DIVISION OF WATER QUALITY

NORTH CAROLINA DEPARTMENT OF
ENVIRONMENT AND NATURAL RESOURCES

PASS/FAIL METHODOLOGY FOR DETERMINING ACUTE TOXICITY IN A SINGLE EFFLUENT

December 2010

Version 3.0

4/21/11

Coleen H. Sullins, Director, Division of Water Quality — Date

PASS/FAIL METHODOLOGY FOR DETERMINING ACUTE TOXICITY IN A SINGLE EFFLUENT CONCENTRATION

It is the intent of this document to provide a method by which significant mortality in a single effluent concentration may be determined. It may be applied to any concentration of effluent, but is specifically designed to test higher effluent concentrations where a measured LC50 may not necessarily protect for acute toxicity. This would occur in situations where the receiving stream or mixing zone is almost entirely effluent. In these instances it is important that short term acute effects not occur. These methods are based upon the standard methodology specified in the EPA guidance document, Methods for Measuring the Acute Toxicity of Whole Effluents to Freshwater and Marine Organisms, Fifth Edition EPA/821-R-02-012, October, 2002. This document is intended to specify exact procedural modifications of the above cited methodology for performing analyses of this type. Unless specifically modified by this procedure, the North Carolina Laboratory Certification/Criteria Procedures document, or by written exception made by NC DWQ, all other test conditions and requirements will be as specified in the EPA acute toxicity document.

METHODOLOGY

The procedure shall be performed as an acute, static, non-renewal toxicity test. *Pimephales promelas*, *Daphnia pulex*, *Ceriodaphnia dubia, Mysidopsis bahia,* or *Menidia beryllina* are commonly used test species. There will be two concentrations utilized in this procedure. The control population will be specified as Treatment 1, and the effluent treatment will be specified as Treatment 2. Typically, the effluent concentration utilized for this procedure will be 90%. The actual effluent concentration at which the test is to be performed, as well as the sample type, test duration, and test species, will be specified through the NPDES Permit or Administrative Letter requirement.

Each treatment will be tested using four identical test vessels, each containing 10 test organisms for a total of 80 test organisms. At the end of the test, all organisms will be observed as being either alive or dead and recorded on the appropriate laboratory forms. If all chemical and physical protocols have been met, data analysis can proceed. If testing protocols have not been met, the analysis must be repeated.

Acute toxicity tests will be conducted at 25.0 ± 1.0°C. Variances may be requested for species which may require alternate temperatures, such as trout species. Control/dilution water for freshwater test species will have a pH between 6.5 and 8.5 S.U. and total hardness as calcium carbonate between 30 and 50 ppm $CaCO_3$ at test initiation and subsequent test solution renewals. Should receiving waters have characteristics outside of stated ranges then alternative pH and hardness ranges may be accepted upon demonstration that the alternate ranges are better suited to testing objectives and quality assurance standards have been met.

Sampling must be performed below the last waste treatment process, including disinfection. There may be no removal of chlorine or any other effluent constituent by either chemical or physical methods prior to testing with the exceptions of allowable filtration of the effluent through 60 um nylon screen or plankton netting and the reduction of excess dissolved oxygen to the saturation level, as per EPA methods.

DATA ANALYSIS

If mortality in the control population exceeds 10%, the test is considered invalid. If mortality in the control population is less than or equal to 10%, the test data are then analyzed to determine whether mortality in the effluent treatment (Treatment 2) is significantly different from that of the control population (Treatment 1). The procedures required for data analysis are contained in Section 11.3 (*Determination of No-Observed-Adverse-Effect Concentration (NOAEC) From Multi-Concentration Tests, and Determination of Pass or Fail (Pass/Fail) For Single-Concentration (Paired) Tests*) of the EPA acute manual referenced above. All statistical analyses are performed using arc sine square root transformed data. This transformation is described on page 88, sub-section 11.3.5.1 ("*Arc Sine Square Root Transformation*").

The assumption of normality of the transformed data should be tested with the Shapiro-Wilk's Test at the 0.01 level of significance, according to sub-section 11.3.6.2 (pages 96-97). Statistical tables used with this analysis are presented in Tables 21 and 22 on pages 90-92 of the acute manual. The F test for equality of variances should also be performed at the 0.01 level of significance, as stated in sub-section 11.3.6.3 on page 97.

Final comparisons of control and treatment responses should be made at the 0.01 level of significance with a t test (for normal data with equal variances), a modified t test (for normal data when variances are not equal), or the nonparametric Wilcoxon Rank Sum Test (for non-normal data). Note: These analyses are to be performed at a different level of significance than that specified in Section 11.3.6 of the EPA manual.

Two sample data sets analyzed according to these procedures are presented below in Examples 1 and 2. In addition, two data situations which cannot be analyzed by the above statistical procedures due to constant rate of response are also presented in Examples 3 and 4.

SPECIAL CONSIDERATIONS

F-test: The F test for equality of variances is not possible when the response proportions of all four replicates within the control or treatment set are equal. This results in a variance of 0 in the denominator of the F equation, which is not valid. When data pass the test for normality, but cannot be tested for equality of variances, the modified t test should be used, since the assumption of equality of variances cannot be made.

Wilcoxon Rank Sum Test: This document uses a critical value of 10 for a one-sided Wilcoxon Rank Sum test for tests with 2 treatments (control and the effluent treatment) and four replicates in each treatment. A critical value of 10 can be found on page 320, Section X.5 “Critical Values for the Wilcoxon Rank Sum Test”, P=0.01, one-sided, n=4, m=4 in the following reference: Beyer, W., ed. 1991. Standard Probability and Statistics, Tables and Formulae, CRC Press, Boca Raton, 503 pp. Some statistical programs do not have a critical value or may use a different value for n=4, m=4 at the 0.01 level of significance. For instance, some statistical programs give a critical value of 9 for n=4, m=4 at the 0.01 level of significance. A critical value of 9 will cause all results to pass, even in cases with high mortality. This is because the lowest 4 ranks will always add up to at least 10. If the statistical program uses a critical value of 9 or does not have a critical value at n=4, m=4, a critical value of 10 may be substituted. Alternatively, a 0.05 level of significance may be chosen rather than the 0.01 level of significance. The 0.05 level of significance may only be chosen for non-normal acute pass fail tests with 4 replicates and two concentrations (control and effluent treatment).

t-value non-calculable: If all replicates of the control (Treatment 1) have identical responses equal to or less than 10 percent mortality and all replicates of the effluent concentration (Treatment 2) have equal responses though greater than that of the control response, the t value is not calculable. This situation is presented in Example 3. The Wilcoxon Rank Sum Test should be used in this case. In the Wilcoxon Rank Sum test, tied values are given an average rank. In Example 3, Treatment 2 has the four lowest values, and they are equal. They will each be given a rank of 2.5, which is the average value of ranks 1, 2, 3 and 4. The test adds the rank sums of all values within each group. Treatment 2 will receive a final sum of 10. This sum is equal to the critical sum of 10, indicating that the responses are statistically different. Therefore, the test result should be reported as "FAIL."

All responses identical: If all replicates in both treatments have identical responses which are equal to or less than ten percent mortality, no statistical test is appropriate. This situation is presented in Example 4 below. When all responses are equal and mortality is equal or less than 10%, the test result should be reported as "PASS."

REPORTING

Data obtained by use of this methodology will be entered on the form AT-2 and sent to the designated address in the NPDES permit or Administrative Letter. Additionally, the results of this testing should be recorded as "PASS" or "FAIL" on the State MR-1 form for the month in which the test was performed. The appropriate parameter code to enter on the MR-1 is specified in the NPDES permit or Administrative Letter Requirement.

EXAMPLE 1:

In this example, data are not normally distributed. Therefore, the appropriate final analysis is the nonparametric Wilcoxon Rank Sum Test, which indicates that the test result is "FAIL."

(10 Organisms/Replicate)

	Control (Treatment 1)				Effluent (Treatment 2)			
Replicate	1	2	3	4	1	2	3	4
% Mortality	0	0	10	0	20	20	30	20

Test for Normality

Shapiro Wilk's W = 0.663

Critical W (@0.01) = 0.749

Result = Not Normally Distributed

Test for Equality of Variance

F=1.97

Critical F (@ 0.01) = 47.47

Result = Variances Equal

t-test

Calculated t value = 5.86

Critical t value (@ 0.01) = 3.14

Result = FAIL; Not appropriate test

Wilcoxon Rank Sum Test

Wilcoxon Rank Sum =10

Critical Sum (@ 0.01) = 10

Result = FAIL

EXAMPLE 2:

In this example, data are normally distributed and have equal variances. The appropriate final analysis is the t test, which indicates that the test result is "FAIL."

(10 Organisms/Replicate)

	Control (Treatment 1)				Effluent (Treatment 2)			
Replicate	1	2	3	4	1	2	3	4
% Mortality	0	0	10	0	50	60	70	50

Test for Normality

Shapiro Wilk's W = 0.807

Critical W (@0.01) = 0.749

Result = Normally Distributed

Test for Equality of Variance

F=1.46

Critical F (@ 0.01) = 47.47

Result = Variances Equal

t test

Calculated t value = 10.37

Critical t value (@ 0.01) = 3.14

Result = FAIL

Wilcoxon Rank Sum Test

Wilcoxon Rank Sum =10

Critical Sum (@ 0.01) = 10

Result = FAIL; Not appropriate test

EXAMPLE 3*:*

In this example, the t-test and F test cannot be performed. Different statistics programs output different results for the Shapiro Wilk's test for normality, but regardless, the t test and the F test cannot be performed due to the constant responses within each group. The appropriate final analysis is Wilcoxon Rank Sum Test, which indicates that the test result is "FAIL."

(10 Organisms/Replicate)

	Control (Treatment 1)				Effluent (Treatment 2)			
Replicate	1	2	3	4	1	2	3	4
% Mortality	0	0	0	0	10	10	10	10

Test for Normality

Shapiro Wilk's W = Varies with statistics program

Critical W (@0.01) = 0.749

Result = Depends on statistical program

Test for Equality of Variance

F=not calculable

t test

Calculated t value = not calculable

Wilcoxon Rank Sum Test

Wilcoxon Rank Sum =10

Critical Sum (@ 0.01) = 10

Result = FAIL

EXAMPLE 4*:*

In this example, no statistical test is appropriate as all responses are the same. There is no variation. Different statistics programs output different results for the Shapiro Wilk's test for normality, but regardless, the t test and the F test cannot be performed due to the constant responses within each group. The Wilcoxon Rank Sum Test will be 18 and will PASS. However, because all results are the same and there is no variation, conducting any statistical test is not necessary nor is any statistical test appropriate. If mortality is less than or equal to 10%, the result is PASS. NOTE: If mortality is equal in every replicate and greater than 10%, the result is an invalid test due to high control mortality.

(10 Organisms/Replicate)

	Control (Treatment 1)				Effluent (Treatment 2)			
Replicate	1	2	3	4	1	2	3	4
% Mortality	10	10	10	10	10	10	10	10

Test for Normality

Shapiro Wilk's W = Varies with statistics program

Critical W (@0.01) = 0.749

Result = Depends on statistical program

Test for Equality of Variance

F=not calculable

t test

Calculated t value = not calculable

Wilcoxon Rank Sum Test

Wilcoxon Rank Sum =18

Critical Sum (@ 0.01) = 10

Result = PASS

REFERENCES

Beyer, William, ed. 1991. Standard Probability and Statistics, Tables and Formulae, CRC Press, Boca Raton. 503 pp.

North Carolina Division of Water Quality. 2010. North Carolina Biological Laboratory Certification/Criteria Procedures Document, Version 3.0. Revised December 2010.

United States Environmental Protection Agency. 2002. Short Term Methods for Estimating the Acute Toxicity of Effluents and Receiving Waters to Freshwater Organisms. Fifth Edition. EPA 821-R-02-012, 275 pp.