APPENDIX D

FREQUENTLY ASKED QUESTIONS (FAQS)

This page intentionally left blank.

FREQUENTLY ASKED QUESTIONS (FAQS)

Appendix D contains some of the frequently asked questions regarding WET and WET testing. These questions and answers were prepared by and appear on a web site maintained by the Society of Environmental Toxicology and Chemistry (SETAC) (http://www.setac.org). The SETAC WET Expert Advisory Panels provide scientific opinion and training on WET technical issues under a cooperative agreement with EPA (WET Cooperative Agreement No. CX 824845-01-0). EPA's inclusion of these questions and answers in this document is not an endorsement of the Panels' opinions or responses to the FAQs, but rather provides readers with an additional source of information in issues commonly raised with regard to WET and WET testing. This information was prepared in response to questions received by SETAC about WET. It was generated by the WET Expert Advisory Panels (EAP) Steering Committee (SC), all volunteers and all member of the Society of Environmental Toxicology and Chemistry. Each person is considered an expert in some aspect of WET, and the information provide in these FAQs represents the consensus of the Committee's collective expertise at the time this summary was written (Feb., 1999).

This information is intended to stimulate further discussion about WET, WET-related research, and the science underlying WET. The information is not to be construed as representing an official position of SETAC, the SETAC Foundation for Environmental Education, or the U.S. Environmental Protection Agency. Any questions, comments, and requests should be sent to: Society of Environmental Toxicology and Chemistry (SETAC), 1010 North 12th Avenue, Pensacola, FL 32501-3367, Telephone: 850-469-1500, Facsimile: 850-469-9778, e-mail: setac@setac.org. All materials copyright Society of Environmental Toxicology and Chemistry (SETAC), 2000, and may not be used without written permission.[1,2]

Whole effluent toxicity tests rely on the assumption that test organisms used are representative of a normal and healthy population. What indicators of test organism health are utilized in testing programs?

Both subjective and objective (e.g., test acceptability criteria) indicators of organism health are available, some described within the methods manuals. Some national indicators exist which allow comparison of analytical results between laboratories (i.e., the DMRQA program for major NPDES facilities) or regional activities such as State WET certification programs which provide round-robin validation of test practice including organism health (e.g., North Carolina's Biological Laboratory Certification program). Other national programs like the National Environmental Laboratory Accreditation Program (NELAP) are being followed by the WET EAP SC. Commonly used indicators of organism health are the required reference toxicity analyses and individual test acceptability criteria. Tests properly utilizing randomization procedures along with required and suggested quality control standards retain many built-in checks of typical organism response.

What are the definitions of acceptability criteria for reference toxicant tests?

Reference toxicant tests should meet the same test acceptability criteria as those of compliance test. With regard to assessment of organism health and the overall test practice, USEPA has recommended that routine reference toxicant tests be performed to establish a CUSUM or cumulative summation chart of testing results. Normal results should lie within plus or minus two standard deviations of the cumulative mean value

---

[1] Reprinted with permission of SETAC.

[2] Note that the terms, abbreviations, and acronyms used in this appendix may differ from their usage throughout the rest of this document. EPA consciously chose not to edit this SETAC-supplied information so that the actual nomenclature and terminology as used by SETAC on their web site would be reflected here.

of point estimate endpoints. Values falling outside of those ranges should result in careful scrutiny of the data and testing systems. Data produced during these "out of control" conditions should be considered suspect.

How does increasing the difference in test concentration dilutions affect the prediction of response?

Better resolution around threshold effect concentrations provide better input to mathematical models to predict point estimations of effect and reduce uncertainty in hypothesis tests of effect. Reducing the distance between effluent dilutions should be encouraged. There may be some confusion about USEPA's specification of dilution series in these cases. The methods specify a minimum set of dilutions, i.e., no wider than 0.5 dilution between concentrations. No limitations on added concentrations within that range exist. Experimental design should account for concentrations of concern and should attempt to maximize resolution in that range. Test design should maximize test concentrations around the effect concentration of concern, i.e., the instream waste concentration or limited concentration of a discharging facility, in order to minimize the need for interpolation of effects between tested concentrations.

What are the different types of variability in whole effluent toxicity tests?

Variability is inherent in any analytical procedure. The precision of a method describes the closeness of agreement between test results obtained from repeated testing of a prescribed method. WET test precision can be categorized by: 1) intratest (within-test) variability, 2) intralaboratory (within-laboratory) variability, and 3) interlaboratory (between-laboratory) variability. Intratest variability can be attributed to variables such as the number of treatment replicates, the number of test organisms exposed per replicate, and the sensitivity differences between individual organisms (i.e., genetic variability). Intralaboratory variability is that which is measured when tests are conducted under reasonably constant conditions in the same laboratory (e.g., reference toxicant or effluent sample tested over time). Sources of intralaboratory variability include those factors described for intratest variability, as well as differences: 1) in test conditions (e.g., seasonal differences in dilution water quality, differences in environmental conditions), 2) from test to test in organism condition/health, and 3) in analyst performance from test to test. Interlaboratory variability reflects the degree of precision that is measured when the same sample or reference toxicant is analyzed by multiple laboratories using the same methods. Variability measured between laboratories is a consequence of variability associated with both intratest and intralaboratory variability factors, as well as differences allowed within the test methods themselves (e.g., source of dilution water), technician training programs, sample and organism culturing/shipping effects, testing protocols, food quality, and testing facilities.

Two general categories of variability are of greatest concern: 1) analyst experience, and 2) test organism condition/health. The experience and qualifications of the analyst who actually performs the toxicity test in the laboratory will dictate how well the culture and test methods are followed and the extent to which good judgment is exercised when difficulties/issues arise in the process of conducting the test, analyzing the data, and interpreting the results. Improper utilization of WET methods can have a substantial impact on test result variability. Guidance for specific test conditions and standard methods to control many causes of variability are found in the USEPA (U.S. Environmental Protection Agency) methods manuals (USEPA 1993, USEPA 1994a, USEPA 1994b, USEPA 1995). Strict adherence to these methods can greatly reduce variability.

USEPA. 1993. Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms. 4th ed. Weber C.I., editor. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-90/027F. 293 p.

USEPA. 1994a. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to marine and estuarine organisms. 2nd ed. Klemm, D.J., Morrison, G.E., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors.

Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/003. 341 p.

USEPA. 1994b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. 3rd ed. Lewis, P.A., Klemm, D.J., Lazorchak, J.M., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/002. 341 p.

USEPA. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. Chapman, G.A., Denton, D.I., Lazorchak, J.M., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/R-95-136. 661 p.

What specific factors influence WET test variability?

There are a number of factors that can meaningfully influence the variability of test results. These factors include, but are not limited to, those listed below.

Sample Characteristics

The nature of the sample collected can have a significant influence on the outcome of a WET test. Care must be exercised to collect the most representative sample possible during the time frame of interest. Sample volume can influence the outcome of a toxicity test. For example, if the sample-to-container-wall ratio is small, or if the sample-container contact time is especially long before the sample is refrigerated; certain particulate-active constituents such as zinc (Chapter 5 in Grothe et al. 1996), polymeric substances, charged materials, or hydrophobic chemicals in a sample can interact with the container. Samples too small in volume may also increase the potential of collecting a non-representative fraction of a non-homogenous sample stream. The type of sample (i.e., grab or composite) may influence the outcome of a WET test and contribute to variability. Grab samples may hit or miss toxicity spikes thus possibly increasing the variability between samples taken at different times at the same outfall. Composite samples will average concentrations over the entire collection period, possibly smoothing peaks and valleys of toxicity in variable water media. The various USEPA method manuals review the importance of using appropriate sample types for different types of effluents. Storage and handling can affect the toxicity and variability of samples. The general assumption is that the toxicity of a sample is most likely to decrease with holding time due to factors such as biodegradation, hydrolysis, and adsorption. These factors are minimized by “cold” storage and shipment on ice as well as test initiation within the specified USEPA guidelines. Water samples for WET testing may be manipulated in a variety of ways to comply with special requirements or circumstances. This applies, for example, when freshwater effluents are discharged to a saline receiving stream and marine or estuarine organisms are used for testing. Care must be taken, in this case, that ionic strength and composition are within levels tolerated by the specific test organisms or results may not be representative of actual toxicity or comparable between labs.

Abiotic Conditions

Abiotic conditions can strongly influence the variability of WET test results. For that reason, most of the abiotic conditions that should be standardized during WET testing (DO, light, hardness, alkalinity, etc.) are specified in protocols contained in the USEPA methods manuals. While these factors may not be problematic sources of variability within tests, they may be of major concern across tests (both within and among laboratories). Very small ranges of temperatures are specified for WET testing. Test solution pH can influence the bioavailability and toxicity of chemical constituents, such as some metals (e.g., Cu, Zn) and ammonia. Careful use of dilution waters, salinity adjustments, aeration, feeding, and other factors causing shifts in pH will help to reduce variability.

Exposure

In WET testing, we seek a balance between realistically mimicking exposure scenarios and evaluating effluents with sufficient testing while controlling testing costs. Variability in test results can be greatly influenced by the method of exposure chosen (i.e., static, static renewal, and flow-through). For example, tests of samples with nonpersistent toxicants or with chambers with high loading rates will be influenced to a greater degree using a static design rather than a flow-through design. As the number of variables which influence test results increases, overall test variability increases unless those variables are controlled. However, flow-through tests are much more costly than static tests. The number of concentrations and dilution series may influence variability of the test results. Point estimate models will more precisely estimate the statistical endpoint if the test concentrations are near the actual LCx (concentration that is lethal to x percent of organisms), ECx (concentration that affects x percent of organisms), or ICx (concentration that inhibits response by x percent). In contrast, as the NOEC approaches the concentration at which effects begin to be observed (i.e., LOEC), estimates may show greater variation. Many NPDES permits include a test dilution that is consistent with the Instream Waste Concentration (IWC) based upon dilution in the receiving system. The minimum number of tested dilutions recommended can be increased, particularly in the range of expected effects (if known), in order to improve resolution of the acute or chronic endpoint. Costs of increased dilutions testing are incremental to the cost of a typical test, but such testing is cost effective in cases where small changes in organism responses may affect compliance.

The WET endpoint is a function of test duration, in most cases (percent mortality after a period of time, for example). Test duration can be a function of the endpoint that is to be assessed. In at least one situation, the *C. dubia* survival and reproduction test, exposure duration is governed by the amount of time needed for 60 percent of the control organisms to produce a third brood (up to 8 days), at which time the test is repeated if the control performance is not acceptable (USEPA 1994b). The timing for test termination can therefore vary between 6 and 8 days. This introduces the possibility of intertest variability in terms of both number of young produced and test sensitivity due to exposure duration. The cost of reducing test duration variability is small; the corresponding reduction in test results variability could, however, be significant.

Sample Toxicity

The exposure-response relationship can be affected by the sensitivity of the test species to the individual and combined chemicals of a sample as well as the concentrations of those chemicals in that sample. Testing of samples which exhibit high slopes in their concentration-response curves at the test statistical endpoint (LCx, ECx, and ICx) tends to provide less variable (intratest and inter-test) results than tests of samples exhibiting low slopes in their concentration-response curves. The sensitivity of different species to any single chemical or mixture of chemicals can also be quite different, even when all variables are held constant. For example, rainbow trout are approximately an order of magnitude more acutely sensitive to cadmium than daphnids (USEPA 1985a) while daphnids are approximately 2.5 times more acutely sensitive to chlorine than rainbow trout (USEPA 1985b). Herbicides (e.g., atrazine) are more acutely toxic to plants than fish (Solomon et al. 1996). This is why vertebrates, invertebrates, and plants are recommended for testing effluents in the NPDES program.

Food

Food quality can vary in a number of ways. Organisms whose diets vary in nutritional quality and size, before and during testing, may respond differently to the same sample under identical test conditions. For example, brine shrimp nauplli that are less than 24 hours old are required in all tests using these organisms as food to maintain the nutritional quality of the nauplii and to keep their size at the optimum for consumption by test organisms. The YCT and algal diet for *C. dubia* should contain specific concentrations of solids and algal cells as outlined in the manual. The quantity of food available can affect dissolved oxygen and pH levels within a test chamber and act as a substrate for the absorption and adsorption of toxic chemicals from the tested sample, thus reducing bioavailability.

Dilution Water

Optimally, the dilution water should replicate the quality of the receiving water. However, if the objective of the test is to estimate the absolute toxicity of the sample (effluent), which is the primary objective of NPDES permit-related toxicity testing, then a synthetic (standard) dilution water is used (USEPA 1993, USEPA 1994a, USEPA 1994b). If the objective is to estimate the toxicity of the sample in uncontaminated receiving water, then the test may be conducted using non-toxic receiving water. Dilution water quality can affect the toxicity of effluent, surface water, and stormwater dilutions by modifying the bioavailability of toxic chemicals in the sample. In addition, parameters such as TDS (hardness, salinity, conductivity), turbidity, DO, pH, micronutrients, and bacteria counts can impact test organism physiology, sensitivity, and biological response. Therefore, test variability at all levels can be affected by variability in dilution water quality. Synthetic dilution water quality can also vary with the age of the prepared water in relation to the exposure of test organisms and with the source and quality of the base water.

Organism History and Handling

Perhaps one of the most important considerations in controlling WET variability is an organism's pretest history of health and maintenance, which consists of four factors: collection, culture, acclimation, and handling specific to the test. Organism history can be evaluated through charting performance of laboratory controls with a reference toxicant over time. All practical attempts should be made to avoid use of field-collected animals for WET testing. The most common sources of test organisms for WET tests are in-house cultures and/or organism suppliers. Organisms to be tested, whether field-collected or cultured, may require acclimation to test conditions. Variation in acclimation practices between tests can result in the use of organisms of varying sensitivity between tests. The importance of analyst technique is most pronounced when the analyst handles organisms before and during the test.

Randomization

Results will be variable in all analytical techniques, not just WET, despite all efforts to eliminate and reduce sources of variability. The randomization approach used to assign test replicates within an incubator or water bath and the approach used to assign test organisms to test replicates are attempts to evenly distribute this variability within the testing environment and between organisms. All test methods include procedures for randomization which must be followed.

Organism Numbers

The number of organisms exposed in a toxicity test has a direct and calculable bearing on the ability of that test to detect and estimate effects resulting from that exposure. Generally, as the total number of organisms increases in a test, the ability to detect effects (i.e., statistical power in a hypothesis test) and the certainty in point estimates increases. Differences in number of organisms per replicate and treatment can be due to the loss of individuals or replicates through analyst errors or to the death or lack of response of all organisms in one or more replicates. The former reduces power or effect-estimate certainty (point estimate confidence intervals) by reducing sample size. The latter may reduce power or effect-estimate certainty by increasing variation in response relative to other replicates and treatments. Intra- and interlaboratory variability can include the factors discussed above, as well as possible differences in study design (total number of organisms and total number of replicates).

Organism Age and Quality

The recommended ages of test organisms for established protocols have two general considerations: (1) relative physical sensitivity of different life stages to the test conditions, independent of the challenges of a toxicant and, (2) relative sensitivity of different life stages to toxic constituents. Young organisms are often considered more sensitive to toxic and physical stressors than their older counterparts. For this reason, the use of early life stages, such as first instars of daphnids and juvenile mysids and fish, is recommended for all tests.

The effects of organism age on WET variability are potentially greatest between tests and between laboratories where age differences may be greater. As examples, all *C. dubia* used in a reproduction test must be within 8 hours of age but can be up to 24 h old; and fathead minnow larvae used in the growth test must be within 24 hours of age in a single test but could range between 1 to 2 days depending on whether the organisms are cultured in-house or shipped from an off-site culture facility. In the acute tests with fathead and sheepshead minnows, the age difference between tests can range from <24 h to 14 d.

Grothe, D. R., K. L. Dickson, and D. K. Reed-Judkins, eds.1996. Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, SETAC Press, Pensacola, FL, USA. 340 p.

Solomon, K.R., D.B. Baker, R.P. Richards, K.R. Dixon, S.J. Klaine, T.W. LaPoint, R.J. Kendall, J.M. Giddings, J.P. Giesy, L.W. Hall, Jr. and W.M. Williams. 1996. Ecological risk assessment of atrazine in North America surface waters. Environ. Toxicol. Chem. 15:31-76.USEPA. 1985a. Ambient water quality criteria for cadmium - 1984. EPA 440/5-84-032. Office of Regulations and Standards, Washington, DC.

USEPA. 1985b. Ambient water quality criteria for chlorine - 1984. EPA 440/5-84-030. Office of Regulations and Standards, Washington, DC.

USEPA. 1993. Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms. 4th ed. Weber C.I., editor. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-90/027F. 293 p.

USEPA. 1994a. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to marine and estuarine organisms. 2nd ed. Klemm, D.J., Morrison, G.E., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/003. 341 p.

USEPA. 1994b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. 3rd ed. Lewis, P.A., Klemm, D.J., Lazorchak, J.M., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/002. 341 p.

How can WET variability be quantified?

Intratest Variability

Intratest variability is the variability of the responses (survival, growth, or reproduction), both among and between concentrations of the test material for a given test. Hypothesis test intratest variability is derived for an individual test by pooling the variability at each concentration including the control to obtain an estimate of the random error for the test. The intratest variability is used to determine the amount of difference from the control that can be detected statistically. When adjusted for the control mean, the minimum significant difference (MSD) represents the amount of difference expressed as a percentage of the control response (MSD%). Intratest variability for the point estimate approach is also represented by an estimate of the random error for the test, the mean square error (MSE). The MSE is one component in the calculation of confidence intervals for a point estimate, thus the width of a 95 percent confidence interval provides an indication of the magnitude of the intratest variability.

The intratest variability is the foremost single measure used to indicate the statistical sensitivity of a WET test analyzed with the hypothesis test approach. Statistical sensitivity, in this case, equates to a test's ability to distinguish a difference between an exposure concentration and the control. Controlling or reducing the amount of variability within a single test will increase the power of the test and therefore the ability of the test to detect responses that differ from the control response (decrease MSD). Increased power will also increase certainty in the determination of a difference from controls, which is important to regulators and the regulated community. However, minimal variability in all treatments of a test may lead

to such high statistical power that detected differences may not be biologically significant. Such tests should be interpreted with caution. Although there is no specific guidance from the USEPA on statistical versus biological significance, various States and USEPA Regions have developed some guidelines (e.g., see SETAC FAQ on addressing variability). Close attention to the factors described under the FAQ on factors affecting variability will tend to decrease heterogeneity among replicates and decrease intratest variability. In addition, increasing the number of replicates will also lead to an increase in the sensitivity of the test by decreasing the MSD.

Intratest variability is also important in representing the uncertainty associated with point estimates of toxicity. As the 95 percent confidence intervals of the point estimate increases, the uncertainty in that estimate of the statistical endpoint increases. The confidence intervals for chronic endpoints are directly influenced by the variability of response between replicates in each treatment and the model used to interpolate the point estimate. The confidence intervals for acute test results using a point estimate approach, however, are not influenced by variability between replicates but by the characteristics of the dose-response relationship. As discussed before, the certainty in point estimates is also a function of the dilutions tested and their proximity to the actual statistical endpoint being calculated. One will get a better estimate of the LC50 (tighter confidence intervals) if dilutions are tested near the concentration which actually results in 50 percent mortality.

Evaluation of a number of existing data sets by members of the Pellston workgroup (Sessions 3 and 4) (Grothe, et al, 1996) seemed to indicate that, for most WET test methods, MSDs of <40 percent were achievable. MSD's for most methods examined ranged from 18 percent to 40 percent. The consensus of the workgroup is that an additional study is necessary to determine the acceptable level of intratest variability for each USEPA recommended toxicity method, although some participants proposed that sufficient data exists to select MSD criteria. In the proposed study, data would be used to establish variability limits from laboratories that document data quality and adhere to USEPA method guidelines. Study data from each assay evaluation would include expected CVs, MSD, MSD%, MSE, and American Society for Testing and Materials (ASTM, 1992) "h" and "k" statistics. The "h" statistic represents a measure of the reproducibility between laboratories while the "k" statistic represents the repeatability within laboratories. Distributions of these values would be examined to determine criterion levels for intratest variability, and probabilities of laboratories exceeding the criterion levels would be calculated. The direct advantages of an acceptability criterion for intratest variability are 1) establishing a minimum protection level, 2) setting the power of a test to detect a toxic sample for each method, and 3) decreasing intra- and interlaboratory variability. Acceptability criteria will also allow users of WET data to better evaluate test acceptability, laboratory performance, and program effectiveness.

Intertest and Interlaboratory Variability

The scientific community familiar with analytical procedures, not just WET, recognizes that tests performed on presumably identical materials in presumably identical circumstances do not typically yield identical results. An indication of a test method's consistency is its repeatability and its reproducibility with repeatability defined as the variability between independent test results obtained from the same laboratory in a short period of time and reproducibility defined as the variability between test results obtained from different laboratories.

Several measures of repeatability and reproducibility have been proposed. The simplest of these is the intra- and interlaboratory CV (standard deviation (s) of repeated test results, divided by the mean (m) of the repeated test results, multiplied by 100 (CV = (s/m) x 100). The intralaboratory CV is generated by test results from repeated tests performed in the same laboratory, while the interlaboratory CV is obtained from test results from several different laboratories. The use of the CV removes from consideration the units of the measurement and allows the analyst to compare variability of different types of test methods (i.e., WET tests with analytical chemistry tests). It also allows analysts to compare tests that use different scales of measurement.

However, CVs alone cannot be used as diagnostic tools to help identify unusual test values or outliers. Since the CV is a function of the standard deviation of a set of test results, the measure suffers from the same problems associated with standard deviations, and there is no common agreement on what is an acceptable standard deviation. For instance, the range of test values is an easier descriptive statistic to understand. In addition, the value of the standard deviation is affected by extreme values in the data set; single large or small test values inflate the standard deviation. The CV also ignores the 95 percent confidence intervals (uncertainty) associated with each point estimate and can only be calculated for point estimates. CVs are not appropriate for hypothesis test endpoint comparisons since the effect levels are fixed by the choice of test concentrations.

Quality Management Considerations. Reference toxicant tests are typically used to monitor a laboratory's performance. Charting the performance of a laboratory's controls relative to its reference toxicant test results is a good way to track the laboratory's performance and to identify when the laboratory's performance is not acceptable. The width of a control chart's limits is an indication of a laboratory's capability to reproduce the desired endpoints of a reference toxicant test. However, control chart limits are a function of the reference toxicant, test species, test type (acute or chronic) and biological endpoint (survival, growth, etc.). These factors must be considered before drawing conclusions regarding laboratory performance. Performance on reference toxicant tests as recorded by control charts should be a criterion that is used by permittees in selecting which laboratories to use for WET tests.

Laboratories with very wide control limits, and/or many points outside of the control limits, should investigate problems related to the quality of the data being produced. Laboratories should monitor at a minimum, using control charts, the calculated endpoints for each test type/species combination. Laboratories can also monitor the control treatment mean response for survival, growth, and reproduction. In addition, laboratories can chart the control treatment replicate variance, or standard deviation. Reference toxicant tests are very important to track analyst technique and the health and condition of the test organisms. It is particularly important when performing these tests (as with all compliance toxicity tests) that the analysts precisely follow the published test methods, without deviation between tests.

ASTM-American Society for Testing and Materials. 1992. Standard practice for conducting an interlaboratory study to determine precision of a test method, E691-92. In: *Annual Book of ASTM Standards*, Vol. 14.02. Philadelphia, PA.

Grothe, D. R., K. L. Dickson, and D. K. Reed-Judkins, eds.1996. Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, SETAC Press, Pensacola, FL, USA. 340 p.