BUREAU OF WATER

Options for Data Analysis of Whole Effluent Toxicity Testing Required by NPDES Permits

2nd edition

September, 2001

Scope and Purpose

Toxicity tests, especially whole-effluent toxicity (WET) tests, are used in measuring compliance with National Pollution Discharge Elimination System (NPDES) permits and assessing biological risk in polluted aquatic systems. Toxicity test methods to be used for compliance with NPDES permit requirements, unless specified otherwise in the permit (SC Reg 61-9.122 Part C), are described in <u>Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms, 4th ed.</u>, EPA/600/4-90/027F, with undated addenda (Acute Manual), <u>Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Water to Freshwater Organisms, 3rd ed.</u>, EPA/600/4-91/002 (Freshwater Chronic Manual), and <u>Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Water to Marine and Estuarine Organisms, 2nd ed.</u>, EPA/600/4-91/003 (Saltwater Chronic Manual). These documents were incorporated by reference into Chapter 40, Part 136 of the Code of Federal Regulations, "Guidelines Establishing Procedures for the Analysis of Pollutants" at Federal Register vol. 60, no. 199, p. 53529, October 16, 1995. Errors in the promulgated toxicity test methods were corrected by incorporation of <u>Errata for the Effluent and Receiving Water Toxicity Testing Manuals: Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms; Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms; and Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms</u>, EPA/600/R-98/182 at Federal Register vol. 64, no. 21, p. 4975, February 2, 1999.

In addition to promulgated methods, EPA has issued the following guidance documents. These do not change the regulations, but provide EPA's interpretation of the regulations:

1.   <u>Whole Effluent Toxicity (WET) Control Policy</u>, EPA/833/B-94/002 relates WET testing to NPDES permitting practices, including reasonable potential determinations, compliance schedules and limits.

2.   <u>Technical Support Document for Water Quality-based Toxics Control</u>, EPA/505/2-90/001 presents methods for integrating toxicity test data into a comprehensive permitting program to achieve water quality standards.

3.   Clarifications of flexibility in the promulgated methods were described in a memorandum dated April 10, 1996, from Tudor T. Davies, Director of the Office of Science and Technology.

4.   Clarifications and guidance on WET requirements in NPDES permits were described in a memorandum dated July 21, 1997, from Mr. Davies and Michael B. Cook, Director of the Office of Wastewater Management.

5.   A less technical discussion of WET test method variability, published as part of a settlement agreement, is <u>Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing (40 CFR Part 136)</u>, (EPA 821-B-00-004, July 2000) (Method Guidance), which is available as a portable document file at http://www.epa.gov/OST/WET/wetguide.pdf.

6.   A technically detailed clarification of selected issues and additional guidance to regulators, permittees and laboratories as part of a settlement agreement were provided in <u>Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program</u> (EPA 833-R-00-003, June 2000) (Understanding Variability) which is available at http://www.epa.gov/owm/npdes.htm. This document identifies sources of test variation and discusses how to minimize and address it in the NPDES program.

According to SC regulation 61-9, <u>Water Pollution Control Permits</u>, tests must be performed

according to methods in 40 CFR 136 or methods described or referenced in the permit. While the acute and chronic manuals provide some methods for WET data analysis, Chapter 11 of the Acute Manual and Chapter 9 of the Chronic Freshwater Manual recognize that other analyses may be more appropriate, depending on circumstances, and gives the permittee and regulatory agency broad discretion in this area. Method Guidance and Understanding Variability provide some suggestions for reducing method variability through the use of test acceptability criteria (TAC's) that are more stringent than those in regulations, and through excluding from NPDES reporting test endpoints (statistics) that vary greatly from expectations (invalidation).

Variation in toxicity test results can be reduced by reducing the variance in each test. Within-test variance has two components. The first is due to deviation of an individual response from the estimate of expected response. This is often expressed by the synonymous terms random error and residual error. The second is due to differences between test results. While this is influenced by random error, it is further controlled by test design, which includes number and level of concentrations, number of test organisms and their proportional allocation to each concentration, and statistical analysis. TAC's attempt to control random error through more consistent test methods and reduce test result variability by rejecting tests with control groups that vary inordinately from expected performance and tests with random error that greatly exceeds expectations.

Optional Test Acceptability Criteria for WET tests in NPDES permits

TAC's are numeric criteria that a test must meet to be considered acceptable for NPDES reporting (a valid test). A TAC must be mandatory and unconditional for an exception to automatically invalidate a test. The TAC's in the Freshwater Acute and Chronic Manuals are mandatory, but may be exceeded on the condition that review of the data demonstrates that test quality was not compromised. The recent EPA guidance documents allow the Department to offer optional TAC's. If the permittee elects to use these optional TAC's in determination of permit compliance, the TAC's must apply uniformly to all tests and the NPDES permit must reference the TAC's to be used.

The optional TAC's for any chronic *C. dubia* tests are:

1.   At least 80% of control organisms must produce 3 broods.

2.   The test endpoint is calculated using data only from the first 3 broods,

3.   Control group reproduction must be less than $2\bar{X}-15$, where $\bar{X}$ is laboratory average control reproduction. (This corrects a theoretical bias ($E(\bar{X}_c-\bar{X}_t)>E(\mu_c-\mu_t)$) by restoring symmetry to the distribution of control averages. The upper tail of the distribution of control averages is truncated by the amount truncated by the lower limit of 15 on control group averages.) The laboratory shall calculate an overall control group reproduction average from control groups of all valid tests.

The optional TAC for multi-concentration tests using any test species is:

4.   The coefficient of variation (CV) for the control group reproduction must be less than the 90th percentile for the test species from Table 3-6 in Understanding Variability or established by the laboratory performing the test.

The optional TAC's for single concentration tests using any test species are:

5.   The coefficient of variation (CV) for the control group must be between the 10th and 90th percentiles for the test species from Table 3-6 in Understanding Variability or established by the laboratory performing the test.

6.   The percent minimum significant difference (PMSD) as defined in Section 6.4.1 of Understanding Variability must be between the 10th and 90th percentiles for the test species from Table 3-6 in Understanding Variability or established by the laboratory performing the test.

Options not completely described above must be detailed in the permit.

Expected failure rates

The Department recognizes that setting α levels creates the expectation that $\alpha*100\%$ of tests for an effluent will violate limits if the effluent is never toxic. If N tests are performed over the life of a permit, the expected number of violations, $E(N_v)$ is $\alpha*N$. Since this is an expectation, the actual number of violations will not generally be $E(N_v)$. A limit violation depends on whether the test endpoint, % effect at CTC, is above or below the limit, so violation data are Bernoulli random variables. Since effluents are tested with replacement (you can collect and test all you want), violations are assumed to be binomially distributed if the effluent is never toxic. If the α levels are treated as the parameter P of a binomial distribution, an approximate sampling distribution of the estimator of $E(N_v)$, $\hat{N}_v = \dfrac{number\ of\ actual\ violations}{number\ of\ actual\ tests}$, can be established, which is necessary to characterize variation in the actual number of violations. If we let $E(N_v)=P*N$ given a binomial distribution with known parameters P and N, then quantiles of the distribution of $\hat{N}_v$ can be approximated from $N_{v(UCL,1)} = \dfrac{\sqrt{\chi^2_{(UCL,1)}E(N_v)}}{E(N_v)}$ by solving $\chi^2_{(UCL,1)} = \dfrac{(N_{v,(UCL,1)} - E(N_v))^2}{E(N_v)}$, since $\hat{N}_v$ has an approximate χ^2 distribution. UCL is the desired percent upper confidence limit divided by 100. The only unknown value now is χ^2, which is easily obtained from the inverse of the chi-square distribution as the value corresponding to 0.95 of the area of the distribution at 1 degree of freedom for a 95% UCL.

The probability that N_v effluent samples will violate limits after N tests are performed is a different but related question. This is answered by treating α, N_v and $N-N_v$ as parameters of a negative binomial distribution and calculating probabilities. Both approaches produce results that correspond closely, but not exactly. The main source of differences is use of continuous distributions to characterize discrete random variables. A lesser source is that these distributions are approximate. Continuity corrections are not recommended, as they cause the distributions to approximate distributions of hypergeometric distribution averages, which is appropriate for sampling without replacement.

Since chronic tests involve 2 simultaneous comparisons with a limit, one for survival and one for reproduction, these variables are jointly distributed. Correlation must be accounted for in any joint distribution. Measurements of correlation are not available, but expectations can be computed under 2 assumptions, that of independence and that of complete dependence, that represent no correlation and complete correlation. When an effluent is not toxic, the results are independent. Dependency is complete when all survival and reproduction results in terms of meeting the criteria are the same. An example is when an effluent kills all test organisms before they reproduce, although dependency may arise from other conditions. Assuming independence, each value obtained is counted as a separate test (each toxicity test is treated as 2 results), and the α level is the sum of the probabilities of events that would cause a violation, or $P(f,f)+P(p,f)+P(f,p)=\alpha^2+2*\alpha*(1-alpha)$, since the probability of failing is α and the probability of observing a joint occurrence of values of independent random variables is the product of their individual probabilities. Assuming dependence, each test is treated as one value (one result) and the probability of violation is α. The true probability that N_v effluent samples will violate limits after N tests if the effluent is not toxic will be between the probability assuming independence and that assuming dependence, but should be closer to the

independence case. **These expectations do not extend to an effluent that is toxic at one or more samplings. Meeting these expectations does not change compliance status of any test. Expectations may be used to aid interpretation of compliance patterns over time.**

Single concentration test design, analyses, and limits

Single concentration (pass/fail) tests are most appropriate for a discharger whose only interest is to assess environmental risk at an effluent concentration related to a specific effluent flow and minimum receiving water flow. The endpoint of this test is survival after 48 h in an acute test, and survival and reproduction during the release of the first 3 broods of neonates in a chronic test. The test consists of a control group of organisms that are reared in laboratory culture water and a test group in effluent diluted with culture water to a concentration specified in the permit. The test concentration (TC) may be higher than effluent concentrations predicted from data or models. TC refers to both acute and chronic test concentrations. Analysis is performed according to Appendix H of the Freshwater Chronic Manual. In NPDES permits, limits are based on outcomes of tests of two statistical hypotheses. One is that control survival is less than or equal to test survival against the alternative that control survival is greater than test survival. The other is that control reproduction is less than or equal to test reproduction against the alternative that control reproduction is greater than test reproduction. If either null hypothesis is rejected, the test is reported as a fail. Limits are on the maximum, so failure of any test during a monitoring period is a limit violation.

The $''$ levels for these tests are selected according to 40 CFR 136 and Method Guidance, Chapter 2. To test at $''=0.01$, the PMSD$_{''}$ of an individual test must be between the 10th and 90th percentiles for the test species from Table 3-6 or the laboratory-specific PMSD$_{''}$. $''$ is reported on the DMR attachment. Figure 2.2 of Method Guidance is a flowchart of a process that may be used to determine whether increased replication is needed in order to achieve the PMSD$_{''}$ needed to test at $''=0.01$ (PMSD$_{0.01}$).

Multiple concentration test designs, regressions, and limits

Multiple concentration (MC) tests are appropriate for estimating biological effect over a range of effluent and stream flow conditions. Survival and reproduction are biological effects that are estimated for *C. dubia* from MC test data. Limits are placed on biological effect at the TC, which is (1-TC group performance/control group performance)*100. MC data are analyzed by regression, which consists of finding an equation that adequately describes the data. The equation is then used to estimate biological effect at any concentration within the range of tested concentrations, including TC. When the data are regressed, estimated control and TC group performance will not generally be equal to group means, although the differences should be slight. Regression has not been formally recommended by EPA, but is discussed as a subject for further research in Section 5.5 of Understanding Variability. Chapter 4 of Method Guidance is a readable introduction to MC test data patterns that describes some violations of the monotonic dose-response assumed for NOEC's. Unfortunately, Chapter 4 does not discuss statistical testing of monotonicity, and does not provide techniques for modeling non-monotonic data. To address these problems, the Department allows us of regression for analyzing MC tests. A permittee may use the 3-parameter logistic regression (3PLR) model described below or propose other regression models and methods. If another model is proposed, the model and its underlying assumptions must be agreed upon by all parties prior to permit issuance or modification. The model will be specified in the NPDES permit.

Test design considerations and options.

Test design is specified in the NPDES permit. A dilution water control group is required in any design. The Acute and Freshwater Chronic Manuals require a minimum of 5 concentrations containing 10 test organisms each and suggest effluent concentrations of 100%, ((100+TC)/2)%, TC%, (TC/2)%, (TC/4)%. This design may not meet the objectives of the Department or permittee, especially at TC's above

80%. When TC is 100%, there are 3 concentrations of 100%, which defeats the intent of multiple concentrations. A more relevant design can be derived by picking a high concentration such as the TC and a low concentration such as 50% and determining a geometric series of concentrations. Given a low concentration L, a high concentration H, and n concentrations, the concentration factor is $F = \left(\dfrac{H}{L}\right)^{\frac{1}{n-1}}$ and the ith concentration in a geometric series is $C_i = L * F^{i-1}$. In the case where a discharger hopes to demonstrate that it has no reasonable potential to exceed or contribute to an exceedence of narrative stream biological criteria, 100% may be a useful high concentration.

Maximum likelihood regression

Data from toxicity tests which include multiple concentrations of effluent may be used to predict biological function over a range of effluent concentrations. The statistical technique for deriving a descriptive relationship between a response variable, such as biological function, and a ratio or interval scale explanatory variable, such as effluent concentration, is known as regression. Draper and Smith (1981) provide a comprehensive treatment of regression and understandable introductions to matrix algebra and nonlinear regression. This section gives a brief explanation of maximum likelihood regression concepts with an example that is recommended for use in NPDES permits, although other models may be used. Some criteria for assessing other non-linear models will be goodness of fit to several sets of data, degree of non-linearity of parameters, and ability of the model to reach a convergent solution. The formulas are given to assist in using existing statistics software, but do not give enough detail to enable one to do the calculations by hand or to write a program.

A regression model is a combination of a relational model and a probability distribution. The relational model is an approximation of the relationship between the response and explanatory variables, and the probability distributional is an approximation of the errors, which are the differences between estimated and actual responses. A method for combining the relational model and probability distribution must be selected. The most desirable method in most cases is maximum likelihood. There are an infinite number of possible regression models, but a few desirable models can be selected for any problem based on widely accepted statistical principles. The regression model that results from combining the relational model and distribution is solved, or fitted to the data, by computing estimates for parameters. These are denoted by Greek letters. This process is sometimes referred to as "finding a line of best fit." The models discussed here are not lines, and selection of the distribution and fitting method that define best fit are based on rigorous statistical theory, but the concept is the same. The parameter estimates determine the particular regression model that fits a set of data and the predictions of biological impairment at selected effluent concentrations.

The well-known least squares regression refers to fitting a polynomial relational model to data that are normally distributed with equal variances at all levels of the explanatory variable. Nonlinear regression usually refers to iteratively reweighted least squares regression of a nonlinear model. The maximum likelihood method generalizes regression to nonlinear models and non-normal probability distributions. Probit analysis is a maximum likelihood regression in which mortality is related to substance concentration using the cumulative normal distribution as a relational model and the binomial probability distribution. Seber and Wild (1989) give a comprehensive treatment of nonlinear regression, including maximum likelihood regression, and discuss nonlinear regression of sigmoid models. The maximum likelihood method can be used to combine relational models and probability distributions other than the ones discussed here. The maximum likelihood method is the preferred method for regression because it has many desirable properties, and is most often used in statistics problems.

Maximum likelihood derives its name from the strategy of maximizing the likelihood of a sample by

5

choosing the appropriate value for a parameter estimate. The likelihood of a sample is the product of the sample probabilities. These are obtained from probability distributions, such as the binomial and Poisson. Since the data values are fixed, the likelihood of a sample is maximized when a particular value for the distributional parameter is chosen. In a regression, the parameter is replaced by the relational model, which contains parameters, so the likelihood is maximized by the right choice of regression parameters. The maximum likelihood estimates for the parameters are obtained by taking the natural log of the product of the probability for each response value, replacing its parameters with the relational model and differentiating with respect to each of the regression parameters. When set equal to zero, these partial derivatives are called the normal equations, solution of which yields maximum likelihood estimates of the regression parameters. The resulting parameter estimates minimize the loss function, which is the negative logarithm of the probability distribution. Hence, the "regression of best fit" is the one which produces the lowest value of the loss function. In a least-squares regression, the data are assumed to be normally distributed, which makes the loss function the sum of squares. The maximum likelihood parameter estimates are those that minimize the sum of squares.

After the regression is solved, asymptotic variances of the parameter estimates can be calculated. Asymptotic indicates that variances are correct for large samples, but may be biased for small samples. Unlike in least-squares regression, the population variance is not F^2 since the count distributions do not contain F^2. Instead, variance is a function of the mean. Most formulas and algorithms for calculating variance assume that the variance-covariance matrix is to be multiplied by the residual mean square, which is the estimate of F^2. To obtain asymptotic variance estimates, F^2 must be set equal to 1 so that it does not affect the variance-covariance matrix.

In a linear model, a critical step in obtaining parameter estimates is to invert the Fisher matrix, which is a function solely of the design (number of organisms and concentrations). The inverted Fisher matrix is the variance-covariance matrix. A resulting feature of linear models is that parameter estimates are computed directly from the design and results. Non-linear models do not have a true Fisher matrix. They have a substitute called the Jacobian matrix, a matrix of first derivatives of the relational model with respect to the parameters, which is used in the solution process like the Fisher matrix. This substitution is needed because the Fisher matrix a non-linear model is unknown. The Jacobian is a close approximation that allows approximate parameter estimates. Likewise, inverting the Jacobian matrix yields an approximate variance-covariance matrix. Another matrix that is sometimes encountered in fitting nonlinear models is the Hessian matrix. This is a matrix of second derivatives of the relational model with respect to the parameters. While the Jacobian matrix is always used, only some fitting algorithms use the Hessian matrix. For small problems, like usual WET tests, using the Hessian usually speeds convergence. In larger problems, where the Hessian is extensively reevaluated, computer resources can be exceeded and convergence can be slowed. Some programs allows the user to specify a particular algorithm, while others ask only whether to use the Hessian or second derivatives in fitting. For WET tests, usually the best approach is to use the Hessian in fitting.

Sigmoid relational models

Survival data range from 0 to 100%, and reproduction data range from 0 to a positive number. When increasing concentrations of a substance may cause increased biological impairment, impairment is negligible up to some concentration, after which the rate of change of impairment rapidly increases until it becomes nearly constant. Above a concentration where impairment is nearly complete, the rate of change of impairment decreases rapidly until impairment again increases negligibly as concentration increases. Such a pattern is well described by a sigmoid relational model, which is illustrated in Figure 1.
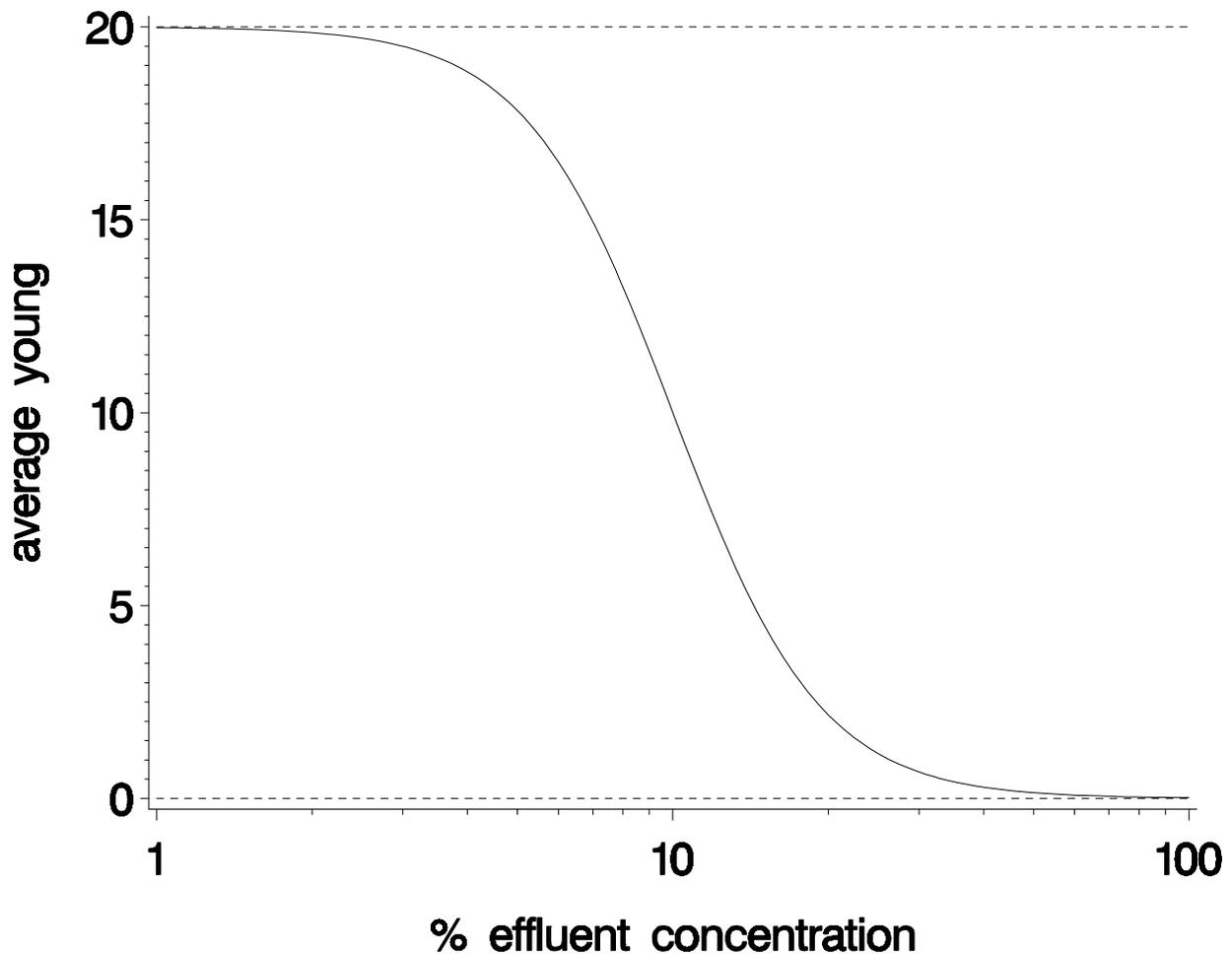
6

Figure 1. A sigmoid dose-response regression model. Horizontal dashed lines denote asymptotes.

Sigmoid models range between upper and lower asymptotes, represented by the horizontal dashed lines. Asymptotes are approached but never reached, although the regression appears to merge with the asymptotes. A sigmoid model which is particularly useful for reproduction data is the **3-parameter logistic**

model $Y = \begin{cases} \dfrac{\gamma}{1+e^{-\beta*(\log_{10}(X)-\alpha)}}, & X>0 \\ \gamma, & X=0 \end{cases}$ where X is effluent concentration and Y is reproduction. It is

symmetric about ", the inflection point, which is the concentration which causes 50% impairment (EC $_{50}$). (is the value of the upper asymptote and mean control reproduction, and $ is the steepness of the curve. High (large positive) and low (large negative) values of $ cause the curve to be steep, while the regression would be a horizontal line if $ were zero. While there are other equations that will produce this relationship, this one is recommended because it has desirable estimation properties. Ratowsky (1990) describes this and other nonlinear regression models.

Historically, toxicologists have focused on death as a measure of acute toxicity, while demographic ecologists have focused on survival. Survival and mortality data are counts of organisms which survived or died during a toxicity test. A useful model for mortality data is the **spontaneous response logistic**

$$Y = \begin{cases} \gamma + \dfrac{100 - \gamma}{1 + e^{-\beta*(\log_{10}(X) - \alpha)}}, & X > 0 \\ \gamma, & X = 0 \end{cases}$$, which occurs when an adjustment for spontaneous mortality (Finney, 1978)

is applied to the 2-parameter logistic model, which is a 3-parameter logistic where gamma is replaced by 100%. γ is percent mortality which occurs in the absence of effluent, so 100-γ is the estimate of percent control survival. When γ=0, the spontaneous response logistic reduces to the 2-parameter logistic.

The Department has found it more convenient to limit reduction in survival rather than increase in mortality, partly to be consistent with other limits. The survival counterpart to the spontaneous response

logistic is the **3-parameter logistic** used for reproduction $$Y = \begin{cases} \dfrac{\gamma}{1 + e^{-\beta*(\log_{10}(X) - \alpha)}}, & X > 0 \\ \gamma, & X = 0 \end{cases}$$. The difference

is that γ is control percent or proportion survival. The spontaneous response is (100-γ)%.

It is important to note that survival can be modeled as either proportion or percent survival or mortality. Proportion is a number between 0 and 1 inclusive, and percent is proportion*100. Whether proportion or percent is modeled depends on how survival or mortality data is input. Inputing survival or mortality data as proportion (# alive/n) leads to an answer in proportion, while inputing the data as percent (# alive*100/n) leads to an answer in percent.

The binomial probability distribution for survival and mortality data

The random variable is a Bernoulli random variable, because it can take only one of two values, live or dead. While there are practical limits on the number of individuals that can be cultured in a laboratory, the test organisms are assumed to be part of the population of all individuals. Each laboratory culture is not considered a separate population. Under this assumption, a Bernoulli random variable is considered to be binomially distributed, so the binomial distribution is the preferred distribution to assume for survival data. The probability that the number of individuals surviving, *s*, will be a specified value *S* is given by the

binomial probability distribution $P(s = S) = \dfrac{n!}{s!(n-s)!} p^s (1-p)^{n-s}$ where *n* individuals are tested and *p* is the

proportion of the population that will survive. The probability that *d=n-s* individuals will die is 1-*P*(*s*).

The Poisson probability distribution for count data

Reproduction data are counts of young organisms produced during the toxicity test. If all test organisms in an effluent concentration released the same number of young during a test, there would be no need for statistical analysis. To get an accurate fit of the relational model and some measure of its appropriateness, it is necessary to model the random error, which is the unexplained variation between individuals. For counts, the Poisson distribution may be appropriate. A special property of the Poisson is that the sample average equals the sample variance, which is sometimes a reasonable assumption for counts. Estimation is simple, since only the sample average is computed. For these reasons, the Poisson is a recommended distribution to assume for reproduction data. The probability that the number of young, y_i, released by an adult will equal some specified value *Y* is given by the **Poisson probability distribution**

$$P(y = Y) = \dfrac{e^{-\lambda} \lambda^y}{y!}$$ where λ is the population mean.

Potential problems with fitting nonlinear models

In a linear model, derivatives of the loss function are solved for the parameters algebraically. A distinguishing feature of non-linear models is that the derivatives cannot be solved algebraically, so numerical methods must be used. Numerical methods are trial and error methods. For example, $5 = x\,e^x$ cannot be solved for x. However, an approximate value can be found by trying some reasonable guesses for x and evaluating the right hand side, then making improved guesses based on those results. One is a reasonable starting point, since it would evaluate to about 2.7. Two will produce a number much larger than 5, so x is between 1 and 2. Numerical methods used in non-linear regressions differ from this example mostly in that they can guess a step size and direction. Step size is the difference between the previous guess and current guess, and direction is whether the step is greater or less than the previous guess. Repeated guessing is called iteration, and the approximate solution is called a convergent solution. Convergence is achieved when a small change in parameter guesses causes a sufficiently small change in the value of the loss function.

A common problem in nonlinear models is failure to converge. While linear models can be unsolvable due to poor design, certain data patterns can prevent convergence in nonlinear estimation. Other causes can be lack of fit of the model to the data, poor initial parameter estimates, or use of an inappropriate fitting algorithm. Lack of convergence in these cases can sometimes be overcome by supplying different initial parameter values.

For convergence failure, an analysis program may print a message such as "The Jacobian (or Hessian) was singular," but usually returns parameter values when the process was stopped. Since obtaining values for all parameters and their variances requires inversion of the Jacobian, and a singular matrix cannot be inverted, the regression cannot be done. This problem is usually caused by an uninformative test. Graphs of these results make it intuitively obvious that the data do not provide much information about the EC50 and degree of dose response (curvature). When applying the 3PL model to toxicity data, these non-convergent tests fall into the following categories:

1)   **Lack of effect**. This is usually expressed as ludicrous values for the EC50 (denoted **"**) and curvature (denoted **$**), such as an EC50 outside the range of tested values and very high (>1000) or very low or negative (<1) curvature. The resulting model will be a straight line across all concentrations at the control performance level. This implies that the effluent was not toxic and results should be reported as 0% effect at the test concentration.

2)   **Low (<10) curvature or EC50 above the highest tested concentration**. This may represent slight toxicity, lack of fit to the model, or variance problems. It does not provide evidence for a dose-response pattern and results should be reported as 0% effect at the test concentration.

3)   **High curvature within the range of concentrations**. This occurs when there is complete survival up to some concentration, then complete mortality at all higher concentrations. The regression will report very high curvature with the EC50 between two concentrations. In this case, the value of the curvature parameter can be assumed to be 1,000 and the EC50 can be assumed to be $[X_i * X_{i+1}]^{0.5}$ where X_i is the highest concentration at which mean reproduction or survival is greater than 50% of control, and X_{i+1} is the next highest concentration. Effect at a concentration can be calculated and reported using these assumed values.

A convergence problem that is often hard for either the user or algorithm to detect is convergence on a local rather than global minimum of the loss function. A good analogy is a ball rolling down a hill. Gravity (the algorithm) causes the ball to roll down the hill (the loss function). If the ball rolls into a hole in the side of the hill, it stops (convergence). The only way to start the ball rolling again is to pick it up and let it roll again (start with different initial parameter estimates). To keep a hole from causing problems, let the ball roll from a different place on the hill (start with several initial parameter estimates). Sometimes, convergence on a local minimum has little consequence for the parameter estimate solutions. This is when

the loss function has a wide, flat valley. Occasional, this can cause substantial errors (when a hole is much higher than the bottom of the hill). It is recommended that users try several sets of initial values even if the algorithm is converging to avoid this trap.

A second source of fitting failure is for the algorithm to converge on something other than a minimum in the loss function, usually a saddle point. Existence of a saddle point arises from problems other than an uninformative test. Some programs generate a message like "The Hessian (or Jacobian) was not positive definite," a matrix property that indicates a minimum was reached. This is usually caused by an error in data entry or in assignment of columns or variables to be dose, response, model, etc.

A third problem is the inability of the algorithm to find an appropriate step size. A step is an evaluation of the loss function at the current guesses of the parameter values, and a step size is simply the amount of change in the guesses from one step to the next. Fitting a nonlinear problem depends in part on stepping far enough to get a detectable change in the loss function, but not so far that the algorithm winds up further from the global minimum than it started from. Most algorithms decrease step size as a minimum is approached, and new step size is frequently half of the previous step size. While this strategy often works, there are situations where the steps get too small too fast. Since the algorithm is unable to detect a decrease in the loss function if the step is too small, step size halvings can prevent convergence. When this happens, the software may display a message about step size, such as "Step size failed to improve" or "Failed to converge after many step halvings." In these cases, convergence can often be achieved by allowing no or only a few step halvings. In rare cases, adjustment to the initial step size itself can be beneficial.

SAS/STAT® PROC NLIN and JMP

Two products that are known to be capable of performing maximum likelihood regressions are JMP and SAS/STAT®, both available from SAS Institute, Inc. JMP is a menu-driven, spreadsheet interface program with most of the functionality of SAS. The data, model, and loss and weight functions are entered in columns of the spreadsheet. The analysis is obtained by selecting the NONLINEAR FIT option under the ANALYZE menu, and identifying the columns. A SAS program for performing analysis is available upon request.

Comparison of the "Modifications Documents" to current regulations and practices

Some older NPDES permits reference "South Carolina Procedures for Pass/Fail Modification of the *Ceriodaphnia* Survival and Reproduction Test, September 1988", "South Carolina Procedures for Pass/Fail Modifications of the *Ceriodaphnia* 48-hour Acute Toxicity Test and *Ceriodaphnia* Survival and Reproduction Test, May 1989" and "South Carolina Procedures for Screening Modifications of the *Ceriodaphnia* 48-hour Acute Toxicity Test and *Ceriodaphnia* Survival and Reproduction Test, May 1989", hereafter referred to as the "Modifications Documents." Some of these permits additionally reference a "superseding document". These test modifications were not contained in 40 CFR 136. The present document supersedes the "Modifications Documents" and describes elements of toxicity testing that are not specifically addressed in the regulations. Tests performed under a permit which **specifically** references "superseding document" or the present document should now be performed according to the promulgated methods as explained in the present document. It is the permittee's responsibility to ensure that all analyses conform to conditions of the NPDES permit, and state and federal regulations.

The "Modifications Documents" contained effluent sampling type requirements, specified that 20 organisms would be used in chronic tests, that one organism would be put in each test vessel in acute tests, that Fisher's exact test would be used to analyze all mortality data, and contained a 2-page form to be used for submittal of test data. Promulgation of toxicity test methods in 40 CFR 136 made the "Pass/Fail Modifications" unnecessary, and sample types are now specified in NPDES permits.

40 CFR 136 requires a minimum of 10 organisms per concentration for chronic tests, regardless of number of concentrations. For acute tests, 40 CFR 136 specifies putting five organisms in each test vessel, and using four vessels for a total of 20 organisms per concentration. Mortality data from acute tests are analyzed according to a flowchart procedure in 40 CFR 136 that includes tests of assumptions and hypotheses. For chronic tests, reproduction data are analyzed according to a flowchart procedure in 40 CFR 136 that includes tests of assumptions and hypotheses, and mortality data are analyzed with Fisher's exact test. Toxicity test data beyond that reported on the DMR are reported on DHEC form 3420, which is pre-printed and issued to the permittee along with the DMR.

References

Draper N, Smith H. 1981. *Applied Regression Analysis, 2nd ed.* John Wiley & Sons, New York, NY.

Finney DJ. 1978. *Statistical Method in Biological Assay*, 3rd ed. Macmillan Publishing Co., Inc., New York, NY.

Ratowsky DA. 1990. *Handbook of Nonlinear Regression Models*. Marcel Dekker, Inc., New York, NY.

Seber GAF, Wild CJ. 1989. *Nonlinear Regression.* John Wiley & Sons, New York, NY.