test for testing the homogeneity of variance. Use of one-sided or two-sided test is not indicated in 87 studies. Only one study indicated use of non-parametric test.

Qualitative data

Since urinalysis data were classified into many grades, chi-square test was used to analyze these data in most of the studies. For macro- and microscopic pathological findings, Mann-Whitney's *U* test, Fisher's test and Chi-square test were used. Most of the studies did not indicate the alpha. Only the pathological findings of 3 studies were examined for dose-relationship (Table 2).

Use of a one-sided test was more common than a two-sided test in the case of analysis of both quantitative and qualitative data (Table 3).

DISCUSSION

National Toxicology Program, USA published technical reports of long-term carcinogenicity studies and short-term toxicity tests carried out with more than 500 substances in rat and mouse (NIH, 2006). Most of these studies used the statistical tools almost similar to the ones currently used to analyze the data obtained from the toxicity tests of agricultural chemicals and medical drugs (Kobayashi *et al.*, 2000b).

On examination of 122 studies, it was found that complex and easy courses of analytical techniques were used for the analysis of the quantitative data. These tools may be classified into 4 different categories. Five tools (tool

numbers; 4, 5, 8, 16 and17) are the advanced type of the algorithm, similar to the one developed by Yamazaki *et al.* (1981). These tools include Scheffé's test, non-parametric type Dunnett's and Scheffé's tests with very low power. Six tools (tool numbers; 3, 7, 9, 10, 12 and 15) are again advanced type of algorithm developed by Sano and Okayama (1990), which can be used even if the number of animals in the groups are different. Use of the non-parametric Dunnett type test with low power is also seen in few studies. Mann-Whitney's *U* test was also used (tool number; 9) in 14 studies in order to retain the power. Three tools (tool numbers; 2, 6 and 11) are an improved version of non-parametric type Dunnett's test ('joint type') and Steel's test ('separate type'). Dunnett's or Scheffé's tests is independently used for 3 tools (tool numbers; 1, 13 and 14). Though use of Scheffé's test has the advantage of comparison of groups in various combinations, for example, control+mid dose *vs*.high dose, low dose+mid dose *vs*. high dose, etc., it has extremely low detection power. Hence, this test is not widely used in recent years.

Yoshimura (1987) used Bartlett's test to analyze the difference in distribution of variance among the groups, where number of animals in the group is more than 10. The power of Bartlett's test decreases when the number of animals in the group is less.

Dunnett's test is the expanded version of *t*-tests, hence, it becomes *t*-test when two groups are analyzed by Dunnett's test. Therefore, for the comparison of two groups either Dunnett test or *t*-test can be used.

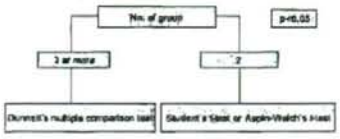
The most important purpose of applying statistical analysis in toxicity studies is to know whether the items estimated in the experimental group has increased or decreased compared to the control. Therefore, a one-sided test is used. Detection rate of two-sided test is half of the one-sided test, hence it is important to mention in the study report whether a one-sided or two-sided test is used. It may

Table 1. Classification of number of studies based on the statistical tools used for the analysis of quantitative data.

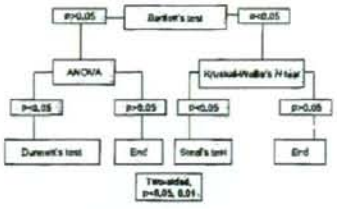
Tool. No.	Description of statistical tools	Number of studies
1	Dunnett's test: Three groups or more; Student or Aspin-Welch's *t*-test: Two groups	5
2	Bartlett's test, ANOVA, Dunnett's test, Kruskal-Walis's *H* test, Steel's test	7
3	Bartlett's test, ANOVA, Dunnett's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test: Three groups or more; Student or Aspin-Welch's *t*-test: Two groups	9
4	Bartlett's test, ANOVA, Dunnett's test, Scheffé's test, Kruskal-Walis's *H* test, Non-para type Dunnett's test, non-parametric type Scheffe's test: Three groups or more; Student or Aspin-Welch's *t*-test: Two groups	10
5	Bartlett's test, NOVA, Dunnett's test, Duncan's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test	9
6	Bartlett's test, Dunnett's test, Steel's test	20
7	Bartlett's test, Dunnett's test, non-parametric type Dunnett's test	10
8	Bartlett's test, ANOVA, Dunnett's test, Scheffé's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test, non-parametric type Scheffé's test	23
9	Bartlett's test, ANOVA, Dunnett's test, Kruskal-Walis's *H* test, Mann-Whitney's *U* test	14
10	Bartlett's test, ANOVA (p=0.10), Dunnett's test, Kruskal-Walis's *H* test (p=0.10), Mann-Whitney's *U* test, When compared with control setting (p=0.05)	1
11	Bartlett's test, Dunnett's test, Steel's test	3
12	Bartlett's test, ANOVA, Dunnett's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test: Three groups or more; Student's *t*-test or Mann-Whitney's *U* test: Two groups	1
13	Dunnett's test: Three groups or more; *t*-test or Mann-Whitney's *U* test: Two groups	4
14	Dunnett's or Scheffé's tests: Three groups or more; *t*-test or Mann-Whitney's *U* test: Two groups	1
15	Bartlett's test, ANOVA, Dunnett's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test	3
16	Bartlett's test, ANOVA, Dunnett's test, Jaffé's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test, non-parametric type Jaffé's test	1
17	Bartlett's test, ANOVA, Dunnett's test, Scheffé's test, Kruskal-Walis's *H* test, non-parametric type Dunnett's test, non-parametric type Scheffé's test: Three groups or more; Student's *t*-test: Two groups	1
	Jonckheere's trend test (Not included in the number of tools)	8
	Total	122

Statistical tools used in short-term toxicity studies.

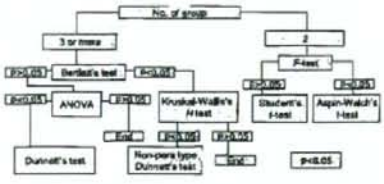
Tool No. 1, use rate;5/122



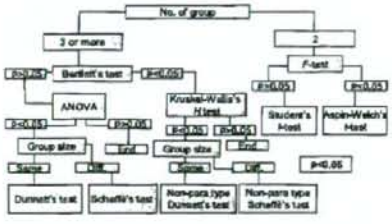
Tool No. 2, use rate;7/122



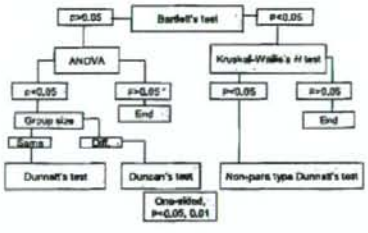
Tool No. 3, use rate;9/122



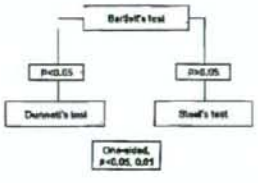
Tool No. 4, use rate;10/122



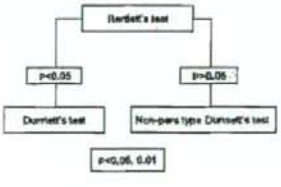
Tool No. 5, use rate;9/122



Tool No. 6, use rate;20/122



Tool No. 7, use rate;10/122



Tool No. 8, use rate;23/122

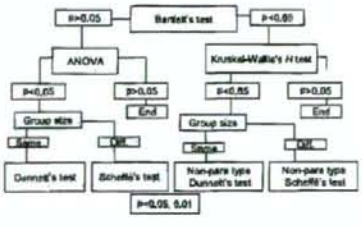


Fig. 1. Classification of number of studies based on the statistical tools used for the analysis of quantitative data.

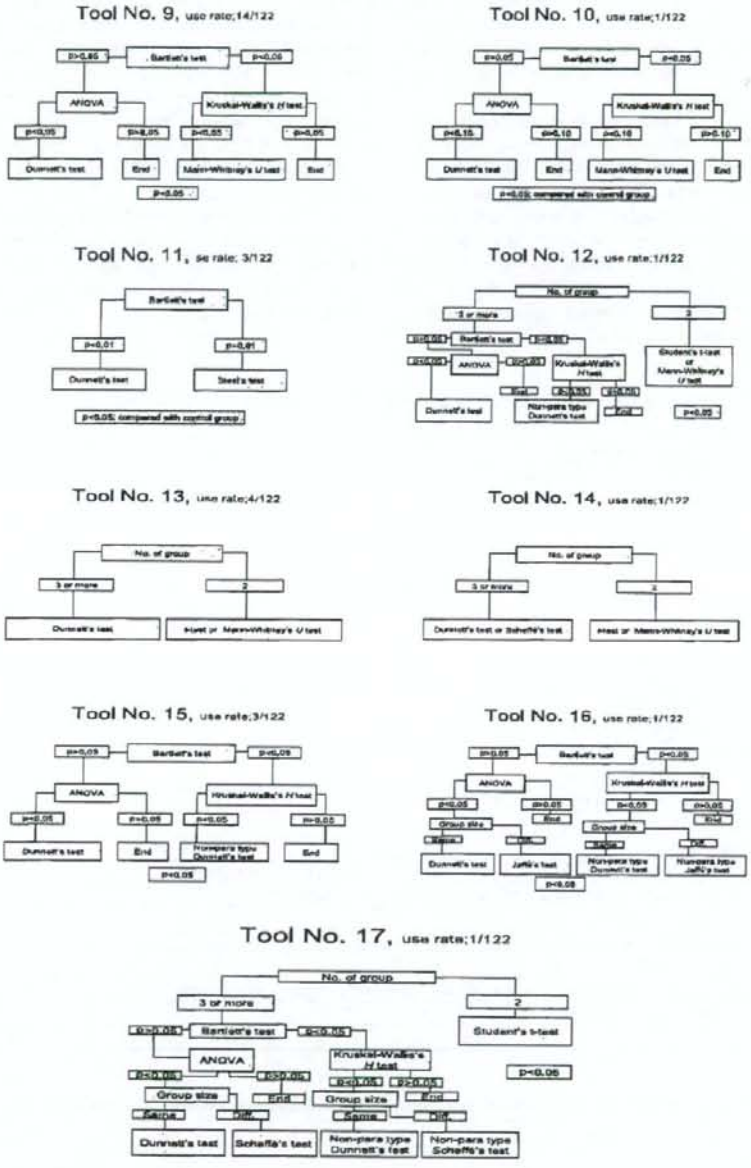


Fig. 1. Continued.

be noted that use of ANOVA causes the error of the second kind. Because of this, some of the recent studies skipped ANOVA in the decision tree and straight away used the statistical tools for *post hoc* comparison (Sumida *et al.*, 2006; Nagano *et al.*, 2006).

For the analysis of qualitative data, chi-square and Fisher's tests do not seem to be appropriate, though Fisher's test is slightly more sensitive than the chi-square test. These two tests do not detect a significant difference between a finding in the dosage group and control group, when all the animals (5/5) show the finding in the dosage group and 2 animals in the control group (2/5). On the other hand, Mann-Whitney's *U* test, which converts the scores into numerical values, detects a significant difference, when the finding in the dosage group is 5/5 and control group is 2/5. Therefore, Mann-Whitney's *U* test has better sensitivity to analyze qualitative data than the chi-square and Fisher's tests. Trend test like Jonckheere test can be used to determine no observed adverse effect level/ no observed effect level (NOAEL/NOAL) in the twenty-eight-day repeated dose oral toxicity tests. The statistical tools used, especially in the case of non-parametric tests, to determine the NOAEL/NOAL may be clearly elaborated in the study report.

We propose Dunnett's test for the analysis of quantitative data obtained from twenty-eight-day repeated dose oral toxicity tests in rodents and for qualitative data, Mann-Whitney's *U* test. For both tests, one-sided test with p=0.05 may be applied.

Table 2. A classification of number of studies based on the statistical tools used for the analysis of qualitative data.

Tool. No.	Description of statistical tools		Number of studies
1	Scored data	Frequency data	6
	Mann-Whitney's *U* test (two-sided, p<0.05)	Fisher's test (one-sided, p<0.05)	
2	Urinalysis	Pathological findings	7
	Cumulated Chi-square test (two-sided, p<0.05, p<0.01)	Mann-Whitney's *U* test (two-sided, p<0.05, p<0.01)	
3	Urinalysis	Pathological findings	13
	Cumulated Chi-square test (p<0.05)	Mann-Whitney's *U* test (two-sided, p<0.05)Fisher's test (one-sided test, p<0.05)	
4	Pathological findings		26
	Fisher's test (one-sided test, p<0.05)		
5	Pathological findings		19
	Chi-square test (p<0.05)		
6	FOB, urinalysis and differential leucocytes		15
	Kruskal-Wallis's *H* test, Mann-Whitney's *U* test (p<0.05)		
7	Urinalysis and pathological findings		9
	Mann-Whitney's *U* test (two-sided, p<0.05, p<0.01)		
8	Pathological findings		1
	Fisher's test		
9	FOB, sense function test and macroscopic and microscopic findings of pathology		1
	Wilcoxon rank-sum test, Fisher's test and Mann-Whitney's *U* test (p<0.05, p<0.01)		
10	Pathological findings		4
	Nonparametric type Dunnett's test or non-parametric type Scheffe's test, and Cochran-Armitage's trend test		
11	FOB, sense function test and macroscopic and microscopic findings of pathology		21
	No statistical tool mentioned		
Total			122

Table 3. Use of one-sided or two-sided test for short-term repeated dose administration toxicity studies with rats.

Data	One-sided	Two-sided	No mentioned	Total
Quantitative	22	13	87	122
Qualitative	34	22	70	126

ACKNOWLEDGMENT

Research described in this paper was supported by Grant (Project name: Development of Hazard Assessment Techniques Using Structure-activity Relationship Methods) from New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

Dunnett, C.W. (1964): New tables for multiple comparisons with a control. Biometrics, September, 482-491.

Finney, D.J. (1995): Thoughts suggested by a recent paper: Questions on non-parametric analysis of quantitative data (Letter to editor). J. Toxicol. Sci., **20**, 165-170.

Hamada, C., Yoshino, K., Matsumoto, K., Nomura, M. and Yoshimura, I. (1998): Three-type algorithm for statistical analysis in chronic toxicity studies. J. Toxicol. Sci., **23**, 173-181.

Inaba, T. (1994): Problem of multiple comparison tests for evaluation of enzyme inhibitor XI. Bull. Jap. Soc. Biopharm. Stat., **40**, 33-36.

Jonckheere, A. R. (1954): A distribution-tree k-sample test against ordered alternatives. Biometrika, **41**, 133-145.

Kobayashi, K., Watanabe, K. and Inoue, H. (1995): Questioning the usefulness of the nonparametric analysis of quantitative data by transformation into ranked data in toxicity studies. J.Toxicol. Sci., **20**, 47-53.

Kobayashi, K. (1997a): A comparison of one- and two-sided tests for judging significant differences in quantitative data obtained in toxicological bioassay of laboratory animals. J. Occup. Health, **39**, 29-35.

Kobayashi, K., Ohori, K., Kobayashi, M. and Takeuchi, H. (1997b): Choice of method for statistical analysis of quantitative data obtained from toxicological studies. San Ei Shi, **39**, 86-92 (in Japanese).

Kobayashi, K., Ohori, K., Kanamori, M. and Takeuchi, H. (1997c): Comparison to Dunnett and *t*-test for two groups. Jap. Soc. Occup. Health, Tokai Branch, pp.56-57 (No. 203).

Kobayashi, K., Miura, D., Ohori, K., Kobayashi, M., Kanamori, M. and Takeuchi, H. (1997d): Statistics for grade data of pathological findings. 43th. Tokai-koshueisei-gagai, (No. 3-03).

Kobayashi, K., Miura, D., Watari, N., Yamamoto, T., Shoji, A., Kitajima, S., Shiya, M., An, R., Sho, H. and Takasaki, K. (1998): A view of Bartlett test for small sized animal setting. Shizuoka Experimental Animals Research Association, **25**, 8-13 (in Japanese).

Kobayashi, K., Kanamori, M., Ohori, K. and Takeuchi, H. (2000a): A new decision tree method for statistical analysis of quantitative data obtained in toxicity studies on rodent. San Ei Shi, **42**, 125-129 (in Japanese).

Kobayashi, K. (2000b): Trends of the decision tree for selecting hypothesis-testing procedures for the quantitative data obtained in the toxicological bioassay of the rodents in Japan. J. Environ. Biol., **21**, 1-9.

NIH(2006): http://ntp.niehs.nih.gov/index.cfm?objectid=084801F0-F43F-7B74-0BE549908B5E5C1C

MHLW (2006): http://wwwdb.mhlw.go.jp/ginc/html/db1-j.html

Nagano, K., Umeda, Y., Senoh, H., Gotoh, K., Arito, H., Yamamoto, S. and Matsushima, T. (2006): Carcinogenicity and chronic toxicity in rats and mice exposed by inhalation to 1, 2-dichloroethane for two years. J. Occup. Health, **48**, 424-436.

OECD (1995): Organization for Economic Cooperation and Development. OECD Guideline for the Testing of Chemicals. 407 Adopted 27.07.95. Repeated Dose 28-day Oral Toxicity Study in Rodents. Geneva.

Sakaki, H., Igarashi, S., Ikeda, T., Imamizo, K., Omichi, T., Kadota, M., Kawaguchi, T., Takizawa, T., Tsukamoto, O., Terai, K., Tozuka, K., Hirata, J., Handa, J., Mizuma, H., Murakami, M., Yamada, M. and Yokouchi, H. (2000): Statistical method appropriate for general toxicological studies in rats. J. Toxicol. Sci., **25**, 71-98 (in Japanese).

Sano, M. and Okayama, Y. (1990): Programs of Dunnett's multiple comparison test by one and two sided with high accuracy table. Bull. Jap. Soc. Biopharm. Stat., **32**, 21-44 (in Japanese).

Sumida, K., Saito, K., Ueda, K., Otsuka, M., Tsujimura, K., Miyaura, H., Sekijima, M., Nakayama, K., Kawano, Y., Kawakami, Y., Asamoto, M. and Shirai, T. (2006): Optimization of an animal test protocol for toxicogenomics studies (II). J.Toxicol. Sci., **32**, 47-56.

USEPA (2000): United States Environmental Protection Agency. Health Effects Test Guidelines. OPPTS 870. 3050. Repeated Dose 28-Day Oral Toxicity Study in Rodents. Washington, D.C., USA.

USFDA (2000): United States Food and Drug Administration. Short-Term Toxicity Studies with Rodents. Toxicological Principle for the Safety Assessment of Food Ingredients. Redbook 2000. Rockville, USA.

Yamazaki, M., Noguchi, Y., Tanda, M. and Shintani, S. (1981): Statistical method appropriates for general toxicological studies in rats. J. Takeda Res. Lab., **40** (3/4), 163-187 (in Japanese).

Yoshimura, I. and Ohashi, S. (1992): Statistical Analysis for Toxicology Data. pp23, Chijin-shokan, Tokyo (in Japanese).

Yoshimura, I. (1987): In "Statistical analysis of toxicological data", pp45, Scientist, Tokyo (in Japanese).

Mutation Research 655 (2008) 1–3

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test
I. Theoretical aspects

E. Lorge [a,*], M. Hayashi [b], S. Albertini [c], D. Kirkland [d]

[a] Servier Group, Drug Safety Assessment, BP 43255, 45403 Fleury-les-Aubrais, France
[b] Division of Genetics and Mutagenesis, NIHS, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan
[c] F. Hoffmann-LaRoche AG, Department of Toxicology, CH-4070 Basel, Switzerland
[d] Covance Laboratories Limited, Otley Road, Harrogate HG3 1PY, UK

ARTICLE INFO

Article history:
Received 28 January 2008
Received in revised form 28 May 2008
Accepted 3 June 2008
Available online 14 June 2008

Keywords:
Cytotoxicity
Genotoxicity
In vitro micronucleus test
Mouse lymphoma cells
Population doubling
Relative increase in cell count
Increased cell count

ABSTRACT

A decrease in the cytokinesis-block proliferation index (CBPI) or replication index (RI) is routinely used to determine cytotoxicity of a test compound and therefore the choice of its appropriate test concentration for the *in vitro* micronucleus (MN) test conducted in the presence of cytochalasin B. As a number of laboratories prefer to conduct the *in vitro* MN test in the absence of cytochalasin B, it is important that selected test concentrations, based on cytotoxicity, should be similar to what they would have been if cytochalasin B had been used, and should be relevant of a true cytotoxicity. By using models to analyse the dynamics of the cell cultures with and without cytochalasin B we have compared different methods for evaluation of cytotoxicity, and demonstrate that relative decrease in population doubling or relative increase in cell counts are the most appropriate measures of cytotoxicity to compare with reduction in CBPI or RI.

1. Introduction

The overall cytotoxicity of an agent in a cell culture results in both cytostasis and cell death. Cytostasis is not restricted to inhibitors of cell division but may also be the consequence of many cytotoxicity pathways leading to a delayed cell cycle. As cytostasis may result from effects on cell division and may also be cell death-related, this component of cytotoxicity could be predominant and therefore should be accurately taken into account in the overall cytotoxicity.

In the *in vitro* micronucleus test, when using cytochalasin B, two main methods for measurement of cytotoxicity have been recommended [1,2]. One is based on CBPI (Cytokinesis-Block Proliferation Index), where:

$$\mathrm{CBPI} = \frac{\text{no. mononucleated cells} + 2 \times \text{no. binucleated cells} + 3 \times \text{no. multinucleated cells}}{\text{total number of cells}}$$

CBPI is determined in treated ($\mathrm{CBPI_T}$) and control ($\mathrm{CBPI_C}$) cultures, and the amount of cytostasis induced by the treatment is determined as follows:

$$\%\,\text{cytostasis} = 100 - 100\left[\frac{\mathrm{CBPI_T} - 1}{\mathrm{CBPI_C} - 1}\right].$$

Another recommended measure is the replication index (RI). This directly gives the extent of cell replication in treated cultures relative to control as follows:

$$\mathrm{RI} = \frac{(\text{no. binucleated cells} + 2 \times \text{no. multinucleated cells})/\text{total number of cells treated cultures}}{(\text{no. binucleated cells} + 2 \times \text{no. multinucleated cells})/\text{total number of cells control cultures}} \times 100$$

Thus, the percentage of cytostasis = 100 − RI.

Many laboratories prefer to conduct the *in vitro* micronucleus test in the absence of cytochalasin B, so the above measures of CBPI and RI cannot be used. To avoid irrelevant positive results due to excessive cytotoxicity or other artifacts [3], it is important that selection of the top concentration (based on inducing a certain level of toxicity) should be similar in the absence or presence of cytochalasin B.

* Corresponding author. Tel.: +33 2 38 23 86 11.
E-mail address: elisabeth.lorge@fr.netgrs.com (E. Lorge).

1383-5718/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2008.06.003

where population doubling = [log(post-treatment cell number/initial cell number)]/log 2 or relative increase in cell counts (RICC):

$$\text{RICC} = \frac{\text{increase in number of cells in treated cultures (final - starting)}}{\text{increase in number of cells in control cultures (final - starting)}} \times 100$$

We decided to compare these two measures and the more traditional relative cell counts, with CBPI and RI in several different kinetic models. In the examples below, we show that kinetic evaluation of cell growth by RPD or RICC is preferable to simple cell counts and is the equivalent of CBPI-derived parameters used in the presence of cytochalasin B. This modeling exercise was paralleled by the publication of real data using different cytotoxicity measurements in the accompanying paper [5].

2. Modeling of cytotoxicity with and without cytochalasin B

In Fig. 1 we have drawn some theoretical examples of different types of cytotoxic effect in the presence and absence of cytochalasin B. In the first example, we simply study the situation where 50% of the cells have failed to divide. In the second example the effects of a higher concentration of the test substance, where 80% of the cells failed to divide, are examined. In the third example, in addition to inhibition of cell division, some cells have also died. Table 1 compares the CBPI-derived replication index in the presence of cytochalasin B with RPD, RICC and relative cell counts in the absence of cytochalasin B.

In all three examples, relative cell counts give a much higher estimate of survival (i.e. lower estimate of cytotoxicity) than RPD, RICC, CBPI or RI. In the second and third example, the use of relative cell counts would have led to the selection of too high concentrations, on the basis of 50% cytotoxicity, whereas RPD, RICC, CBPI and RI would all have resulted in selection of a lower concentration, since treatment with the concentration in the examples caused toxicity in the range 62.5–90%. In the first and second example, RPD and RICC (58.5 and 50%, respectively) are similar to both RI and CBPI (50%), whilst relative cell counts indicate a much higher toxicity (75%). In the third example, where cell death occurs, RPD and RICC derive from RI. This is due to the fact that, in the absence of cytochalasin B, RPD and RICC, do not distinguish individual cells that have divided. On the other hand, this apparent discrepancy is also due to the fact that RI is related to the total number of harvested cells, in this example decreased by two cells that died, whilst RPD and RICC are related to the initial number of treated cells, which remains constant. These are the limitations of both measurements and both methods.

These theoretical case studies show that relative cell counts are the least favourable cytotoxicity measurement to use in the *in vitro* micronucleus assay. This is not surprising as, in the number of cells counted on harvest, the loss of cells due to cytotoxicity is offset by the cell gain due to cell division occurring during treatment. In fact, at 50% reduction in cell counts too few cells have divided in the cultures to allow a meaningful evaluation of micronuclei in the absence of cytochalasin B.

In conclusion, simple cell counts were shown to underestimate cytotoxicity and, therefore, should not be recommended for the evaluation of cytotoxicity in the absence of cytochalasin B, as this method is likely to lead to the selection of excessively high top concentrations, susceptible to induce irrelevant positive results. Instead it is recommended that reduction in population doubling or reduction in relative increase in cell counts be used as cytotoxicity measures in the absence of cytochalasin B.

Conflicts of Interest

None.

References

[1] M. Kirsch-Volders, T. Sofuni, M. Aardema, S. Albertini, M. Fenech, M. Ishidate Jr., S. Kirchner, E. Lorge, T. Morita, H. Norppa, J. Surralles, A. Vanhauwaert, A. Wakata, Report from the in vitro micronucleus assay working group, Mutat. Res. 540 (2003) 153–163.
[2] M. Kirsch-Volders, T. Sofuni, M. Aardema, S. Albertini, D. Eastmond, M. Fenech, M. Ishidate Jr., S. Kirchner, E. Lorge, T. Morita, H. Norppa, J. Surralles, A. Vanhauwaert, A. Wakata, Corrigendum to "report from the in vitro micronucleus assay working group", Mutat. Res. 564 (2004) 97–100.
[3] D. Kirkland, M. Aardema, L. Henderson, L. Mueller, Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity, Mutat. Res. 584 (2005) 1–256.
[4] M.D. Fellows, M.R. O'Donovan, Cytotoxicity in cultured mammalian cells is a function of the method used to estimate it, Mutagenesis 22 (2007) 275–280.
[5] M.D. Fellows, M.R. O'Donovan, E. Lorge, D. Kirkland, Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test. II. Practical aspects with toxic agents, Mutat. Res. 655 (2008) 4–21.

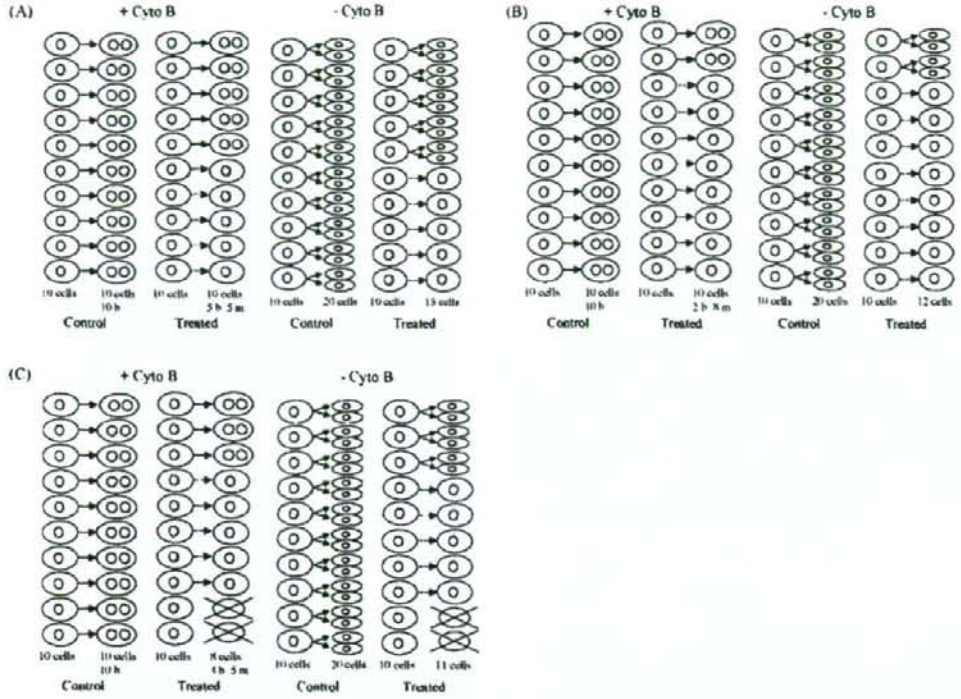


Fig. 1. Theoretical examples of cytotoxicity. (A) First example, inhibition of division in 50% of cells. (B) Second example, inhibition of division in 80% of cells (high concentration). (C) Third example, inhibition of cell division and cell death. b: binucleated; m: mononucleated.

Several different methods to measure cytotoxicity in the absence of cytochalasin B are available, but no clear recommendations have been made other than general statements in the IWGT report [1]. A similar effort of clarification has already been made in case cytochalasin B is used [2], but it is also required when cytochalasin B is not used. One commonly used method is the determination of relative cell counts (final counts in treated cultures relative to final counts in control cultures). However, this static measurement does not reflect all types of cytotoxicity, and therefore may lead to selection of excessively high concentrations, which may lead to irrelevant *in vitro* positive results [3]. Also, cytotoxicity may be different according to the method with which it is evaluated [4], so it is important to assess different measurements for their appropriateness. There are methods that can be used in the absence of cytochalasin B that better reflect the ability of the cultures to proliferate, for example relative population doubling (RPD):

$$\text{RPD} = \frac{\text{no. of population doublings in treated cultures}}{\text{no. of population doublings in control cultures}} \times 100,$$

Table 1
Comparison of different indicators of cell proliferation (and by deduction, indicators of cytotoxicity) in the presence or absence of cytochalasin B

	+CytoB		−CytoB		
	CBPI	RI	Population doubling	Increase in cell counts	Cell counts
	Cytokinesis-block proliferation index	Replicative index			
First example (Fig. 1A)					
Value in control	2.0	1.0	1.0	10.0	20.0
Value in treated	1.5	0.5	0.6	5.0	15.0
Relative cytostasis or cytotoxicity[a]	50.0%	50.0%	41.5%	50.0%	25.0%
Second example (Fig. 1B)					
Value in treated	1.2	0.2	0.3	2.0	12.0
Relative cytostasis or cytotoxicity[a]	80.0%	80.0%	73.7%	80.0%	40.0%
Third example (Fig. 1C)					
Value in treated	1.4	0.4	0.1	1.0	11.0
Relative cytostasis or cytotoxicity[a]	62.5%	62.5%	86.2%	90.0%	45.0%

[a] Calculated according to the formulae in the text.

承　諾　書

平成２１年　　月　　日

研究代表者　　広瀬　明彦　殿

機関名　国立医薬品食品衛生研究所

所属機関長　職　名　所長

氏　名　西島　正弘　　印

次の職員が、平成２１年度厚生労働科学研究費補助金の交付を受けて、化学物質リスク研究事業に係る次の課題の研究分担者として調査研究を実施することを承諾します。

研　究　課　題　名　　ナノマテリアルの健康影響評価手法の総合的開発および体内動態を含む基礎的有害性情報の集積に関する研究

職　　名	フ　リ　ガ　ナ 氏　　名
室　長	ｻﾄｳ　ｶｵﾙ 佐藤　薫

作成上の留意事項
１．所属機関長の印は、職印を使用すること。
２．研究課題名は、交付申請書と同じ研究課題名により記入すること。
３．日本工業規格Ａ列４番の用紙を用いること。

承　諾　書

平成２１年　　月　　日

研究代表者　　広瀬　明彦　殿

機関名　国立医薬品食品衛生研究所

所属機関長　職　名　所長

氏　名　西島　正弘　　印

次の職員が、平成２１年度厚生労働科学研究費補助金の交付を受けて、化学物質リスク研究事業に係る次の課題の研究分担者として調査研究を実施することを承諾します。

研　究　課　題　名　　ナノマテリアルの健康影響評価手法の総合的開発および体内動態を含む基礎的有害性情報の集積に関する研究

職　　名	フ　リ　ガ　ナ 氏　　名
室　長	ﾎﾝﾏ　ﾏｻﾐﾂ 本間　正充

作成上の留意事項
１．所属機関長の印は、職印を使用すること。
２．研究課題名は、交付申請書と同じ研究課題名により記入すること。
３．日本工業規格Ａ列４番の用紙を用いること。