Whole Effluent Toxicity Testing Guidance and Test Review Criteria

Revised June 2016
Publication no. WQ-R-95-80

DEPARTMENT OF ECOLOGY
State of Washington

Publication and Contact Information

This report is available on the Department of Ecology's website at
https://fortress.wa.gov/ecy/publications/SummaryPages/9580.html

For more information contact:

Water Quality Program
P.O. Box 47600
Olympia, WA 98504-7600

Washington State Department of Ecology -
http://www.ecy.wa.gov/programs/wq/wet/

o Headquarters, Olympia	(360) 407-6000
o Northwest Regional Office, Bellevue	(425) 649-7000
o Southwest Regional Office, Olympia	(360) 407-6300
o Central Regional Office, Yakima	(509) 575-2490
o Eastern Regional Office, Spokane	(509) 329-3400

Cover photo: Rainbow trout (*Oncorhynchus mykiss*)

To request ADA accommodation including materials in a format for the visually impaired, call The Water Quality Program at Ecology, 360-407-6401. Persons with impaired hearing may call Washington Relay Service at 711. Persons with a speech disability may call 877-833-6341.

Whole Effluent Toxicity Testing Guidance and Test Review Criteria

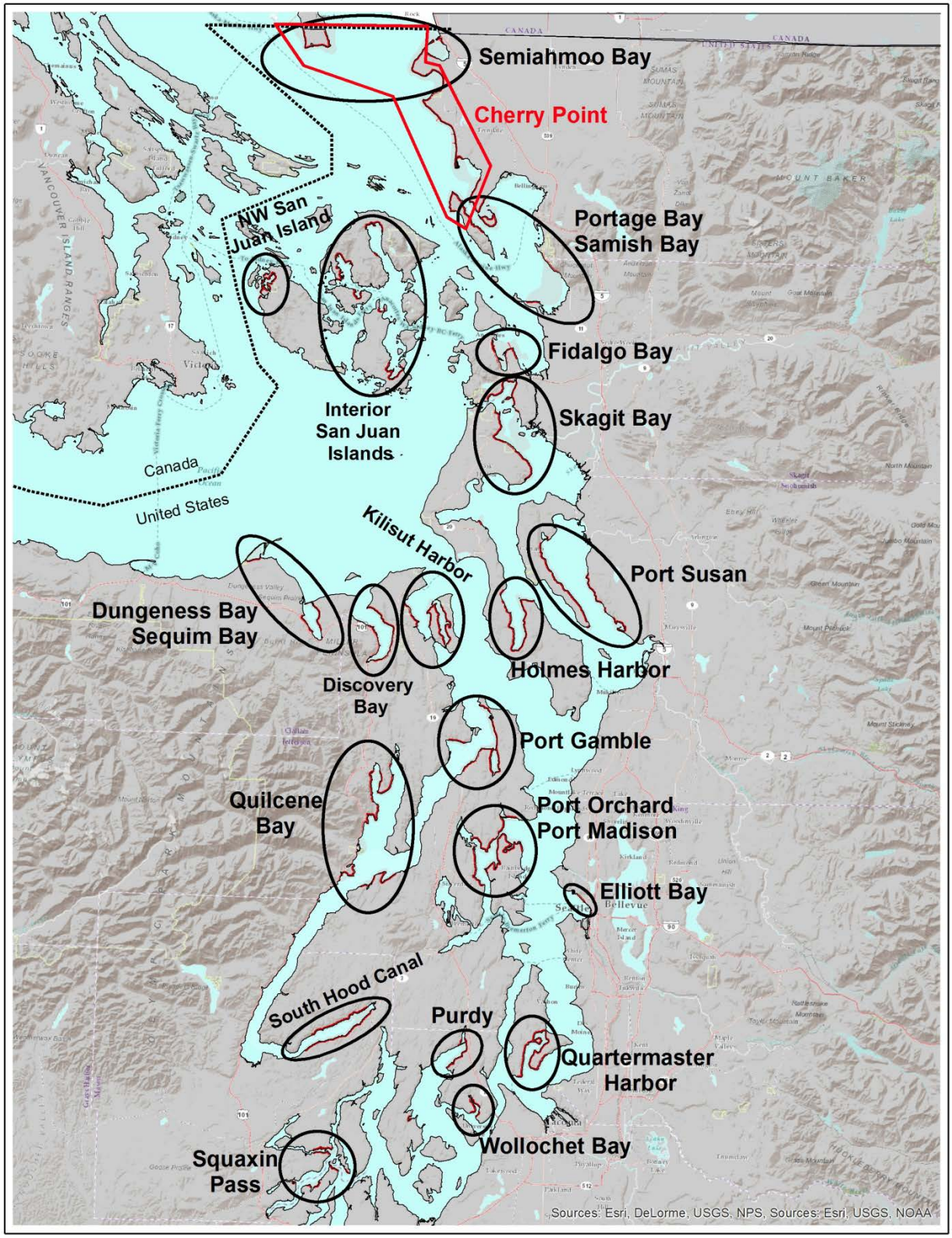
by
Randall Marshall

Water Quality Program
Washington State Department of Ecology
Olympia, Washington 98504-7710

Daphnid photos courtesy of
Rowe, C.L. and Hebert, P.D.N. 1999. **Cladoceran Web Site**. University of Guelph. Canada.
http://www.cladocera.uoguelph.ca/

Daphnia magna

Daphnia minnehaha



Documented Herring Spawning Grounds from *2012 Washington State Herring Stock Status Report,* Washington Department of Fish and Wildlife
http://wdfw.wa.gov/publications/01628/wdfw01628.pdf

Table of Contents

Whole Effluent Toxicity Testing Regulatory Guidance

Introduction

Whole effluent toxicity (WET) is the toxicity of an effluent sample measured directly with a toxicity test in order to assess the total toxic effect of all pollutants. WET testing is necessary because the Environmental Protection Agency (EPA) cannot develop water quality criteria for every toxic pollutant possible in wastewater discharges. The expense would be huge to try to measure every possible chemical. WET testing is also the only method for assessing the toxic interaction of all pollutants in a discharge. The approach is called "whole effluent toxicity" to contrast it with measuring the individual concentrations of the multiple toxic chemicals in an effluent for a one-by-one comparison to a water quality criterion if there happens to be one. Because WET testing often discovers effects from unknown toxicants, the toxic effects cannot be reduced until the toxicants involved have been identified.

The authority for WET activities in Washington State comes from RCW 90.48.520, Chapter 173-205 WAC, and the state water quality standards (Chapter 173-201A WAC). Interested persons can access these laws and rules online at:

http://leg.wa.gov/LawsAndAgencyRules/Pages/default.aspx

RCW 90.48.520 directs the Department of Ecology (Ecology) to require all known, available, and reasonable methods to control effluent toxicity in order to improve water quality regardless of whether or not it already meets minimum water quality standards.

We implemented this directive by writing Chapter 173-205 WAC so that the need for an acute WET limit is determined by survival in 100% effluent. As long as 100% effluent consistently has 65% survival or better, no acute WET limit will be required. The 65% survival cutoff is based upon data showing that a large majority of WET tests have from 65% to 100% survival in 100% effluent. A significant minority of tests show 0% survival in 100% effluent, but few tests show survival between 65% and 0%. Dilution is a factor in setting acute WET limits but is not considered in whether to require them. Our intention is to create an incentive to completely eliminate acute WET.

Chronic WET tests are expensive and have too many diverse sublethal endpoints to justify an attempt to completely eliminate chronic WET. Our goal for chronic WET testing is solely to maintain compliance with water quality standards.

The Ecology WET webpage (www.ecy.wa.gov/programs/wq/wet/index.html) posts various documents providing guidance or explanations about WET. These documents include permittee guidance, integrated ambient monitoring reports, WET program evaluation report, herring toxicity test development report, discussion of possible causes for the Cherry Point herring decline, and discussion of WET test statistical analysis. The webpage also contains links to important EPA information such as toxicity testing manuals and toxicity identification evaluation (TIE) guidance.

WET testing requirements in NPDES permits

National Pollutant Discharge Elimination System (NPDES) permits describe how provisions in chapter 173-205 WAC apply to each individual permittee. Labs should follow the instructions in a client's permit. It is important that labs get a copy of the toxicity testing pages of a permit in order to provide the best service. Permits can be acquired from a permittee or online at:

www.ecy.wa.gov/programs/wq/permits/paris/paris.html

Permit language

Permit language for WET consists of a series of steps in a regulatory process. The step to follow next will depend on the results of the previous step. The permit might contain two sets of instructions, but only require that one set be followed depending on circumstances. This permit language avoids the expense of modifying permits, but will require careful reading and planning ahead by labs and permittees.

WET limits

The state's water quality standards prohibit toxicity past the edge of an approved mixing zone. Therefore, WET limits are based on the concentration of effluent at the edge of an approved mixing zone during critical conditions. Critical conditions are situations when the effect of the effluent is greatest such as during low river flow. The concentration of effluent existing at the edge of a mixing zone during critical conditions is called the *critical effluent concentration*. Compliance with a WET limit means demonstrating no toxicity in a sample of effluent diluted to equal the acute or chronic critical effluent concentration (ACEC or CCEC). The ACEC is used to test for compliance with an acute WET limit and is the concentration of effluent during critical conditions at the edge of the acute mixing zone. (The ACEC is also used in determining the need for a chronic WET limit as described below.) The CCEC is used to test for compliance with a chronic WET limit and is the concentration of effluent during critical conditions at the edge of the chronic mixing zone. More information on WET limits is available in Appendix K.

Effluent characterization to determine the need for WET limits

Effluent characterizations last for one year and are used to determine whether WET limits are needed. After effluent characterization is complete, a permittee might receive an acute WET limit, a chronic WET limit, both, or no WET limit. Each effluent sample during effluent characterization is tested with all WET test species listed in the permit. This "multiple species" testing provides an assessment of effluent sample toxicity to a variety of aquatic organisms.

Permittees who cannot meet the WET performance standards defined in WAC 173-205-020 will receive WET limits. For acute toxicity, the performance standard is no test result showing less than 65% survival in 100% effluent. The acute toxicity performance standard also includes a provision for a median of at least 80% survival in 100% effluent across all of the acute tests conducted during effluent characterization. No effluent has ever had a median survival less than 80% during effluent characterization, so this part of the performance standard has no consequence worthy of discussion.

For chronic toxicity, the performance standard is no statistically significant difference in test organism response between the control and a test concentration equal to the ACEC. Toxicity at the ACEC is used in determining the need for a chronic WET limit because the number of tests conducted during effluent characterization is too small to predict toxicity at the CCEC over the life of the discharge. Significant

chronic toxicity at the ACEC is used to indicate a reasonable potential for significant chronic toxicity someday at the CCEC (where the chronic limit is set).

If a mixing zone has not been established for the discharge at the time of permit writing, the ACEC will not be known during effluent characterization. When the ACEC is unknown, WET testing during effluent characterization will determine the NOEC (no observed effect concentration) and LOEC (lowest observed effect concentration). The NOEC and LOEC will be compared to the ACEC, when it becomes known, to determine if a chronic WET limit is needed. If the ACEC is still unknown at the end of effluent characterization, then effluent characterization will be extended, but only one WET test will be conducted on each sample ("single species" testing). See Appendix L for definitions of NOEC and LOEC.

It is in the best interest of a permittee to include the ACEC in the dilution series if it is known. If a test is conducted without the ACEC, it is possible that the resulting LOEC and NOEC will bracket the ACEC. The usual policy is to consider the ACEC to be toxic in that situation since the ACEC is higher than the NOEC.

However, the ACEC in that situation is also below the LOEC and it is possible that it would not have shown toxicity if included in the testing. The percent minimum significant difference (PMSD) can be used as the effect level in a point estimate (IC_{PMSD} or EC_{PMSD}) and the result compared to the absent ACEC (or CCEC) to estimate if it might have been significantly different from the control. We will use the PMSD in this way when the ACEC or CCEC was inadvertently or accidentally absent from the test concentration series and ended up bracketed by the NOEC/LOEC.

Effluent characterization is also used to establish a baseline toxicity level based upon point estimates such as the LC_{50}, EC_{50}, or IC_{25}. See Appendix L for definitions of these point estimates. Point estimates will not be used in determining compliance, but will provide an effluent toxicity baseline and trends over time. WET tests conducted for effluent characterization must have a dilution series of five effluent concentrations and a control in order to provide point estimates.

Monitoring for compliance with WET limits

A permittee complies with a WET limit when the hypothesis testing procedure in Appendix H of EPA/600/4-89/001 (Fisher's Exact Test for analyzing survival in the *Ceriodaphnia* chronic test) shows no statistically significant difference in response between a control and the ACEC or CCEC. (An outdated EPA manual is referenced here and in permits because WAC 173-205-070 specifies the version that was current at the time of its writing.) Appendix H of EPA/600/4-89/001 is the same as Appendix H in EPA-821-R-02-013 (newest freshwater chronic manual) and Appendix G in either the current East Coast or West Coast marine chronic manuals. The 2002 EPA acute manual describes the single comparison hypothesis testing procedure in Section 11.3 (The flowchart is in Figure 12.) A statistically significant difference in test organism response (*alpha* = 0.05) at the ACEC might mean an acute WET limit violation or a failure to meet the chronic toxicity performance standard. A significant difference at the CCEC would mean noncompliance with any chronic WET limit in place at that time.

A compliance monitoring test for an acute WET limit is conducted with the ACEC (the limit), 100% effluent (the performance standard), and a control. A compliance monitoring test for a chronic WET limit is conducted with the CCEC (the limit), the ACEC (the performance standard), and a control. The permittee may instruct the lab to run a full dilution series to provide more information for review of test quality, and we recommend five effluent concentrations and a control.

Response to noncompliance with a WET limit

If a permittee fails a compliance test for a WET limit, then additional testing is immediately required to assess and confirm the continuing presence of toxicity. WAC 173-205-090 requires WET testing of four weekly samples following noncompliance with an acute WET limit and three monthly samples following noncompliance with a chronic WET limit. If any of these additional WET tests fails to comply with the limit, then the permittee must submit a toxicity identification/reduction evaluation (TI/RE) plan.

Looking for toxicity changes when there is no WET limit

Permittees without WET limits must both routinely check for changes in effluent toxicity and specifically evaluate any known facility or process change which might increase effluent toxicity. Extra effort is needed in the absence of regular compliance monitoring to make sure that effluent toxicity has not increased above the level found during effluent characterization.

WAC 173-205-060 contains a list of changes at the facility or in the discharge which will trigger another effluent characterization unless a permittee uses toxicity testing and/or chemical analysis to show toxicity did not increase. Other triggers for new characterizations include:

- When there is no WET limit, permits usually require a set of WET test results to be submitted with each subsequent permit application. If any of these tests fails to meet the performance standard, then another effluent characterization will be required in the new permit. See WAC 173-205-030(8) and 173-205-060(3)(a).
  - Some POTWs may need to perform additional WET testing to meet an EPA requirement for WET test results to be reported in Part E of the 2A permit application form. This EPA requirement is separate from state requirements pursuant to chapter 173-205 WAC, but testing done to meet state requirements may be used to meet EPA permit application requirements and vice versa.
- As an alternative to the end of permit term toxicity check previously described, a permit might require rapid screening testing to catch any sudden event at the facility which would result in a toxic discharge that would otherwise go unnoticed. A rapid screening test is a single dilution (plus a control) toxicity test conducted on 100% effluent or the ACEC in order to detect unanticipated increases in toxicity. Rapid screening tests are cheaper and quicker than standard WET tests. (See Appendix F for a list of rapid screening tests.) Whenever a permittee fails a rapid screening test, WAC 173-205-120 requires an immediate retest with standard WET tests. The results of these WET tests conducted in response to rapid screening tests will be evaluated to determine the need for a new WET characterization in the next permit or the need for immediate action to reduce toxicity. Compliance with WET limits will not be measured with rapid screening tests. See WAC 173-205-120(2) and 173-205-060(3)(b).

Excessive time to produce a test report

Chapter 173-205 WAC contains time limits on permittee responses to toxicity test results. Labs should be careful not to add to permittee difficulties by taking too long to produce a test report. When serious toxicity occurs, labs should contact the permittee immediately.

Expression of permittee toxicity test results

Even though chapter 173-205 WAC describes decisions made on the basis of single comparison hypothesis testing, NOECs and LOECs based on multiple comparisons are saved in our database. When

the NOEC/LOEC is far from the ACEC or CCEC, performing a single comparison of the ACEC or CCEC against the control is not really necessary. (If the LOEC is just above the ACEC or CCEC and the effect at the ACEC or CCEC is similar to the effect at the LOEC, a single comparison of the ACEC or CCEC versus the control will be performed.)

Point estimates for the 15%, 25%, and 50% effect levels of each test endpoint with a statistically significant difference from the control are also stored in our database. Both weight and biomass are calculated for the 7-day survival and growth tests (See Appendix C).

Options for permittees

Effluent screening tests

WAC 173-205-050(1)(f) allows Ecology to approve a request from a small business or a publicly owned treatment works (POTW) discharging less than 0.5 mgd to conduct effluent screening tests. Acute effluent screening tests use only a control and 100% effluent, and chronic effluent screening tests use only a control and the ACEC. If an effluent screening test shows toxicity, the permittee must resample and conduct a full dilution series test. Two tests (an initial screening test plus a follow-up full test) will be more expensive than one full dilution series test. Effluent screening tests are also *strongly discouraged* because they prevent consideration of the anomalous test criteria in Appendix D and could cause avoidable and expensive regulatory consequences for permittees.

Full dilution series tests

Because Chapter 173-205 WAC allows WET tests to sometimes be conducted with less than a full dilution series, it also makes clear that permittees may choose to conduct any WET test using a full dilution series. WET tests conducted using a full dilution series of at least five effluent concentrations and a control provide the best information for evaluating the quality of WET test results. A full dilution series protects permittees by allowing anomalous test results to be identified more easily (See Appendix D.). Anomalous WET tests will not be used for compliance determinations. The ACEC or CCEC may be included in a dilution series as an extra concentration or as a substitute for one (the closest) of the five concentrations in the series.

Notification of an anomalous test result

WAC 173-205-090(1)(d) allows a permittee to avoid the cost of additional testing when noncompliance with a WET limit is believed to be due to an anomalous WET test result. A laboratory should be able to inform a permittee of any anomalous WET test result associated with apparent noncompliance. (See Appendix D, Identifying Anomalous WET Tests.) The permittee then sends Ecology notification with the test report that the test might be anomalous and that the permittee intends to take only one additional sample for toxicity testing. The notification must identify the reason for considering a compliance test result to be anomalous. If Ecology agrees that the test causing noncompliance was anomalous, the test on the additional sample replaces the anomalous test and the permittee is saved the cost of further testing.

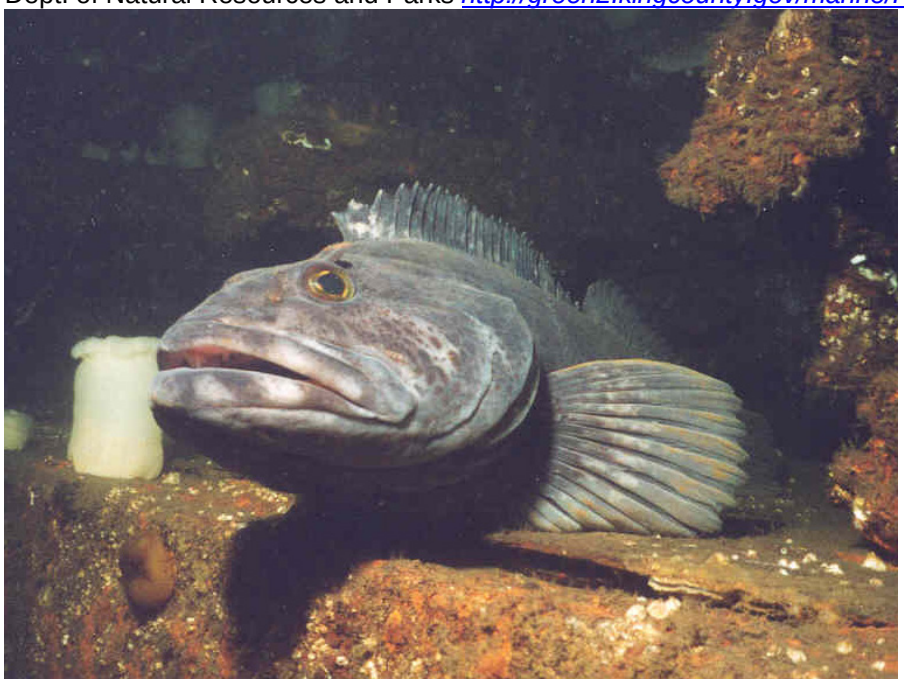
Accredited labs

Labs performing testing to meet these requirements must be accredited. To find an accredited lab follow these instructions:

1. Go to: https://fortress.wa.gov/ecy/laboratorysearch/.

2. Click on the gray button that says "**Search by Analyte**."
3. Click on the down arrow by "**Analyte**."
4. A list will appear of all of the chemical, physical, and biological analyses done by accredited labs. The list is very long.
5. Scroll to the species (scientific) name for the organism needed for toxicity testing and select it by clicking. Do not worry if you also need other toxicity test species because most labs do toxicity testing using many of the common test species and methods.
6. Click on the gray button that says "**Select by Analyte**."
7. A list of labs will appear. Clicking on an individual lab name will provide contact information and a list of the specific toxicity tests accredited for the lab.

Lingcod (*Ophiodon elongatus*) courtesy of Richard Jack, King County Dept. of Natural Resources and Parks *http://green2.kingcounty.gov/marine/Photo*



Whole Effluent Toxicity Test Review

Introduction

Invalid WET tests happen when the lab does not follow the test method or when the results do not meet the validation criteria in the test method. Permittees are obligated to look for invalid tests because the permit requires the submission of valid test results. Ecology will also review WET tests for validity. Chapter II. *Whole Effluent Toxicity Test Review* and Chapter III. *Test Species and Specific Test Conditions* of this document contain the test validity criteria.

The test review criteria in this document have been reviewed by labs and other interested parties. The document will be revised in response to any persuasive comment from a lab. Labs should keep in mind that, although most samples are nontoxic, all tests must be conducted and reported so that we can make regulatory decisions with confidence.

Please direct any questions to Randall Marshall at 360-407-6445 or randall.marshall@ecy.wa.gov.

Sample handling

Transfer and storage

Sample transfer must be documented with signed and dated chain-of-custody forms and a copy put in the test report. For composite samples, the sample date is the end date of the compositing period and the end date/time must be reported on the chain-of-custody form by the sampler consistent with the Pacific Time Zone. Labs should note in the test report if the sample and test occurred in different time zones or if daylight savings time has gone on or off between sampling and test initiation.

Chain-of-custody forms must accompany all samples unless:

1. A person from the testing lab does the sampling, delivers the sample personally to secure storage at the lab, and the test report documents this procedure; or

2. Personnel who are all employees of the permit holder are the only ones conducting the sampling, transportation, and toxicity testing, and a responsible person from the organization signs a page in the test report stating that the result is an honest and accurate reflection of the toxicity of the sample.

Chain-of-custody forms must contain the name and address of the discharger, outfall number, date and time that the sample is taken, the name of the sampler, the type of sample (grab or composite, compositing method and duration, effluent or stormwater, etc.), and the number and volume of sample containers. The chain-of-custody form must describe the type of sample container.

The sampler's signature must be in the first "relinquished by" blank. Each person subsequently taking physical custody of the sample must sign the next "received by" blank and then the next "relinquished by" blank when the sample is given to someone else. This sequence of signing is repeated until the sample is secure at the testing lab. Every signature must have a date and time, and each pair of "relinquished by" and "received by" signatures must have the same date and time (within a couple of

minutes to allow for differences in clocks). The use of a courier is the only circumstance when the "relinquished by" and "received by" signatures can have significantly different times.

Couriers do not need to sign the "received by" blank on the chain-of-custody form if the cooler containing the samples was packed by the sampler and has been locked or sealed with a seal that is initialed and dated by the sampler and cannot be removed without the removal being obvious (i.e. evidence tape). The name of the courier company and the method for locking or sealing the cooler must be identified on the chain-of-custody form. The sampler signs and dates (including time) the "relinquished by" blank on the chain-of-custody form and immediately locks or seals it in the cooler with the samples. Immediately upon receipt at the testing lab, a responsible person inspects the cooler to make sure that locks or seals are intact, opens the cooler, removes the chain-of-custody form, signs and dates (including time) it, and places the sample containers in secure storage at 0°- 6°C. When a courier is used, all signers to the chain-of-custody form are testifying to the proper condition of the cooler, lock, and seal unless otherwise noted.

One chain-of-custody form should accompany the sample throughout its travels. When a second lab is subcontracted to perform some of the tests on a sample originally received at the primary testing lab, the required chain-of-custody procedure is:

1. The sampler completes all information pertinent to sampling and transportation on the chain-of-custody form and signs relinquishing the sample. The chain-of-custody form is locked or otherwise sealed in the cooler if a courier is used.

2. The primary lab opens the cooler immediately upon receipt, signs the "received by" line on the chain-of-custody form, and makes a copy for its own test report.

3. The primary lab notes on the original chain-of-custody form the number and volume of containers placed in the cooler for the second lab, notes the method of transportation, signs the second (or next) "relinquished by" line, and locks/seals the form in the cooler with the sample.

4. The second lab opens the cooler immediately upon receipt, signs the next "received by" line on the chain-of-custody form, and makes a copy for its test report.

5. The completed original chain-of-custody form is returned to the primary lab to be kept in its records.

Holding time and temperature

Composite samples are chilled to 0°- 6°C while being collected and grabs immediately following collection. Maximum time from sample collection to first use is 36 hours. When not preparing test solutions, labs must store samples at 0°- 6°C (The preferred temperature is 4º C.) in the dark with minimal headspace or under a nitrogen atmosphere. Test solution renewals may be made up to 72 hours from collection. (Renewals may be made up to 84 hours after collection for a 96-hour acute test.) To be able to calculate holding times, the date and time of test initiation must be clearly recorded and reported on bench sheets.

Sample temperature must be measured at lab receipt and recorded on the chain-of-custody form or the initial water chemistry form. Tests conducted on samples received too warm will be rejected. Temperature requirements are related to the time to reach a lab:

- If a sample is received at the testing lab within one hour after collection and is immediately refrigerated at the lab or used in a test, it must have a temperature between 0° C and 20° C.
- If the sample is received at the testing lab within 4 hours after collection, it must be between 0° C and 12° C.
- All other samples must be between 0° C and 6° C.
- If any part of a sample has frozen, it cannot be used or the test will be rejected.

The original sample may be used for test solution renewal at 48 hours in an acute test if stored at 0°- 6°C in the dark with minimal headspace. If a 7-day chronic test is already underway using an initial sample that arrived in a timely manner and was a good temperature at receipt and the second or third samples arrive past the holding time, use the process in the *Deviations from Protocols and Acceptability Criteria* section to find out whether to continue the test or not. Properly stored samples which had an acceptable temperature at receipt and met the holding time are preferred over subsequent samples which arrived too warm.

If a chronic test requiring daily renewal will be conducted on an intermittent discharge (especially stormwater) which does not allow collection of three separate samples over 7 days, then sufficient sample must be collected during the available discharge events to provide daily renewals for the remainder of the test. The extra sample must be collected in a separate container with no headspace. It must be stored at 0°- 6°C until used. If necessity requires, the maximum 72-hour holding time may be exceeded if only one discharge was available for sampling.

High quality glass containers provide superior sample preservation for organic chemicals but cannot be collapsed like a cubitainer in order to minimize headspace and associated volatilization or oxidation. Minimization of head space is important. All glass containers should be filled to the top with sample. A sample should be collected into two or three glass containers of an adequate size for daily renewal. The use of a nitrogen headspace with glass sample containers is strongly recommended.

Filtration

No filtration of samples is allowed unless the necessity for filtration has been documented. Justification for filtration is limited to the observation of organisms which would attack, be confused with test organisms, or otherwise interfere with the test. Most samples do not contain indigenous organisms that would attack or be confused with test organisms. Many labs rarely filter samples and have no problems with toxicity tests. Unless the test report contains good justification, tests on filtered samples will be rejected.

If a lab can demonstrate that a particular effluent contains organisms which interfere with toxicity testing, then samples of that effluent may be filtered. A good demonstration would be to conduct a toxicity test with one set of replicates of 100% effluent filtered and another set unfiltered. If there is a difference in test results and organisms are identified in the raw sample or filter backwash, then filtration of that effluent has been justified. This demonstration need only be made once for an effluent discharge and then all future samples may be filtered. The demonstration is not necessary before filtering surface water samples or samples from treatment lagoons if the lagoon is part of a biological treatment system or has been colonized by aquatic plants.

Filter pore diameters must be no smaller than necessary to remove unwanted organisms. Pore diameters must never be smaller than 60 µm as specified in the test method (except for the Phytoplankton Growth Inhibition Test which is 1.6 µm; for more details see the list of test conditions and review criteria for the *Phytoplankton Growth Inhibition Test*).

Sample aeration

Aeration of samples is not allowed unless justified by dissolved oxygen (DO) measurements at deleterious concentrations. DO measured at concentrations below 4.0 mg/L (6.0 mg/L for rainbow trout) justify aeration.

The preparation of test solutions and filling of test chambers will almost always adequately remove supersaturation without aeration. The replicates for the 100% effluent concentration should be prepared first so they can equilibrate while the containers for the rest of the effluent dilution series are prepared. If this procedure occasionally does not work, then the test chambers should be briefly aerated. If this procedure often fails to work for samples from a discharge, then document the problem and request permission to routinely aerate samples from the discharge prior to test setup.

Dechlorination

WET tests conducted on effluent samples which are dechlorinated under any circumstance other than allowed by WAC 173-205-080(3) or the NPDES permit cannot be used for regulatory determinations and must be repeated. We prefer that samples for WET testing of chlorinated effluents be taken prior to the chlorinator if the ACEC is below 25% and the discharge can meet water quality-based effluent limits for chlorine. Otherwise, WET testing must be performed on an unmodified sample of final effluent. See Appendix I **Chlorine Toxicity** for more explanation.

Sample hardness adjustment

Permits might allow sampling of receiving water at the same time as effluent so that the hardness of the effluent can be adjusted to match the hardness of the receiving water. The receiving water sample will only be used for hardness determination and can be small. If receiving water is inaccessible for sampling near the discharge or remote from the effluent sampling location (e.g. industrial stormwater discharged to a storm sewer), then stream monitoring data from the Department of Ecology or USGS may be used to determine a typical receiving water hardness. Another good source for a target hardness would be the hardness used by the permit writer to calculate metals water quality criteria.

We only recommend hardness adjustment for stormwater or effluent samples which have less than 50 mg/L hardness. Labs must adjust the hardness by adding the specified proportion of the reagents marked in red in Appendix J. The reagents and proportions are also listed in Table 3 of Section 7 of EPA-821-R-02-013 (freshwater chronic manual) or Table 7 of Section 7 of EPA-821-R-02-012 (acute manual). A hardness between the ranges in the EPA tables may be prepared as long as the same proportions are used. Avoid as much as possible raising pH. Let the hardness-adjusted sample age for a few hours if the sample holding time allows.

Sample pH adjustment

If the sample pH is outside of the range 6.0 to 9.0, then the permittee is likely to be in violation of a technology-based effluent limit for pH and could also be violating state water quality standards. *Permittees should be immediately alerted to a potential problem if this occurs*. Samples outside of this range will be rare.

We do not allow addition of acids or bases to samples. In principle, no substance should be introduced into the sample unless absolutely necessary for a successful toxicity test. Acids and bases might themselves be toxic or enhance the toxicity of other substances. Every effluent sample must be tested without pH adjustment regardless of initial pH.

Parallel testing of pH adjusted and unadjusted sample concentrations will have little regulatory consequence. If the adjusted and unadjusted portions agree (both are toxic or nontoxic), then the unadjusted alone would have had the same outcome as parallel testing. If the adjusted is toxic and the unadjusted is nontoxic, the unadjusted will be considered the most reliable because the acid or base will be assumed to have created artifactual toxicity. If the adjusted is nontoxic and the unadjusted is toxic, then there is a good indication of a pH effect or pH-influenced toxicity, but this information, even though useful in a TI/RE, would not alter the determination based on the unadjusted sample that the effluent was toxic.

General test conditions

Randomization

A critical assumption in the statistical analysis of toxicity data is independence among observations. Independence of observations is especially critical for the parametric hypothesis test procedures (Dunnett's, Bonferroni's, and Student's t-tests). Randomization of test chambers is the method provided in the EPA test manuals for achieving independence of observations. Randomization of test chambers must be standard practice for labs conducting toxicity tests for permittees in this state. Randomization must be documented in the standard operating procedure (SOP) approved as a part of accreditation of the lab for the test method. True randomization must be documented involving the use of random numbers to assign test container positions. Failure to do so might cause test results to be rejected. (See Appendix A of any EPA chronic toxicity test manual or Section 11.1.6 of the EPA acute manual.)

Appropriate negative controls

Negative controls serve two important functions in toxicity tests:

- **Establishing test validity -** A control provides a measure of test organism health and laboratory technique in order to establish the validity of the test result. Every toxicity test must have a control that accomplishes this function.

- **Providing a standard for comparison in hypothesis testing -** The control measures test organism response under nontoxic conditions. Test organism response in the effluent concentrations can then be statistically compared to the control response to determine if any effluent concentration is considered to be toxic.

For a test to be acceptable, it is important that controls are:

- Nontoxic laboratory or natural water,
- The same water used to dilute the sample and create a concentration series, and
- Subjected to the exact same test conditions as all other test concentrations.

In order to use one control in testing more than one sample, a lab must demonstrate in the SOP approved as a part of accrediting the lab for the test method that all of these important conditions are being met. Randomization of the control along with the test containers from all samples is especially important (See Appendix A of any EPA chronic toxicity test manual or Section 11.1.6 of the EPA acute manual.). Every test container for every sample sharing a control should be handled as if part of one large test with all activities occurring within the same space and time. Implementation of this procedure must be documented for all tests sharing one control. Failure to do so will cause test results to be rejected.

One misuse of a control which will certainly result in rejection of the toxicity test result is running extra replicates in the control and only using the results from the replicates with the best performance. Controls must be handled the same as other test concentrations. Failure to do so will cause rejection of the test.

Brine controls and sources of salinity

The dilution water control is the control used in comparisons with the effluent concentrations and must meet acceptability criteria. A brine control is used to assess brine toxicity. When hypersaline brine is used, it has a concentration gradient in the same direction as the effluent. Without the use of a brine control, brine toxicity could be mistaken for effluent toxicity because the concentration-response relationship would be in the same direction. If test organisms in the brine control do less well than in the dilution water control and the difference is statistically significant, then any test result showing adverse effects at concentrations of regulatory concern will be suspect and the test would need to be repeated on a fresh sample. For the purpose of effluent monitoring in Washington State, brine and dilution water controls are not pooled.

Salinity for any test must be from either a high grade commercial sea salt or a hypersaline brine prepared from clean natural seawater. If artificial salts are used to provide salinity to a freshwater effluent sample, the salts should be added to both the sample and the deionized water used to prepare dilution water to achieve a salinity around 30 ppt.

Acceptable start counts

The control of experimental conditions is lessened when the number of test organisms is not equal in the test chambers at the beginning of a test. Unequal start counts will cause either the amount of food/animal to be unequal or the amount of food/test solution volume to be unequal. Unequal start counts will also mean that either test organism loading is unequal or test solution volumes are unequal.

EPA statistics are based on the assumption of equal numbers of test organisms in each replicate at the start of a test. Small deviations (one or two test organisms) from equality will not be a big problem, but larger differences will put the validity of statistics in doubt. Labs should always recount (verify) the number of organisms in each replicate container at the beginning of a test.

If test organisms are lost due to a documented accident, then the start count should be appropriately reduced. Accidents are specific events observed (and sometimes caused) by people. Examples of accidents include spilling, siphoning, or crushing test organisms. If aeration is necessary during a test, then any test organism found stranded on the side of the test chamber, caught in the test solution's surface tension, or entrained in an air bubble can be assumed to be a victim of an accident.

Test organism cannibalism, stranding on the side of the test chamber (unless due to aeration or agitation of the test chamber during handling), or simple disappearance of a test organism are not documentable accidents and do not justify adjusting start counts. Test organism weakness or death often precedes cannibalism, stranding, or disappearance. Cannibalism should be controlled by regular feeding, and stranding can be minimized by avoiding supersaturation or excessive shaking of test chambers.

Tidying-up data or improving control performance by adjusting start counts will increase statistical sensitivity by reducing variation across replicates and is unfair to permittees. Toxicity tests with large or

numerous differences in start counts will be rejected and returned to the permittee. No more than 10% of chambers in a test can vary in organism start count or it will be rejected.

Acute toxicity test duration

WAC 173-205-050(1)(c) requires the duration of an acute toxicity test to be 48 hours for an invertebrate and 96 hours for a fish. If a permit has not specified acute test duration, then WAC 173-205-050(1)(c) should be followed or the toxicity test results might be rejected. The duration of a WET test should be within ± 2 hours of the duration specified in *Test Species and Specific Test Conditions* section or the test will be rejected.

Appropriate test termination

All tests must be continued for the full duration specified in the permit or test method. If all test organisms die in every test concentration, the control must still be continued for the full duration in order to produce acceptable test results. It is acceptable to terminate a test early which, if continued, would not meet the requirements of the permit or test protocol as long as the effluent is resampled and an acceptable test result produced as soon as possible. An explanation of the reasons for early termination must accompany the report for the test on the new sample. The time and date of test termination must always be reported. Dual endpoint tests (See *Dual Endpoint Tests* section) must have the time of the 48-hour count (mysids) or the 96-hour count (fish) recorded as the end of the acute test.

Suppression of pH rise

Control of pH rise in test solutions may be accomplished by holding test chambers in a CO_2 atmosphere. More frequent test solution renewals might also control pH drift. If aeration is needed to maintain DO levels and pH control is also needed, then try aerating with CO_2-supplemented air. Another strategy for maintaining both DO and pH without aeration would be to construct boxes that allow a slow flow of a CO_2-supplemented air over test chambers. Supplementation with a small amount of O_2 in addition to CO_2 may also be necessary if the sample has an oxygen demand. References:

Mount DR, Mount DI. 1992. A simple method of pH control for static and static-renewal aquatic toxicity tests. *Environ Toxicol Chem* 11: 609-614.

Elphick JR, Bailey HC, Hindle M, Bertold SE. 2005. Aeration with CO_2-supplemented air as a method to control pH drift in toxicity tests with effluents from wastewater treatment plants. *Environ Toxicol Chem* 24: 2222-2225.

Aeration of test chambers

Sample aeration is preferred to test solution aeration. (See II.B.4. **Sample Aeration**) Use of an oxygen-enriched headspace would be preferable to aeration in maintaining adequate DO because it is nonintrusive to the test solutions. Aeration must:

- Be kept to the minimum intensity and duration necessary to maintain desired DO levels until test solution renewal or test termination.
- Not be initiated more than once in the same test if it can be avoided.
- Be initiated as soon as DO begins dropping steeply and need not wait until DO has dropped below acceptable levels.
- Be done equally in all test chambers regardless of DO concentration.

EPA manuals

1. *Acute testing*: EPA-821-R-02-012, available online at http://www.epa.gov/cwa-methods/whole-effluent-toxicity-methods (1st tab)
2. *Freshwater chronic testing*: EPA-821-R-02-013, available online at http://www.epa.gov/cwa-methods/whole-effluent-toxicity-methods (2nd tab)
3. *Saltwater chronic testing with East Coast organisms*: EPA-821-R-02-014, available online at http://www.epa.gov/cwa-methods/whole-effluent-toxicity-methods (3rd tab)
4. *Saltwater chronic testing with West Coast organisms*: EPA/600/R-95/136, available online at http://nepis.epa.gov/Exe/ZyPDF.cgi/1000409M.PDF?Dockey=1000409M.PDF

Permits reference an outdated EPA manual (EPA/600/4-89/001) because WAC 173-205-070 (compliance monitoring) was written when that version was current.

Water chemistry gradients

A toxicity test on a low hardness sample diluted with moderately hard dilution water will have a hardness gradient with hardness decreasing as the effluent concentration increases. *Ceriodaphnia* reproduction can be reduced by lowering hardness and produce a pattern that mimics a toxic response (an "adverse effect" that increases with effluent concentration). Labs should consider preparing a lower hardness synthetic water for use in testing low hardness effluents, especially for permittees without mixing zones (ACEC and CCEC = 100% effluent).

Sample chemistry (hardness, alkalinity, total dissolved solids, dissolved organic carbon, etc.) will interact with toxicants such as metals. Water chemistry gradients in a test may produce complex concentration-response relationships. A chemical's concentration at lower effluent concentrations might be below the toxic threshold. At higher effluent concentrations water chemistry (hardness or dissolved organic carbon) might keep the chemical not toxic. However, in-between effluent concentrations could still be toxic. This situation shows why we look for both anomalous concentration-responses as defined in Appendix D and water chemistry gradients based upon the water chemistry measurements described in the *Water Quality Measurements* section. It also illustrates why we keep test results in a database where we can look for previous occurrences of unusual concentration-response relationships.

Sporadic mortalities

Sporadic mortalities are deaths of test organisms that are not related to sample toxicity and do not fit a good concentration-response relationship. These sporadic mortalities sometimes cause a flat concentration-response relationship with low but nearly equal proportions alive which resemble an infection rate, not toxicity. In other cases, sporadic mortalities are confined to a few test chambers scattered throughout the test as if susceptible individual test organisms in those chambers were becoming infected and then spreading the pathogen within the chamber. Large standard deviations in proportion alive result when affected and unaffected test chambers occur in the replicates at a concentration. Any of the three anomalous test criteria in Appendix D would be consistent with tests having sporadic mortalities.

Assuming that the test organisms came from a good source, the pathogens which infect them can come from inside the lab, from a composite sampler, or from the sample itself. These pathogens can often be observed as filaments or patches on test organisms. An alert lab will notice whether diseases are killing test organisms and look for a source. If sporadic mortalities occur with a few clients, then the source of pathogens is likely their effluents or composite samplers. If sporadic mortalities occur for all clients, in

controls, or in reference toxicant tests, the pathogen source is likely in the lab. If test organism mortalities are lower in higher test concentrations, it is possible that the sample is toxic to the pathogen.

Some effluents are associated with sporadic mortalities more often than others. Noncontact cooling water has the highest frequency of sporadic mortalities. Pathogens in noncontact cooling water might originate in the source water body but might also be enhanced by biofouling in pipes or on other surfaces within the plant. Surface water samples can also have sporadic mortalities. Naturally occurring pathogens from a water body are likely the cause of sporadic mortalities in surface water samples.

Sporadic mortalities are a common and preventable cause of anomalous tests. Chemical contamination of containers or equipment can also cause unexpected mortalities. Labs should give extra attention to:

- Proper glassware cleaning and rinsing so that all residues are removed.
- Using only food grade disposable cups and changing supplier when there is a problem.
- Not skipping quality control steps such as soaking test containers in clean water overnight before using in a test.
- Running acute tests with fathead minnows or daphnids at 20° C instead of 25° C.
- Keeping samples at 0°- 6° C from the moment of collection until used in the test.
- Regularly cleaning incubators and their contents. General lab cleanliness helps.
- Covering test chambers to prevent airborne transfer of microbes.
- Keeping the lab free of mosquitoes, chironomids, and other flies which breed in water.
- Using enough sterile pipettes or other equipment for transferring test organisms so that cross contamination between replicates does not occur.
- Routinely rinsing temperature, pH, and DO probes before using in another test chamber.
- Ensuring that composite samplers are clean and have new tubing before sampling.
- Not feeding unhatched *Artemia* cysts and empty exoskeletons to fathead minnow larvae.
- Thoroughly cleaning and disinfecting of *Artemia* hatcheries.

The most promising technique for controlling pathogens in fathead minnow chronic tests is described in the article by Downey, *et al* in the following reference. The best feature of the technique is that it does not modify samples as does ultraviolet (UV) disinfection or filtration. It changes the test setup to 2 fish in each of 10 replicate chambers per concentration. This simple change was the most successful method for controlling pathogens in the study. A susceptible test organism can only spread the pathogen to one other test organism in such a test configuration.

Permittees and labs involved in testing samples of noncontact cooling water, stormwater, or wastewater treated in a pond or lagoon are encouraged to try the technique if needed. Test conditions are listed just after the usual list of fathead minnow chronic test conditions. Labs will spend more effort on the extra test chambers but the minimum number of fish needed for a test will be 25% less. Note that Fisher's Exact Test must be the nonparametric alternative procedure for analyzing survival because a start count of 2 guarantees lots of ties in ranked values and therefore a lack of sensitivity from Steel's or Wilcoxon.

If the alternative setup does not work, a lab may ask for permission to routinely UV disinfect samples from that permittee. The UV exposure should be kept to a minimum and this minimum effective exposure is often less than that reported in the following papers. One lab here in the Northwest routinely disinfected noncontact cooling water and river water using UV exposure for about 2 minutes (1 or 2 passes through the unit depending on turbidity). The lab demonstrated the effectiveness of this UV exposure time both by comparing sporadic mortalities in treated and untreated samples and by taking before and after disinfection plate counts of bacteria. We realize that minimum exposure times

will vary depending on the effluent characteristics and abilities of the ultraviolet sterilizer, but because of the potential for UV light to change toxicity up or down, labs must first demonstrate the inadequacy of a short exposure time before being allowed to increase beyond three minutes UV exposure duration.

Exposure of an effluent to UV light can induce toxicity, especially if polycyclic aromatic hydrocarbons are present. If a lab can demonstrate that UV disinfection changes toxicity relative to untreated effluent and that the alternative test setup doesn't work, then filtration through a 0.45 μm filter to remove pathogens is allowed. References:

Theisen, Daniel D., Donald D. Stansell, L. Curry Woods III, 1998: Disinfection of Nauplii of *Artemia franciscana* by Ozonation. *The Progressive Fish-Culturist:* Vol. 60, No. 2, pp. 149-151.

Grothe, Donald R., Daniel E. Johnson, 1996: Bacterial Interference in Whole-Effluent Toxicity Tests***—Short Communication.*** *Environ Toxicol Chem* 15: 761–764.

Kszos, Lynn Adams, Arthur J. Stewart, James R. Sumner, 1997: Evidence that Variability in Ambient Fathead Minnow Short-Term Chronic Tests is Due to Pathogenic Infection**.** *Environ Toxicol Chem* 16: 351–356.

Downey, Patrick J., Kari Fleming, Richard Guinn, Norris Chapman, Patricia Varner, John D. Cooney, 2000: Sporadic Mortality in Chronic Toxicity Tests Using *Pimephales promelas* (Rafinesque): Cases of Characterization and Control**.** *Environ Toxicol Chem* 19: 248–255.

Test analysis

Failure of EPA statistical flowcharts

A WET test must be repeated if the flowcharts for determining NOECs in the EPA toxicity test manuals cannot be followed due to a low number of replicates. The problem will happen when test results from less than four replicates have data that are not normally distributed or that have unequal variances. The minimum number of replicates required in the EPA manuals are sometimes too low for reliably determining LOECs and NOECs or for single comparison hypotheses testing. The flow charts for single comparison hypothesis testing can be found in Appendix H of the EPA freshwater chronic manual, Appendix G of the marine chronic manuals, and in Figure 12 of the acute manual. Be aware of the EPA recommendation to use the Kolmogorov "D" statistic to replace Shapiro-Wilk's Test when $n > 50$.

Four replicates can be inadequate for determining an NOEC when replicate numbers are unequal and test data are not normally distributed or have unequal variances. Labs intending to run more replicates in the control than in the effluent concentrations should consult the table of critical values for Wilcoxon's Rank Sum test to determine the minimum number of replicates needed for the test concentrations. The accidental loss of a test chamber in a typical test of five test concentrations and a control will also cause replicate numbers to be unequal and four replicates to be inadequate if the nonparametric hypothesis test (Wilcoxon's Rank Sum) must be used.

The minimum number of replicates need not be increased to compensate for the occasional loss of test chambers because it will rarely necessitate rejection of a test for failing the EPA statistical flowchart. If a test chamber has been accidentally lost from a test using four replicates per concentration and a nonparametric hypothesis test is required, the concentration-response relationship can be examined to see if the concentration losing a replicate can be excluded from the analysis because it has healthy test organism performance nearly equal to adjacent concentrations and the control. If the ACEC and CCEC have been included in the concentration series of a test losing a test chamber and have at least three

replicates remaining at the end of the test, then single comparison hypothesis testing can be used to compare the ACEC or CCEC to the control to provide the needed regulatory determination.

If a lab increases the number of effluent concentrations in a test series beyond five, the EPA flowcharts for determining NOECs may not work. Adding extra concentrations to the series improves the ability of a test to measure toxicity and calculate point estimates. Unfortunately, the extra concentrations also raise the minimum number of replicates required to five or higher for using Steel's Many-one Rank Test or Wilcoxon's Rank Sum Test to determine the LOEC/NOEC.

An important point to note here is that labs are free to perform and report statistics in any way they feel is appropriate to meet a client's needs. After we get the report, we will recalculate the statistics as described in this document and only insist that the test be conducted and data recorded so that we can perform our own calculations.

Controlling Type I and Type II errors

Power Standards

Variability among replicates can prevent large differences in response between the control and an effluent concentration from being detected as statistically significant. To reduce the potential for Type II errors (false negatives) when variability is high, WAC 173-205-020 lists statistical power standards for both acute and chronic tests. The power standards act as a safety net to prevent large effects from being declared nontoxic.

An acute toxicity test must be able to detect a difference of 30% at the ACEC as statistically significant or it must be repeated with an increased number of replicates. A chronic toxicity test must be able to detect a difference of 40% at the ACEC or CCEC as statistically significant or be repeated with an increased number of replicates. Appendix E shows examples. Power standards do not apply to Fisher's Exact Test.

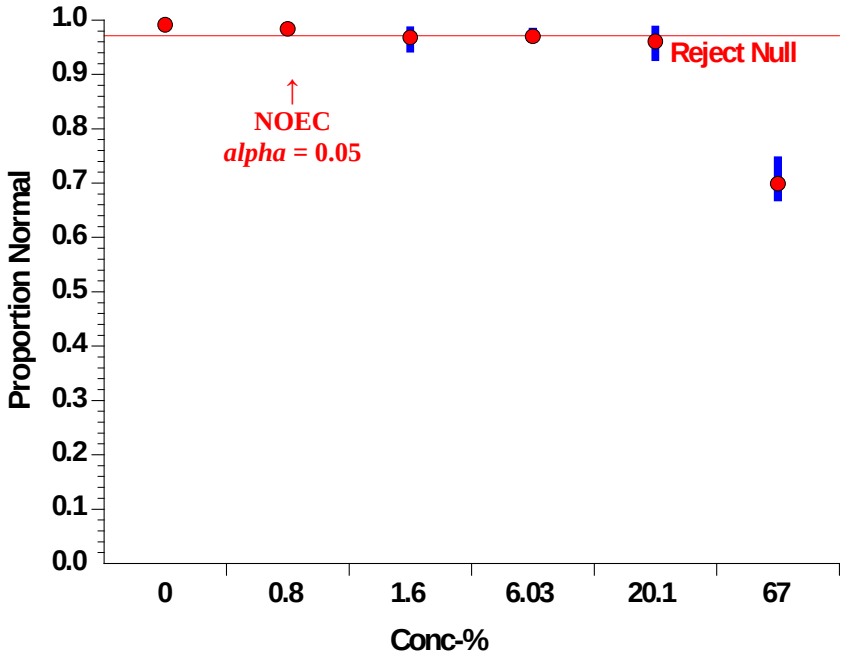
The EPA West Coast Manual now includes minimum significant difference (MSD) criteria for some tests. These MSD criteria have been incorporated in the test condition lists in the *III. Test Species and Specific Test Conditions* section.

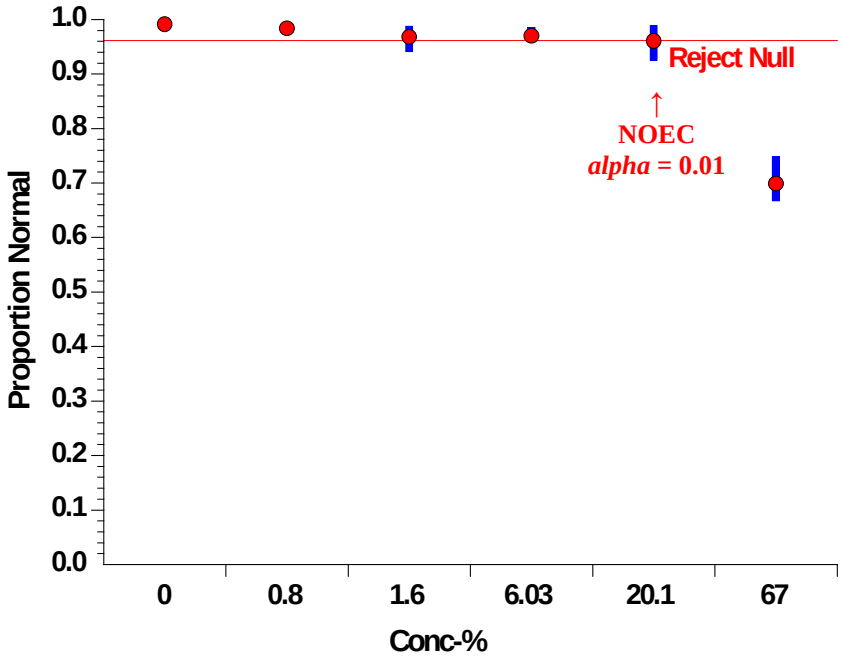
Changing alpha for small differences in response

A Type I error for a WET test is concluding that the effluent is toxic when it actually isn't. These false positives occur when a difference in response that is solely due to chance is large enough to be identified as significant by statistical analysis. It is extremely unlikely though that chance would cause all test organisms in the control to live and all test organisms in the effluent to die. In addition, when everything lives in both the control and effluent, no hypothesis test will ever find a significant difference no matter how many times that it is run. Somewhere in between these two extremes, the Type I error rate approaches alpha as the measured difference in response becomes smaller and more likely to be due to chance. Decreasing alpha for smaller differences in response keeps the Type I error rate from ever being 1/20 tests.

To reduce WET limit violations (and anomalous concentration-response relationships) due to statistical significance that is a Type I error, we lower alpha when differences in test organism response are small. Alpha is an approximation of the maximum Type I error rate. To reduce Type I errors and the number of interrupted concentration-response relationships, we set alpha = 0.01 for small differences in test organism response. Alpha will be lowered from 0.05 to 0.01 if a 10% difference in an acute test is

significant or a 20% difference in a chronic test is significant. The following graphs illustrate the improvement in test results by changing alpha.





Our WET database was queried for NOECs (alpha = 0.05) for fathead minnow survival and growth, *Ceriodaphnia* survival and reproduction, and phytoplankton cell density endpoints. Those tests with an interrupted concentration-response relationship were counted. An interrupted concentration-response relationship occurs when a concentration has a statistically significant difference from the control but one or more higher effluent concentrations do not. It is reasonable to consider these interrupted concentration-response relationships to be examples of statistical false positives (Type I errors) since the higher effluent concentrations would have more of whatever toxicants were present.

The tally gave percentages of interrupted concentration-responses that were close to the 5% predicted by setting alpha = 0.05. 42 out of 724 (5.8%) 7-day survival NOECs were identified as having an interrupted concentration-response. 52 out of 764 (6.8%) sublethal endpoint NOECs were identified as having an interrupted concentration-response. These results suggest that the incidences of statistical significance with an interrupted concentration-response can mostly be explained as Type I errors.

Looking for a concentration-response relationship

Type I errors can often be spotted by examining the concentration-response relationship. The lower effluent concentrations in a WET test are typically nontoxic and have a flat concentration-response at generally the same level as the control. However, test organism response at any of these nontoxic concentrations rarely falls exactly on this level line. When the response at one of these concentrations falls below the line, then a hypothesis test might identify it as a statistically significant reduction in response. A generally flat concentration-response relationship in the vicinity of the failing concentration and an absence of statistical significance at higher concentrations allow a reviewer to discount the anomalous statistical significance and avoid a Type I error (See Appendix D).

The four 10s in the control and four 9s in the effluent problem

Changing alpha cannot help when the control has four replicates with complete survival and one of the effluent concentrations has lost exactly one test organism in each of its replicate chambers. Steel's test will always find this to represent a statistically significant reduction in test organism survival from the control. Steel's test has no significance values at alpha = 0.01 for most tests with 4 or 5 replicates.

When higher concentrations show decreasing survival with increasing concentration, the statistical significance of 80% or 90% survival is acceptable. However, when the test has a flat overall concentration-response relationship (nearly equal test organism response), it is difficult to say that 80% or 90% survival is an adverse effect caused by the effluent when the test methods say that it is acceptable for there to be 90% survival in the nontoxic control of an acute test or 80% survival in the nontoxic control of a chronic test.

Test rejection is the only response available for four 10s in the control and four 9s at the ACEC or CCEC from a test conducted for WET limit compliance monitoring. However, Appendix D contains remedies other than test rejection which can be used when the ACEC or CCEC is not involved in a test conducted for compliance monitoring.

Outliers

Labs may identify outliers if they choose to do so using an appropriate statistical procedure and submit the test results with the outliers both excluded and included. EPA recommends Gentleman Wilk's A statistic or Dixon's test. Comprehensive Environmental Toxicity Information System (CETIS) uses Grubbs test, which is acceptable if data are normally distributed. If outliers are to be excluded, then they should be identified at both low and high ends of test organism performance. If a 7-day survival result is identified as an outlier, then any associated sub-lethal combined endpoint having a survival component (e.g. biomass) must also have its result excluded.

An important function of the WET database is to provide an accurate record of test performance as well as of effluent toxicity. The exclusion of outliers will hide some important features of test performance. Most labs are likely to not look for outliers and include results from all test chambers in calculations, and this is also how we will be recording most test results. Outlier exclusion is allowed after considering the following:

- A lab suspects but cannot document that a physical factor (e.g. contaminated glassware) was responsible for one or two aberrant values and wishes to officially exclude the results from those test chambers.
    - Outlier identification is not necessary before excluding the results from one or two test chambers with a documented physical accident or an observation of infected test organisms.
- Outlier identification may be attempted to improve the concentration-response relationship of a test which might be rejected for being anomalous.
- Outlier identification may be used to meet the power (statistical sensitivity) standards when the pooled variance has been adversely affected by one or two values. Otherwise, outlier identification should not be used to suppress test variability and bias hypothesis testing.
- Tests having more than two test chambers with aberrant results will be analyzed with all test chambers included and test acceptability based upon the results.

NOEC expression

When the lowest effluent concentration tested ends up as the LOEC (has a statistically significant difference from the control), the NOEC must be expressed as < that lowest concentration. When the highest effluent concentration tested ends up as the NOEC (has no statistically significant difference from the control), the LOEC should be expressed as > that highest concentration.

All test concentrations should be used in calculating the NOEC/LOEC for a sublethal endpoint regardless of whether statistically significant for survival.

Dual endpoint tests

Labs sometimes provide clients with an acute test result from a 7-day chronic test. This is called "dual endpoint testing." In a dual endpoint test, the 48-hour survival counts for mysids or the 96-hour survival counts for fish from 7-day chronic tests are used as the final counts in an acute test. Permittees should always be informed by the lab that dual endpoint testing will deprive them of the advantages of a separate acute test run at a cooler temperature, without daily renewals, and using older test organisms. Dual endpoint tests must have the date and time of the 48-hour count (mysids) or the 96-hour count (fish) recorded as the end of the acute test. Dual endpoint tests are acceptable with fathead minnows (*Pimephales promelas*), silverside minnows (*Menidia beryllina*), topsmelt (*Atherinops affinis*), and the Atlantic mysid (*Americamysis bahia*). The required chronic test conditions take precedence over the acute in a dual endpoint test.

Acute tests derived from 7-day *Ceriodaphnia* chronic tests are not acceptable because this chronic test fails to meet the minimum number of test organisms required in an acute test. The setup of the *Ceriodaphnia* chronic test (1 daphnid per chamber) makes Fisher's Exact Test the only option for analyzing survival causing a loss in sensitivity.

Reference toxicant tests

Long and careful consideration have led to the conclusion that the only reference toxicant test results useful for interpreting WET test results are from concurrent reference toxicant tests. A reference toxicant test has little relevance to a WET test result from which it is separated in time by as much as 15 to 30 days. Concurrent reference toxicant testing is the only method that produces a true positive control for a toxicity test. No reasonable person would use a negative control separated by 15 to 30 days from a toxicity test. For this reason, we feel justified in seeking a more relevant and cost-effective method for

relating lab performance to the interpretation of an individual WET test result. See "Ongoing Control Mean and CV Reporting" section for more information.

However, routine reference toxicant testing is irreplaceable as an ongoing evaluation of intralaboratory variability. If labs use the same reference toxicant and a similar dilution series, then interlaboratory comparisons are also possible. Labs trying a new test method or with a concern about a routine test gone awry sometimes seek reference toxicant test results from other labs. Reference toxicant test results reflect both test organism sensitivity and lab technique. When assessed over time, reference toxicant test results form the basis for test method quality control by a lab. Monthly reference toxicant testing is appropriate for this quality control task.

The minimum reference toxicant testing needed to meet our interpretation of the requirements in the EPA manuals (both Sections 4.7 and 4.16) is:

- One reference toxicant test per month for every acute and 7-day (short-term) chronic test species used routinely (more than once per month).
    - Because an acute test result can be determined during a 7-day chronic test, acute and chronic reference toxicant testing for a fish or mysid can be combined.
- Concurrent reference toxicant testing is required for:
    - All non-routine tests (test performed once per month or less).
    - All tests conducted with bivalves or echinoderms.
    - Testing with plants.
    - Tests using organisms or brood stock collected from the wild.
- A batch of bivalve or echinoderm tests may be covered by a single concurrent reference toxicant test if gametes from the same spawning are used.
- Labs getting test organisms from a supplier that does not routinely conduct reference toxicant testing must conduct a reference toxicant test on each shipment of organisms.
- Reference toxicant test results from test organism suppliers cannot substitute for the reference toxicant testing required to demonstrate ongoing lab performance. Organisms tested with reference toxicants by suppliers have not been packaged and shipped and the dilution water and other test conditions will differ between the labs.

All labs must conduct ongoing control charting based on reference toxicant testing and report the results, whether within control limits or not, in the report for each effluent or ambient water test. Acceptability of an effluent or ambient water toxicity test result will be based on a variety of the considerations described in this document and reference toxicant test results being within control limits is only one consideration that is not by itself conclusive.

The Ecology staff person with primary responsibility for reference toxicant testing requirements is the toxicologist for the Environmental Laboratory Accreditation Program. This person reviews standard operating procedures (SOPs) and quality assurance manuals for toxicity tests and performs on-site evaluations as part of the lab accreditation process. For bioassay labs to maintain accreditation they must perform at least one reference toxicant test every six months. Except for ASTM E 1022 and E 1706 and other bioaccumulation/bioconcentration tests, this requirement applies to all effluent, sediment, soil and dangerous waste characterization type bioassay methods for which accreditation is sought. Even if a lab does not conduct any tests on environmental samples using a particular species/method within a six-month period, it must perform a reference toxicant test as an accreditation requirement at least every six months. If you have questions regarding accreditation of bioassay labs, contact Ecology's Lab Accreditation Unit at (360) 871-8844 or e-mail alan.rue@ecy.wa.gov.

Ongoing control mean and CV reporting

For the sublethal endpoint from each chronic WET test in a report, the lab must report the mean and coefficient of variation (CV) from the last twenty (or however many are available) dilution water controls, regardless of test material. The means and CVs may either be in the form of a table or a QC-plot. The information will serve to spot tests with unusual control response affecting statistical sensitivity. Labs are encouraged to also use this information to spot trends and aid in quality control for culturing, dilution water preparation, and testing.

Water quality measurements

Purpose

Water quality measurements are important for verifying control of test conditions and can aid in test interpretation. The following Parameters and Schedule must be followed and the list notes when water quality measurements affect test acceptability. CETIS entries should include on the Samples tab the sample measurements marked in the following section with an asterisk (*****). Other sample measurements need not be entered into CETIS.

Parameters and schedule

Temperature*:* Experience has shown that inadequate monitoring and maintenance of temperature contribute to poor control performance and to test variability. Failure to adequately measure and control temperature will cause test reports to be rejected. Continuous temperature monitoring is required by Section 4.6.1 of all of the EPA manuals in at least one representative location inside the environmental chamber and the records kept for audits.

In the EPA test manuals' instructions for each test type, EPA requires temperature measurement in at least one test chamber at each concentration and the control at the beginning of each test and daily thereafter. These test-specific sections also require that temperature be measured in a sufficient (but unquantified) number of test vessels at the end of the test to determine temperature variation within the environmental chamber. In addition, some of these same test-specific sections repeat the requirement from Section 4.6.1 for continuous temperature monitoring and increase the minimum number of locations to two.

It would be better if temperature measurements were done based on location within the test setup rather than on test concentration. One temperature measurement at each concentration will not be enough for drawing conclusions about temperature effects on concentration-response if container positions are randomized. Depending upon the randomization, selecting one chamber per test concentration might not provide temperature measurements at all locations of concern (edges, corners, middle) within the test setup. Extra effort is needed to get a temperature measurement at each test concentration while maintaining randomization and keeping technicians as blind as possible to test chamber identity.

There is little advantage to using surrogate test chambers for temperature measurement. Surrogate test chambers obviously cannot be used for other parameters such as dissolved oxygen or pH. Temperatures can be different in test chambers relative to surrogate chambers because of differences in evaporation

rate, the activities of test organisms, or the color of solutions. Surrogate test chambers require extra effort and must be cleaned regularly and refilled with clean water.

Temperature measurements to at least a tenth of a degree must be made at the beginning of a test, daily during the test (before renewal if solutions are renewed that day), and at test termination by one of the following options (the first two are preferred):

1. Temperature may be monitored in at least one test chamber per concentration. Use the randomized chamber positions to select test chambers for temperature measurements that are in representative locations near corners, edges, and the middle of the test setup. Record the test chamber selections in the test report.
2. Temperature may be measured in at least six locations (one on each corner and two near the middle of each half) within the test setup without regard to concentration. Record the location selections in the test SOP for your lab.
3. Surrogate test chambers for temperature monitoring must be similar in size, shape, and volume of contents to test chambers. At least three surrogate test chambers should be used (on each shelf if an incubator with one in the middle, one on a corner at the front, and one on the opposite corner at the back. Temperature measurements must also be made in test solutions at test termination at a minimum of six locations (one on each corner and two near the middle) within the test setup in order to assess the variability of temperature between test chambers and to provide for comparison to the surrogate chambers.

Dissolved oxygen is measured once per day in at least one test chamber at each effluent concentration (including the control) at a minimum and often enough to detect any drop in DO before test organisms are adversely affected. DO must be measured in samples and in one test chamber at each effluent concentration at test initiation to determine if levels are adequate (or to detect supersaturation). DO should be checked again several hours later to see if it has dropped sufficiently to cause concern. If DO does not drop significantly, then it may be measured once per day (before and after test solution renewal). DO measurements are required in order to justify aeration of the sample or test chambers. Test results will be rejected if aeration is done when not justified or if DO is allowed to persist at levels lower than that specified in the test method.

pH* is measured in samples and in one test chamber at each concentration (including the control) at the beginning of a test, daily during the test (before and after renewal if solutions are renewed that day), and at test termination. pH must be measured to a tenth of a unit.

Total ammonia* is measured in all samples which might contain ammonia (all municipal effluents and any industry with the potential for ammonia). Care should be taken that permittees do not have to pay for a toxicity identification evaluation to discover that ammonia was the cause of noncompliance.

Total residual chlorine* is measured in all samples which might contain chlorine (all municipal effluents and any industry with the potential for chlorine). Chlorine must also be measured in each batch of dilution water if prepared from tap water. Care should be taken that permittees do not have to pay for a toxicity identification evaluation to discover that chlorine was the cause of noncompliance.

Conductivity* is measured in all samples and in one test chamber at each test concentration (including the control) at the beginning of a test using freshwater organisms, prior to test solution renewal, and at test termination.

Salinity is measured in the dilution water control, the brine control, and in at least one test chamber at every effluent concentration at the beginning of a test using saltwater organisms, prior to each test solution renewal, and at test termination. Reports will be rejected if salinity is not maintained within accepted ranges.

Total hardness* and ***Total alkalinity**** are measured in all samples. They are both also measured in the dilution water for all tests using freshwater organisms. See Appendix J.

Bivalves and Echinoderms: DO, pH, temperature, and salinity are measured in each test concentration and the control at the beginning of the exposure period for all bivalve and echinoderm tests. Temperature must be monitored continuously in at least two locations in the test setup or measured and recorded daily in two locations. Temperature should be measured in a sufficient number of test chambers at the end of the test to determine temperature variation between environmental chambers.

Deviations from protocols and acceptability criteria

In order to have an imperfect test result accepted, a lab must contact Randall Marshall at 360-407-6445 or randall.marshall@ecy.wa.gov during or immediately following the test. If acceptance has been given, the lab must document the test deviations *and the telephone conversation or e-mail exchange* in the test report.

Imperfect test acceptance is generally based upon:

- Protocol deviations are minor and not likely to mask toxicity.
- Valid statistical calculations can be performed as described in the "Test Analysis – Failure of EPA Statistical Flowcharts" section.
- The test results show no significant toxicity.
- Control acceptability criteria failures are accompanied by robust and consistent organism performance at all other test concentrations.

Untimely arrival of samples for 7-day chronic tests causes most requests. The rules for accepting 7-day tests when the sample conditions in the "Sample Handling – Holding Time and Temperature" section are not met are:

- If the first sample arrives past the 36-hour holding time or is too warm, then the test must be rescheduled and started with another good sample.
- If the second sample arrives late, the test will be accepted if daily renewals were continued using the first sample and the second sample arrives with a good temperature.
- If the third sample arrives late, the test will be accepted if daily renewals were continued using the second sample and the third sample arrives with a good temperature.
- A test is not acceptable if any sample is first used more than 72 hours after being taken. (The EPA maximum sample holding time was once set at 72 hours.)
- A test is not acceptable if both the second and third samples arrive late.
- A test is not acceptable if any sample arrives late and is more than 8° C at receipt.
- A test is not acceptable if any sample arrives late and the test result shows effluent toxicity at levels of regulatory concern for the discharge.

Check for completeness of report

Labs must attach a readable copy of all WET test bench sheets and chain-of-custody forms to the test report. The bench sheets must include toxicological and water chemistry data. The bench sheets must record counts of number alive (not percentages or number dead) in order to be acceptable. Start counts must be clearly recorded on the bench sheet. The WET test report must include computer printouts of test data and summary results of statistical analyses. The full details of the statistical analyses do not need to be printed and included in reports. If a lab is sending a CETIS export of test data, then a scanned copy of the paper report may be included on the disk or uploaded as described below and an extra paper copy is not needed.

The test report must contain all of the information needed for comparison with the requirements in this document. Labs using lists rather than narratives tend to produce more accurate and easy to read reports. Presenting the information once makes accuracy and consistency easier to achieve. Test organism source, age, and unusual conditions (lethargy, hyperactivity, spots or filaments, discoloration, excessive ventilation, etc.) should be reported. Special circumstances such as treatment system upsets known to exist at the time of sample collection must be reported.

The report must contain a description and justification of any dechlorination procedure used. The stoichiometric calculations for determining the proper amount of dechlorinating agent must be included in the test report. The report must contain a description and justification of any filtration, aeration, hardness adjustment, UV disinfection, or pH control procedure used. Each test report must contain a section where deviations from test protocols are listed or their absence noted.

Electronic submission of test data

We use CETIS for analyzing and storing WET data. CETIS is an MS Access application produced by Tidepool Scientific Software (https://tidepool-scientific.com/) which has the ability to create an export database file that other CETIS users can import. Any lab which has CETIS should use this feature to prepare electronic data submissions. We do not require labs to use CETIS. If you do not have CETIS, please put raw test data into an MS Excel compatible spreadsheet.

Any lab which gets CETIS is eligible to receive an export file of their own test records from our database so they can have complete Washington State client records. Labs who get CETIS should contact us (randall.marshall@ecy.wa.gov or 360-407-6445) for a list of standardized entries into fields. This will ensure compatibility with us and other labs.

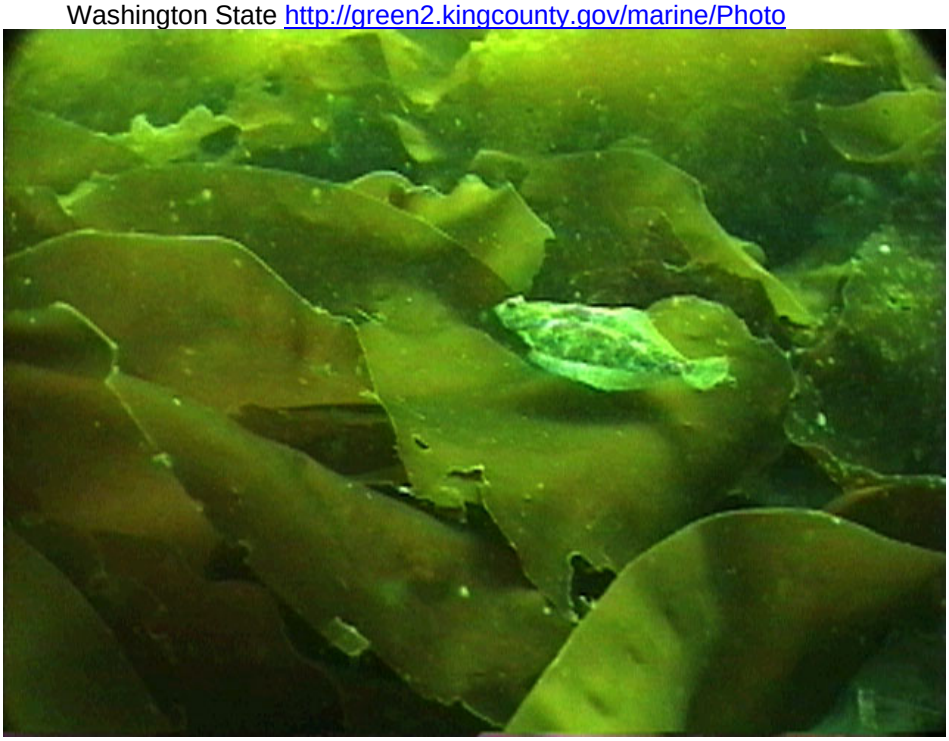
Labs have begun scanning the paper report described in the previous section and including the PDF with the CETIS export. This is more convenient for us and saves paper. It also allows permittees to submit WET tests results online. Our preferred reporting method is now to directly upload the report PDF and CETIS export or MS Excel with raw data onto PARIS, our permittee compliance website (www.ecy.wa.gov/programs/wq/permits/paris/portal.html). Your client is responsible for the upload, but may need assistance. Folks with questions about PARIS uploads should visit: www.ecy.wa.gov/programs/wq/permits/paris/contacts.html or call 1-800-633-6193 Option 3.

If necessary, you or your client may e-mail (randall.marshall@ecy.wa.gov) the test reports and results. Please remember to change the extension of the export file (xxx.mdb) to something else (xxx.mdm for example) so MS Outlook does not block the file.

Feedback to permittees and labs

CETIS produces a special report called the WET evaluation. We provide a WET evaluation for each test where we make statements on test quality and lab performance. We also note effluent toxicity and how it relates to permit compliance. These WET evaluations will soon be posted online for viewing by permittees and the public. The name of the lab along with its contact information are clearly identified on each evaluation.

Sanddab (*Citharichthys stigmaeus*) over *Laminaria saccharina*
courtesy of King County, Dept. of Natural Resources and Parks, Washington State http://green2.kingcounty.gov/marine/Photo



Test Species and Specific Test Conditions

Acute toxicity tests and species

Because EPA has not provided an acute toxicity test for plants, effluents can only be tested for acute toxicity with a fish and an invertebrate. Acute toxicity tests with fish are 96-hour static-renewal tests. Acute toxicity tests with invertebrates are 48-hour static tests. EPA has developed the freshwater WET testing program around the use of fathead minnows for fish testing. If Ecology decides to require acute WET testing with rainbow trout (*Oncorhynchus mykiss*) in order to provide direct protection of salmonids, it is possible that the permit will also require fathead minnow testing so that any TI/RE can be performed with fathead minnow.

If the effluent itself is freshwater, freshwater species will typically be used for acute WET testing regardless of receiving water salinity. If the effluent is too saline for freshwater organisms, the permit will require acute testing with topsmelt (*Atherinops affinis*) and a mysid (*Americamysis bahia*, formerly *Mysidopsis bahia*). Testing with saltwater organisms might be used instead of hardness adjustment to prevent exaggerating toxicity when a low hardness wastewater is discharged to marine water. If salinity adjustment is needed, artificial sea salts must be used in acute toxicity testing because WAC 173-205-050 requires that the response in 100% effluent be used to determine the need for an acute toxicity limit or a new effluent characterization.

The species for acute testing now include Pacific herring (*Clupea pallasii*). Herring for testing are available from January to June and may be produced either by hatching natural spawn or conducting in-lab fertilization using gonads excised from wild-caught running ripe herring. Availability peaks February through April.

Saanich Inlet Pacific herring (*Clupea pallasii)* photo courtesy of VENUS Project/University of Victoria http://www.venus.uvic.ca/

Table of Required Acute Toxicity Test Conditions

test organism	test type	chamber size	solution volume	# organisms per chamber	# replicates	age	temperature	aeration	feeding
Ceriodaphnia dubia	48-hr static	minimum 30 mL	minimum 15 mL	minimum 5	minimum 4	< 24 hrs	20° ± 1°C	if DO < 2.0 mg/L	for at least 2 hrs prior to test start; none after
Daphnia pulex/magna	48-hr static	minimum 30 mL	minimum 25 mL	minimum 5	minimum 4	< 24 hrs	20° ± 1°C	if DO < 1.0 mg/L	for at least 2 hrs prior to test start; none after
Pimephales promelas	96-hr static-renewal (at 48 hrs)	minimum 250 mL	minimum 200 mL	minimum 10	minimum 4	1- 14 days, 24 hr age range	20° ± 1°C or 25° ± 1°C	if DO < 4.0 mg/L	prior to start and 2 hrs prior to renewal
Oncorhynchus mykiss	96-hr static-renewal (at 48 hrs)	minimum 5 L	minimum 4 L	minimum 10	minimum 4	15 - 30 days after swim-up[1].	12° ± 1°C	if DO < 6.0 mg/L	none within 12 hours of test start
Menidia beryllina	96-hr static-renewal (at 48 hrs)	minimum 250 mL	minimum 200 mL	minimum 10	minimum 4	9 - 14 days, 24 hr age range	20° ± 1°C or 25° ± 1°C	if DO < 4.0 mg/L	prior to start and 2 hrs prior to renewal
Atherinops affinis	96-hr static-renewal (at 48 hrs)	minimum 500 mL	minimum 200 mL	minimum 5	minimum 4	7 - 15 days, 24 hr age range	20° ± 1°C	if DO < 4.0 mg/L	prior to start and 2 hrs prior to renewal
Americamysis bahia	48-hr static-renewal (at 24 hrs)	minimum 250 mL	minimum 200 mL	minimum 10	minimum 4	1 - 5 days, 24 hr age range	20° ± 1°C or 25° ± 1°C	if DO < 4.0 mg/L	prior to start and daily 2 hrs prior to renewal
Clupea pallasii	96-hr static-renewal (at 48 hrs)	minimum 400 mL	minimum 200 mL	minimum 10	minimum 4	≤ 48 hours post-hatch	12° ± 1°C	if DO < 5.0 mg/L	none

NOTE: All of the conditions in this table and in the general test conditions below must be documented in each test report.

General Conditions

- The approved test manual is EPA-821-R-02-012. See the *Supplemental saltwater chronic toxicity tests* section for the method reference for herring toxicity tests.
- Dual endpoint tests must meet conditions in chronic manual to have a valid chronic result.
- Illumination must be for 16 hours at 10 - 20 $\mu E/m^2/s$ (50 - 100 ft-c) followed by 8 hours of darkness.
- Holding time is 36 hours maximum prior to test initiation. The original sample (up to 84 hours old) may be used for renewals at 48 hours if held at 0-6° C in the dark.
- Controls must have at least 90% survival or the test should be repeated as soon as possible on a fresh sample.
- The salinity should be 30‰ for all acute tests with marine organisms.
- *Americamysis bahia* was formerly *Mysidopsis bahia*.
- See Appendix A for a discussion of trout age determination.
- *Menidia beryllina* or any other fish or mysid may be fed daily as long as an 80% renewal of test solution follows 2 hours after each feeding.

Freshwater chronic toxicity tests

Chronic WET test selection is fairly simple for discharges to freshwater. EPA recommends testing with a fish, an invertebrate, and a plant and has provided only one of each for freshwater chronic WET testing (fathead minnow, *Ceriodaphnia dubia*, and *Pseudokirchneriella subcapitata*). WAC 173-205-050(1)(a) requires that effluents with a risk for aquatic toxicity be tested at a minimum for toxicity to a fish, an invertebrate, and if appropriate, a plant. Permits for discharges to freshwater will contain standard requirements for the use of fathead minnow and *Ceriodaphnia* in chronic toxicity tests. The fathead minnow chronic test measures survival and growth. The *Ceriodaphnia* chronic test measures survival and neonate production.

The Phytoplankton (*Pseudokirchneriella subcapitata*) Growth Inhibition Test is considered an optional chronic toxicity test. The Phytoplankton Growth Inhibition Test is sometimes less sensitive than tests with fish and invertebrates and can suffer from various effects (nutrients or residual flocculants) which can mask or confuse the measurement of effluent toxicity. However, any clearly toxic response in an effluent test using *Pseudokirchneriella subcapitata* is a good indication of toxicity to plants, and the Phytoplankton Growth Inhibition Test will sometimes be required.

Fathead minnow (*Pimephales promelas*) courtesy of Konrad Schmidt, **Fishes of Minnesota.** James Ford Bell Museum of Natural History. University of Minnesota. http://hatch.cehd.umn.edu/research/fish/fishes/

All conditions in the following tables for the freshwater chronic toxicity tests must be met and reported for each test. Three separate samples are required for daily renewals in 7-day chronic tests.

Ceriodaphnia survival and reproduction

Test species: *Ceriodaphnia dubia*

Approved test method: EPA-821-R-02-013, method 1002.0

Test type: 7-day static-renewal (> 90% renewal of test solution in each test chamber daily by transfer of test organism to another container with fresh test solution)

Temperature: 25° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 $\mu E/m^2/s$ (50 - 100 ft-c) followed by 8 hours of darkness.

Test chamber size: 30 mL (minimum)

Test solution volume: 15 mL (minimum)

Age of test organisms: < 24 hours and within an 8 hour age range

Number of organisms/chamber: 1 from a female with ≥ 8 neonates in the 3rd or subsequent broods

Number of replicates/concentration: 10 (minimum)

Feeding: 0.1 mL YCT and 0.1 mL algal suspension daily

Aeration: None unless DO < 2.0 mg/L and then optional at lab discretion using a very low bubbling rate

Test duration: The duration of exposure is expressed in terms of time (7 days) for the survival endpoint and in terms of life cycle (3 broods) for the reproduction endpoint. Final survival counts must be taken at the end of 7 days. Final counts of neonate production should be taken immediately upon production of the third brood by 60% of the surviving control organisms. The third brood will usually occur on the 6th, 7th, or 8th day. The maximum test duration allowed is 8 days as long as test solutions are renewed on each full day. Tests may not be continued beyond the third brood in order to get 15 neonates/surviving adult in the control.

Endpoints: Number of survivors at 7 days and number of neonates per female at 3 broods (# neonates per concentration divided by the # females at test initiation)

Control performance criteria: ≥ 80% survival in the control
An average of 15 neonates per surviving adult in the control
≥ 60% of the surviving control organisms producing 3 broods

Other test acceptability criteria:	≤ 10% males in the surviving test organisms over all test concentrations ≤ 20% males in the surviving test organisms in the ACEC, CCEC, or LOEC

Specific concerns

All surviving *C. dubia* producing no neonates in the test must be examined to determine gender and the results of the determination reported unless reproduction has been nearly eliminated in a test concentration and this fits an expected concentration-response relationship. It is understood that very young *C. dubia* can be difficult to sex and any *C. dubia* that dies in the first two days of the test may be excluded from calculations for reproduction if gender is difficult to determine and it is one of no more than two mortalities in a concentration. Otherwise, difficult to sex young *C. dubia* must be considered to be female and included in all calculations.

Each successive brood from 1 to 4 tends to increase in neonate count from 50% to 75% over the previous brood. Differences in the number of broods or in the neonate totals due to differences in age or the timing of counting are a big source of variability. The test method requires that all of the *C. dubia* used in a test be less than 24 hours old and be within 8 hours of the same age. Because of the very short lifecycle of *C. dubia*, this restriction cannot completely eliminate these age-related differences in reproduction. The test method also says that all observations at test termination should be completed within 2 hours or the last containers counted might have produced significant numbers of neonates after the first containers received final counting. Labs must therefore strive to keep differences in age and the timing of counting to as small as possible and never exceed the limits in the test method.

Neonate counts are made at 24-hour intervals and will not occur for many females at a time between broods. A daily count may include neonates from only a partial brood or from two separate broods. A skilled technician is needed to tell the difference between broods in order to properly judge when 60% of the surviving control organisms have produced 3 broods. Judging brood occurrence requires experience, a good stereomicroscope, and sufficient time.

The tendency toward reduced neonate production in some females but not in others when the culture condition is borderline goes beyond the normal variation in individual test organism response. It is analogous to testing with organisms from two different life stages, each with its own baseline for the response being quantified. Labs must monitor culture health and reproduction daily, renew cultures both on a regular schedule and when needed, and immediately replace poorly performing batches of food and water.

Temperature inequalities that exceed the ± 1° C in the test method can influence the rate of neonate production for different containers.

Sources of error are unavoidable in any test and the proper solution is to distribute them randomly to avoid bias and invalid conclusions. For the *C. dubia* reproduction endpoint to be valid, all of the listed sources of error plus any others must be randomly distributed throughout the test.

Neonates from a single brood female must be placed in all test chambers assigned the same replicate number so that they appear only once in each test concentration (blocking by known parentage). *The process for achieving blocking by known parentage must be described in the report for each test*. The technique recommended by EPA seems to be to place cups into a test board and assume that each column is a test concentration and that each row contains replicates that have the same replicate number but are each from a different concentration. Test solutions are added accordingly. The cups in each row are then placed into a random order. One neonate from the same brood female is added to each cup in that row but not to cups in any other. All of the cups are then randomized together and the test conducted. Differences due to parentage are then evenly distributed among concentrations. Replicates can be compared at the end to see if there are differences due to parentage.

Fathead minnow survival and growth

Test species: *Pimephales promelas*

Approved test method: EPA-821-R-02-013, method 1000.0

Test type: 7-day static-renewal (80% renewal of test solution in each test chamber daily)

Temperature: 25° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Test chamber size: 500 mL (minimum)

Test solution volume: 250 mL (minimum)

Age of test organisms: < 24 hours (< 48 hours if shipped)

Number of organisms/chamber: 10

Number of replicates/concentration: 4 (minimum)

Feeding: 0.1 g wet weight (approximately 1,000 *Artemia* nauplii) per container 3 times daily at 4-hour intervals (4 times/day at 2.5- to 3.0-hour intervals is acceptable) or 0.15 g wet weight (approximately 1,500 *Artemia* nauplii) per container twice daily at 6 hour intervals: no food in final 12 hours

Aeration: none unless DO < 4.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate
Total weight of survivors divided by the initial count (biomass)
Total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 80% survival in the control
Average dry weight ≥ 0.25 mg per surviving fish in the control

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C. for more explanation.

Fathead minnow survival and growth (alternate version for samples having pathogens)

Test species: *Pimephales promelas*

Approved test method: EPA-821-R-02-013, method 1000.0

Test type: 7-day static-renewal (test organisms transferred to fresh chambers daily)

Temperature: 25° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Test chamber size: 30 mL (minimum)

Test solution volume: 20 mL (minimum)

Age of test organisms: < 24 hours (< 48 hours if shipped)

Number of organisms/chamber: 2

Number of replicates/concentration: 10 (15 preferred)

Feeding: 0.02 g wet weight (approximately 200 *Artemia* nauplii) per container 3 times daily at 4 hour intervals (4 times/day at 2.5- to 3.0-hour intervals is acceptable) or 0.03 g wet weight (approximately 300 *Artemia* nauplii) per container twice daily at 6 hour intervals: no food in final 12 hours

Aeration: None unless DO < 4.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate; total weight of survivors divided by the initial count (biomass); total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 80% survival and average dry weight ≥ 0.25 mg per surviving fish

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C. for more explanation.

The biomass endpoint is affected by stray deaths due to the low number of fish per concentration which causes a 5% reduction on average in mean biomass for each death. Because within concentration variability will cause a loss in statistical sensitivity or an anomalous concentration response, we prefer 15 replicates.

Phytoplankton growth inhibition

Test species: *Pseudokirchneriella subcapitata* (formerly *Raphidocelis subcapitata* and *Selenastrum capricornutum*)

Approved test method: EPA-821-R-02-013, method 1003.0

Test type: 96-hour static (nonrenewal)

Temperature: 25° ± 1°C

Illumination: Illumination must be continuous at 86 ± 8.6 μE/m^2/s (400 ± 40 ft-c or 4306 lux) and equally distributed over all test chambers.

Test chamber size: 125 mL flask

Test solution volume: 50 mL

Age of stocking solution: 4 to 7 days

Number of organisms/chamber: 10,000 cells/mL

Number of replicates/concentration: 4

Endpoints: Endpoints are cell density, fluorescence, or absorbance (680 nm is preferred but 750 nm as in the EPA manual is also acceptable). Control performance is verified by counting cells under a microscope.

Control performance criteria: An average of 1,000,000 cells/mL at end of test with variability not exceeding 20% coefficient of variation.

Other test acceptability criteria:

A concurrent reference toxicant test must be conducted with each batch of tests.

Only reconstituted water with 1 ml of stock nutrient solution per liter may be used as dilution water. Up to 1 mL of stock nutrient solution per liter of test solution should also be added to the sample so that an even distribution of nutrients between test concentrations and the control is achieved to the best degree practicable. The use of EDTA in the stock nutrient solution is now required. Hardness gradients are to be avoided as much as possible. Continual shaking by a mechanical shaker is required. Filters for samples must have a pore size no smaller than 1.6 μm.

Reference

Geis, S.W., K.L. Fleming, E.T. Korthals, G. Searle, L. Reynolds, D.A. Karner. 2000. Modifications to the Algal Growth Inhibition Test for Use as a Regulatory Assay. *Environmental Toxicology and Chemistry*. 19(1): 36–41.

Supplemental freshwater chronic toxicity tests

The following rainbow trout test methods have been used in the evaluation of stormwater treatment chemicals (See Appendix G.). Because they do not qualify under WAC 173-205-050(1)(d), they cannot be used for effluent characterization or compliance monitoring. They can also be used in permits as monitoring tools for receiving waters and trigger TI/REs if needed.

Environment Canada trout embryo viability

Test species: *Oncorhynchus mykiss*

Approved test method: E Test in Environment Canada EPS 1/RM/28 and: Canaria, E.C., J.R. Elphick, and H.C. Bailey. 1999. A Simplified Procedure for Conducting Small Scale Short-Term Embryo Toxicity Tests with Salmonids. Env. Toxicol. 14, 301-307.

Test type: 7-day static-renewal (80% renewal daily and all settled material removed from contact with embryos)

Temperature: 14° ± 1°C

Illumination: Dark with dim or red light only during test solution renewals

Test chamber size: 1 L or 4 L

Test solution volume: 500 mL or 3 L

Age of test organisms: Maximum 30 minutes after fertilization

Number of organisms/chamber: 30

Number of replicates/concentration: 4 (minimum)

Feeding: None

Aeration: Continuous gentle aeration (< 100 bubbles/minute)

Test duration: 7 days

Endpoints: Development Rate - number of viable embryos / number of survivors
Survival Rate – number of survivors / start count

Control performance: ≥ 70% embryo Development Rate and ≥ 70% embryo Survival Rate

Other: Pool milt but keep roe from each of four females separate and use in one replicate at each test concentration (block by parentage) so results from poor quality eggs in one replicate can be excluded, if needed, to meet control performance criteria.

Rainbow trout 7-day survival and growth

Test species: *Oncorhynchus mykiss*

Approved test method: Lazorchak, J.M. (lazorchak.jim@epa.gov) and M.E. Smith. 2007. Rainbow Trout (*Oncorhynchus mykiss)* and Brook Trout (*Salvelinus fontinalis*) 7-day Survival and Growth Test Method. Arch Environ Contam Toxicol. 53, 397–405.

Test type: 7-day static-renewal (80% renewal of test solution in each test chamber daily)

Temperature: 15° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Test chamber size: 500 mL (minimum)

Test solution volume: 400 mL (minimum)

Age of test organisms: 15-25 days post-hatch and 2-6 days post-swimup and actively feeding

Number of organisms/chamber: 5

Number of replicates/concentration: 4 (minimum)

Feeding: 0.5 mL brine shrimp slurry (3000-3500 nauplii) of < 24-hour old *Artemia* twice daily; none on last day. Food left for 2 hours before test solution renewal.

Aeration: none unless DO < 6.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: survival rate
total weight of survivors divided by the initial count (biomass)
total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 90% survival in the control
average dry weight in control at the end of the test must be at least 1.5 times the average initial dry weight

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C for more explanation.

Standard saltwater chronic toxicity tests

Permits for discharges to saltwater or brackish water will contain standard requirements for the use of a fish and a mysid in chronic toxicity tests measuring survival and growth. Permits will instruct permittees to use the West Coast fish (topsmelt, *Atherinops affinis*) and East Coast mysid (*Americamysis bahia*) for chronic toxicity testing. *Holmesimysis costata* (the West Coast mysid) has been tested only four times in this state and hasn't been used since 1998 because it can't be cultured. Silverside minnows (*Menidia beryllina*) may be substituted for topsmelt when topsmelt are temporarily unavailable.

Mysidopsis bahia has been changed to *Americamysis bahia*. The test title has retained the old name. Labs do not need to attempt the fecundity endpoint with *Americamysis bahia*.

Labs can use brine in chronic toxicity testing with saltwater organisms, and the highest effluent concentration in the test will be around 70%. Brine cannot be used in acute tests (or dual endpoint tests) because we need the survival response in 100% effluent.

All conditions in the following tables for the standard saltwater chronic toxicity tests must be met and reported for each test. These conditions take precedence over the Table of Required Acute Toxicity Test Conditions when conducting dual endpoint tests.

Mysis relicta courtesy of NOAA, Great Lakes Environmental Research Laboratory

West coast mysid survival and growth

Test species: *Holmesimysis costata*

Approved test method: EPA/600/R-95/136

Test type: 7-day static-renewal (75% renewal of test solution in each chamber at 48 and 96 hours)

Temperature: 13° ± 1°C (No mysids allowed originating from south of Pt. Conception)

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Salinity: 30 or 34 ± 2‰

Test chamber size: 1000 mL (minimum)

Test solution volume: 200 mL (minimum)

Age of test organisms: 3 - 4 days post hatch

Number of organisms/chamber: 5

Number of replicates/concentration: 5 (minimum)

Feeding: Twice daily (20 *Artemia* nauplii/mysid at each feeding); no food on day 7

Aeration: None unless DO < 4.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate
Total weight of survivors divided by the initial count (biomass)
Total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 75% survival and average dry weight ≥ 0.40 mg per surviving mysid

Reference toxicant acceptability criteria: MSD < 40% (survival) and 50 µg (growth)survival and growth NOECs < 100 µg/L in a zinc sulfate reference toxicant test

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C. for more explanation.

East coast mysid survival and growth

Test species: *Americamysis bahia* (formerly *Mysidopsis bahia*)

Approved test method: EPA-821-R-02-014, method 1007.0

Test type: 7-day static-renewal (90% renewal of test solution in each test chamber daily)

Temperature: 26° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Salinity: 30 ± 2‰

Test chamber size: 8 oz plastic disposable cups or 400 mL glass beakers (minimum)

Test solution volume: 150 mL (minimum)

Age of test organisms: 7 days

Number of organisms/chamber: 5

Number of replicates/concentration: 8 (minimum)

Feeding: Twice daily (75 *Artemia* nauplii/mysid at each feeding) with 8 - 12 hours between feedings

Aeration: None unless DO < 4.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate
Total weight of survivors divided by the initial count (biomass)
Total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 80% survival in the control
Average dry weight ≥ 0.20 mg per surviving mysid in the control

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C for more explanation.

Topsmelt survival and growth

Test species: *Atherinops affinis*

Approved test method: EPA/600/R-95/136

Test type: 7-day static-renewal (75% renewal of test solution in each test chamber daily)

Temperature: 20° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Salinity: 30 or 34 ± 2‰

Test chamber size: 600 mL (minimum)

Test solution volume: 200 mL (minimum)

Age of test organisms: 9 - 15 days post-hatch

Number of organisms/chamber: 5

Number of replicates/concentration: 5 (minimum)

Feeding: Twice daily (40 *Artemia* nauplii/fish at each feeding) morning and afternoon; no food on day 7

Aeration: None unless DO < 4.0 mg/L; aerate all chambers with < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate
Total weight of survivors divided by the initial count (biomass)
Total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 80% survival and average dry weight ≥ 0.85 mg/surviving fish

Reference toxicant acceptability criteria:

Copper chloride is the only acceptable reference toxicant. The survival LC50 must be < 205 µg/L Cu. The PMSD must be < 25% for survival and < 50% for biomass. The results should also be used for QC as discussed in the "Reference Toxicant Tests".

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C for more explanation.

Inland silverside survival and growth

Test species: *Menidia beryllina*

Approved test method: EPA-821-R-02-014, method 1006.0

Test type: 7-day static-renewal (80% renewal of test solution in each test chamber daily)

Temperature: 25° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Salinity: 30 ± 2‰

Test chamber size: 600 - 1000 mL

Test solution volume: 500 - 750 mL

Age of test organisms: 7 - 11 days

Number of organisms/chamber: 10 - 15 as long as each test chamber contains the same number and test chamber sizes and test solution volumes toward the larger end of the acceptable range are used for larger numbers of fish

Number of replicates/concentration: 4

Feeding: 0.10 g wet weight *Artemia* nauplii once per day per replicate through day 2; 0.15 g wet weight per replicate on days 3 - 6; no food on day 7

Aeration: None unless DO < 4.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate
Total weight of survivors divided by the initial count (biomass)
Total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 80% survival in the control
Control average dry weight ≥ 0.50 mg per surviving fish

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan count should be blank for each replicate with zero survival. See Appendix C for more explanation.

Supplemental saltwater chronic toxicity tests

The bivalve embryo-larval development test will be placed into permits along with fish and mysid tests when there is a risk of toxicity to sensitive larval life stages of marine organisms. This test is appropriate for discharges to ecosystems of special importance or fragility. The echinoderm development test is an alternative to the bivalve development test and may be more sensitive to polycyclic aromatic hydrocarbons (PAHs).

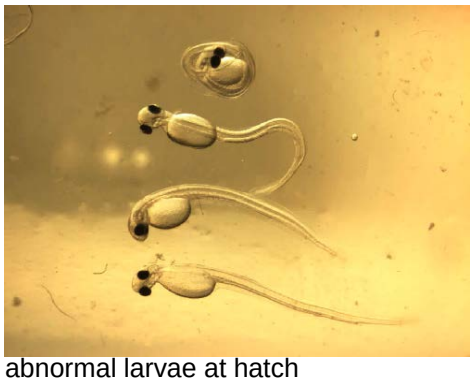
The combination of sensitivity with a very short duration is unique to the echinoderm fertilization test. The test takes only 40 minutes to run. Small volumes of effluent can be tested successfully and one spawning yields enough material for several tests. When an economy of scale is achieved, the fertilization test can be inexpensive.

If shallow and rocky receiving water contains or should contain kelp beds, then the *Macrocystis* germination and growth test might be required.

The Shannon Point Marine Center (SPMC) of Western Washington University has developed three Pacific herring (*Clupea pallasii*) toxicity tests. The three tests are a 16-day embryo development test, a 4-day acute test with yolk sac larvae, and a 7-day larval survival and growth test. These methods do not meet the conditions in WAC 173-205-050(1)(d) and cannot be used for compliance monitoring, but have been required to assess the potential impacts of wastewater on herring early lifestages. The herring tests have also been used in the evaluation of ballast water biocides (See Appendix H.). Copies of the test methods and other reports are available from Dr. Paul Dinnel at SPMC (padinnel@aol.com or 360-293-2188). The reference for the herring test methods is:

Dinnel, P.A., D.P. Middaugh, N.T. Schwarck, H.M. Farren, R.K. Haley, R.A. Hoover, J. Elphick, K. Tobiason, R.R. Marshall. 2011. Methods for Conducting Bioassays Using Embryos and Larvae of Pacific Herring, *Clupea pallasii*. Arch Environ Contam Toxicol 60: 290–308

Normal larvae at hatch



abnormal larvae at hatch

Photos above are from Paul Dinnel.

All conditions in the following tables for the supplemental saltwater chronic toxicity tests must be met and reported for each test.

Bivalve survival and development

Test species: *Crassostrea gigas* or *Mytilus sp.* (*M. trossulus, M. galloprovincialis, M. californianus*)

Approved test method: EPA/600/R-95/136

Test type: Static (nonrenewal)

Temperature: 20° ± 1°C for oysters, 15° or 18° ± 1°C (16° ± 1° if already the lab's standard temperature) for mussels

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Salinity: 30 ± 2‰

Test chamber size: 30 mL

Test solution volume: 10 mL

Age of test organisms: < 4 hours after fertilization

Number of organisms/chamber: 150 - 300

Number of replicates/concentration: 4

Aeration: None in test chambers; the sample may be aerated if DO < 4.0 mg/L

Test duration: 48 hours (up to 54 hours in order to achieve complete development)

Endpoints: Proportion alive and proportion normal

Combined proportion normal/alive is used for comparing dilution water and brine controls.

For more discussion of bivalve development endpoints, see Appendix B.

Test acceptability criteria:

Bivalve development tests will be evaluated for compliance with the following test acceptability criteria rather than the list in item 16 in Table 4 of the EPA manual. The test will be reviewed for compliance with all other conditions and procedures specified in the EPA manual and in Section 13 of ASTM E 724.

A test is acceptable if ≥ 70% of oyster or mussel embryos introduced into the dilution water control grew into live larvae with completely developed shells at the end of the test.

A test is acceptable if the percent minimum significant difference for development is < 25%.

Unless all embryos are counted in each test chamber at the beginning of the test to get a true start count, the estimated initial count is derived from the mean of the counts of at least 6 extra test chambers prepared exactly as the control test chambers using a procedure that randomly distributes their preparation throughout the setting up of all the test chambers.

The coefficient of variation should be ≤ 15% for the embryo counts on the minimum of 6 subsamples taken from the stocking solution at the beginning of the test in order to estimate an initial count. If the 15% coefficient of variation is exceeded, the test report must note this fact and warn to use the test result with caution. Tests will not be rejected solely for exceeding the 15% coefficient of variation. Tests might be rejected if an imprecise initial count results in more than just a few survival proportions > 1 in the test.

A concurrent reference toxicant test must be conducted with each batch of tests.

Echinoderm fertilization

Test species: *Strongylocentrotus purpuratus* or *Dendraster excentricus*

Approved test method: EPA/600/R-95/136

Test type: static (nonrenewal)

Temperature: 12° ± 1°C

Salinity: 30 ± 2‰

Test chamber size: 16 × 100 mm or 16 × 125 mm disposable culture tubes

Test solution volume: 5 mL

Age of test organisms: < 4 hours after collection of gametes

Number of spawners: Gametes pooled from ≤ 4 males and ≤ 4 females (≤ 6 female sand dollars).

Number of organisms/chamber: Approximately 1,120 eggs and ≤ 3,360,000 sperm

Number of replicates/concentration: 4

Aeration: None in test chambers; the sample may be aerated if DO < 4.0 mg/L

Test duration: 40 minutes (20 minutes exposure of sperm; 20 minutes with eggs)

Endpoints: Fertilization of eggs (elevation of the fertilization membrane)

Test acceptability criteria:

- A test is acceptable if ≥ 70% of eggs in the control are fertilized. Control fertilization percentages close to 100% are to be avoided if possible.
- A test is acceptable if the minimum significant difference is < 25%.
- Fertilization at the NOEC must be within 80% of control fertilization.
- A concurrent reference toxicant test must be conducted with each batch of tests.
- Both dilution water and effluent egg blanks should have essentially no eggs with elevated fertilization membranes.
- The density of the final sperm stock must be ≤ 33,600,000/mL and one of these options met:

  Option 1, trial fertilization used - The sperm count for the final sperm stock must not exceed double the target density determined from the fertilization trial test

used to determine the sperm density that will provide about 80% to 100% fertilization without oversperming. 90% to 95% fertilization is the ideal range.

Option 2, sperm/egg ratio kept ≤ 500:1 - confirmation of a sperm stock density of ≤ 5,600,000/mL

Option 3, use any reasonable sperm stock density and run two extra sets of controls (a high and a low density control) - the high density control (0.2 mL sperm stock) must have at least 5% higher fertilization than the low density control (0.05 mL sperm stock).

Echinoderm survival and development

Test species: *Strongylocentrotus purpuratus* or *Dendraster excentricus*

Approved test method: EPA/600/R-95/136

Test type: Static (nonrenewal)

Temperature: 15° ± 1°C

Illumination: Illumination must be for 16 hours at 10 - 20 µE/m^2/s (50 - 100 ft-c) followed by 8 hours of darkness.

Salinity: 30 ± 2‰

Test chamber size: 30 mL

Test solution volume: 10 mL

Age of test organisms: ≤ 1 hour after fertilization

Number of organisms/chamber: Approximately 250 fertilized eggs in 0.25 mL of egg solution

Number of replicates/concentration: 4

Aeration: None in test chambers; the sample may be aerated if DO < 4.0 mg/L

Test duration: 48 to 96 hours and ended at the point when at least 80% of controls have reached the pluteus stage.

Endpoints:

- Proportion alive and proportion normal
- Combined proportion normal/alive is used for comparing dilution water and brine controls.
- The endpoint of the echinoderm development test should be the same as the endpoint for the bivalve development test. For a discussion of the calculation of the bivalve development endpoint, see Appendix B.

Test acceptability criteria:

- A test is acceptable if ≥ 80% of larvae in the control have developed normally.
- A test is acceptable if the percent minimum significant difference for development is < 25%.
- Unless all embryos are counted in each test chamber at the beginning of the test to get a true start count, the estimated initial count is derived from the mean of the counts of at least 6 extra test chambers prepared exactly as the control test chambers using a

procedure that randomly distributes their preparation throughout the setting up of all the test chambers.

- The coefficient of variation should be ≤ 15% for the embryo counts on the minimum of 6 subsamples taken from the stocking solution at the beginning of the test in order to estimate an initial count. If the 15% coefficient of variation is exceeded, the test report must note this fact and warn to use the test result with caution. Tests will not be rejected solely for exceeding the 15% coefficient of variation. Tests might be rejected if an imprecise initial count results in more than just a few proportions > 1 in the test.
- A concurrent reference toxicant test must be conducted with each batch of tests.

Sand dollar (*Dendraster excentricus*) courtesy of
King County, Dept. of Natural Resources and Parks
http://green2.kingcounty.gov/marine/Photo/Individual/61/186?photoId=89

Macrocystis germination and growth

Test species: *Macrocystis pyrifera*

Approved test method: EPA/600/R-95/136

Test type: Static (nonrenewal)

Temperature: 15° ± 1°C

Illumination: Illumination must be for 16 hours at 50 ± 10 μE/m^2/s equally distributed over all test chambers followed by 8 hours of darkness.

Salinity: 34 ± 2‰

Test chamber size: 600 mL

Test solution volume: 200 mL

Age of test organisms: < 2.5 hours after sporophylls begin releasing zoospores

Number of organisms/chamber: 7,500 zoospores/mL

Number of replicates/concentration: 5

Aeration: None unless DO < 4.0 mg/L; aerate all chambers and use < 100 bubbles/minute.

Test duration: 48 hours

Endpoints: Percent of zoospores with germination tubes at least one spore diameter in length

Average length of 10 germination tubes randomly selected from each test chamber

Test acceptability criteria: ≥ 70% germination of zoospores in the control
≥ 10 μm average germ tube length in the control

Reference toxicant acceptability criteria: NOEC is < 35 μg/L Cu in a concurrent copper chloride reference toxicant test.

The MSD is < 20% relative to the control for both germination and germ tube length in the copper chloride reference toxicant test.

Pacific herring embryo viability

Test species: *Clupea pallasii*

Test method: Dinnel, P.A., D.P. Middaugh, N.T. Schwarck, H.M. Farren, R.K. Haley, R.A. Hoover, J. Elphick, K. Tobiason, R.R. Marshall. 2011. Methods for Conducting Bioassays Using Embryos and Larvae of Pacific Herring, *Clupea pallasii*. Arch Environ Contam Toxicol 60: 290–308

Test type: 16-day static-renewal (100% renewal of test solution in each test chamber daily with renewals ceasing on day 8)

Organism source: In-lab fertilization using gonads excised from wild-caught running ripe herring (available January to June)

Temperature: 12° ± 1°C

Illumination: Dim light (about 20-50 lux or 2-5 foot candles)

Salinity: 16 ± 2‰

Test chamber size: Glass Petri dishes, 100 mm diameter x 15 mm depth

Test solution volume: 30 mL

Age of test organisms: Maximum 60 minutes after fertilization

Number of organisms/chamber: About 20

Number of replicates/concentration: 4 (minimum)

Feeding: None

Aeration: None

Test duration: 16 days or when all live eggs have hatched, whichever occurs first
The duration of exposure to may vary. Exposure to effluent would be 7 days and environmental samples would be for the full test duration.

Endpoints: Normal Survival – number of normal hatchlings / number of beginning eggs
Heart Rate (optional) – Average heart rate for a random selection of embryos at day 7 or 8. Heart rate dose-response might be an indication of toxicant identity; increasing KCl concentrations decrease heart rate and increasing SDS concentrations increase heart rate. Contact Paul Dinnel for a report.

Control performance criteria: ≥ 70% embryo Normal Survival

Pacific herring 7-day survival and growth

Test species: *Clupea pallasii*

Test method: Dinnel, P.A., D.P. Middaugh, N.T. Schwarck, H.M. Farren, R.K. Haley, R.A. Hoover, J. Elphick, K. Tobiason, R.R. Marshall. 2011. Methods for Conducting Bioassays Using Embryos and Larvae of Pacific Herring, *Clupea pallasii*. Arch Environ Contam Toxicol 60: 290–308

Test type: 7-day static-renewal (75% renewal of test solution in each test chamber daily)

Organism source: In-lab hatching of natural spawn or in-lab fertilization using gonads excised from wild-caught running ripe herring (available January to June)

Temperature: 12° ± 1°C

Illumination: Illumination should be for 16 hours at 600 – 1,200 lux (about 60 - 120 foot candles) followed by 8 hours of darkness

Salinity: 30 ± 2‰

Test chamber size: 400 mL (minimum) Glass beakers are preferred.

Test solution volume: 200 mL (minimum)

Age of test organisms: 6-8 days post-hatch and feeding successfully

Number of organisms/chamber: 15-20 initially, culled to 10 healthy feeding larvae at day 1

Number of replicates/concentration: 4 (minimum)

Feeding 0.5 mL brine shrimp slurry (2,000 nauplii) of < 24-hour old *Artemia* at test initiation; 1,000 nauplii on Days 1 through 6; none on last day

Aeration: None unless DO < 5.0 mg/L; aerate all chambers and use < 100 bubbles/minute

Test duration: 7 days

Endpoints: Survival rate; total weight of survivors divided by the initial count (biomass); total weight of survivors divided by the final count (weight)

Control performance criteria: ≥ 80% average survival and average dry weight must be at least 1.3 times the average initial dry weight

Data entry: Because biomass can be zero, total weight equals tare weight for each replicate with zero survival. Because division by zero is undefined, the pan

count should be blank for each replicate with zero survival. See Appendix C for more explanation.

This page is purposely left blank

Appendix A: Rainbow trout age discussion

Fish age criteria standardize toxicity testing to occur at a sensitive stage of the fish's life cycle. We were concerned that the age of rainbow trout has been determined differently from lab to lab because the point in the fish's life cycle representing day 1 is not always the same.

The EPA protocol for the acute rainbow trout test sets an age requirement for the fish of 15 to 30 days old. There has been some uncertainty, however, at what point in the life cycle is day 1. This issue was researched through consultations with fish biologists, labs, and EPA. Little agreement exists about the upper end of the sensitive age range for rainbow trout testing, and many believe that EPA might be too restrictive on the upper age. There is general agreement, however, that testing should not begin until after the yolk sac is completely absorbed and the fish are actively feeding. Swim-up is believed to be the least ambiguous event to use in timing the readiness of trout for testing.

In accordance with the findings of these consultations, Ecology intends to evaluate rainbow trout acute test fish age criteria as follows:

- Ecology will enforce the EPA age range of 15 to 30 days old. Fish age will be determined using swim-up as day 1. Labs must express rainbow trout age as days after swim-up.

- The fish should be held at 12 ± 1°C after reaching the swim-up life stage. This ensures that fish age and condition are consistent.

The test fish should be the same age and from the same source. Because of individual development rate variation, test fish will be considered to be at a stage in their life cycle when 80% of the batch have achieved that stage. Rainbow trout development is temperature dependent. 12°C is the preferred rearing temperature, but trout may be held at a lower temperature prior to swim-up.

The life cycle stage definitions are:

Hatch: When the fish (alevins) have broken out of the egg casing, but are inactive, remain mostly on the bottom, do not feed, and live off the attached yolk sac.

Swim-up: Around 3 weeks from hatch, the fish emerge from the relatively inactive bottom dwelling stage and actively move up and remain in the upper water column. The fish have begun feeding but still have some yolk sac.

Appendix B: Bivalve development test endpoint discussion

The EPA 1995 bivalve test contains an adjusted combined normal/alive proportion calculation where the # normal for each replicate is divided by the larger of the initial or final count. Because the initial count is based on a mean of embryo counts from subsamples of the stocking solution, the final count or the number normal for some replicates can exceed the initial count. The EPA adjustment avoids the generation of proportions greater than 1 but is also an attempt to increase test sensitivity. Because 1 is the highest meaningful proportion (all test organisms have lived or developed normally), it is customary to consider a proportion greater than 1 to be equal to 1. The adjustment is unnecessary to increase test sensitivity. The bivalve development test is very sensitive and the adjusted combined endpoint does little to increase its sensitivity.

The adjusted combined endpoint calculation incorporates a bias toward lower combined proportion normal results achieved by arbitrarily dividing by the larger of the two available numbers. If the final count is greater than the initial count, it is assumed to be due to subsampling error in determining the initial count and the final count is used in the denominator. However, the same type of subsampling error can also cause initial counts to be greater than final counts. Picking the initial count to use in the denominator when it is larger than the final count implies that the difference between initial and final counts is due to lower final survival even though it might be simple subsampling error. This situation may also violate the independence of observation assumption required for valid parametric hypothesis testing.

The simplest approach is to use separate proportion alive and proportion normal endpoint calculations. The database will then contain information on which effluents affect development more than survival and which effluents do the opposite. Both patterns have been seen and the information might someday be useful or enlightening. Combined endpoints obscure such differences and are often no more sensitive.

Data indicate that mussel controls perform as well as oyster controls so the control performance acceptability criterion is now the same (≥ 70%normal survival) for both oysters and mussels.

Unless all embryos are counted in each test chamber at the beginning of the test to get a true start count, an estimated initial count is derived from the mean of the counts from at least 6 extra test chambers prepared exactly as the control test chambers at random times throughout the setting up of the other test chambers. These extra chambers will be used at the beginning of the test to estimate an initial count and assess pipetting and counting technique. The coefficient of variation must be ≤ 15% for the embryo counts on these subsamples. If the 15% coefficient of variation is exceeded, the test report must warn to use the test result with caution. Tests will not be rejected solely for exceeding the 15% coefficient of variation.

Terminology

Initial count = the mean of a minimum of 6 subsamples taken from the stocking solution
Normal = number of larvae at the end of the test with completely developed shells (See the test method for a more complete description.)
Abnormal = number of larvae at the end of the test with incompletely developed shells

Final count = # normal + # abnormal
Proportion alive = final count ÷ initial count
Proportion normal = # completely developed ÷ final count
Combined proportion normal/alive = # completely developed ÷ initial count

Appendix C: Combined survival and growth endpoint discussion

EPA changed the growth calculation for the 7-day survival and growth tests. Instead of dividing the final weight by the number of surviving organisms at test end, EPA manuals now instruct to divide by the number of organisms at test initiation. The new endpoint calculation results in a combined survival and growth number and is usually called "biomass", even though it is still expressed as mg per fish or mysid as if it was a growth measurement.

When all test organisms survive, the original growth calculation and the biomass calculation will give the same result. If any concentration has some mortalities, the biomass calculation will increase the magnitude of the adverse effect relative to the first growth calculation. It usually increases variability across replicates as well. Increased variability decreases statistical sensitivity and results in about equal sensitivity for the growth and biomass endpoints. If the control has mortalities, its biomass number will be reduced and test concentrations will often show a smaller apparent biomass reduction than they would using the first growth calculation. The virtues and vices of the biomass calculation tend to cancel one another out resulting in little difference in test results overall. Published EPA data show no increased test sensitivity from the biomass endpoint. Department of Ecology data on the 7-day survival and growth tests using three different species of test organisms also show no increased sensitivity from the biomass endpoint and an increased tendency toward anomalous tests as described in Appendix D.

We accept the increased test variability with the combined endpoint. However, when sporadic mortalities occur, the variability becomes unacceptable. Tests that have a standard deviation for proportion alive above 0.2 in any concentration (unless it occurs near the threshold of a good concentration-response relationship) will be analyzed for the original growth endpoint.

We will also switch back to the original growth endpoint when the biomass endpoint results in an anomalous concentration-response relationship which would cause test rejection in accordance with Appendix D and the original growth endpoint produces a good concentration-response relationship in the same test.

Zero weights make no sense for the original growth calculation (weight/final count) since zero weights can happen only if everything died and division by zero is undefined. Zero weights are ecologically meaningful and mathematically necessary for biomass calculations. Biomass calculations will be inaccurate if containers with complete mortality are not included. Biomass IC_{25} point estimates are especially erroneous if containers with zero weights are not included. The correct data entry and calculation instructions are included in this document in the list of specific test conditions for every survival and growth test method.

Reference:

Pickering, Q., J. Lazorchak and K. Winks. 1996. Subchronic sensitivity of one-, four-, and seven-day old fathead minnow (*Pimephales promelas*) larvae to five toxicants. *Environmental Toxicology and Chemistry.* 15:353-359.)

Appendix D: Identifying anomalous WET tests

Introduction

WAC 173-205-070(5)(c) states that anomalous WET test results will be identified and not used for compliance determinations. WAC 173-205-090(1)(d) describes the process for a permittee to notify Ecology that noncompliance with a WET limit may have been caused by an anomalous WET test result and avoid the expense of unnecessary extra WET testing. The notification must include the reason for considering the test result to be anomalous. If Ecology agrees with the permittee's reason for considering the test result to be anomalous, the additional monitoring required by WAC 173-205-090(1) will be avoided. A list of criteria at the end of these guidelines contains the considerations that Ecology will use in deciding if WET test results are anomalous.

Text of WAC 173-205-090(1)(d)

WAC 173-205-090(1)(d) If the permittee believes that the compliance test failure will be identified by the Department (Ecology) as an anomalous test result in accordance with WAC 173-205-070(5)(c), the permittee may send the Department notification with the compliance test result that the compliance test result might be anomalous and that the permittee intends to take only one additional sample for toxicity testing and wait for notification from the Department before completing the additional monitoring required in this subsection.

(i) The notification must identify the reason for considering the compliance test result to be anomalous.
(ii) The permittee shall take the additional sample and retest as soon as possible after receiving the compliance test result.
(iii) The additional test result shall replace the compliance test result upon determination by the Department that the compliance test result was anomalous.
(iv) The permittee shall complete all of the additional monitoring required by this subsection as soon as possible after notification by the Department that the compliance test result was not anomalous.
(v) If the additional sample fails the compliance test, then the permittee shall proceed without delay to complete all of the additional monitoring required by this subsection.

The difference between invalid tests and anomalous test results

Invalid WET tests occur when the lab does not follow the test protocol or when the results do not meet the test acceptability criteria in the test protocol. Permittees and labs are obligated to look for invalid tests because the permit requires that the test protocol be followed. Permittees and labs only need to identify potentially anomalous test results within the context of WAC 173-205-090(1)(d).

Anomalous test results happen when the lab appears to have conducted the WET test in accordance with the test protocol, but the results are considered unreliable according to anomalous test identification criteria. There is no requirement for permittees or labs to attempt an identification of anomalous WET test results. All valid WET test results must be submitted whether the test is regarded as anomalous or not. A notification of an anomalous test result does not by itself imply any fault on the part of the permittee or lab, but frequent anomalous tests can be an indication of poor lab technique or poor condition of test organisms.

The anomalous test identification criteria are a common sense approach to making WET test results fair and enforceable. The anomalous test criteria will be used during test review to intervene with human judgment when statistics seem to be reaching the wrong conclusion about effluent toxicity. Their underlying principle is the definition of the NOEC as the highest effluent concentration showing no statistically significant difference from the control, along with an expectation for a concentration-response relationship typical for toxicity under the conditions of the test. Finney (1978) noted that the adverse effect increases with dose for almost all useful assays. He also noted that non-monotonic dose-response relationships do occur and are useful for the information that they provide. Our criterion for a meaningful non-monotonic concentration-response relationship is that it happens more than once with the same effluent. Our database gives us the ability to look for these.

Reference:

Finney, David J. 1978. *Statistical Method in Biological Assay*. Third Edition. Charles Griffin & Company, London. 43 pp.

The permittee must resample and conduct another WET test after being notified by Ecology of rejection of an anomalous test result. However, identifying a test as anomalous does not necessarily mean rejection of the test and a requirement to repeat. If a test result meets one of the criteria for anomalous test identification but has no statistically significant toxicity at concentrations of regulatory concern (ACEC or CCEC), then the test need not be repeated unless other factors contribute to a decision to reject the test.

The main purpose for conducting WET tests with at least five effluent concentrations in a series is to allow concentration-response to be evaluated and anomalous tests discarded. The identification of anomalous tests is a valuable tool for reducing false positives. A concentration-response relationship where response increases with concentration is a good identifier of toxicity as opposed to other sources of organism stress such as disease. Lab error, test organism disease, or variability in test organism response will rarely produce a good concentration-response relationship.

Different toxicity tests have different expectations for a good concentration-response relationship. The proportional (quantal) endpoints (survival, fertilization, development) have steeper concentration-response relationships than do the non-proportional (non-quantal) endpoints such as growth or neonate production. Tests with both survival and a sub-lethal endpoint sometimes have stepwise effects thresholds: a sub-lethal effect threshold followed by a survival effect threshold at a higher concentration. The anomalous test definitions must be considered in light of the expectations for the different toxicity tests and endpoints.

A suspected anomalous concentration-response relationship will be compared to past results from the same discharge in order to ensure that it has not occurred before and might therefore be meaningful.

Notification of an anomalous test result

A permittee benefits from notifying Ecology of an anomalous test result only when there is noncompliance with a WET limit. The notification allows the permittee to delay additional monitoring while Ecology evaluates the test result. The notification will also help Ecology determine sooner that the test result is anomalous and does not represent a WET limit violation. Using the following criteria, a lab can inform a permittee of a likely anomalous WET test result that resulted in noncompliance with a WET limit. A permittee can then send Ecology notification with the compliance test report that the test result might be anomalous and that the permittee intends to conduct only one additional WET monitoring test. If the additional sample fails to comply with the WET limit or Ecology determines the first test result to not be anomalous, then the permittee must proceed without delay to complete all of the additional monitoring. Otherwise, the permittee is not required to conduct the rest of the additional monitoring. The additional test result replaces the compliance test result upon determination that the compliance test result was anomalous.

Anomalous test results criteria

1. A WET test result is anomalous if it shows a statistically significant difference in response between the control and the ACEC or CCEC, but no statistically significant difference in response at one or more higher effluent concentrations. The lack of statistical significance must be associated with a lower adverse effect at the higher effluent concentration.
2. A WET test is anomalous if there is a statistically significant difference in response between the control and the ACEC or CCEC which together with nearby concentrations of effluent have a zero slope and appear to be nontoxic (performance is typical of healthy test organisms). Sometimes a test meeting this criterion has a control that seems to not belong to the concentration-response relationship because of exceptionally good performance.
3. A WET test is anomalous if the standard deviation for proportion alive equals or exceeds 0.3 at the ACEC or the CCEC unless the partial mortality occurs close to the threshold in a good concentration-response relationship.

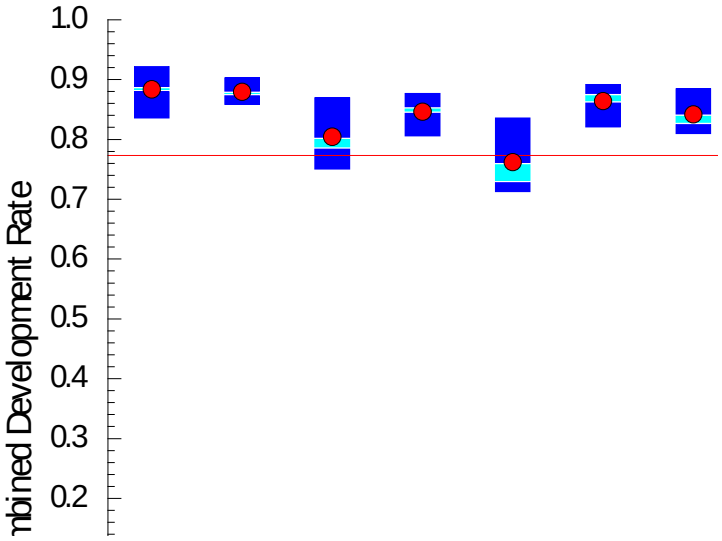
Reactions to anomalous concentration-response other than test rejection

- Change *alpha* from 0.05 to 0.01 when the significant difference in an acute test is less than 10% or the significant difference in a chronic test is less than 20%.
- Use the higher LOEC/NOEC pair unless statistical significance at the ACEC or CCEC force rejection of a test performed for WET limit compliance monitoring.
- Examine past tests performed on the same discharge and accept the test result if similar non-monotonic concentration-response relationships are found.
- Do comparisons based on raw counts of normal bivalve larvae.
- Switch to a combined proportion normal (# normal/initial count) when final counts are larger than initial counts more frequently in effluent concentrations than in controls. Another description is a test result having large numbers of “abnormal larvae” in all effluent concentrations and a flat concentration-response for proportion normal. The excess “abnormal larvae” are most likely misidentified particles from the sample.
- Do comparisons based on raw counts when most proportions are 1 (or > 1).

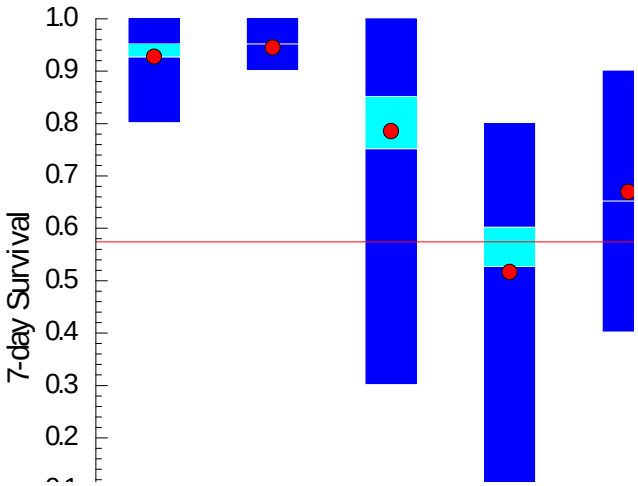
- Check for effects due to parentage and exclude from analysis all replicates containing offspring performing differently from those having other parents.
- Exclude statistically confirmed outliers or use a nonparametric analysis when exceptionally high or low responses in a few test chambers are creating the anomalous concentration-response relationship.
- Switch from the biomass (weight/initial count) to the original growth (weight/pan count) endpoint calculation.

Example Anomalous Concentration-Responses

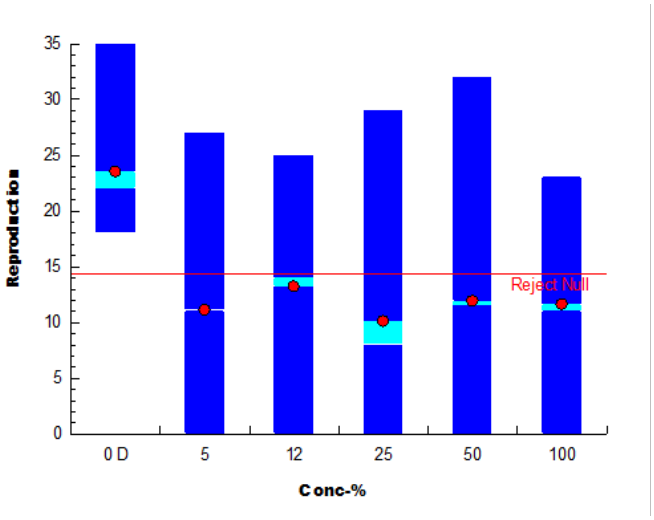
Example Test for Anomalous Test Criterion 1 (also Criterion 2)
Bivalve development test on an industrial effluent



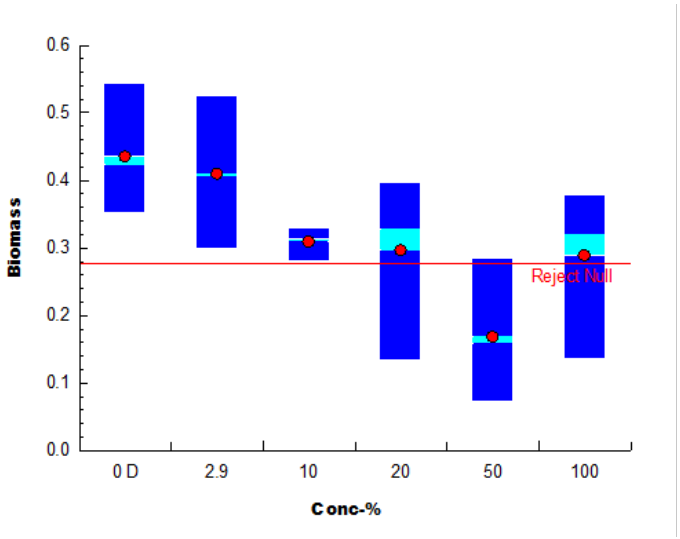
Example Test for Anomalous Test Criterion 3 (also Criterion 1)
Fathead minnow chronic test on an industrial effluent



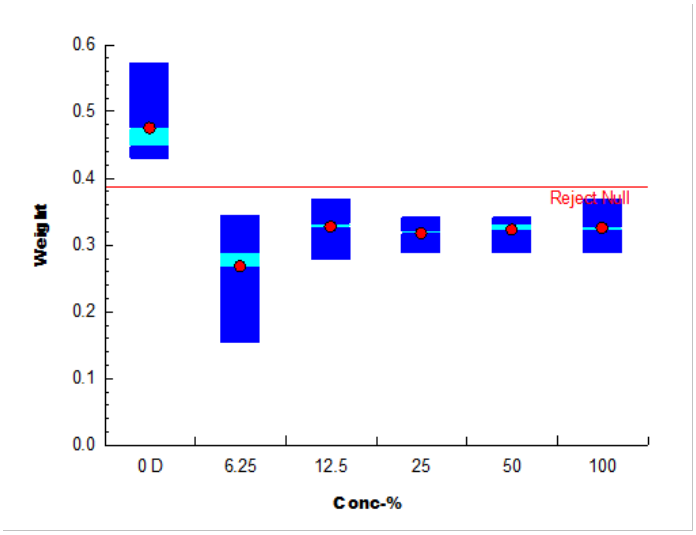
Example Test for Anomalous Test Criterion 2
Ceriodaphnia chronic test on an industrial effluent



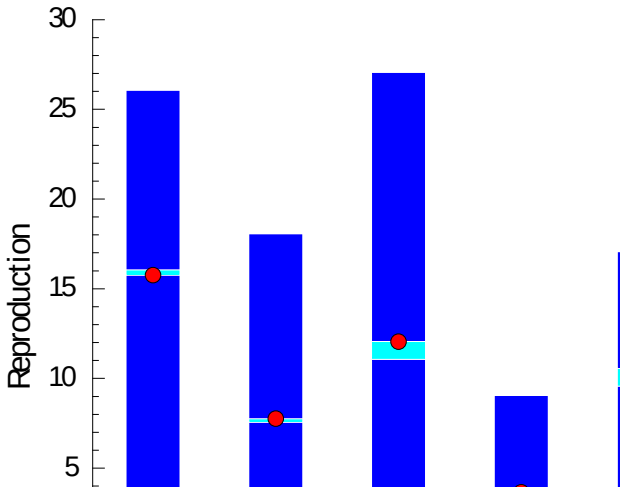
Example Test for Anomalous Test Criterion 1
Fathead minnow chronic test on an industrial effluent



Example Test for Anomalous Test Criterion 2
Fathead minnow chronic test on a POTW effluent



Example Test for Anomalous Test Criterion 1
Ceriodaphnia chronic test on an industrial effluent



Appendix E: Example calculations for the power standards

ACEC	Fathead minnow- number surviving				
	replicate 1	replicate 2	replicate 3	replicate 4	mean of replicates
25% effluent	6	4	8	7	6.25
Control	Fathead minnow- number surviving				
	replicate 1	replicate 2	replicate 3	replicate 4	mean of replicates
lab water	9	10	9	9	9.25

1. Subtract the mean survival across the replicates in the ACEC from the mean survival across the replicates in the control. A negative result counts as passing.

$$9.25 - 6.25 = 3.00$$

2. Divide this difference between the mean survivals by the mean survival across the control replicates.

$$3.00 \div 9.25 = 0.32$$

3. Multiply the result by 100 and express as a percent difference in survival.

$$0.32 \times 100 = 32\% \text{ difference in response}$$

4. If the percent difference in survival is ≤ 29%, then the WET test has met the power standard.

$$\text{The 32\% difference in response is} > 29\%$$

The WET test has not met the power standard and must be repeated. (Assuming that the WET test did not violate the WET limit; the power standards are not an issue for WET tests that violate WET limits.)

The percent minimum statistical difference (PMSD) from statistical comparison of effluent concentration means versus control mean may be used instead of the example calculation.

CCEC	Fathead minnow- average weight/larva (mg)				
	replicate 1	replicate 2	replicate 3	replicate 4	mean of replicates
5% effluent	0.529	0.554	0.425	0.373	0.470
Control	Fathead minnow- average weight/larva (mg)				
	replicate 1	replicate 2	replicate 3	replicate 4	mean of replicates
lab water	0.560	0.636	0.613	0.452	0.565

1. Subtract the mean of the responses across the replicates in the CCEC from the mean of the responses across the replicates in the control. A negative result counts as passing.

$$0.565 - 0.470 = 0.095$$

2. Divide this difference between the mean responses by the mean response across the control replicates.

$$0.095 \div 0.565 = 0.168$$

3. Multiply the result by 100 and express the product as a percent difference in response.

$$0.168 \times 100 = 16.8\% \text{ difference in response}$$

4. If the percent difference in response is ≤ 39%, then the WET test has met the power standard.

A 16.8% difference in response is < 39%; the WET test has met the power standard.

The percent minimum statistical difference (PMSD) from a statistical comparison of effluent concentration means versus control mean may be used instead of the example calculation.

Appendix F: Rapid screening tests and species

Acute rapid screening tests

Rapid screening tests for acute toxicity are expected to have a maximum mortality proportion of 0.20 in 100% effluent. The mortality proportion is calculated by subtracting the number of test organisms living in 100% effluent at the end of the test from the number of test organisms living in the control and dividing the result by the number of test organisms living in the control (Abbott's correction). The 100% effluent test concentration and the control must have equal numbers of test organisms.

Rotifer

The rotifer (*Brachionus sp.*) method is ASTM E 1440-91. The test is a 24-hr acute test using rotifers hatched from cysts. Tests with organisms hatched from cysts are less expensive because no time or materials are consumed by maintaining a culture. The rotifer acute test can be used in freshwater or saltwater although the details of the saltwater version are not yet established in this document.

24-hour EPA acute screening tests

The 24-hour EPA acute tests are conducted using the same EPA manual and species that were used for effluent characterization. The same reference toxicant tests as for the 48-hour and 96-hour EPA acute toxicity tests may be used for the 24-hour rapid screening test versions.

Chronic rapid screening tests

Bacterial Bioluminescence Test (*Standard Methods* 8050)

Rotifer chronic test

The rotifer test measures the intrinsic rate of population increase. Measuring the intrinsic rate of population increase simultaneously evaluates both mortality and fecundity. Because it starts with rotifer cysts, uses small volumes of effluent, and takes only two days, it should be less expensive than EPA chronic tests. The chronic rotifer test method is Standard Methods – 8420 along with:

Snell, Terry W. 1992. A 2-d Life Cycle Test with the Rotifer *Brachionus calyciflorus*. *Environ. Toxicol. Chem.* 11:1249-1257.

Echinoderm fertilization test

The echinoderm fertilization rapid screening test method is: EPA/600/R-95/136.

Brachionus acute toxicity test

Test species: *Brachionus calyciflorus*

Approved test method: ASTM, E 1440 - 91

Test type: Static (nonrenewal)

Temperature: 25° ± 1°C

Illumination: None

Test chamber size: 2.5 mL

Test solution volume: 1.0 mL

Age of test organisms: < 2 hours post-hatch

Number of organisms/chamber: 5

Number of replicates/concentration: 4 (minimum)

Number of concentrations: 5 (plus a control)

Feeding: None

Aeration: None

Test duration: 24 hours

Endpoints: Mortality

Control performance criterion: ≥ 90% survival

Reference toxicant acceptability criteria: A coefficient of variation < 0.85 for last 20 copper sulfate reference toxicant tests

Brachionus 2-day chronic reproductive test

Test species: *Brachionus calyciflorus*

Approved test method: Standard Methods - 8420

Test type: Static (nonrenewal)

Temperature: 25° ± 1°C

Illumination: None

Test chamber size: 0.5 mL to 2.0 mL

Test solution volume: 0.3 mL to 1.0 mL

Age of test organisms: < 2 hours post-hatch

Number of organisms/chamber: 1

Number of replicates/concentration: 8 (minimum)

Number of concentrations: 5 (plus a control)

Feeding: No pretest feeding. Provide 1.0×10^6 *Pseudokirchneriella subcapitatum* cells/rotifer in the test solutions at test initiation (before introduction of rotifers).

Aeration: None

Test duration: When $r \geq 0.7$ at 48 hours or 50 hours

Endpoints: R, the intrinsic rate of population increase

Because r is calculated using the natural logarithm of the final count, it cannot be used if no rotifers survive. To overcome this disadvantage, the raw final counts will also be used in test statistics just as the neonate totals are for the *C. dubia* chronic test.

Control performance criterion: $r \geq 0.7$

Reference toxicant acceptability criteria: A coefficient of variation < 0.85 for last 20 potassium dichromate reference toxicant tests

Appendix G: Evaluation of proposed stormwater treatment chemicals

Introduction

Being chemically active, treatment chemicals can be toxic to aquatic organisms. Aquatic toxicity will make some treatment chemicals unsafe to use while other chemicals may be safe. A simple measurement of toxicity will not be much help in making the risk assessment needed to determine treatment chemical acceptability. Relative toxicity is not an effective tool since a less toxic chemical may be less safe than a more toxic one in actual use. Toxicity testing of proposed treatment chemicals should be done so that the results can be related to intended use conditions and potential environmental exposures. Approval can then be based upon the likelihood of environmental exposure at an adverse concentration rather than just on relative toxicity.

Toxicity testing of proposed treatment chemicals

The purpose of treatment chemical toxicity testing is to determine the safety margin. The safety margin is the difference between the threshold of toxicity to the most sensitive species tested and the discharge concentration. Many of the better treatment chemicals have a toxic threshold above the intended use concentration (the dose to the treatment system). The difference between the toxic threshold and intended use concentration should be considered the safety margin for the better chemicals because it shows an extra degree of safety over comparing the threshold to the discharge concentration.

The concentration series for each toxicity test must include the intended use and discharge concentrations (estimated or measured) and the toxic threshold. A range-finding test may be necessary to estimate the toxic threshold before performing a definitive toxicity test to determine the safety margin. If the range-finding tests clearly identify a most sensitive species, then definitive testing may focus solely on that test species. If the range-finding tests show that an adequate safety margin can be determined from testing in clear water, then testing should be done on the treatment chemical dosed to clear water as another extra safety measure.

If removal by the treatment process is needed to bring a treatment chemical's concentration below its toxic threshold, or if the treatment chemical concentration at discharge is unknown, the safety needs to be determined by testing the toxicity of turbid water treated with the chemical in a series of concentrations representing a range of possible intended use concentrations. In order to better define the reaction between dose and turbidity, separate tests will need to be done at typically low (50 NTU), medium (200 NTU), and high (500 NTU) initial turbidities using a range of intended use concentrations appropriate for each turbidity. Turbidity must be created using soil samples from the proposed or a typical treatment site. Before conducting the bench scale tests on turbid water, toxicity testing of the treatment chemical diluted with lab water may be used to find the most sensitive species for use in determining the safety margin.

Point estimation methods are best for determining toxic thresholds because they can interpolate between concentrations and avoid overestimating the safety margin by using the LOEC or underestimating the safety margin by using the NOEC. The 25% effect level is used by EPA as an estimate of the toxic threshold. The IC_{25} should be used for growth or reproduction and the

EC_{25} used for the survival or development endpoints. If an EC_{25} cannot be calculated, then the MATC (geometric mean of the NOEC and LOEC) should be used or the test should be repeated with more concentrations around the anticipated toxic threshold.

Species and toxicity tests for proposed treatment chemicals

Treatment Chemical Toxicity Test Choices for Different Discharge Circumstances

receiving water condition	toxicity test	method
salmonid or other fish passage	Rainbow Trout 96-hour Acute	EPA - 2019.0
	Fathead Minnow 96-hour Acute	EPA - 2000.0
juvenile salmonid or other fish rearing or habitat	Rainbow Trout 7-day Survival & Growth	USEPA - NERL SOP
	Fathead Minnow 7-day Survival & Growth	EPA - 1000.0
	Daphnid 48-hour Acute	EPA 2002.0 or 2021.0
salmonid or other fish spawning	Rainbow Trout Embryo	EPS 1/RM/28
	Fathead Minnow Embryo-Larval Survival & Teratogenicity	EPA - 1001.0
lake	*Ceriodaphnia dubia* Survival and Reproduction	EPA - 1002.0
	alternate - *Americamysis bahia* 7-day Survival & Growth	EPA - 1006.0
marine water	Topsmelt 7-day Survival & Growth	EPA/600/R-95-136
	Americamysis bahia 7-day Survival & Growth	EPA - 1006.0
sensitive marine habitat	Bivalve Embryo-Larval Survival & Development	EPA/600/R-95-136

A proposed treatment chemical might be tested solely with the two 96-hour fish acute toxicity tests, but then it would be considered for use in only a few circumstances such as highly channelized urban streams or drainage ditches. In order to be considered for even moderately widespread use, a proposed treatment chemical must be tested at a minimum with the daphnid 48-hour acute test, the rainbow trout 7-day survival and growth test, and the fathead minnow 7-day survival and growth test. (The rainbow trout and fathead minnow 96-hour acute tests would be less sensitive and do not need to be run if the 7-day fish tests are used.) If the proposed treatment chemical is to be considered for use where fish might spawn, then the two fish embryo toxicity tests must also be run. If the proposed treatment chemical is to be considered for use where it could be discharged to a lake, then the *Ceriodaphnia dubia* survival and reproduction test should be run in addition to whichever fish tests apply. The *Americamysis bahia* 7-day survival and growth test is an acceptable alternate to *Ceriodaphnia* if the food needed by the test organisms during the test would be removed by a treatment chemical and confuse the results. If a treatment chemical is intended for use where it could be discharged to typical marine waters, then the topsmelt 7-day survival and growth test and the *Americamysis bahia* 7-day survival and growth test must be run. If the treatment chemical is intended to be discharged near especially important marine waters (Marine sanctuaries, National parks, National Wildlife refuges, National Wilderness areas, Outstanding National Resource waters, or Washington Department of Natural Resources aquatic reserves), then the bivalve embryo-larval survival and development test must be run along with the topsmelt and mysid 7-day tests.

Intended use plan

Once the toxicity testing has been done and the safety margin or the appropriate dosing relative to turbidity has been determined, then a plan must be written describing how the treatment chemical will be used in a treatment system in order to be both effective and safe. The plan must address how treatment system monitoring will ensure proper dosing and maintain the safety

margin during routine use. Treatment chemicals used for suspended solids removal are also removed during the process by binding to the solids. If the treatment chemical is properly dosed, there will be little or no treatment chemical carry-over, but the method for maintaining proper dosing as suspended solids levels vary must be described. In addition to routine use, the plan must address over-dose prevention and spill control. Approval will be based on the confidence generated by the plan that the safety margin for the treatment chemical will always be maintained. The approved intended use plan will also serve as the basis for treatment system inspections either by the operator or agency inspectors.

If the toxic threshold is above the intended use concentration (the dose to the treatment system), then monitoring of the chemical dose relative to the treatment system discharge volume should be adequate to demonstrate maintenance of the safety margin. If the toxic threshold is below the intended use concentration, then removal of the treatment chemical by binding to solids needs to be considered in determining conditions for achieving a safety margin at different solids loadings. If the concentration of the treatment chemical can be reliably measured in the discharge, then regular monitoring of the discharge concentration is an effective method for demonstrating safe conditions for using treatment chemicals with toxic thresholds below the intended use concentration.

Chitosan acetate evaluation

The complete suite of toxicity tests has been performed on six treatment chemical formulations relying on chitosan acetate to reduce suspended solids in stormwater. In each case, the most sensitive endpoint was either trout survival (96-hour or 7-day) or mysid 7-day growth. The lowest trout survival thresholds (LC25) have all been just above 1 mg/L chitosan acetate and the mysid biomass thresholds (IC25) have all been just below 1 mg/L chitosan acetate. The intended discharge concentration identified in the intended use plans for the treatment chemical formulations is a maximum of 0.2 mg/L chitosan acetate. Approvals require regular monitoring of the chitosan acetate discharge concentration at the treatment plant.

The entire suite of toxicity tests no longer needs to be done on any treatment chemical formulation with chitosan acetate as its sole active ingredient because the most sensitive endpoints for chitosan acetate have been established. In addition, toxic thresholds for these endpoints are at least 4 times the maximum discharge concentration, and the chitosan acetate discharge concentration can be routinely measured at the treatment plant. Intended use plans for discharges to an urban stream only need to include results from 96-hour rainbow trout acute survival testing. Intended use plans for discharge to all freshwater streams only need results from rainbow trout 7-day survival and growth testing. Adding results from the mysid 7-day survival and growth test will expand the scope to include lakes and ordinary marine waters. Adding results from bivalve embryo-larval survival and development testing will expand the scope to include especially important marine waters.

Cationic polymers have an affinity for fish gills. Chitosan is cationic. The mortality threshold in rainbow trout tests is sharp, indicating the potential for fish kills if chitosan acetate is not properly dosed or is spilled into surface water.

Anionic PAM evaluation

In the November 28, 2008 Federal Register (FR) announcement of the proposed rule on Effluent Limitations Guidelines and Standards for the Construction and Development Point Source Category, EPA says that polymer use is widespread in water treatment and that no studies have shown adverse impacts to water quality from properly dosed anionic polymers used to treat stormwater. The December 1, 2009 FR announcement of the final rule expanded the confidence in polymer treatment to include passive treatment systems. The reason for this confidence is that the typical polymer dose needed for treatment will be well below the toxic threshold. Polymers having a sufficient safety margin between a treatment dose and a higher treatment chemical toxic threshold are inherently low risk when properly dosed.

Anionic polyacrylamides (PAM) have an extensive track record of safety in soil stabilization and suspended solids removal. Toxic thresholds are well above typical anionic PAM treatment doses of 10 to 40 mg/L. For this reason, many government agencies in North America have simply set safety criteria for anionic PAM and allow use of any formulation meeting those criteria. These agencies include the Virginia Department of Conservation and Recreation, Wisconsin Department of Natural Resources, US Department of Transportation, US Department of Agriculture-National Resources Conservation Service, Toronto and Region Conservation Authority, Michigan Department of Environmental Quality, and Washington Department of Ecology. The anionic PAM criteria from these government agencies are the source for the criteria in this document.

The Toronto and Region Conservation Authority published a review of the toxicity of anionic PAM formulations in January 2014. All water-based and granular anionic PAM formulations had LC50s (median lethal concentration) for fish and invertebrates in excess of 100 mg/L except for some daphnid tests, which may have had interference from viscosity or binding of test organism food. Nevertheless, there were toxicity test results using daphnids that reported no adverse effects from a week long exposure at 100 mg/L. An LC50 for lake trout exceeded 600 mg/L of granular anionic PAM. Kerr *et al* (2014) found rainbow trout gill morphology and survival to be unaffected by anionic PAM concentrations up to 100 mg/L. The median lethal concentration (LC50) and the concentration causing a 25% effect are usually close for treatment chemicals. Therefore, the LC50s are adequate for showing anionic PAM doses to be about an order of magnitude below toxic thresholds.

Any anionic PAM formulation meeting the criteria listed in this document does not need the toxicity testing described in this appendix. Meeting these acceptability criteria ensures that the formulation will not cause aquatic toxicity if properly dosed. Documentation of the process for determining and delivering the proper dose is required and a part of the evaluation.

Criteria for acceptability of an anionic PAM formulation:

1. Water-based or dry anionic PAM formulation (no petroleum distillate or other emulsions)
2. Food grade (less than 0.05% free acrylamide monomer or approval by the National Sanitation Foundation for drinking water treatment)
3. Linear (not cross-linked)
4. Molecular weight 6 – 24 mg/mol

5. Charge density 8% – 35%
6. Jar testing or equivalent often enough to ensure dosing that just meets treatment needs and minimizes carry-over of PAM. Failure to maintain jar testing or equivalent will cause removal from the list of acceptable formulations. PAM carry-over must be kept to the minimum achievable and never exceed 30 mg/L.

References

Kerr, J.L., J.S. Lumsden, S.K. Russell, E.A. Jasinska and G.G. Goss. 2014. Effects of Anionic Polyacrylamide Products on Gill Histopathology in Juvenile Rainbow Trout (*Oncorhynchus mykiss*). Environ Toxicol Chem 33: 1552–1562

Toronto and Region Conservation Authority (TRCA), 2014. *Polymer Backgrounder: The Nature, Efficacy and Safety of Polymers for Erosion and Sediment Control*. Toronto, Ontario.

Appendix H: Establishing ballast water biocide environmental safety

Purpose of ballast water biocide toxicity testing procedures

State law (RCW 77.120.030) gives responsibility to the Washington Department of Fish and Wildlife (WDFW) for setting and implementing standards for approval of ballast water treatment systems. The Department of Ecology assists WDFW on environmental safety issues and setting conditions on the discharge of biocide-treated ballast water. This document describes the testing needed by the Department of Ecology in order to advise WDFW on treatment system approval conditions.

All ballast water treatment systems need testing to verify effectiveness in inactivating or removing zooplankton, phytoplankton, and bacteria. Testing to determine efficacy must include toxicity testing if the treatment system relies on a biocide to be effective. Proposed ballast water biocides will also need toxicity testing to establish conditions needed to protect receiving waters from toxic effects after ballast water discharge. Efficacy testing should focus on resistant organisms, while environmental safety toxicity testing should focus on sensitive organisms. Toxicity testing for biocide efficacy evaluation cannot be used for assessing environmental safety and vice versa.

Biocide toxicity is beneficial as long as it does not act against nontarget species or outside of its intended location. Because ballast water will contain unwanted nonindigenous species and native species which could be disease carriers, there is no such thing as a nontarget species for a ballast water biocide. However, if ballast water is excessively toxic at the time of discharge into the environment, receiving water organisms might suffer unacceptable harm. Toxicity testing is used to determine the potential for a biocide to harm sensitive receiving water organisms and then set conditions for safe use.

The following toxicity tests cannot reasonably be done on ships. The tests require special skills and facilities which are lacking on commercial vessels. In addition, the tests take too long for a ship operator waiting to discharge ballast. The goal of the testing described in this document is not routine monitoring of ballast water toxicity just prior to discharge but to establish procedures that vessel operators can follow to prevent any environmental threats. Inspectors can also check to make sure these safe operating conditions are maintained. Accredited labs must do the testing for this evaluation.

Treatment systems using biocides which are added or generated during treatment must be evaluated with toxicity testing to determine safe operating conditions. Solely physical ballast water treatment methods such as filtration, centrifugation, ultraviolet irradiation, or oxygen stripping are assumed to not need toxicity testing.

The testing for all ballast water biocides should include:

- 96-hour acute toxicity test with silverside minnows (*Menidia beryllina* in EPA-821-R-02-012).

- 48-hour acute toxicity test with a mysid (*Americamysis bahia* in EPA-821-R-02-012).
- 48-hour bivalve (*Mytilus sp*.) or 72-hour echinoderm (*Strongylocentrotus purpuratus* or *Dendraster excentricus*) embryo-larval survival and development test (EPA/600/R-95-136).
- 72-hour growth inhibition test with a marine diatom (*Skeletonema costatum* according to ISO 10253).

This suite of tests appears to meet or exceed the current international recommendations.

If more than one ship during the same week will be discharging ballast water containing the same biocide at the same pier or at an adjacent pier in the same port, then testing must include:

- 7-day survival and growth test with silverside minnows (*Menidia beryllina* in EPA-821-R-02-014).
- 7-day survival and growth test with a mysid (*Americamysis bahia* in EPA-821-R-02-014).

If the 7-day chronic tests are conducted, the acute tests with the same species are unnecessary. The bivalve or echinoderm survival and development test must still be conducted even if the 7-day chronic tests are used.

If treated ballast water will be discharged in the Cherry Point area (Bellingham north to the border with Canada) or another bay or shoreline with similar importance for Pacific herring, surf smelt, sand lance, or rock sole reproduction, then toxicity testing must include:

- 48-hour giant kelp germination and growth test (*Macrocystis pyrifera* in EPA/600/R-95-136).
- 7-day Pacific herring (*Clupea pallasii*) survival and growth test.
- 16-day Pacific herring embryo survival and normal development test.

Consult *Critical Spawning Habitat for Herring, Surf Smelt, Sand Lance, and Rock Sole in Puget Sound, Washington* (WDFW, 2000) for guidance on these critical areas. Call 360-902-2700 for a copy of the document. Contact Randall Marshall at randall.marshall@ecy.wa.gov or 360-407-6445 for guidance on labs capable of performing the herring tests.

General testing and reporting instructions:

1. Biocides lacking a steep concentration-response relationship should not be proposed and are not likely to be found acceptable.
2. Biocide solutions should be stored in the dark at 8 ± 1ºC with minimal headspace. The storage container should be made out of a substance to which the biocide would not adsorb or react.
3. If the biocide concentration can be analyzed reliably in the toxicity testing lab, then it should be done in one test chamber (or a surrogate) at each test concentration at the beginning and end of the test. If it can't be measured reliably, then the concentrations

should be assumed to be as prepared. If the concentration cannot be easily measured, do not attempt the option for toxicity testing to maintain moderate biocide toxicity.

4. Test solutions are renewed at 48 hours in 96-hour acute tests and daily for 7-day chronic tests.
5. Test solutions in acute tests should be renewed with the original biocide solution prepared at the start of the test.
6. 7-day chronic tests should be started with freshly prepared biocide solution and renewed with biocide solution that is freshly prepared twice (on days 3 and 5) during the 7 days of the test.
7. Herring tests which last longer than 7 days must have test solutions renewed with biocide solutions prepared as in 6. above for 7 days and then with clean seawater thereafter.
8. Herring for testing are not available outside of January to June. Please plan ahead.

Test reports must meet the reporting requirements in the EPA toxicity testing manuals and describe test conditions such as test chamber size, solution volume, temperature, dilution water source, exact test start time, exact test end time, etc. Test reports must contain a readable copy of all hand-written bench sheets. The bench sheets must include both the toxicological and water chemistry data for the biocide tests. The bench sheets must contain actual counts (not percentages) in order to be acceptable. Start counts must be clearly recorded on the bench sheet. The test report must include computer printouts of test data and statistical analyses. Test organism source, age, and unusual conditions must be reported. Each test report must contain a section where all deviations from test protocols must be accurately listed or the absence of such deviations noted. The results of range-finding tests must also be included if they were used to determine the most sensitive species prior to definitive testing.

Testing and reporting must support one of the following three options depending on the treatment system and biocide. Performing more than one of these options may make sense. A combined strategy, for example, could allow approval conditions for biocide neutralization if discharge must happen earlier than the approved minimum time needed for biocide degradation. Please anticipate all circumstances before choosing one or more of the options:

Biocide toxicity testing to demonstrate zero toxicity at ballast water discharge

The most convenient ballast water treatment systems will use biocides that begin as highly toxic and gradually become nontoxic due to volatilization, reaction, or degradation during a voyage. The ship operator will not need to measure biocide concentration in the ballast water at discharge but will only need to document the biocide dose and the time elapsed from treatment to discharge. Toxicity testing will be used to determine the minimum time allowed between treatment and discharge. The steps in this determination are:

1. Determine a maximum biocide concentration which will be used.
2. Add the biocide to seawater up to that concentration and hold under conditions that are as close as possible to actual use in a ballast tank.
3. These storage conditions would include $8 \pm 1^{o}C$ in the dark.

4. The surface area to volume ratio of the storage container should be as small as practical in order to resemble a ballast tank. The storage container should be made out of a substance to which the biocide would not adsorb or react.
5. Separate toxicity tests must be conducted on each sample drawn from the storage container every four hours over a 24-hour time period chosen so that the estimated time for the disappearance of toxicity is near the end of the 24 hours.
6. A pretest may be needed to estimate the time for disappearance of biocide toxicity. The pretest may also determine the most sensitive species so only that species needs testing with all of the samples. Concentration series are not required since a time series is the testing goal.
7. An alternate approach is to prepare separate volumes of biocide-treated water at 4-hour intervals and subsequently initiate all of the toxicity tests at the same time on samples drawn from each volume.
8. An example of an efficient method for the alternative approach is to set up a schedule similar to the following:
    a. Prepare biocide-treated water at 4:00, 8:00, 12:00, 16:00, and 20:00 on one day.
    b. At 16:00 on a subsequent day predicted to be the beginning of the last 24 hours of toxicity, use the treated water prepared at 4:00, 8:00, 12:00, and 16:00 to start tests.
    c. Start tests at 12:00 on the next day using the treated water prepared at 12:00, 16:00, and 20:00, and then do the daily renewals, etc. as soon as possible for the tests begun on the first day.
    d. Conduct daily renewals, etc. at 14:00 for the duration of the testing.
9. Exercise caution in setting an alternative testing schedule so that the minimum time needed before ballast water discharge is not overestimated and ship operators are not needlessly inconvenienced.
10. The earliest sample to produce no statistically significant toxicity relative to a concurrent control test will be the indicator of the minimum time needed before discharge.

Toxicity testing to verify biocide neutralization

A ballast water biocide that would be protective of the environment and convenient for ship operators would be one that would stay toxic during the voyage and could be neutralized just prior to discharge. The neutralizing chemical might also have some toxicity but this toxicity would be less than the toxicity of the biocide that it neutralizes. Toxicity testing should focus on determining the margin between the concentration of the neutralizing chemical which effectively eliminates biocide toxicity and the toxic threshold of the neutralizing chemical itself. Approval will be based upon the demonstrated ability for a ship operator to manage and measure the biocide and neutralizing chemical to stay within this margin. Any toxic reaction byproducts produced by the neutralization will be discovered during the toxicity testing and could affect approval.

The general process is:

1. A range-finding test may be necessary to estimate the toxic threshold before performing a definitive toxicity test. If the range-finding tests clearly identify a most sensitive species from the species listed, then definitive testing may focus solely on that test species.

2. The 25% effect level is used by EPA as an estimate of the toxic threshold. The IC_{25} should be used for growth and the EC_{25} for the survival and development endpoints.
3. If point estimates with 25% effect levels are not possible, the NOEC may be used if the percent minimum significance difference (PMSD) is no greater than 25%. *Alpha* should be 0.05 unless the PMSD would be less than 20%, in which case it may be changed to 0.01.
4. No test should have less than four replicates.
5. Once the toxicity testing has been done, then a plan must be written and submitted describing how the neutralizing chemical will be used to be both effective in removing biocide toxicity and stay below its toxic threshold.
    a. If the toxic threshold is above the maximum concentration needed to neutralize the biocide, then monitoring of the chemical dosing relative to the ballast water discharge volume should be adequate to demonstrate maintenance of the safety margin.
    b. If the toxic threshold is below the maximum concentration needed to neutralize the biocide, then a removal or reaction rate needs to be determined for relevant concentrations of both the neutralizing chemical and biocide. Careful monitoring of neutralizing chemical dosing relative to biocide dosing will be needed for maintaining environmental safety.
    c. If the concentration of the biocide and neutralizing chemical can be reliably measured onboard on a routine basis, chemical monitoring can be used for demonstrating safety when the threshold is below the maximum concentration. A small residual of neutralizing chemical can be an indication of effective biocide removal.

Toxicity testing to maintain moderate biocide toxicity at ballast water discharge

We reluctantly accept that ballast water may need to be discharged while moderately toxic under some circumstances. *Moderately toxic* means that the dilution in the immediate vicinity of the vessel is sufficient to eliminate toxicity. The circumstances when we will consider approving a toxic ballast water discharge are:

1. The biocide cannot be neutralized and
    a. Residual toxicity at discharge is necessary to meet very low treatment standards for ubiquitous microorganisms such as bacteria, which rebound by feeding on dead organisms in ballast tanks.
    b. Residual toxicity is needed because ballast tank sediments harbor communities of living organisms out of reach of biocide exposure.
2. Ballast water must be discharged before the intended time for biocide degradation has passed and the biocide cannot be neutralized.
3. The biocide neutralizing system malfunctions and ballast water must be discharged.

The process for establishing a maximum biocide discharge concentration under these circumstances is:

1. Because they are selected to be very toxic, biocides tend to have steep concentration-response relationships. A steep concentration-response relationship means that the difference between a concentration that is toxic and one that is nontoxic can be small. Toxicity testing can be used to determine the smallest difference that exists between the toxic and nontoxic concentrations of a biocide so that a target discharge concentration can be set such that ballast water toxicity will be eliminated very quickly after discharge.
2. The toxicity testing to determine the maximum allowable residual toxicity will need to focus test concentrations around the toxic threshold for lethality to fish or mysids or for combined survival and development in bivalves/echinoderms.
3. A range-finding test may need to be done first in order to find the general vicinity of the toxic threshold. If the range-finding testing clearly determines a most sensitive species, then definitive testing may be conducted using only that species.
4. The tests must have a series of at least five concentrations based on a dilution factor of $\geq$ 0.5, and the toxic threshold should ideally be in the middle to upper part of the lower half of the concentration series where spacing between concentrations is small.
5. A few partial responses are greatly desired since they will allow calculation of an LC_{25}.
6. At least four replicates must be run at each concentration for every test.
7. If the lowest concentration tested shows a statistically significant effect relative to the control, then the testing needs to be repeated with an improved concentration series.
8. The report must also propose an onboard method for accurately measuring either biocide concentration or a meaningful surrogate in order to verify that the biocide is near the target discharge concentration just prior to beginning ballast water discharge.
9. Candidates for the maximum discharge concentration include in order of preference:
    a. A target discharge concentration set at two times the LC_{50} as long as the LC_{50} is no more than three times the LC_{25} or a target discharge concentration set at three times the LC_{50} as long as the LC_{50} is no more than two times the LC_{25}. If the bivalve test is the most sensitive, the EC_{50} and EC_{25} for combined survival and development will be used as described for the LC_{50} and LC_{25}.
    b. A target discharge concentration set at three times the LOEC if the LOEC and NOEC are at the lower end of a series of at least five concentrations prepared using a dilution factor $\geq$ 0.5.

Representatives for the treatment system and lab should contact the WET Coordinator before beginning testing to make sure that results will support approval decisions:

Randall Marshall, WET Coordinator
360-407-6445 or randall.marshall@ecy.wa.gov

Appendix I: Chlorine Toxicity

WET testing is not a good tool for regulating chlorine toxicity. The holding time for WET samples gives chlorine a longer time to react with organics or dechlorinating agents than occurs in the receiving environment. Chlorine is very volatile and the steps taken to remove the supersaturation occurring when cold samples are warmed to test temperature will also remove chlorine. Chlorine concentrations can be reduced significantly as test solutions are prepared and poured into test chambers or during subsequent aeration to maintain oxygen levels in test solutions. Such a hit-or-miss situation is unfair to dischargers and labs who minimize holding times and sample handling and find chlorine toxicity more often than dischargers and labs who are not as careful. In addition, the hit-or-miss detection of chlorine toxicity using WET tests is obviously not as protective as monitoring chlorine directly and comparing the results to the water quality criteria for chlorine.

When chlorine is added to freshwater, the solution will contain two forms of free chlorine: hypochlorous acid (HOCl) and the hypochlorite ion (OCl^-). If the effluent also contains ammonia, then the addition of chlorine will result in two forms of combined chlorine: monochloramine and dichloramine. Municipal effluents usually contain all four of these forms of chlorine in some proportion, and taken together they are known as "total residual chlorine" (TRC) and the EPA analytic method for TRC detects them in combination. Because saltwater contains bromide, the addition of chlorine to saltwater will also form hypobromous acid (HOBr), hypobromous ion (OBr^-), and bromamines. The term for the combination of chlorine and bromine compounds formed by the addition of chlorine to saltwater is "chlorine-produced oxidants" (CPO) and the EPA method for measuring TRC also detects them.

The water quality criteria for chlorine in freshwater are based on TRC and the criteria for saltwater are based on chlorine-produced oxidants (CPO). Both are measured, however, as total residual chlorine. The water quality criteria for chlorine in freshwater are: 19 μg/L (acute) and 11 μg/L (chronic). The criteria for saltwater are 13 μg/L (acute) and 7.5 μg/L (chronic). These criteria were calculated by U.S. EPA based on many toxicity tests on many species from both freshwater (33 animal species from 28 genera) and saltwater (24 animal species in 21 genera). Aquatic plants were less sensitive than aquatic animals and were not included in the calculations. Levels of TRC and CPO degrade very rapidly in water. In order to compensate for the degradation of TRC, CPO and their associated toxicity, U.S. EPA conducted the toxicity testing in the development of the water quality criteria for chlorine using flow-through systems with continuous introduction and monitoring of TRC during the test. The water quality criteria for chlorine better protect surface waters from chlorine toxicity than the WET tests required in permits because they are based on toxicity testing that is much more sensitive than the static or static-renewal tests used for effluent monitoring.

Other organochlorines formed by the chlorination of a complex effluent will not be detected by the method for total residual chlorine, but will also not affect WET. Scientists in the Environmental Assessment Program (EAP) of the Department of Ecology evaluated 16 POTW effluents sampled between February 1988 and August 1991 for 14 chlorinated organic compounds that were detected by chemical analysis. Only four of these chlorinated organic compounds appeared to be formed by effluent chlorination based on the observation that their

concentrations were higher in the effluent than in the influent. These were chloromethane and three trihalomethanes (bromodichloromethane, dibromochloromethane, and chloroform). The four chlorinated organics presumed to be formed by effluent chlorination were orders of magnitude below water quality criteria for aquatic life protection in every sample. These chlorinated organics in POTW effluent that are not detected when TRC is measured are also very unlikely to contribute to WET.

40 CFR 122.44(d)(1)(v) allows us to use chlorine limits instead of WET testing to regulate chlorine toxicity because our state has narrative water quality criteria for toxicity. To avoid the hit-or-miss detection of chlorine toxicity by WET testing, and to avoid encouraging excessive use of dechlorinating agents by POTWs which already control chlorine well enough to meet water quality standards at the edge of a mixing zone, we prefer that samples for WET testing be taken before the chlorinator for chlorinated discharges which can meet water quality-based effluent limits for chlorine and have an ACEC below 25% effluent. If a permit requires dechlorination of samples or if a permit requires sampling prior to the chlorinator and this is physically impossible, then the sample should be dechlorinated using a stoichiometrically determined amount of sodium thiosulfate or sulfur dioxide. The calculations for determining the amount of dechlorinating agent must be included in the test report. Because of the effluent-dominated receiving water condition when the ACEC is 25% effluent or higher, permits will encourage extra control on chlorine through WET testing of an unmodified sample of final effluent.

Appendix J: Hardness and Alkalinity Constituents

Hardness (permanent & temporary) and Alkalinity Constituents

CONSTITUENT	hardness	alkalinity
calcium chloride	P	
calcium sulfate (27.3%)	P	
calcium carbonate	T	X
calcium bicarbonate	T	X
calcium hydroxide		X
magnesium chloride	P	
magnesium sulfate (27.3%)	P	
magnesium carbonate	T	X
magnesium bicarbonate	T	X
magnesium hydroxide		X
sodium carbonate		X
sodium bicarbonate (43.6%)		X
sodium hydroxide		X
potassium carbonate		X
potassium bicarbonate		X
potassium hydroxide		X

Red = a reagent in EPA synthetic water preparation. EPA synthetic water also includes potassium chloride (1.8%). The percentages used in EPA synthetic water are in parentheses.

Temporary (T) hardness is removed by boiling while permanent (P) hardness is not. The temporary hardness concentration can be estimated to be the lesser of either the total hardness concentration or the alkalinity concentration.

Even though both hardness and alkalinity are expressed as mg CaCO3/L, their concentrations in a sample can be quite different and be an indication of water chemistry. If the hardness concentration is higher, then the hardness constituents in the table above not shared with alkalinity are involved. If the alkalinity concentration is higher, then the alkalinity constituents in the table above not shared with hardness are involved.

Appendix K: Purpose of WET Testing and Statistics

WET testing and water quality criteria

Monitoring effluents for toxicity is a necessary activity. We need a regulatory strategy for detecting, identifying, and eliminating toxic substances or combinations of substances in effluents that would otherwise be missed. Effluents thoroughly characterized chemically and considered safe can still be toxic due to unknown constituents. Low flows will eventually occur, and even if control of effluent toxicity has been adequate for the last few years, such controls must anticipate dry weather that will occur on average only once per decade or so. WET testing is done to discover effluent toxicity at levels of concern for future low flow events and to identify unknown toxicants. WET testing is not performed to match receiving water conditions at the time of discharge. WET tests are conducted under standard conditions set to optimize test organism performance and reliability.

WET water quality criteria

EPA's water quality criteria have acute and chronic points of compliance. The chronic point of compliance is located at the edge of a mixing zone where receiving water must be suitable for long-term habitation, even during low flow conditions. Short-term chronic and critical life stage tests apply at the chronic point of compliance and preserve the suitability of receiving water for the complete life cycles of aquatic species. Inside of the mixing zone closer to the discharge is the acute point of compliance where there must be no lethality. WET tests only assessing short-term survival apply at the acute point of compliance.

EPA regulations allow either numeric or narrative water quality criteria for WET. Washington State has narrative criteria for WET. Compliance with narrative criteria is determined by comparing the response at an effluent concentration representing the point of compliance to the response in a nontoxic control to see if any differences in response are statistically significant.

Measuring compliance with narrative WET criteria is equivalent to the numeric criteria recommended by EPA in the March 1991 Technical Support Document (TSD) for Water Quality-based Toxics Control (EPA/505/2-90-001 available at: http://www.epa.gov/npdes/pubs/owm0264.pdf).

Acute toxic units (ATU) are determined by dividing the LC_{50} into 100. The acute toxicity detection limit is LC_{50} = 100% effluent (1.0 ATU). The TSD sets the acute criterion at 0.3 ATU to reflect the typical ratio of an LC_1 to an LC_{50} and be more protective. This criterion is less than the detection limit of 1.0 ATU. When the receiving water dilution allowance is insufficient to raise 0.3 ATU above the detection limit, EPA's recommendation is to change to a criterion of 1.0 ATU calculated by dividing the NOEC into 100. Two acute criteria and calculations are unnecessary since a criterion of 1.0 ATU calculated using the NOEC will always work.

The TSD proposes a chronic criterion of 1.0 chronic toxic units (CTU) calculated by dividing the IC_{25} into 100. EPA bases the validity of the 25% effect level for regulating chronic toxicity on its equivalence to the NOEC. If comparability to the NOEC is the standard for validity, it is

simpler to just use the NOEC. The TSD offers the alternative of dividing the NOEC into 100 for calculating CTU. As explained earlier in this document, using the NOEC is the only method for calculating ATU under all circumstances. Acute and chronic criteria are both 1.0 toxic unit if NOECs are used.

Further reasoning shows that toxic units and NOECs are unnecessary. Dropping toxic units simplifies both criteria and is expressed as: NOEC ≥ the effluent concentration at the point of compliance. Further simplification is possible by dropping the NOEC so that the criteria become no statistically significant toxicity at the point of compliance. This is also how compliance with narrative criteria is determined.

WET statistics

We need to be able to make regulatory decisions based upon single test results. WET tests are too expensive and difficult to use in generating monthly averages. For this reason, WET requirements usually are applied as daily maximums.

Point estimates are best when comparing toxicity test results to discover trends or to determine which substance is more toxic or which organism is more sensitive. Point estimates, however, cannot account for measurement variability in making a regulatory decision based upon a single test result. Point estimates often have confidence intervals, but these confidence intervals extend in both directions and require a judgment on which side to err.

Hypothesis testing allows control of both false positives and false negatives. Each hypothesis test takes into account variance across replicates before determining that differences in means are statistically significant. *Alpha* approximates the false positive rate when the values being compared are close. *Alpha* can be chosen based upon the size of the difference between means in order to minimize compliance failures that are not due to toxicity.

As the difference between values being compared increases, the false positive rate decreases without intervention, and the false negative rate becomes a concern. The false negative rate will tend to exceed the false positive rate because the null hypothesis is no toxic effect (burden of proof is on the demonstration of toxicity). To minimize undetected toxicity, the number of replicates can be increased when statistical sensitivity is inadequate.

Yellow shore crab (*Hemigrapsus oregonensis*) zoea photo courtesy of Kevin Li
Purple shore crab (*Hemigrapsus nudus*) adult photo courtesy of Shannon Beauford
King County - http://green2.kingcounty.gov/marine/Photo

Appendix L: Standard Toxicological Expressions

EC_{25}	25% effect concentration, approximate concentration causing a 25% reduction in the number of organisms having a healthy response (determined by linear regression of quantal data)
EC_{50}	50% effect concentration, approximate concentration causing a 50% reduction in the number of organisms having a healthy response (determined by linear regression of quantal data)
IC_{25}	25% inhibition concentration, approximate concentration causing a 25% reduction in a biological function of the test organisms (determined by linear interpolation of nonquantal data)
IC_{50}	50% inhibition concentration, approximate concentration causing a 50% reduction in a biological function of the test organisms (determined by linear interpolation of nonquantal data)
LC_{25}	25% lethal concentration, approximate concentration causing a 25% reduction in the number of surviving organisms (determined by linear regression on proportion survived)
LC_{50}	50% lethal concentration or median lethal concentration, approximate concentration causing a 50% reduction in the number of surviving organisms (determined by linear regression or Spearman-Kärber method on proportion survived)
LOEC	Lowest observed effects concentration (the lowest concentration in a series of concentrations having a statistically significant reduction in survival or a sublethal response relative to the control)
MATC	Maximum acceptable toxicant concentration (geometric mean of the NOEC and LOEC as an approximation of the true toxic threshold)
MSD	minimum significant difference (smallest reduction in response relative to the control which would be determined to be statistically significant)
NOEC	No observed effect concentration (the highest concentration in a series of concentrations showing no statistically significant reduction in survival or a sublethal response relative to the control)
Non-quantal	Data derived by measuring a property of the test organisms and then averaging to produce a number such as the weight per fish or number of neonates per female
PMSD	Percent minimum significant difference (smallest reduction in response relative to the control which would be determined to be statistically significant expressed as a percent reduction from the control)
Quantal	Data derived by counting the organisms which survived, were fertilized, or developed normally and dividing by the total to produce a proportion
Sub-lethal	Adverse effects other than death such as reduced growth, reproduction, development, or fertilization
Survival	Remaining alive and responsive to stimulation

Appendix M: Acronyms

ACEC	*Acute critical effluent concentration* (concentration of effluent estimated to occur at the edge of the acute mixing zone during critical low flow - See Appendix K – Purpose of WET testing and Statistics.)
ASTM	American Society for Testing and Materials
ATU	*Acute toxic units* (See Appendix K – Purpose of WET testing and Statistics.)
CCEC	*Chronic critical effluent concentration* (concentration of effluent estimated to occur at the edge of the full mixing zone during critical low flow - See Appendix K – Purpose of WET testing and Statistics.)
CETIS	*Comprehensive Environmental Toxicity Information System* produced by Tidepool Scientific Software (https://tidepool-scientific.com/)
CPO	*Chlorine-produced oxidants* (See Appendix I – Chlorine Toxicity.)
CTU	*Chronic toxic units* (See Appendix K – Purpose of WET testing and Statistics.)
CV	*Coefficient of variation* (standard deviation divided by the mean)
DO	Dissolved oxygen
EPA	United States Environmental Protection Agency
mgd	Million gallons per day
MS Access	Microsoft Access database software program
NPDES	National Pollutant Discharge Elimination System
POTW	Publically Owned Treatment Works
QC-plot	Quality control plot
SOP	Standard operating procedures
TDS	Total dissolved solids
TI/RE	*Toxicity identification/reduction evaluation* (responding to the detection of effluent toxicity by trying to identify the chemical constituents causing the toxicity and/or by trying to make changes to the facility which reduce effluent toxicity to acceptable levels)
TRC	*Total residual chlorine* (See Appendix I – Chlorine Toxicity.)
USGS	United States Geological Survey
UV	Ultraviolet
WAC	Washington Administrative Code
WET	*Whole effluent toxicity* (using living organisms to measure directly the toxicity of an effluent sample rather than estimating its toxicity based upon the measured concentrations of multiple individual chemical constituents)