SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 19, No. 1, pp. 113–117, 2000
Printed in the USA
0730-7268/00 $9.00 + .00

Annual Review

APPLICATIONS AND STATISTICAL PROPERTIES OF MINIMUM SIGNIFICANT DIFFERENCE-BASED CRITERION TESTING IN A TOXICITY TESTING PROGRAM

QIN WANG,† DEBRA L. DENTON,‡ and RAKESH SHUKLA*§
A†Biostatistics and Epidemiology, Abt Associates Clinical Trials, 55 Wheeler Street, Cambridge, Massachusetts 02138, USA
‡U.S. Environmental Protection Agency, Region IX, 75 Hawthorne Street WTR-5, San Francisco, California 94105-3901
§Division of Epidemiology and Biostatistics, Department of Environmental Health, University of Cincinnati Medical Center, Cincinnati, Ohio 45267-0056, USA.

(*Received* 4 *January* 1999; *Accepted* 17 *July* 1999)

Abstract—As a follow up to the recommendations of the September 1995 SETAC Pellston Workshop on Whole Effluent Toxicity (WET) on test methods and appropriate endpoints, this paper will discuss the applications and statistical properties of using a statistical criterion of minimum significant difference (MSD). We examined the upper limits of acceptable MSDs as acceptance criterion in the case of normally distributed data. The implications of this approach are examined in terms of false negative rate as well as false positive rate. Results indicated that the proposed approach has reasonable statistical properties. Reproductive data from short-term chronic WET test with *Ceriodaphnia dubia* tests were used to demonstrate the applications of the proposed approach. The data were collected by the North Carolina Department of Environment, Health, and Natural Resources (Raleigh, NC, USA) as part of their National Pollutant Discharge Elimination System program.

Keywords—Aquatic toxicity testing Minimum significant difference Statistical power

INTRODUCTION

The U.S. Environmental Protection Agency (U.S. EPA) and individual states in the United States are in the process of implementing Clean Water Act Section 101(a)(3), which prohibits the discharge of toxic pollutants in toxic amounts by requiring dischargers to conduct acute and chronic toxicity testing with effluents and receiving waters. One recommendation of a recent Pellston conference on whole effluent toxicity testing (WET) was to explore how minimum significant difference (MSD) and tests of bioequivalence can be used to define statistical quality control criteria [1]. To address some concerns in the application of hypothesis testing, the upper MSD limits as an additional test acceptability criterion for toxicity tests has been included in the chronic west coast marine methods manual [2]. Thursby et al. [3] have also suggested revision of the current acceptability requirements and recommended some changes of acceptance criteria based, in part, on an empirical database of MSDs. In this paper, we propose a statistical basis of selecting the limits of MSDs in a given WET test and then evaluate its statistical properties.

Current statistical approach and test design

The current statistical approach in WET testing is the test of the null hypothesis H_0: $\mu_c - \mu_i \leq 0$ versus a toxic effect alternative, H_1: $\mu_c - \mu_i > 0$, where μ_c and μ_i are the true population means of response of the control group (0% effluent) and the effluent concentration group, respectively. The results from the tests that fail to reject the H_0 often lead to the inference that a suspected toxicant does not deleteriously affect test organisms. A lack of sufficient statistical power in a toxicity test implies that a statistically nonsignificant result cannot be taken to imply the absence of a real toxic effect. The difficulty is seen clearly when noting that, if a toxin is in fact present, the statistical result that fails to find such a toxic effect is easier to obtain with a small study than with a large one. Before drawing the conclusion of no significant toxic effect, it is very important to consider the power of the test. Since in practice toxicity tests may often have low power and it is not unusual to find tests that have 30% power or less, current statistical tests may provide inadequate protection against drawing the wrong conclusion when the concentration does have a toxic effect [4].

As noted by Hayes [4], power of the toxicity tests can be improved by increasing sample size, increasing the false positive rate (α), decreasing the variability in response, or limiting the analysis to detection of large differences between the treatment and the control. Using limited sample size and fixed α (0.05 or 0.01) level, toxicity tests often have poor power to detect small, but biologically significant, effect. Denton and Norberg-King [5] examined several U.S. EPA test methods using the current experimental design and based the 75th percentile of MSD values at a power of 80% to determine the effect size. They recommended that power analysis can provide useful information to regulators upon which to establish both test sensitivity (decrease the rate of false negatives) and minimize the issue of excessive power (possible false positive rates).

Minimum significant difference-based criterion testing

As a means of increasing the power of the statistical tests, the test methods require that toxicity tests meet some minimum standards in order to be valid and acceptable. For example, survival in the control must be 80 or 90% for some tests to be acceptable. The implications of defining test acceptability in terms of control survival rates on sample size requirements

* To whom correspondence may be addressed (rakesh.shukla@uc.edu).

This work has not been subjected to U.S. Environmental Protection Agency review. Therefore, it may not necessarily reflect the views of

cussed elsewhere [6]. Some test methods require that the observed MSD must be below a setting upper limit to be acceptable [5]. The inclusion of MSD limits as acceptance criteria for toxicity tests has recently received attention in the literature [1–3,5].

METHODS

The essential points for the MSD-limits approach proposed in literature can be summarized as follows. (1) Select an upper limit of acceptable MSDs. (2) Compare the observed MSD of a test to the upper limit established in step 1. Only when the observed MSD is less than the upper limit is the test considered acceptable. (3) If the test is acceptable according to step 2, then and only then perform the formal statistical testing at the prespecified α level. If the test fails to achieve significance, conclude the effluent concentration has no toxic effect on test organisms; otherwise, declare the concentration is toxic to test organisms.

We call this three-step process minimum significant difference-based criterion testing (MSDBCT). An important design issue for implementing MSDBCT is the selection of the upper limit in step 1. Some criteria for selecting upper limits based on the historical databases of MSDs have been suggested in the literature [3,5].

RESULTS

Since the purpose of MSDBCT is to ensure the power of the toxicity tests, it is necessary to evaluate the power of any proposed MSDBCT. In this article, we propose a way to establish the upper limit of MSDBCT that takes the statistical power directly into consideration. We discuss the statistical rationale of the proposed MSDBCT and investigate the statistical properties of the proposed MSDBCT according to the false negative rate and the false positive rate. We also illustrate three examples from a WET testing database to show the applications of the proposed MSDBCT.

In toxicity tests, one is interested in testing the null hypothesis of no difference, H_0: $\mu_c - \mu_t \leq 0$. A level α test for the above null hypothesis is usually conducted by the Student's t statistics. Assume for simplicity that equal numbers of organisms, say n organisms, are to be used in each group. Also assume we wish to detect a biologically significant level of mean difference in response, θ (>0), between two groups with a power of at least $1 - \beta$. To ensure this power requirement being satisfied, we propose an upper limit (MSD_{max}) of acceptable MSDs as

$$MSD_{max} = \frac{t_{1-\alpha,r}\theta}{\delta_{\alpha,\beta,r}}\sqrt{\frac{\chi^2_{1-\pi,r}}{r}} \quad (1)$$

where $t_{1-\alpha,r}$ ($\chi^2_{1-\pi,r}$) denotes the $1 - \alpha$ ($1 - \pi$) percentile of the central t (χ^2) distribution with r ($=2n - 2$) degrees of freedom and $\delta_{\alpha,\beta,r}$ is the noncentrality parameter of the noncentral t distribution associated with significance level α, power $1 - \beta$, and degrees of freedom r. Tables 1 and 2 contain some $t_{1-\alpha,r}$, $\chi^2_{1-\pi,r}$, and $\delta_{\alpha,\beta,r}$ for selected α, β, π, and r. (Detailed versions of these tables can be obtained from many textbooks, such as [7].)

Let $\bar{x}_c$ and $\bar{x}_t$ denote sample means of response from the control and the effluent groups, respectively. Then the MSDBCT we propose consists of the following steps: (1) For specified choices of α, β, θ, π, and r, find the constants $t_{1-\alpha,r}$, $\chi^2_{1-\pi,r}$, and $\delta_{\alpha,\beta,r}$ from Tables 1 and 2 or elsewhere and then

Table 1. $1 - \alpha$ ($1 - \pi$) percentile of the t (χ^2) distribution with r degrees of freedom

	$t_{1-\alpha,r}$			$\chi^2_{1-\pi,r}$		
r	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\pi = 0.10$	$\pi = 0.05$	$\pi = 0.01$
6	1.44	1.94	3.14	10.65	12.59	16.81
18	1.33	1.73	2.55	25.99	28.87	34.81
22	1.32	1.72	2.51	30.81	33.92	40.29

observed MSD from the test result as $t_{1-\alpha,r}s\sqrt{2/n}$, where s is the pooled within-test sample standard deviation. Compare the observed MSD with MSD_{max} in step 1. If MSD $\leq$ MSD_{max}, go to step 3; otherwise, consider the test as unacceptable and stop. (3) Conclude that there exists no significant toxic effect if $\bar{x}_c - \bar{x}_t \leq$ MSD; otherwise, declare that the particular effluent concentration has significant toxic effect on the test organisms.

In words, the MSDBCT states that, if the observed MSD of a test is less than or equal to an established upper limit (MSD_{max}) and the null hypothesis of no difference in response between the control and the effluent concentration is not rejected by the usual Student's t test, then and only then conclude that there is no toxic effect.

To apply the above MSDBCT to toxicity tests, the α, β, π, and θ must be chosen by the regulatory agency. For the purpose of illustrating the method on real examples, let us suppose the regulatory agency specified a priori the 95/25 rule for conducting the toxicity testing: If the mean response in effluent concentration group is not statistically significantly different from the mean response in the control group at the 0.05 level ($\alpha = 0.05$) and if there is at least 95% power ($\beta = 0.05$) for detecting a 25% difference from the control mean ($\theta = 0.25\mu_c$), then a nontoxic result is concluded.

Examples

Table 3 contains the reproductive data of three selected toxicity tests from short-term chronic WET test with *Ceriodaphnia dubia* tests. The data were collected by the North Carolina Department of Environment, Health, and Natural Resources (Raleigh, NC, USA) as part of their National Pollutant Discharge Elimination System program.

The pooled control mean through all the tests of this database is 24.52 [8] so that we can reasonably assume that the unknown population mean of the control is $\mu_c = 24.52$.

The sample size n is 12 for each of the three tests in Table 3; therefore, the degrees of freedom $r = 22$ for each test. According to the above 95/25 rule, $\alpha = 0.05$, $\beta = 0.05$, and $\theta = 0.25\mu_c = 0.25 \cdot 24.52 = 6.13$. If we choose $\pi = 0.05$, then we obtain $t_{0.95,22} = 1.72$, $\chi^2_{0.95,22} = 33.92$, and $\delta_{0.05,0.05,22} = 3.40$ from Tables 1 and 2.

Table 2. The noncentrality parameter of the noncentral t distribution

r	α	β	$\delta_{\alpha,\beta,r}$	α	β	$\delta_{\alpha,\beta,r}$	α	β	$\delta_{\alpha,\beta,r}$
		0.10	2.77		0.10	3.33		0.10	4.74
6	0.10	0.05	3.17	0.05	0.05	3.75	0.01	0.05	5.25
		0.01	3.91		0.01	4.55		0.01	6.21
		0.10	2.62		0.10	3.04		0.10	3.91
18	0.10	0.05	3.00	0.05	0.05	3.42	0.01	0.05	4.31
		0.01	3.70		0.01	4.13		0.01	5.05
		0.10	2.61		0.10	3.02		0.10	3.85
22	0.10	0.05	2.98	0.05	0.05	3.40	0.01	0.05	4.24

Table 3. Reproduction data from a whole effluent toxicity testing with *Ceriodaphnia*

	Test A		Test B		Test C	
	Control	Effluent	Control	Effluent	Control	Effluent
1	44	15	26	26	33	26
2	30	0	31	21	27	24
3	17	14	23	18	31	29
4	22	23	27	20	33	27
5	33	26	15	30	32	27
6	12	26	28	26	31	27
7	14	37	33	28	28	27
8	39	13	35	23	22	27
9	24	19	13	22	32	26
10	23	23	29	25	32	35
11	17	27	24	23	34	22
12	39	7	30	25	24	29
$\bar{x}$	26.17	19.17	26.17	23.92	29.92	27.17
s	10.69	10.00	6.66	3.42	3.82	3.13

Example 1: Apply the MSDBCT to test A in Table 3.

1. Compute the MSD_{max} according to Equation 1 as

$$MSD_{max} = \frac{t_{0.95,22}\theta}{\delta_{0.05,22}}\sqrt{\frac{\chi^2_{0.95,22}}{r}} = \frac{1.72 \times 6.13}{3.40}\sqrt{\frac{33.92}{22}} = 3.85$$

2. Calculate the observed MSD as

$$MSD = t_{1-\alpha,r}s\sqrt{\frac{2}{n}} = 1.72\sqrt{\frac{11 \times 10.69^2 + 11 \times 10^2}{22}}\sqrt{\frac{2}{12}} = 7.26$$

Since MSD > MSD_{max}, we conclude this test is unacceptable and stop.

Based on a failure to attain a significant result ($\bar{x}_c - \bar{x}_t = 7 < MSD = 7.26$) from the Student's t test, the test A found no toxic effect. Since the estimated post hoc power of detecting a mean difference of 6.13 for test A (assuming that the observed significant difference [S] can be taken to represent the population significant difference [σ]) is only 41%, it is doubtful that the nonsignificant result from the Student's t test implies a nontoxic effect. The proposed MSDBCT prevents the Student's t test with a low power from declaring a nontoxic effect in this example.

Example 2: Apply the MSDBCT to test B in Table 3.

1. Compute the MSD_{max} as in Example 1: $MSD_{max} = 3.85$.
2. Calculate the observed MSD as

$$MSD = 1.72\sqrt{\frac{11 \times 6.66^2 + 11 \times 3.42^2}{22}}\sqrt{\frac{2}{12}} = 3.71$$

Since MSD < MSD_{max}, this test is acceptable and we go to step 3.

3. $\bar{x}_c - \bar{x}_t = 2.25$, which is less than 3.71 (the observed MSD). Declare a nontoxic effect on the test organisms.

Based on a failure to attain a significant result ($\bar{x}_c - \bar{x}_t = 2.25 < MSD = 3.71$) from the Student's t test, test B found no toxic effect. Since the estimated power of detecting a mean difference of 6.13 for test B is 87%, it is unlikely that the null hypothesis of no difference (a nontoxic effect) would have been wrongly accepted by the Student's t test. The proposed MSDBCT leads to the same conclusion (a nontoxic effect) as the Student's t test with a high power does in this example.

Example 3: Apply the MSDBCT to test C in Table 3.

1. MSD = 3.85.

$$MSD = 1.72\sqrt{\frac{11 \times 3.82^2 + 11 \times 3.13^2}{22}}\sqrt{\frac{2}{12}} = 2.45$$

Since MSD < MSD_{max}, this test is acceptable. We go to step 3.

3. $\bar{x}_c - \bar{x}_t = 2.75$, which is larger than 2.45 (MSD). Declare a toxic effect on test organisms.

Based on a significant result ($\bar{x}_c - \bar{x}_t = 2.75 > MSD = 2.45$) from the Student's t test, test C detected a toxic effect. Test C has a very high estimated power (99.5% for detecting a mean difference of 6.13). As can be expected, the proposed MSDBCT leads to the same conclusion (a toxic effect) as the Student's t test with a high power does in this example.

Statistical rationale for the MSDBCT approach

To test H_0: $\mu_c - \mu_t \leq 0$ versus H_1: $\mu_c - \mu_t > 0$ at a significance level α and to ensure a power $1 - \beta$ to detect a biologically significant difference $\mu_c - \mu_t = \theta$ between the control and the effluent concentration, the standard power approach requires

$$p\{\bar{x}_c - \bar{x}_t > MSD \mid \mu_c - \mu_t = \theta\} \geq 1 - \beta \qquad (2)$$

where $MSD = t_{1-\alpha,r}s\sqrt{2/n}$. This power depends on unknown population variance σ^2, which is estimated by the sample variance s^2. It can be shown that the inequality in Equation 2 is guaranteed if the true population variance σ^2 satisfies the inequality

$$\sigma^2 \leq \frac{n\theta^2}{2(\delta_{\alpha,\beta,r})^2} = \sigma^2_{max} \qquad (3)$$

Since we do not know the true σ^2, we cannot compare it with σ^2_{max}. Schuirmann [9] suggests using $s^2 \leq \sigma^2_{max}$ as a reasonable alternative to $\sigma^2 \leq \sigma^2_{max}$. Using this approach, the false negative rate of the test (the probability of wrongly declaring no toxic effect when $\mu_c - \mu_t = \theta$) turns out to be noticeably smaller than the specified nominal level β if the degrees of freedom is finite. This is undesirable since the false positive rate (the probability of wrongly declaring toxic effect when $\mu_c - \mu_t \leq 0$) will be increased well above nominal level α [9]. A better approach is to frame the question in the following hypothesis testing scenario:

$$H_0'\text{: } \sigma^2 \leq \sigma^2_{max} \qquad H_1'\text{: } \sigma^2 \geq \sigma^2_{max}$$

where σ_{max}^2 is a known constant as defined in Equation 3.

With usual assumptions, rs^2/σ^2 has a chi-squared distribution with r degrees of freedom. Therefore, we reject H_0' if and only if

$$\frac{rs^2}{\sigma^2_{max}} > \chi^2_{1-\pi,r} \quad \text{or} \quad s^2 > \frac{\chi^2_{1-\pi,r}\sigma^2_{max}}{r}$$

else we retain H_0'. This is a uniformly most powerful unbiased test for testing H_0' at the significance level π [7].

Thus, an improved version of ensuring the power requirement in Equation 2 consists of accepting the null hypothesis H_0, declaring no toxic effect, if

1. the H_0' (the hypothesis of at least $1 - \beta$ power to declare a toxic effect if $\mu_c - \mu_t = \theta$) is not rejected, i.e.,

$$s^2 \leq \frac{\chi^2_{1-\pi,r}}{r}\sigma^2_{max} \qquad (4)$$

Since $MSD = t_{1-\alpha,r}s\sqrt{2/n}$ and $\sigma^2_{max} = (n\theta^2)/[2(\delta_{\alpha,\beta,r})^2]$, Equation 4 is equivalent to saying

$$MSD < \frac{t_{1-\alpha,r}\theta}{\sqrt{\chi^2_{1-\pi,r}}} \qquad (5)$$

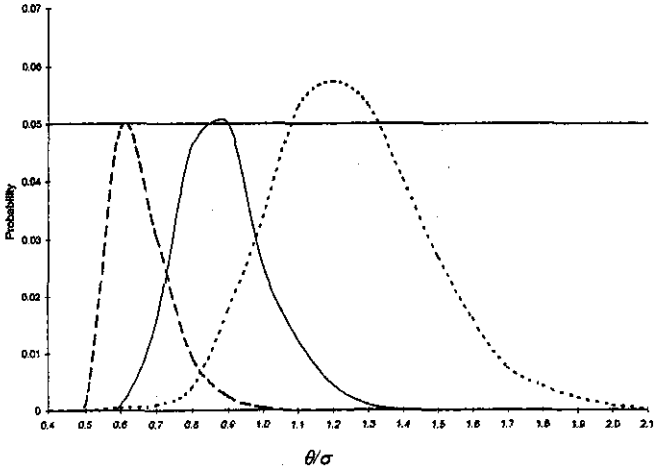

Fig. 1. Probability of declaring a nontoxic effect when $\mu_c - \mu_t = \theta$ for the minimum significant difference-based criterion testing (MSDBCT) as a function of θ/σ with degrees of freedom 22 (–.–), 50 (———), and 100 (– – –).

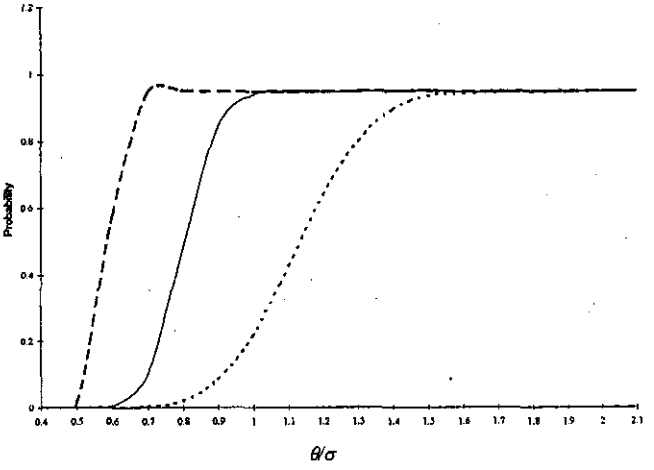

Fig. 2. Probability of declaring a nontoxic effect when $\mu_c - \mu_t = 0$ for the minimum significant difference-based criterion testing (MSDBCT) as a function of θ/σ with degrees of freedom 22 (–.–), 50 (———), and 100 (– – –).

2. The hypothesis H_0 (no difference in response between the control and the effluent concentration) is also not rejected, i.e.,

$$\bar{x}_c - \bar{x}_t \leq \text{MSD} \quad (6)$$

then we conclude that there exists no significant toxic effect.

Combining Equations 5 and 6, we obtain the MSDBCT as suggested in the preceding sections.

Statistical properties of the MSDBCT

Given $\alpha = \beta = \pi = 0.05$, Figure 1 provides the probabilities of declaring a nontoxic effect when $\mu_c - \mu_t = \theta$ (the false negative rates) at various degrees of freedom r for the MSDBCT approach. It is seen that the probabilities depend on θ/σ. The probabilities rise to the peaks of about 0.056, 0.051, and 0.050 at θ/σ = 1.19, 0.85, and 0.61, corresponding to degrees of freedom 22, 50, and 100, respectively.

Unlike the Schuirmann's [9] approach (as given by $s^2 \leq \sigma^2$ and $\bar{x}_c - \bar{x}_t \leq$ MSD), the maximum false negative rate of the MSDBCT approach is much closer to the nominal level of 0.05. This is true for each degree of freedom.

Figure 2 plots the probabilities of declaring a nontoxic effect when $\mu_c - \mu_t = 0$ (1 − false positive rates) at various degrees of freedom r for the MSDBCT approach. It is seen that the probabilities are an increasing function of θ/σ.

DISCUSSION

In whole effluent toxicity testing, regulators have searched for ways to ensure that the power of the statistical test is sufficient to detect toxicity. The technique employed to control the power is setting some test acceptability criteria. These criteria are designed to provide adequate power to detect a biologically significant difference from the control [1,2,5]. Traditional acceptance criteria focus primarily on response (survival rate) in a control. Bailer and Oris [6] have examined the implications of defining test acceptability in terms of control survival rates on sample size requirements for detecting a stated decrement in survival. The inclusion of MSD limits as additional criteria for toxicity tests has begun recently [1–3,5]. This additional criterion is an improvement over traditional data. However, the effect of establishing an upper limit of acceptable MSDs to the toxicity tests has not been carefully examined. To the best of our knowledge, our paper is the first and only report to evaluate the performance of the MSDBCT according to the false negative (and false positive) rates actually achieved by the MSDBCT.

The MSDBCT proposed in this paper demonstrates that the *inclusion of MSD limits as acceptability criteria in whole effluent toxicity testing can be made scientifically and statistically sound provided one chooses the MSD limit* (MSD_{max}) according to Equation 1. The actual false negative rate of the proposed MSDBCT can be held very close to the nominal level. The actual false positive rates can also be evaluated, as shown in Figure 2. The approach of MSDBCT must perform satisfactorily for normally distributed endpoints. How this method performs when the assumptions are grossly violated remains to be shown. It may, however, be safe to state that, for mild to moderate deviations from the stated assumptions, the proposed method should perform reasonably well.

The proposed MSDBCT can prevent the Student's t test with a low power from declaring a toxic test as nontoxic. On the other hand, the proposed MSDBCT will not change the result (significant or nonsignificant) from the Student's t test with a high power.

An important design question for implementing the proposed MSDBCT is the choice of the value θ (a difference from the control desired to be detected as statistically significant) and the corresponding power requirement. Thursby et al. [3] have suggested ways to select the value θ (called the threshold value in their paper). In our examples, we have chosen θ as 25% of the control mean and required 95% power for the test to detect this difference. If we had required 80% power to detect this difference, MSD_{max} computed according to Equation 1 for the examples discussed in our paper would have been 5.09 instead 3.85.

CONCLUSION

The MSDBCT can be applied to multiple concentration situations, but it does not address excessive statistical power problems for an individual test. The bioequivalence testing

methods including both freshwater and marine test species be examined for both the MSDBCT and bioequivalence testing approach before consideration for application in the WET testing program.

Acknowledgement—This work was supported, in part, by cooperative agreement EPA CR820480 between the University of Cincinnati and the U.S. Environmental Protection Agency. We express our sincere thanks to Larry Ausley for providing us with the data.

REFERENCES

1. Chapman GA, et al. 1996. Methods and appropriate endpoints. In Grothe DR, Dickson KL, Reed-Judkins DK, eds, *Whole Effluent Toxicity Testing: An Evaluation of Methods and Predictions of Receiving System Impacts.* Society of Environmental Toxicology And Chemistry, Pensacola, FL, USA, pp 97–99.
2. Chapman GA, Denton DL, Lazorchak JM. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to West Coast marine and estuarine organism. EPA 600/95/136. U.S. Environmental Protection Agency, Cincinnati, OH.
3. Thursby GB, Heltshe J, Scott KJ. 1997. Revised approach to toxicity acceptability criteria using a statistical performance assessment. *Environ Toxicol Chem* 16:1322–1329.
4. Hayes JP. 1987. The positive approach to negative results in toxicology studies. *Ecotoxicol Environ Saf* 14:73–77.
5. Denton DL, Norberg-King TJ. 1996. Whole effluent toxicity statistics: A regulatory perspective. In Grothe DR, Dickson KL, Reed-Judkins DK, eds, *Whole Effluent Toxicity Testing: An Evaluation of Methods and Predictions of Receiving System Impacts.* Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, pp 83–102.
6. Bailer AJ, Oris JT. 1996. Implication of defining test acceptability in terms of control-group survival in two-group survival studies. *Environ Toxicol Chem* 15:1242–1244.
7. Lehmann EL. 1986. *Testing Statistical Hypothesis,* 2nd ed. Wiley, New York, NY, USA.
8. Shukla R, Wang Q, Fulk F, Deng C, Denton D. 2000. Bioequivalence approach for whole effluent toxicity testing. *Environ Toxicol Chem* 19:169–174.
9. Schuirmann DJ. 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* 15:657–680.

APPENDIX

The MSDBCT is equivalent to accepting H_0, thus declaring nontoxicity, if

$$\chi^2 \le \frac{n\chi^2_{1-\pi,r}\theta^2}{2(\delta_{\alpha,\beta,r})^2\sigma^2} \quad \text{and} \quad z \le t_{1-\alpha,r}\sqrt{\frac{\chi^2}{r}} - \frac{\mu_c - \mu_t}{\sigma\sqrt{2/n}}$$

where $\chi^2 = (rs^2)/\sigma^2$ has a central chi-squared distribution with r degrees of freedom and

$$z = \frac{(\bar{x}_c - \bar{x}_t) - (\mu_c - \mu_t)}{\sigma\sqrt{2/n}}$$

has the standard normal distribution. Furthermore, z and χ^2 are statistically independent. From these facts, simulations are performed to obtain all the probabilities plotted in all the figures for the MSDBCT with $\alpha = 0.05$, $\beta = 0.05$, and $\pi = 0.05$.