Selecting Tier III Water Column Toxicity Standards for CDF Discharges: Statistical Alternatives

by Joan U. Clarke, Paul R. Schroeder, and Jeffery A. Steevens

PURPOSE: This technical note describes problems inherent in interpreting water column toxicity tests and in selecting toxicity standards for evaluation of effluent and runoff resulting from upland placement of dredged material in confined disposal facilities (CDFs). The current procedure for comparing mixing zone elutriate concentration calculations with water column toxicity test results (employing $0.01(LC_{50})$ as the water quality standard) is not based on statistical probability or toxicological effect, and can result in unnecessarily restrictive evaluations. Alternative statistical measures are presented that can serve as more realistic and interpretable Tier III water column toxicity standards.

BACKGROUND: CDF effluent is defined in the Clean Water Act (CWA) as a dredged material discharge and is regulated as such under Section 404 of the CWA. The discharge of effluent from a CDF is defined as a dredged material discharge in 33 CFR 323.2 (d) and 40 CFR 232.2 (e):

> The term "discharge of dredged material" means any addition of dredged material into waters of the United States. The term includes, without limitation, the addition of dredged material to a specified discharge site located in waters of the United States and the runoff or overflow from a contained land or water disposal area.

In addition, Section 401 of the CWA provides the States a certification role as to project compliance with applicable State water quality standards (WQS); effluent standards may be set as a condition of the certification.

Any evaluation of potential water column effects must consider the effects of mixing. Section 230.3(m) of the Guidelines defines the mixing zone as follows:

> The term "mixing zone" means a limited volume of water serving as a zone of initial dilution in the immediate vicinity of the discharge point where receiving water quality may not meet quality standards or other requirements otherwise applicable to the receiving water. The mixing zone should be considered as a place where wastes and water mix and not as a place where wastes are treated.

Effluents must be evaluated for toxicity in Tier III of the evaluation protocols (U.S. Environmental Protection Agency/U.S. Army Corps of Engineers (USEPA/USACE) 1998) if there are contaminants of concern (COCs) for which there are no WQS or if there is concern regarding potential interaction of multiple contaminants. Bioassays provide information on the toxicity of contaminants not included in the water quality standards, and indicate possible

interactive effects of multiple contaminants. Tier III evaluates effluent toxicity based on an effluent or runoff elutriate and use of that elutriate as a medium to conduct water column toxicity tests. Tier III toxicity testing assesses the potential toxicity of effluent to appropriate sensitive water column organisms. As with effluent water quality, the results of the water column toxicity tests must be interpreted considering the effects of mixing.

Effluent toxicity in the water column is evaluated by exposing test organisms to a dilution series containing both dissolved and suspended components of the simulated effluent or runoff. Test organisms are added to the exposure chambers and exposed for a prescribed period (usually 96 hr, though some tests (e.g., bivalve larvae) may be run for shorter periods). Surviving organisms are examined at specified intervals and/or at the end of the test. The concentration of test material that produces an effect, if it does so, is determined. Results of the water column toxicity test are currently reported in terms of the 96-hr LC_{50} or 96-hr EC_{50} expressed as a percentage of the suspended dredged material concentration (or 100 percent elutriate). The detailed procedures for conducting the water column toxicity tests for effluent are identical to those for evaluating open-water discharges provided in the Inland Testing Manual (ITM) (USEPA/USACE 1998), except that effluent or runoff elutriate is used for the test medium.

Results of the toxicity tests are interpreted with a mixing model and are used to determine the acceptability of the effluent discharge in the mixing zone. Where limited mortality is observed and an LC_{50} value cannot be calculated, the survival of the organisms is compared to survival in the dilution water through a simple statistical comparison (t-test). When significant mortality is observed, an LC_{50} value is calculated and an application factor of 0.01 is applied to predict the concentration where long-term and sublethal effects are not likely to occur. The 0.01 factor was recommended for persistent and bioaccumulative compounds by the National Academy of Sciences (1972) as a result of the limited chronic exposure and sublethal effects data available and uncertainty associated with predicting these effects from acute (96-hr) toxicity test results. This predictive procedure was derived from five studies published from 1967 to 1971 that described safe to lethal ratios for fish. As such, a water column toxicity standard of $0.01(LC_{50})$ was adopted in the Marine Protection, Research, and Sanctuary Act (MPRSA) regulations for open-water placement of dredged material in the ocean. For consistency, the ITM also adopted $0.01(LC_{50})$ as the recommended standard. Similarly, $0.01(LC_{50})$ was proposed as the standard for effluent discharges in the ITM for the sake of consistency with open-water placements. The use of $0.01(LC_{50})$ as the water column toxicity standard is not required by CWA regulations; the ITM serves only as guidance and therefore any appropriate alternative standard could be selected.

After selecting a water column toxicity standard, the mixing of the discharge with the receiving water is modeled and the concentration of the discharge is predicted as a function of location from the point of discharge. The predictions are examined to determine if, after mixing, the concentration of the discharge is likely to be below the selected water column toxicity standard at all times outside of the mixing zone and therefore in compliance with the CWA regulations.

STATISTICAL ALTERNATIVES TO LC$_{50}$: The median lethal concentration or LC_{50} is the estimated concentration of contaminant, toxicant, or elutriate that would be lethal to 50 percent of the test organisms in a toxicity test. Similarly, the EC_{50} is that concentration producing a

measurable effect of a given sublethal endpoint on 50 percent of a test population. However, multiplying the LC_{50} by an application factor of 0.01 to obtain the acceptable concentration in the water has limited value because the relationship between concentration and mortality is seldom linear and cannot be used to predict long-term or sublethal effects. In addition, use of the application factor does not include any estimate of probable adverse effects and relies on a single arbitrary point value to make a regulatory decision. Furthermore, the 0.01 value was based on a few studies completed by the early 1970's and does not rely on the best available current science.

Alternatives to the LC_{50} have gained popularity in ecotoxicity studies in recent years. These include low-effects concentrations, inhibition concentrations, and the no-observed effects concentration (NOEC). The first two are probabilistic measures calculated using statistical regression models or interpolation techniques. The NOEC is non-probabilistic and is determined using a trend test or means comparison procedure. These alternatives are described below along with their limitations, and are illustrated using example water column toxicity test data (Table 1), taken from the ITM (USEPA/USACE 1998, Appendix D, Table D-3).

Table 1
Number of Survivors in a Hypothetical Water Column Toxicity Test After 96 hr

Replicate[1]	Treatment[2]				
	Dilution Water[3]	100%	50%	25%	12.5%
1	20	6	8	12	17
2	19	7	8	18	17
3	20	9	9	15	18
4	20	5	10	14	16
5	19	8	11	13	18
Total	98	35	46	72	86
Mean	19.6 (98%)	7.0 (35%)	9.2 (46%)	14.4 (72%)	17.2 (86%)
SE	0.24	0.71	0.58	1.03	0.37

[1] 20 organisms per replicate at initiation of test.
[2] Percent concentrations of dredged-material elutriate:
       100% = 1 part elutriate plus 0 part dilution water
       50% = 1 part elutriate plus 1 part dilution water
       25% = 1 part elutriate plus 3 parts dilution water
       12.5% = 1 part elutriate plus 7 parts dilution water
[3] In this example, the dilution water was control (laboratory) water.

No-Observed Effects Concentration (NOEC). The NOEC and its companion, the lowest observed effects concentration or LOEC, currently enjoy considerable popularity in ecotoxicity studies. These response measures are usually determined by statistical hypothesis testing, in which treatment responses are compared with a control. The NOEC is the highest concentration in which there is no significant difference in response from the control, and the LOEC is the lowest concentration in which a significant difference is observed. The NOEC, or sometimes an average of the NOEC and LOEC, is used as a point estimate of the concentration of contaminant, toxicant, or elutriate that may be considered "safe" in that it caused no significant deleterious effects in the test organisms.

The NOEC has traditionally been calculated using a hypothesis-testing multiple comparison procedure, typically Dunnett's test (Dunnett 1955), in which each treatment is compared with the control. However, the NOEC may be higher than the LOEC with some problem data sets having non-monotonic response or high variability in some treatments. Williams' test (Williams 1971, 1972), another multiple comparison procedure, is specifically designed to detect an increasing concentration-response, and thus is more powerful and more sensitive for determining NOEC than Dunnett's test (Gelber et al. 1985). Capizzi et al. (1985) advocated the use of a one-sided, stepwise, linear trend test to estimate NOEC, based on the hypothesis of progressiveness of response with increasing concentration. A variety of trend tests are available, including Cox and Stuart, Cochran-Armitage, Mantel, and the Mann-Whitney test for step trend. Perhaps the simplest trend test is a correlation analysis. All appropriate concentrations are first included in the stepwise trend test. If a statistically significant trend is observed, the highest concentration is excluded and the test is repeated with the remaining concentrations. These steps are repeated until no statistically significant trend is observed, and the resulting highest concentration is the NOEC.

The NOEC has been heavily criticized by a number of authors, including Skalski (1981), Stephan and Rogers (1985), Chapman et al. (1996), Crane and Newman (2000), and Scholze et al. (2001). Among the criticisms that have been raised concerning NOEC:

- NOEC depends on the choice of statistical test, significance level (α or Type I error rate), and data transformation; different tests, significance levels, or transformations could produce different NOECs from the same data.

- Poor experimental design (small sample size, improper spacing of treatment concentrations, large variability) tends to increase NOEC. Far from being a true no-adverse-effects concentration, NOEC is actually a no-statistically-significant-difference concentration and if the power of the statistical test is low, adverse effects occurring at or below the NOEC can easily be missed.

- NOEC must be one of the treatment concentrations, and therefore is totally dependent on the number and spacing of concentrations.

- Information about the concentration-response relationship (e.g., steepness, variability, non-monotonicity) is disregarded.

- Confidence intervals cannot be calculated for NOEC, and therefore it is impossible to compare accuracy of NOECs from different experiments.

In CDF elutriate evaluations, the use of NOEC will often pose a thorny problem when the calculated NOEC is zero, i.e., less than the lowest tested concentration. This, in fact, is the result using the example data from Table 1. The evaluator is then faced with the difficulty of rerunning the toxicity test using lower concentrations of the elutriate, or else accepting an infinitely large mixing zone.

Low-effects Concentration. The low-effects concentration is that concentration of a contaminant, toxicant, or elutriate resulting in a measurable effect on a specified small proportion of a test population, such as LC_{15}, LC_{10}, or LC_5. It may be calculated, along with

4

confidence limits, using a statistical regression model such as probit analysis when the data satisfactorily fit the model. In a CDF elutriate evaluation, the lower 95-percent confidence limit for the LC_{10}, for example, would be interpreted as the dilution of elutriate necessary to have a 95-percent probability of protecting 90 percent of an exposed population from mortality resulting from exposure to the elutriate.

A variety of nonlinear regression models for estimating low-effects concentrations are described in Moore and Caux (1997) and Scholze et al. (2001). SAS statements (SAS Institute Inc. 1989) using several of these models to analyze the example data are presented in Appendix A. The regression-based approach to estimating low-effects concentrations overcomes many of the problems inherent in the use of NOEC (Stephan and Rogers 1985, Chapman et al. 1996). However, the regression approach does have restrictions and limitations:

- For the most commonly used nonlinear regression techniques (probit and logistic), the data must fit a sigmoidal model. Fit is likely to be poor when the concentration-response relationship is weak, when the highest concentrations do not cause effects, when hormesis (stimulating effects) occurs at low concentrations, or in general when the concentration-response relationship is not monotonic.

- There must be one or more treatments with partial effects, e.g. mortality greater than 0 and less than 100 percent.

- The experimental design must be appropriate for regression. Standard protocols for toxicity tests are generally designed for hypothesis testing; treatments may be too few in number and too widely spaced for an adequate regression analysis.

- The regression approach should be limited to interpolation situations due to lack of precision when it is necessary to extrapolate beyond the data to estimate low effects.

- In a study of 198 toxicity data sets, Moore and Caux (1997) found that low effects of less than 10 percent are model-dependent, i.e. different regression models can produce widely different estimates.

- Confidence intervals can be excessively large for low effects of 5 percent or less.

Low-effects concentrations may also be calculated by modifying the linear interpolation method described below, where target survival = 100 minus the low-effects mortality percent.

Inhibition Concentration. While the low-effects concentration approach requires specifying an absolute amount of effect to be considered acceptable (or unacceptable, depending on point of view), the inhibition concentration approach specifies an amount of effect relative to the control. The inhibition concentration ICp is that concentration of a contaminant, toxicant, or elutriate resulting in a specified percent reduction (p) from the control in the endpoint being measured. Although the IC is usually used in sublethal effects tests, IC may also be calculated for mortality when p is a specified percent reduction in survival relative to the control. IC may be estimated by hand using a nonparametric linear interpolation method described in USEPA (1994), providing there are two test concentrations resulting in survival that brackets the target survival, where target survival = control survival * [1 − p/100]. When the linear interpolation method is used, confidence intervals can be calculated using bootstrap resampling of the data set. This is

impractical to do by hand since ICp must be recalculated for a large number of bootstrap resamples. The linear interpolation method imposes a monotonically decreasing structure on the responses and assumes piecewise linearity in response between adjacent concentrations (Bailer et al. 2000).

Inhibition concentration may also be estimated using the same types of regression models as low-effects concentrations, such as the probit model, and will be subject to the same limitations. Bailer et al. (2000) recommend generalized linear models to estimate effective concentrations, because these models can be applied to any response endpoint. When determined in this manner, the estimator is referred to as the relative inhibition concentration or RIp. Stephenson et al. (2000) and Hughes et al. (2001) present several parametric nonlinear regression models that are suitable for the estimation of RIp.

Analysis of example data. The example toxicity data from the ITM, shown in Table 1, were analyzed using several of the methods described above. Results are presented in Table 2. SAS procedure statements are given in Appendix A.

Table 2 Effective Concentration Estimators from Analysis of Example Toxicity Data			
Type of Estimator	**Model**	**Value (% Elutriate)**	**95% Confidence Interval**
LC$_{50}$ * 0.01	Probit Logit Weibit	0.53 0.53 0.56	NA[1] NA NA
NOEC	Dunnett's test Williams' test Trend test (correlation)	0 0 0	NA NA NA

Low-effects Concentration
  LC$_{25}$
  LC$_{20}$
  LC$_{15}$
  LC$_{10}$
	Probit	
		20.9
		16.6
		12.7
		9.1
			
			15.7, 25.6
			11.8, 21.0
			8.4, 16.7
			5.5, 12.6

Inhibition Concentration
  IC$_{25}$
  IC$_{20}$
  IC$_{15}$
  IC$_{10}$
	Linear Interpolation	
		23.7
		19.3
		14.9
		10.2
			
			NC[2]
			NC
			NC
			NC

[1] NA = Not available. [2] NC = Not calculated.			

LC$_{50}$ was estimated using the SAS PROBIT procedure with three types of distributions specified: normal (Probit model), logistic (Logit model), and Gompertz (Weibit model). Model fit may be evaluated and compared by the goodness-of-fit χ^2 statistic given in the SAS procedure output. Best fit is evidenced by the lowest nonsignificant χ^2 (P > 0.05). Goodness-of-fit for the three models was as follows:

Probit	$\chi^2 = 1.76$	P = 0.42
Logit	$\chi^2 = 1.90$	P = 0.39
Weibit	$\chi^2 = 3.40$	P = 0.18

This indicates that the Probit model was the best fit. LC_{50} for both the Probit and Logit models was 53 percent elutriate, yielding 0.53 percent elutriate when multiplied by 0.01 for the WQS in the mixing model. Because of the 0.01 multiplier, confidence intervals are not available.

NOECs were estimated using Dunnett's test (SAS GLM procedure), Williams' test, and a stepwise trend test (Spearman's rank correlation in the SAS CORR procedure). Williams' test is not available in SAS although the critical *t* value may be computed using an SAS function. Regardless of the test, the NOEC for the example data was zero, i.e., less than the lowest tested elutriate concentration. Confidence intervals cannot be calculated for NOEC.

Low-effects concentrations were estimated using the SAS PROBIT procedure, Probit model. The estimates may be obtained from the same output table as the LC_{50}. Inverse 95-percent confidence limits (called fiducial limits) are provided. The low-effects concentrations range from 20.9 percent elutriate for LC_{25} to 9.1 percent elutriate for LC_{10}. Lower effects concentrations such as LC_5 or LC_1 can be obtained, but these will be less reliable, especially with data that are more variable or have poorer model fit than the example data.

Inhibition concentrations were estimated by hand calculation as percent decrease from the control survival using the linear interpolation method described in USEPA (1994). Inhibition concentrations ranged from 23.7 percent elutriate for a 25-percent decrease in survival (which equates to 27 percent mortality), to 10.2 percent elutriate for a 10-percent decrease in survival (which equates to 12 percent mortality).

RECOMMENDATIONS: Multiplying LC_{50} by an application factor of 0.01 to calculate the water quality standard in mixing zone determinations has no statistical or biological interpretation, and can result in an unrealistic mixing zone for inland receiving waters. The authors recommend substituting one of the statistical alternatives described above, particularly the low-effects concentration. Suitability of the data and goodness-of-fit of the regression model must be determined for successful use of low-effects concentrations. It will also be necessary to decide which low-effects concentration should be used, considering the limitations of the data and the desired level of environmental protection. The inhibition concentration is also acceptable, but its interpretation is less straightforward than that of the low-effects concentration when the endpoint is survival. The NOEC, despite its widespread popularity in ecotoxicity studies, has a number of statistical and practical limitations that should be given careful consideration before it is adopted for use in CDF evaluations. Upon adoption of a different effects level for a water quality standard for toxicity, the number of treatments and their dilution values used in the water column toxicity testing should be reviewed, and the testing protocols should be revised as warranted.

POINTS OF CONTACT: For additional information, contact Ms. Joan U. Clarke (601-634-2954, *clarkej@wes.army.mil*), Dr. Paul R. Schroeder (601-634-3709, *schroep@wes.army.mil*), Dr. Jeffery A. Steevens (601-634-4199, *steevej@wes.army.mil*) or the Program Manager of the Dredging Operations and Environmental Research Program, Dr. Robert M. Engler (601-634-3624, *englerr@wes.army.mil*). This technical note should be cited as follows:

Clarke, J. U., Schroeder, P. R., and Steevens, J. A. (2002). "Considerations for selection of Tier III water column toxicity standards for CDF effluent and runoff discharges: Statistical alternatives," *Environmental Effects of Dredging Technical Notes Collection* (ERDC/TN EEDP-04-33), U. S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/dots/eedptn.html*

REFERENCES

Bailer, A. J., Hughes, M. R., Denton, D. L., and Oris, J. T. (2000). "An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses," *Environmental Toxicology and Chemistry* 19(1), 141-150.

Capizzi, T., Oppenheimer, L., Mehta, H., Naimie, H., and Fair, J. (1985). "Statistical considerations in the evaluation of chronic aquatic toxicity studies," *Environmental Science and Technology* 19(1), 35-43.

Chapman, P. F., Crane, M., Wiles, J., Noppert, F., and McIndoe, E. (1996). "Improving the quality of statistics in regulatory ecotoxicity tests," *Ecotoxicology* 5, 169-186.

Crane, M., and Newman, M. C. (2000). "What level of effect is a no observed effect?" *Environ. Toxicol. Chem.* 19(2), 516-519.

Dunnett, C. W. (1955). "A multiple comparisons procedure for comparing several treatments with a control," *Journal of the American Statistical Association* 50, 1096-1121.

Gelber, R. D., Lavin, P. T., Mehta, C. R., and Schoenfeld, D. A. (1985). "Statistical analysis." *Fundamentals of Aquatic Toxicology, Methods and Applications*. G. M. Rand and S. R. Petrocelli, eds., Hemisphere Publishing Corporation, New York.

Hughes, M. R., Bailer, A. J., and Denton, D. L. (2001). "Toxicant- and response-specific comparisons of statistical methods for estimating effective concentrations," *Environmental Toxicology and Chemistry* 20(6), 1374-1380.

National Academy of Sciences. (1972). "A report of the committee on water quality criteria," National Academy of Sciences, Washington, DC, 118-123.

Moore, D. R. J., and Caux, P.-Y. (1997). "Estimating low toxic effects," *Environmental Toxicology and Chemistry* 16(4), 794-801.

SAS Institute Inc. (1989). "SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2," SAS Institute Inc., Cary, NC.

Scholze, M., Boedeker, W., Faust, M., Backhaus, T., Altenburger, R., and Grimme, L. H. (2001). "A general best-fit method for concentration-response curves and the estimation of low-effects concentrations," *Environmental Toxicology and Chemistry* 20(2), 448-457.

Skalski, J. R. (1981). "Statistical inconsistencies in the use of no-observed-effect levels in toxicity testing." *Aquatic Toxicology and Hazard Assessment: Fourth Conference, ASTM STP 737*, D. R. Branson and K. L. Dickson, eds., American Society for Testing and Materials, Philadelphia, PA, 377-387.

Stephan, C. E., and Rogers, J. W. (1985). "Advantages of using regression to calculate results of chronic toxicity tests." *Aquatic Toxicology and Hazard Assessment: Eighth Symposium, ASTM STP 891*, R. C. Bahner and D. J. Hansen, eds., American Society for Testing and Materials, Philadelphia, PA, 328-338.

Stephenson, G. L., Koper, N., Atkinson, G. F., Solomon, K. R., and Scroggins, R. P. (2000). "Use of nonlinear regression techniques for describing concentration-response relationships of plant species exposed to contaminated site soils," *Environmental Toxicology and Chemistry* 19(12), 2968-2981.

U.S. Environmental Protection Agency. (1994). "Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms," EPA/600/4-91/002, 3rd ed., U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory-Cincinnati, Cincinnati, OH.

U.S. Environmental Protection Agency/U.S. Army Corps of Engineers. (1998). "Evaluation of dredged material proposed for discharge in waters of the U.S. – Testing Manual (Inland Testing Manual)." EPA-823-B-98-004, Washington, DC.

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

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

APPENDIX A

SAS STATEMENTS FOR LC$_{50}$ AND LOW-EFFECTS CONCENTRATIONS: A data set is first created. Here the data set "A" contains the following variables: CONC (elutriate percent concentration), MORT (number dead), and N (number of organisms per treatment). LOG in the procedure (PROC) statement performs a log transformation of the data. D= in the MODEL statement specifies the distribution, where NORMAL is the Probit model, LOGISTIC is the Logit model, and GOMPERTZ is the Weibit model; LACKFIT requests goodness-of-fit analysis; and INVERSECL requests inverse confidence limits (95 percent fiducial limits) for the estimated effects concentrations.

```
PROC PROBIT LOG DATA=A;
 MODEL MORT/N=CONC /D=NORMAL LACKFIT INVERSECL;

PROC PROBIT LOG DATA=A;
 MODEL MORT/N=CONC /D=LOGISTIC LACKFIT INVERSECL;

PROC PROBIT LOG DATA=A;
 MODEL MORT/N=CONC /D=GOMPERTZ LACKFIT INVERSECL;
```

SAS STATEMENTS FOR DUNNETT'S TEST AND WILLIAMS' CRITICAL *t* VALUE: It is first necessary to create a new data set redefining MORT as number of dead divided by N. DUNNETTU in the MEANS statement specifies a one-tailed Dunnett's test, testing if any treatment is significantly greater than the control; ('0') indicates which concentration is the control. In the PROBMC function specification, 0.95 is the desired left probability of the Williams distribution, 20 is the degrees of freedom (total number of samples minus number of treatments), and 5 is the number of treatments (number of elutriate concentrations including control).

```
DATA B; SET A;
 MORT=MORT/N;

PROC GLM;
 CLASS CONC;
 MODEL MORT=CONC;
 MEANS CONC / DUNNETTU('0');

DATA D;
 T=PROBMC("WILLIAMS",.,.95,20,5);
PROC PRINT;
 TITLE '95% CRITICAL T VALUE FOR WILLIAMS TEST';
```

> ***NOTE:*** *The contents of this technical note are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such products.*