J Occup Health 1997; 39: 29–35

Journal of Occupational Health

A Comparison of One- and Two-Sided Tests for Judging Significant Differences in Quantitative Data Obtained in Toxicological Bioassay of Laboratory Animals

Katsumi KOBAYASHI

Biosafety Research Center, Foods, Drugs and Pesticides (An-Pyo Center) and Department of Public Health, Hamamatsu University School of Medicine

Abstract: A Comparison of One- and Two-Sided Tests for Judging Significant Differences in Quantitative Data Obtained in Toxicological Bioassay of Laboratory Animals. Katsumi KOBAYASHI. Biosafety Research Center, Foods, Drugs and Pesticides— Since there are many ambiguous statements concerning the selection of one- or two-sided tests in the statistical analysis of toxicological data, I examined the rate of appearance of significant differences in the data showing a trend in either a fixed direction or a mixed sided direction compared with the control and the number of significant differences to these two tests by *t* and Dunnett's tests in a long-term chronic/carcinogenicity study conducted at the An-Pyo Center, in addition to referring to the most widely used statistical analyses by mean of the one- or two-sided test in the literature. The results were as follows; (1) Almost all quantitative data (578 out of 700 cases) showed a fixed trend with statistically significant differences ($p < 0.05$) compared with the control value. (2) The number of significant differences obtained with the one-sided test was greater than with the two-sided test in either analysis of the *t* or Dunnett's test; that is, the percentages of the significant differences in the two-sided test were 85 and 86% of those in the one-sided test by means of *t* and Dunnett's test, respectively. (3) The frequency of use of the one-sided tests was very low in both Japanese and international publications. Consequently, the one-sided test may be recommended for statistical analyses of toxicological bioassay data that generally show a fixed trend as compared with the control values, since more rigid evaluation of the data of the chemical effects on the living body and the environment is necessary.
(*J Occup Health 1997; 39: 29–35*)

Received May 23, 1995; Accepted May 24, 1996
Correspondence to: K. Kobayashi, Data Assessment and Information System Department, Biosafety Research Center, Foods, Drugs and Pesticides, Shioshinden, Fukude-cho, Iwata-gun 437-12, Japan

Key words: Statistical analysis, One- or two-sided test, Toxicological bioassay, *t*-test, Dunnett's test

Toxicological experiments with laboratory animals are required for registering new agricultural and medical drugs according to the guidelines provided by the Japan Government[1, 2]. The experiments consist of short-term (2–4 week) experiments to determine administration doses and long-term (78–104 week) experiments to assess chronic toxicity and oncogenicity/carcinogenicity. In general, one experiment involves 2–4 animal groups treated with the drug, one control group, and, if necessary, a few positive control groups. The items obtained from these toxicity studies are very numerous, 800–900 items in total, as shown in Table 1. The number of statistical analyses is 3 to 6 times according to the statistical test methods. The effects of test substance are assessed by decision trees[3] such as multiple range/comparison tests including an intergroup test, i.e., *t*-test or analysis of variance (ANOVA), and differences between the control and treated groups or among the groups are considered to be significant at a probability of 0.05 or less by one-sided or two-sided test.

The two-sided test is used in such cases as (1) only the presence or absence of an intergroup difference is questioned, (2) it is impossible to estimate in advance whether the intergroup difference would be plus or minus, and (3) the experimenter expects differences in both plus and minus, simultaneously. On the other hand, a one-sided test is applied for example when (1) the magnitude of difference is questioned, (2) it is possible to estimate in advance whether the intergroup difference would be plus or minus, or (3) the experimenter expects the one-pattern-difference alone to be important.

Yoshimura and Ohashi[4] recommended the one-sided test because the results of a toxicity study are

Table 1. Example of the number of tests for two-year toxicity study on rats

Item to be tested	Numbers of statistical analyses	Content (including times of analysis)
Body weight	66× 2*	Every week for 26 weeks and thereafter every other week
Food consumption	104× 2*	Every week
Food efficiency	52× 2*	Every week until 52 weeks
Hematology	16 items× 4 times× 2*	26, 52, 78 and 104 weeks after administration
Blood chemistry	20 items× 4 times× 2*	Ditto
Urinalysis	2 items× 4 times× 2*	Ditto
Organ weight	5 organs× 4 times× 2*	Ditto
Organ weight/Body weight		Ditto
Total	860	

*Sexes.

evaluated by the presence or absence of an increase in the mean value in many cases. The Kansai Applied Statistical Association[5] recommended the use of the two-sided test for Dunnett's multiple comparison test. Dunnett[6] stated that the selection of these tests is dependent on the policy of the study director. I report the rate of appearance of the significant differences in the data showing a trend in either fixed direction or mixed sided direction compared with the control and the number of significant differences in these two tests seen in the *t* and Dunnett's tests in a long-term chronic/carcinogenicity study conducted by the An-Pyo Center, in addition to referring to the most widely used statistical analyses by mean of one- or two-sided test in the literature.

Materials and Methods

The ratios of two- to one-sided test values were calculated by the *t*-test and Dunnett's multiple comparison test[7].

I examined the articles on toxicological studies in several Japanese and international journals to determine which test, two- or one-sided test, was used for statistical analysis of quantitative data. The Japanese journals were Iyakuhin Kenkyu (from 1991, Vol. 22, No. 1 to 1993, Vol. 24, No. 3), Pharmacometrics (from 1992, Vol. 47 to 1993, Vol. 48, No. 2), and the Japanese Journal of Cancer Research (from 1992, Vol. 83, No. 1 to 1993, Vol. 84, No. 12). The international journals were Toxicological and Applied Pharmacology (TAP: from 1992, Vol. 112, No. 1, Jan. to 1993, Vol. 119, No. 2, April), Food and Chemical Toxicology (FCT: from 1991, Vol. 29, No. 6 to 1993, Vol. 31, No. 4), and the National Toxicology Program Technical Report Series (NTP/NIH: from 1987 to 1992).

To determine whether the actual data from a toxicity study would show a unidirectional or bidirectional change in each treated group against the control group, the properties of items that showed significant differences in the one-sided *t*-test were examined in a chronic toxicity/carcinogenicity combined study (80 males and 80 females/group× 5 groups; 10 rats in each group sacrificed at 26, 52 and 78 weeks, and all survivors sacrificed by 104 weeks), in which the number of statistical analyses to be tested was the largest and the frequency of measurements was the highest.

The abovementioned study was selected from chronic toxicity/carcinogenicity combined studies conducted by Biosafety Research Center, Foods, Drugs and Pesticides. Significant differences ($p < 0.05$) were determined by one- and two-sided tests for the quantitative data, and the difference in the detection of significant differences was compared by the *t*-tests (Student, Aspin-Welch and Cochran)[3,7] and Dunnett's test[6].

Results

1. Properties of t table for two-and one-sided comparisons

Table 2 shows the distribution tables, for both the one- and two-sided tests at the 0.05 probability level, which are used for the *t*-test and Dunnett's multiple comparison test. As shown in this table, the value in the two-sided test was greater than 1/2 the value in the one-sided test, and it was approximately 0.78 in the *t*-table. According to Dunnett's distribution table, which is used in the establishment of 4 groups for toxicological study, the value in the one-sided test was not 1/2 the value in the two-sided test, but about 0.85.

2. Trend in literature examined

Table 3 shows the results obtained from the literature examined.

(1) In all of the Japanese articles reviewed, only the two-sided test was described, and there was no mention of the one-sided test.

(2) In 87 of 103 articles in the TAPs investigated, there was no indication of what test was

Table 2. Table of t for two- and one-sided comparisons

D. F.	Level at 0.05 probability level			
	t-test		Dunnett's multiple comparison test*	
	Two-sided value	One-sided value	Two-sided value	One-sided value
1	12.706	6.314	—	—
2	4.303	2.920	—	—
3	3.182	2.353	3.867	2.912
4	2.776	2.132	3.310	2.598
5	2.571	2.015	3.030	2.433
6	2.447	1.943	2.863	2.332
7	2.365	1.895	2.752	2.264
8	2.306	1.860	2.673	2.215
9	2.262	1.833	2.614	2.178
10	2.228	1.812	2.568	2.149
.	.	.	.	.
.	.	.	.	.
21	2.080	1.721	2.370	2.021
22	2.074	1.717	2.363	2.016
.	.	.	.	.
.	.	.	.	.
31	2.040	1.696	2.317	1.986
32	2.037	1.694	2.314	1.984
.	.	.	.	.
.	.	.	.	.
41	2.020	1.683	2.291	1.969
42	2.018	1.682	2.289	1.968
.	.	.	.	.
.	.	.	.	.
60	2.000	1.671	2.265	1.952
120	1.980	1.658	2.238	1.934
240	1.970	1.651	—	—
∞	1.960	1.645	2.212	1.916
Ratio**	1 :	0.78	1 :	0.85

*Included 4 groups. **The ratio of mean value for two-sided to one-sided.

Table 3. Present status of two-sided and one-sided tests employed in toxicity studies on experimental animals, Student's t-test and Dunnett's multiple comparison test

Journal	Mentions in literature		
	One-sided test	Two-sided test	No description
Iyakuhin Kenkyu	0	5	25
Pharmacometrics	0	3	58
Jpn. J. Cancer Res.	0	5	20
TAP	4	12	87
FCT	3	15	33
NTP	Nonparametric analysis*		

TAP: Toxicology and Applied Pharmacology. FCT: Food and Chemical Toxicology. NTP: National Toxicology Program. *By one-sided test.

Table 4. Properties of quantitative data from a combined study of chronic toxicity and carcinogenicity. Trend of significant differences between each treated group and the control group

Item	No. of statistical analyses	No. of significant differences compared with control	
		Unidirectional change	Bidirectional change
Body weight	132	130	0
Food consumption	208	156	22
Hematology	88	72	7
Blood chemistry	144	125	8
Urinalysis	16	9	0
Organ weight	56	40	2
Organ weight/body weight	56	46	0
Total	700	578	39

used; 12 used the two-sided test, and 4 used the one-sided test.

(3) In 33 of 51 articles in the FCTs investigated, there was no mention of what test (one- or two-sided test) was used; 15 used the two-sided test, and 3 the one-sided test. In FCT, Drewitt *et al.*[8] used the one-sided *t*-test in the main experiments and the two-sided *t*-test in the preliminary test. Shertzer *et al.*[9] determined significant differences between groups by the one-sided *t*-test when significant differences were observed by ANOVA in a metabolism test with rats.

(4) In NTPs, in a few studies analysis was by mean of Williams[10, 11] and Dunnett's multiple comparison test[6], and the two-sided test was used mainly for body weight and the weight of some organs[12, 13]; clinical chemistry, urinalysis, and hematology data were analyzed, and the nonparametric multiple comparison test[14, 15] was used for differences between the control and each treated group. Hematologic, serum chemical, and relative organ weight data[16] were analyzed by the nonparametric multiple comparison procedures of Dunnett[17] and Williams[10, 11]. Some NTP reports[18–20] were analyzed by only nonparametric multiple comparison tests of Dunn[14] or Shirley[15] in the one-sided test.

3. Properties of each quantitative datum from toxicity study

Table 4 shows the results. No body weight items measured showed a significant difference with a bidirectional change in values higher or lower than those in the control group. Food consumption was measured every week for 104 weeks; a total of 208 items were measured in both males and females. Significant differences in a unidirectional change in relation to the mean value in the control group were observed for 156 items. There were 22 items that showed significant differences with a bidirectional change in values higher and lower than those in the control group. Out of 88 items for hematological examination [HCT, HGB, RBC, MCV, MCH, MCHC, PLT, WBC, NEUT, LYMPH and reticulocyte; a total of 11 items tested 4 times (Weeks 26, 52, 78 and 104) × 2 sexes] , 72 showed significant differences in a unidirectional change in relation to the control group, and 7 items showed significant differences with a bidirectional change in values higher and lower than those in the control group. Out of 144 items for biochemical examination (glucose, total cholesterol, triglyceride, phospholipid, NEFA, BUN, total bilirubin, creatinine, total protein, albumin, sodium, potassium, chloride, calcium, inorganic phosphate, GOT, GPT and ALP; a total of 18 items to be tested 4 times × 2 sexes), 125 showed significant differences in a unidirectional change in relation to the control group, and 8 items showed significant differences with a bidirectional change in values higher or lower than those in the control group. Nine of 16 items for urinalysis (volume and specific gravity, a total of 2 items to be tested × 4 times × 2 sexes), showed significant differences in a single direction in relation to the control group, and there was no item showing a significant difference with a bidirectional change in values higher or lower than those in the control group. Forty of 56 items for determination of organ weight (brain, heart, liver, kidneys, spleen, adrenal glands, testes and ovaries; a total of 7 organs × 4 times × 2 sexes), showed significant differences in a unidirectional change against the control group, and 2 items showed significant differences with a bidirectional change in values higher and lower than those in the control group. Out of 56 items for calculation of the organ weight relative to body weight, i.e., the same as for determination of organ weight, 46 showed significant differences in a unidirectional change against the control group, and there was no item showing a significant difference with a unidirectional change in values higher or lower than those in the control group. These trends, unidirectional change in each treated group against the control group were observed in another toxicological report[21–23] on rodents.

Some results of hematological and biochemical examinations show a unidirectional change in the low-, middle- and high-dose groups, as compared with the control group, but the maximum-dose group occasionally displays the opposite direction. This phenomenon may be due to the extremely unbalanced bio-reaction to a very massive dose of the test substance. The test items in the blood chemistry revealing these phenomena were creatinine, phospholipid, total cholesterol and chloride, and they were relatively often observed during examination in the early stage after administration of the drug. Out of 700 items for all tests, 578 showed significant differences in unidirectional change in relation to the control group, and 39 items indicated significant differences with a bidirectional change in values higher and lower than those in the control group. The ratio of the bidirectional pattern (39) to the unidirectional pattern (578) was 1 to 15.

4. Difference in detection of significant differences between two- and one-sided tests

Table 5 shows the results. In analyses with the *t*-test, the number of significant differences detected by the two-sided test was smaller than that detected by the one-sided test, 85% on average (76–95%).

Table 5. Difference in the number of detected significant ($p<0.05$) differences between the one- and two-sided tests in a combined study of chronic toxicity and carcinogenicity

Item	No. of statistical analyses	*t*-test		Dunnett's Multiple comparison test	
		One-sided	Two-sided	One-sided	Two-sided
Body weight	528	246 (100)	233 (95)	223 (100)	212 (95)
Food consumption	832	349 (100)	279 (80)	235 (100)	189 (80)
Hematology	352	159 (100)	126 (79)	123 (100)	105 (85)
Blood chemistry	576	272 (100)	235 (86)	215 (100)	181 (84)
Urinalysis	64	11 (100)	10 (91)	7 (100)	5 (71)
Organ weight	224	80 (100)	61 (76)	47 (100)	42 (89)
Organ weight/body weight	224	104 (100)	89 (86)	82 (100)	67 (81)
Total	2,800	1,221 (100)	1,033 (85)	932 (100)	801 (86)

( ): % of one-sided value in each test.

In analyses with Dunnett's test, the number of significant differences detected by the two-sided test was also smaller than that detected by the one-sided test, 86% on average (71-95%). In other words, the tendencies of the frequencies of detection by the one-sided and two-sided tests, which were revealed by the *t*-test, were similar to those revealed by Dunnett's test[6]. Moreover, the frequency of detection by the two-sided test was about 85% of the frequency by the one-sided test.

Discussion

The following were typical opinions in the 6 references surveyed with regard to selection of the one- or two-sided test: (1) Shirley *et al.*[15] used the two-sided test for Student's *t*-test and Cochran's *t*-test, and if significant differences were observed in ANOVA, they used the one-sided test in Dunnett's multiple comparison test. (2) Dunnett[6] recommended use of the two-sided test to determine simultaneous upper and lower limits for the difference between the control group and each treated group; he used the one-sided test to determine either the upper or lower limit on the difference between the control group and each treated group. (3) Gad *et al.*[24] explained the significant difference between control and treated groups in body weight by using the two-sided test. (4) Yoshimura *et al.*[4, 7] recommended using the one-sided test, because toxicity is evaluated as present or absent by the presence or absence of an increase in the mean values in toxicological study data in many cases. Additionally, quantitative data obtained in the toxicological test should be used in the one-sided test when a difference compared with the control group is expected, and this difference can be supposed before the experiment. The two-sided test should be employed when it cannot be supposed before the experiment that it provides a unidirectional pattern compared with the control group as a result of the study. The change in hypothesis can be accompanied with great danger of error of the first kind. It is very important to increase the power of detection of the differences among the groups. Generally in toxicity studies, the experimenter can make a decision according to the findings in the same or less biological or pharmacological effects after the preliminary studies. (5) Sakuma[25] emphasized deciding which selection would be suitable for the content of the study, and using a report of the same type of experiment for reference. In screening tests for new drugs, the two-sided test is recommended. The experimenter should not change the hypothesis from a two-sided to a one-sided test after the test has been finished. (6) Nakamura[26] also stated that selection of the tests depends on the purpose and content of the study, and the statistical significance of the data should not be foreseen. (7) Ishii[27] stated that it is necessary to the select properly according to the situation in which the difference between two cases has to be considered to be either plus or minus alone and in which the difference has to be considered to be both plus and minus.

All of these opinions suggest that selection of the one- or two-sided test for a toxicological study on experimental animals depends on the purpose of the study, as determined by the experimenter.

Significant differences in quantitative data in toxicity studies on experimental animals are determined according to the one- or two-sided test. The selection of which test to use is dependent upon the judgment of the study director. Investigation of actual test reports shows that the frequency of use of the one-sided test was very low. Some authors[4, 8] have

reported using the two-sided test for preliminary tests to establish the dose, because it is not known whether each quantitatively determined value would be higher or lower than in the control group and that the one-sided test is or should be used for the main test, and because the values are acknowledged to incline toward a unidirectional change in relation to the control group. On the other hand, in an assessment of the tendency to detect significant differences from the actual levels in the chronic toxicity and carcinogenicity combined test composed of 5 groups, 578 of the 700 test items showed significant differences in unidirectional change in relation to the control group, and 39 showed significant differences with a bidirectional change in values higher and lower than those in the control group. The reasons for the high frequency of a mixture of high and low values in food consumption may be the habits of the animals themselves, scraping out the feed, spilling feed, and contamination with urine. With regard to the above test results, the difference in the number of detected significant differences between the one- and two-sided tests by the *t*-test and Dunnett's multiple comparison test was analyzed. The frequency of detection by the two-sided test was 85 and 86% of that determined by the one-sided tests. These ratios of 0.78 and 0.85 for two-sided to one-sided distribution tables (Table 2) for the two statistical methods were in good agreement.

Significant differences are more apt to be observed in the one-sided test than in the two-sided test. Since the data from a toxicity study become the referential values for calculating the safe dose for medication in humans in the case of drugs and for deciding the daily maximum tolerable dose in the case of agricultural chemicals, strict toxicity checks for humans will be necessary. The one-sided test with high detection power is therefore recommended to be used for the analysis of significant differences in quantitative data from main toxicity studies except for short-term preliminary range finding studies on experimental animals.

References

1) Japanese Ministry of Agriculture, Forestry and Fisheries. Guideline on Toxicology Study Data for Application of Agricultural Chemical Registration, 59 NohSan No. 4200. Tokyo: 1985 (in Japanese).
2) Japanese Ministry of Health and Welfare. Guidelines for Toxicity Studies of Drugs, Notification No. 1–24 of the Pharmaceutical Affairs Bureau. Tokyo: 1989 (in Japanese).
3) Gad S, Weil CS. Student's *t*-test, Chochran *t*-test. In: Wallace HA. ed. Statistics for toxicologists, principles and methods of toxicology. New York: Raven Press, 1982: 286–288.
4) Yoshimura I, Ohashi S. Statistical analysis for toxicology data. Tokyo: Chijin shokan, 1992: 92, 96–98, 110 (in Japanese).
5) Kansai Applied Statistical Association. Statistical analysis for toxicity studies. Toxicology-forum 1987; 10: 543 (in Japanese).
6) Dunnett CW. A multiple comparison procedure for comparing several treatment with a control. J Am Stat Assoc 1955; 50: 1096–1121.
7) Imamizo Y, Koda S, Sano M, et al. *t*-test, Welch's test, Dunnett's test, *t*-tables, Dunnett's tables In: Yoshimura I, ed. Statistical analyses for toxicological data. Tokyo: Scientist, 1987: 22, 30–37, 61–62, 248, 263–264 (in Japanese).
8) Drewitt PN, Butterworth CD, Springall CD, Moorhouse SR. Plasma levels of aluminum after tea ingestion in healthy volunteers. Fd Chem Toxic 1993; 31: 19–23.
9) Shertzer HG, Sainsbury M. Chemoprotective and hepatic enzyme induction properties of indole and indenoindole antioxidants in rats. Fd Chem Toxic 1991; 29: 391–400.
10) Williams DA. A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics 1971; 27: 103–117.
11) Williams DA. The comparison of several dose levels with a zero dose control. Biometrics 1972; 28: 519–531.
12) National Institutes of Health. Toxicology and carcinogenesis studies of quercetin in F344/N rats. National Toxicology Program Technical Report Series, 1992; No. 409: 20.
13) National Institutes of Health. Toxicology and carcinogenesis studies of polysorbate 80 in F344/N rats and B6C3F$_1$ mice. National Toxicology Program Technical Report Series, 1992; No. 415: 19–20.
14) Dunn OJ. Multiple comparisons using rank sums. Technometrics 1964; 6: 106–107.
15) Shirley E. A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics 1977; 33: 386–389.
16) National Institutes of Health. Toxicology and carcinogenesis studies of hexachloroethane in F344/N rats. National Toxicology Program Technical Report Series. 1989; No. 361: 25–26.
17) Dunnett CW. Parewise multiple comparisons in the unequal variance case. J Am Stat Assoc 1980; 75: 796–800.
18) National Institutes of Health. Toxicology and carcinogenesis studies of two pentachlorophenol technical-grade mixtures in B6C3F$_1$ mice. National Toxicology Program Technical Report Series. 1989; No. 349: 47.
19) National Institutes of Health. Toxicology and carcinogenesis studies of 2-chloroacetophenone in F344/N rats and B6C3F$_1$ mice. National Toxicology Program Technical Report Series. 1990; No. 379: 23–24.
20) National Institutes of Health. Toxicology and carci-

nogenesis studies of 2-probenecid in F344/N rats and B6C3F$_1$ mice. National Toxicology Program Technical Report Series. 1991; No. 395: 20.
21) Furukawa S, Nishida I, Kikumori M, et al. Twenty-six-week intravenous toxicity study of montirelin hydrate (NS-3) in rats followed by 9-week recovery test. J Toxicol Sci 1995; 20 Sup. II: 191–213.
22) Leibbrandt EI, Wolfgang GHI. Differential toxicity of cisplatin, carboplantin, and Cl-973 correlates with cellular platinum levels in rats renal cortical slices. Toxicol Appl Pharmacol 1995; 132:245–252.
23) Vaittinen SL, Komulainen H, Kosma VM, et al. Subchronic toxicity of 2-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX) in wistar rats. Fd Chem Toxic 1995; 33: 1027–1037.
24) Gad S, Weil CS. Statistics and experimental design for toxicologists. New Jersey: The Telford Press, 1986: 71–72.
25) Sakuma A. Statistical methods in pharmacometrics I. Tokyo: Tokyodaigaku Shupankai, 1977: 56 (in Japanese).
26) Nakamura G. Practice, statistical analyses. Tokyo: Kaimei-Sha, 1986: 106–107 (in Japanese).
27) Ishii S. Bio-statistical analyses guidance. Tokyo: Baihukan, 1975: 68 (in Japanese).