The Journal of Toxicological Sciences (J. Toxicol. Sci.)
Vol.37, No.2, 255-260, 2012

Original Article

Determination of dose dependence in repeated dose toxicity studies when mid-dose alone is insignificant

Katsumi Kobayashi[1,2], Kalathil Sadasivan Pillai[3], Mathews Michael[3], Kotturathu Mammen Cherian[3], Akihiro Araki[4] and Akihiko Hirose[5]

[1]Chemical Management Center, National Institute of Technology and Evaluation (NITE), 49-10, Nishihara-nichome, Shibuya-ku, Tokyo 151-0066, Japan
[2]Shizuoka Prefectural Agriculture and Forestry College, 678-1, Tomioka, Iwata, Shizuoka 438-8577, Japan
[3]Frontier Life Science Services, A Unit of Frontier Lifeline Hospitals & International Centre for Cardio-Thoracic and Vascular Diseases, Chennai- 600101, India
[4]Technology Development Division, Occupational Health Research and Development Center, Japan Industrial Safety & Health Association, 35-2, Shiba-gochome, Minato-ku, Tokyo 108-0014, Japan
[5]National Institute of Health Sciences,18-1 Kamiyoga-ichome, Setagaya-ku, Tokyo 158-8501, Japan

(Received November 14, 2011; Accepted December 13, 2011)

ABSTRACT — Repeated dose toxicity studies with rodents are regulatory requirements for registering chemical substances like drugs and pesticides with the government regulatory agencies. Usually 4 groups of animals, including a control group, are used in repeated dose toxicity studies. Williams' test, Dunnett's test and Jonckheere's trend test are generally used to evaluate the data obtained from these studies. Selection of a statistical tool is relatively easy, when the data obtained from the groups of animals show a dose-dependency. But, occasionally a significance difference, compared to control, is not seen in the mid-dose group alone, thus losing the dose-dependency. We attempted to find the appropriate statistical tool for analyzing the quantitative data obtained from repeated dose toxicity studies, when the data of the mid-dose group alone do not show a significant difference, compared to control. The commonly used Williams' test to analyse such data has a disadvantage as it assigns an estimated mean value for the mid-dose group, rather than the original mean value, for the analysis. Hence, it is likely that Williams' test may misjudge in establishing a dose dependency, when in reality it does not exists. Therefore, to analyse such data we suggest the use of Dunnett's multiple comparison test, to compare each dose group with the control, followed by Jonckheere's trend test for examining dose dependency.

Key words: Dose related relationship, Dunnett's test, Williams' test, Toxicity study, Jonckheere's trend test

INTRODUCTION

One of the main objectives of conducting repeated-dose toxicity studies is to arrive at NOEL/NOAEL (no observed effect level/no observed adverse effect level) (OECD, 2008), which is an important part of the non-clinical risk assessment (Dorato and Engelhardt, 2005). Most of the regulatory guidelines prescribe that the repeated-dose toxicity studies with rodents should be conducted with a minimum of three treatment doses (low, mid and high doses) and a control (OECD, 2008). The high dose is chosen with the aim to induce toxicity but not death or severe suffering (OECD, 2008; EPA, 2000), whereas the low dose is chosen with the assumption that animals exposed to this dose level will not show any effect of the treatment compared to the control (Kobayashi *et al*., 2010). However, these guidelines do not mention how to choose the mid-dose, except an indication that this dose is required to examine dose dependency. According to Gupta (2007), the mid-dose selection should be considered threshold in toxic response and mechanism of toxicity. Choosing the mid-dose is as critical as choosing the high and low doses in repeated dose toxicity studies, since mid-dose plays a determining role in establishing the dose dependency. It is not uncommon to encounter situations where mid-dose alone shows an insignificant dif-

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

ference compared to control, whereas low and high doses show a significant difference. Williams' test is generally carried out to test dose-related trend (Bretz, 2006). The disadvantage of Williams' test is that it uses an estimated value for the mean rather than the original mean value for the analysis. The present paper discusses interpretation of data obtained from repeated dose toxicity studies anlysed by Williams' test in comparison with Dunnett's and Jonckheere's tests.

RESULTS

Simulated and actual data were used for the present study. The analyses were conducted using Excel 2008 (Tokyo, Japan).

Statistical significance and dosage dependence

When the data show a dose related relationship and a significant difference by Dunnett's test (Dunnett, 1955), the interpretation of them can be done in a straight forward manner. In a 4 group-set repeated dose study seven different situations can be expected (Table 1). Interpretation is relatively easier in situations 1-3, whereas it is difficult for situations 4-7, where further investigation on dose related relationship is required.

Situation 1 given in Table 1 is ranked and given in Table 2. When the data were adjusted to simulate situations 4-7 given in Table 1, dose related relationship analysis by Jonckheere's test (Jonckheere, 1954) did not reveal a significant dose related relationship in these situations (Table 3).

Next the data presented in Table 1 are subjected to

Table 1. Significant difference shown by the treatment groups by Dunnett's test – Possible situations

Test group	●: Significant difference, ○: No significant difference from the control group						
	Situation 1	Situation 2	Situation 3	Situation 4	Situation 5	Situation 6	Situation 7
Control	○	○	○	○	○	○	○
Low dose	●	○	○	●	●	○	●
Mid-dose	●	●	○	○	●	●	○
High dose	●	●	●	○	○	○	●
Investigation	Not required	Not required	Not required	Required	Required	Required	Required
Visual dose related relationships	Yes	Yes	Yes	No	No	No	No

Table 2. Ranking pattern of data of situation 1 given in Table 1 (five rats/ groups)

Control	Rank/group		
	Low dose	Mid-dose	High dose
1	6	11	16
2	7	12	17
3	8	13	18
4	9	14	19
5	10	15	20

Table 3. Dose related relationship analysis by Jonckheere's test of situation 4-7 given in Table 1

Statistic	Situation 4	Situation 5	Situation 6	Situation 7
P value	0.9108	0.6317	0.3682	0.0891
Significance	All situations are not significant			

Table 4. Judging NOEL by Williams' test (situations of Table 1 are reproduced here)

Test group	Significant difference ($P < 0.05$) from control group						
	Situation 1	Situation 2	Situation 3	Situation 4	Situation 5	Situation 6	Situation 7
Low dose	Yes	None	None	Yes	Yes	None	Yes
Mid-dose	Yes	Yes	None	None	Yes	Yes	None
High dose	Yes	Yes	Yes	None	None	None	Yes
NOEL by Williams' test	< Low	Low	Mid	High	High	High	Mid

Williams' test to arrive at NOEL and the NOEL of each situation and is given in Table 4.

Analysis of actual data using Dunnett's and Williams' tests and Jonckheere's trend test

Absolute kidney weights of the rats at the scheduled sacrifice in a combined repeated dose reproduction/developmental toxicity study as per OECD TG 422 (OECD, 1996) is given in Table 5. Analyses of the data using Dunnett's and Williams' tests and Jonckheere's trend test provided different results. Dunnett's test showed a significant difference in low and high dose groups, whereas Williams' test showed in all the groups. No significant difference in any group was seen as per Jonckheere's trend test. According to Williams' test, less than the low dose may be considered as NOEL.

Williams' test can be used even if the numbers of the animals in a group differ about 2 times compared to other groups (Sakaki *et al.*, 2000). However, if the number of animals in a group is extremely less use of Williams' test is not recommended (Williams, 1971, 1972).

Williams' test analyzes the difference of the mean values between each treated group and control, like Dunnett's and Tukey's tests, when the mean value of the

Table 5. Absolute kidney weights of rats at scheduled sacrifice

Item and analysis	Test group			
	Control	Low dose	Mid-dose	High dose
Individual data (g)	2.558	3.269	3.116	2.706
	2.789	3.428	2.791	3.293
	2.764	3.083	2.981	3.535
	2.707	3.532	3.337	3.387
	2.793	3.546	2.432	3.064
	3.041	2.677	2.934	3.102
	3.000	2.822	3.388	3.279
	-	3.656	2.911	-
	-	3.271	2.798	-
	-	3.348	3.208	-
	-	3.031	2.876	-
	-	3.742	2.703	-
Number of animal	7	12	12	7
Mean ± S.D.	2.807 ± 0.167	3.284 ± 0.329	2.956 ± 0.273	3.195 ± 0.269
Bartlett's homogeneity test	$P = 0.4130$ (No heterogeneity)			
Dunnett's test	-	$P = 0.0026$**	$P = 0.5190$	$P = 0.0332$*
Mean value used for Williams' test	2.807	3.284	3.120	3.195
Williams' test	-	$P < 0.05$*	$P < 0.05$*	$P < 0.05$*
Jonckheere's trend test	No significant difference			

*$P < 0.05$ and **$P < 0.01$ from control group.

Table 6. Liver weights of rats in a 4-week repeated dose toxicity study

Test group	Liver weight (g), N = 5, (Sum)	Mean ± S.D.	Result of Dunnett's test	Mean for Williams test	Result of Williams' test
Control	10.7, 11.5, 11.6, 12.0, 11.0 (56.8)	11.36 ± 0.51 (100.0) [a]	-	11.36 (100.0) [a]	-
Low dose	11.6, 12.3, 12.5, 12.3, 12.7 (61.4)	12.28 ± 0.41 (108.1)	$P < 0.05$	12.28 (108.1)	$P < 0.05$
Mid-dose	11.2, 11.5, 11.6, 11.5, 11.5 (57.3)	11.46 ± 0.15 (100.9)	Not significant	11.87 (104.5)	$P < 0.05$
High dose	12.2, 12.5, 12.0, 11.9, 13.0 (61.6)	12.32 ± 0.44 (108.5)	$P < 0.05$	12.32 (108.5)	$P < 0.05$

[a]In % of control

treated groups changes in one direction. The example given in Table 6 does not show a dose-dependability as the mid-dose showed an insignificant liver weight compared to control. When the data were analysed by Williams test, significance in the liver weight is observed in the mid-dose group. The reason for this may be better explained by elucidating the calculation procedure of Williams' test given below:

Calculation procedure of Williams' test:

(1) Control vs. High dose

$$\frac{61.4+57.3+61.6}{5+5+5}=12.02$$

$$\frac{57.3+61.6}{5+5}=11.89$$

$$\frac{61.6}{5}=12.32 \quad \leftarrow \text{This largest value is used.}$$

$$t=\frac{|11.36-12.32|}{\sqrt{0.16375\left(\frac{1}{5}+\frac{1}{5}\right)}}=3.751$$

0.16375; variance in error in ANOVA table.

t value is significant at 5% level (Table 7, Number of groups- 4; D.F.- 16).

(2) Control vs. Mid-dose

$$\frac{61.4+57.3}{5+5}=11.87 \quad \leftarrow \text{This largest value is used.}$$

The average value of mid-dose and low dose group is used.

$$\frac{57.3}{5}=11.46$$

$$t=\frac{|11.36-11.87|}{\sqrt{0.16375\left(\frac{1}{5}+\frac{1}{5}\right)}}=1.993$$

t value is significant at 5% level (Table 7, Number of groups- 3; D.F.- 16).

(3) Control vs. Low dose

$$\frac{61.4}{5}=12.28$$

$$t=\frac{|11.36-12.28|}{\sqrt{0.16375\left(\frac{1}{5}+\frac{1}{5}\right)}}=3.595$$

t value is significant at 5% level (Table 7, Number of groups- 2; D.F.- 16).

The reason for Williams' test showing a significant difference of the weight of the liver, when compared with the control group, is that the test used 11.87 as the mean value of the mid-dose group for the comparison instead of the actual value (11.46).

Table 7. Williams' table

D.F.	Number of groups							
	2	3	4	5	6	7	8	9
15	1.753	1.839	1.868	1.882	1.891	1.896	1.900	1.903
16	**1.746**	**1.831**	**1.860**	1.873	1.882	1.887	1.891	1.893
17	1.740	1.824	1.852	1.866	1.874	1.879	1.883	1.885

Table 8. The change in significant difference pattern as per Jonckheere's trend test when the number of animals/group changes

Number of animals/group[a]	Mean rank				Probability (*P*)
	Control	Low dose	Mid-dose	High dose	
4	7	8	9	10	0.131
8	15	16	17	18	0.209
10	19	20	21	22	0.234
20	39	40	41	42	0.303
30	59	60	61	62	0.336

[a]Number of animals/group is selected in such a way that the mean ranks of the groups are more or less the same.

Jonckheere's trend test

The prerequisite conditions for applying Jonckheere's trend test are that the number of groups should be more than 2 and number of animals in each group should be the same. This trend test has been used in several toxicity tests (Bamberger *et al.*, 2000; Campbell, 2003, Ladics *et al.*, 2008; Stout *et al.,* 2008). As per Jonckheere's trend test, probability level of significance decreases with the decrease in number of the animals in groups (Table 8).

In the example given in Table 9 the order of mean scores of the control, low and mid-dose groups are interchanged to know how these changes would influence Jonckheere's trend test in detecting a significant high dose group. In both the cases where number of animals/group is 5 or 10, Jonckheere's trend test revealed a significant high dose group, indicating that order of scores does not influence the power of Jonckheere's trend test.

We examined the significant difference detection pattern when the number of groups is increased (Table 10). It is evident from the Table that, power for a significant difference increases with the increase of number of groups.

Table 9. Power of Jonckheere's trend test in finding a significant difference in high dose group, when the order of ranks of control, low and mid dose groups interchanged

Test group	Significant difference pattern from control group by Dunnett's test		Mean rank							
			N = five/group				N = ten/group			
Control	○		7.0	8.0	9.0	9.0	14.5	15.5	16.5	16.5
Low dose	○	(None)	8.0	7.0	7.0	8.0	15.5	14.5	14.5	15.5
Mid-dose	○	(None)	9.0	9.0	8.0	7.0	16.5	16.5	15.5	14.5
High dose	●	(Yes)	18.0	18.0	18.0	18.0	35.5	35.5	35.5	35.5
Probability (P)[a]			0.0012**	0.0036**	0.0093**	0.0218*	0.0000**	0.0001**	0.0002**	0.0006**

[a]By Jonckheere's trend test; *$P < 0.05$, **$P < 0.01$.

Table 10. The change in significant difference pattern as per Jonckheere's trend test when the number of groups is increased

Number of group	Five animals/group						Probability (P)
	Control (X0)	Low dose (X1)	Mid-dose (X2)	High dose (X3)	Second top dose (X4)	Top dose (X5)	
3	4.0	5.0	10.5	-	-	-	0.0042
4	7.0	8.0	9.0	18.0	-	-	0.0012
5	9.0	10.0	11.0	12.0	23.0	-	0.0009
6	11.0	12.0	13.0	14.0	15.0	28.0	0.0008

DISCUSSION

In repeated dose toxicity studies, the role of statistics is pivotal for interpreting study data. Williams' test is a useful statistical tool in these studies as it provides information on evidence of toxicity and also the dose level that causes the toxicity (Shirley, 1977). Williams' test is similar to Dunnett, Tukey and Duncan multiple comparison (range) tests as it uses the variance in error term of the ANOVA (Nagata and Yoshida, 1997). But, use of Williams' test is not recommended when the number of animals in the groups is different (Williams, 1972). Williams' test is a closed procedure. If no significant difference between control group and highest dose group is observed, all the other treated groups are considered having no significant difference compared to the control group and no further analysis is carried out. If there is a significant difference in the highest dose group, then the next highest dose level is examined for the significant difference from the control. If this dose group does not show a significant difference, no further analysis is carried out and if it shows a significant difference, the next highest dose level is examined for the significant difference from the control. Thus all the dose groups are sequentially examined.

Jonckheere's trend test is commonly used in toxicology for the analysis of dose related relationship (Neuhäuser *et al*., 1998; Neuhäuser and Hothorn, 1998; Tennekes *et al*., 2004). A dose related relationship is not usually detected either by Dunnett's or Jonckheere's trend test, when the mid-dose alone does not show a significant difference compared to control. Jonckheere's test is sensitive to non-monotonic dose related relationship, whereas Williams' test is powerful against monotonic and non-monotonic dose related relationship (Dmitrienko *et al.*, 2007). Since estimated mean values are used in the calculation procedure of Williams' test, it is likely that this test might show dose related relationship, where it actually does not exist. It also may be noted in this context that, according to Gad and Weil (1986) dose related relationship is not necessarily evident in all the parameters. Therefore, we suggest to examine the data for the difference between each dose group and control by Dunnett's test and then examine the data by Jonckheere's trend test for dose related relationship.

REFERENCES

Bamberger, J.R., Ladics, G.S., Hurtt, M.E., Swanson, M.S. and Brocks, W.J. (2000): Subchronic inhalation toxicity of the chlo-

rinated propane 1,1,1,3,3,3-hexachloropropane (HCC-230fa). Toxicol. Sci., **62**, 155-165.

Bretz, F. (2006): An extension of the Williams trend test to general unbalanced linear models. Comp. Stat. Data Anal., **50**, 1735-1748.

Campbell, M. (2003): Evidence on the developmental and reproductive toxicity of phenol. Environmental health hazard assessment's reproductive and cancer hazard assessment section. Reproductive and cancer hazard assessment section office of environmental health hazard assessment California environmental protection agency.

Dmitrienko, A., Chaung-Stein, C. and D'Agostino, R. (2007): Pharmaceutical statistics using SAS-A practical guide. pp.444, SAS Institute, NC, USA.

Dorato, M.A. and Engelhardt, J.A. (2005): The no-observed-adverse-effect-level in drug safety evaluations: Use, issues, and definition(s). Reg. Tox. Pharmacol., **42**, 265-274.

Dunnett, C.W. (1955): A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc., **50**, 1096-1121.

EPA (2000): United States Environmental Protection Agency. Health Effects Test Guidelines, OPPTS 870.3050, Repeated dose 28-day oral toxicity study in rodents, EPA 712-C-00-366 2000

Gad, S. and Weil, C.S. (1986): Statistics and experimental design for toxicologists. pp. 86, Telford Press, New Jersey, USA.

Gupta, R.C. (2007): Veterinary toxicology-basic and clinical principles. pp.1203, Academic Press, New York, USA.

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

Kobayashi, K., Pillai, K.S., Michael, M., Cherian, K.M. and Ohnishi, M. (2010): Determining NOEL/NOAEL in repeated-dose toxicity studies, when the low dose group shows significant difference in quantitative data. Lab. Anim. Res., **26**, 133-137.

Ladics, G.S., Kennedy, G., O'Connor, J., Everds, N., Malley, L.A., Frame, S.R. Gannon, S., Jung, R., Roth, T. Iwai, H. and Shin-ya, S. (2008): 90-day oral gavage toxicity study of 8-2 fluorotelomer alcohol in rats. Drug and Chemical Toxicology, **31**, 189-216.

Nagata, Y. and Yoshida, M. (1997): Tokeiteki-tajuhikakuho-no-kiso. pp.45-52 and 176-177, Scientist Co., Ltd, Tokyo.

Neuhäuser, M. and Hothorn, L.A. (1998): An analogue of Jonckheere's trend test for parametric and dichotomous data. Biomaterial. J., **40**, 11-19.

Neuhäuser, M., Liu, P.Y. and Hothorn, L.A. (1998): Nonparametric tests for trend: Jonckheere's test, a modification and a maximum test. Biomaterial. J., **40**, 899-909.

OECD (1996): Organization for Economic Co-operation and Development. OECD Guideline for Testing of Chemicals. Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test, No. 422. OECD, France.

OECD (2008): Organization for Economic Cooperation and Development. OECD Guideline for Testing of Chemicals. Repeated dose 28-day oral toxicity study in rodents. No. 407, OECD, France.

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.

Shirley, E. (1997): A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics, **33**, 386-389.

Stout, M.D., Bodes, E., Schoonhoven, R., Upton, P.B., Travlos, G.S. and Swenberg, J.M. (2008): Toxicity, DNA binding, and cell proliferation in male F344 rats following short-term gavage exposures to trans-2-hexenal. Toxicol. Pathol., **36**, 232-246.

Tennekes, H., Kaufmann, W., Dammann, M. and van Ravenzwaay, B. (2004): The stability of historical control data for common neoplasms in laboratory rats and the implications for carcinogenic risk assessment. Reg. Tox. Pharmacol., **40**, 293-304.

Williams, D.A. (1971): A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics, **27**, 103-117.

Williams, D.A. (1972): The comparison of several dose levels with a zero dose control. Biometrics, **28**, 519-531.