



31 March 2005

via e-mail

Water Docket
EPA Docket Center
Environmental Protection Agency
Mail Code 4101T
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

Attention Docket ID No. OW-2004-0037

RE: Comments on National Whole Effluent Toxicity (WET) Implementation Guidance

Ladies and Gentlemen of U.S. EPA:

The following comments are submitted on behalf of the Colorado Wastewater Utility Council (CWUC). The CWUC is a non-profit organization with more than 40 members comprised municipal and special district wastewater treatment agencies dedicated to promoting and attaining the goals of the Clean Water Act. The purpose of the Council is to participate in educational, scientific, and outreach activities that support the needs of its members as well as the general public.

The mission of the Council is to promote professional and responsible environmental protection by supporting legislation and regulations that achieve well-defined benefits while maintaining local flexibility. Collectively, the membership of the Council, which is made up of forty-three public entities, municipalities and special districts, throughout Colorado that treat the wastewater generated by approximately 90% of Colorado's population.

This letter includes the comments on the **National Whole Effluent Toxicity (WET) Implementation Guidance** prepared by Risk Sciences and Chadwick Ecological under the direction of the Colorado Wastewater Utility Council.

All members of the CWUC have a very real, day-to-day, responsibility to meet the goals of the Clean Water Act by installing and operating state-of-the-art wastewater treatment facilities. As such, the Council has a clear understanding of the importance of protecting the environment and the challenges that must be faced to meet that responsibility.

The Colorado Wastewater Utility Council continues to endorse the use of Whole Effluent Toxicity (WET) tests as a valuable tool for protecting water quality in our state. We applaud EPA's decision to publish Implementation Guidance to clarify how best to integrate WET testing in the current NPDES permitting program.

We appreciate the opportunity to comment on the draft Implementation Guidance. We recognize that we are now entering a new era for WET testing. Throughout the long litigation process related to Whole Effluent Toxicity (WET) test methods, the U.S. Environmental Protection Agency repeatedly indicated that many of the issues raised by petitioners were best remedied through the implementation process rather than by modifying the test methods themselves:

*The interpretation and application of [toxicity] test results are part of the implementation policy and are not addressed in this rulemaking...*¹

In December of 2004, the U.S. Court of Appeals-D.C. Circuit concurred with the Agency's view and drew a sharp distinction between adopting a method of measurement and relying actual measurements from that method in regulatory practice:

There is an important distinction between the validity of a test method and the validity of a particular result from the test when it is used to determine compliance with permit conditions. Even EPA's calculations [show] WET tests will be wrong some of the time. ...Nothing we have written ... forecloses consideration of the validity of a particular test in an enforcement action. That issue is not before us. The case involves only the validity of the WET test methods... we are concerned here only with test methodology, not the results of particular tests in the field. Our decision does not endorse the validity of any test result in the future, nor does it foreclose a defense that the result is wrong. Those issues are simply not presented in this judicial review of rulemaking."² (emphasis added)

In other guidance documents, EPA warns that:

*The interpretation of the results of the analysis of the data from any of the toxicity tests described in this manual can become problematic because of the inherent variability and sometimes unavoidable anomalies in biological data.*³

¹ USEPA. Whole Effluent Toxicity: Guidelines Establishing Test Procedures for the Analysis of Pollutants - Supplementary Information Document. October 2, 1995 @ pg. 28.

² Edison Electric Institute, et al v. Environmental Protection Agency; Case No. 96-1062; Dec. 10, 2004 pgs. 8-9

³ USEPA. Short-Term Methods for Estimating the Chronic Toxicity of Effluent and Receiving Water to Freshwater Organisms, Fourth Ed. EPA-821-R-02-013. October, 2002. Section 9.4.1.1 @ pg. 39

The mere existence of analytical variability does not invalidate the WET method or any particular test result. However, federal courts have ruled that the possibility of measurement error...

*...deprives the agency of the power to find a violation of the standards, in enforcement proceedings, where the measured departure from them is within the boundaries of the probable measurement error.*⁴

We recognize that EPA has published several other guidance documents intended to explain how to minimize and account for test variability. Nevertheless, in our day-to-day experience, we still find that test variability is the single greatest barrier to more effective use of WET test results. No problem is more vexing than that which occurs when two separate bioassay laboratories analyze identical split samples and report contradictory conclusions as to whether the effluent is toxic or not. This occurs despite the fact that both laboratories are required to adhere rigorously to EPA's WET variability guidance documents.

Therefore, if the WET test methods are to remain useful as regulatory tools, the Implementation Guidance must inform state and federal permitting authorities on how to account for residual measurement error in a manner that complies with previous court rulings. Specifically:

- 1) When a large number of toxicity test results are used to evaluate reasonable potential the permitting authorities must account for the number of Type-I errors that are expected to occur. EPA guidance recommends using a 95% confidence level when conducting statistical analyses of WET test data. Consequently, on average, 5% of all toxicity tests performed will falsely indicate the presence of toxicity in a non-toxic sample.⁵

While the risk of error is relatively low for any single test, it is very high when a large number of statistical analyses are performed. For example, if a discharger conducts chronic toxicity tests on two species (fish & invertebrate) once per quarter, a total of 80 statistical analyses will be performed over the 5 year life of a typical NPDES permit. The probability of passing all 80 tests is less than 2% even if the discharge is chemically identical to the non-toxic control water used by the laboratory. Table 1 illustrates the cumulative risk of error for this sampling scenario.

The formula for calculating the data shown in Table 1 can be found in any standard college textbook on statistics. The formula can also be verified by performing simple Monte Carlo simulations. The probability of observing an exact number of errors (X) in a large number of analyses (N) when the risk of error for any single test (P) is only 5% is:

⁴ *Amoco Oil Co. v. EPA*, 501 F.2d 722 (D.C. Cir. 1974)

⁵ This statistical phenomena is discussed in relation to reference toxicant tests in USEPA. Short-Term Methods for Estimating the Chronic Toxicity of Effluent and Receiving Water to Freshwater Organisms, Fourth Ed. EPA-821-R-02-013. October, 2002. Section 4.16.5 @ pg. 16 but is true for any random sample of non-toxic waters also.

$$\left(\frac{X!}{((N-X)! * X!)}\right) * P^X * (1 - P)^{N-X}$$

The cumulative risk of error rises dramatically as the number of statistical analyses increases. Table 2 illustrates the cumulative risk of error when monthly chronic toxicity tests are performed on three species (fish, invertebrate & plant) during a normal 5-year permit term. Such a monitoring program requires 300 statistical analyses to be performed.

Table 1: Cumulative Probability of Type-I Errors in Quarterly Chronic Test Regime

# of Type-1 Errors Observed in 80 Statistical Analyses	Probability of Observing EXACTLY as many Type-1 Errors	Probability of Observing AT LEAST as many Type-1 Errors
0 in 80	1.7%	
1 in 80	8.6%	98.3%
2 in 80	14.4%	91.4%
3 in 80	19.8%	76.9%
4 in 80	20.0%	57.2%
5 in 80	16.0%	37.1%
6 in 80	10.6%	21.1%
7 in 80	5.9%	10.5%
8 in 80	2.8%	4.7%
9 in 80	1.2%	1.8%

Note: In a quarterly test regime, 20 samples will be analyzed over a 5-year permit term. 80 statistical analyses are required because two different biological endpoints (survival and sublethal growth or reproduction) are evaluated for both the fish and invertebrate test species. 5 years x 4 quarters x 2 species x 2 endpoints = 80 statistical analyses total.

A plain English explanation of Table 1:

If 1,000 permittees were to initiate quarterly chronic toxicity tests using two species on a non-toxic effluent, only 17 would pass all 80 statistical analyses. The other 983 non-toxic effluents would fail at least one toxicity test regardless of the actual water quality. In addition, 572 of the permittees would report AT LEAST four false toxicity test failures during the 5 year monitoring period. And, 12 of the 1,000 dischargers would observe 9 false failures - nearly 2 per year.

The CWUC understands that some level of test imprecision is inevitable when biological organisms are used to evaluate water quality (see memorandum for Chadwick Ecological Consultants, Inc. attached as Appendix-A to this Comment Letter). However, there remains considerable variability in test results even after dischargers and laboratories enact all of EPA's previous recommendations to minimize such problems.

To the extent some errors are irreducible, and beyond the discharger's control, it is essential that state authorities recognize that fundamental truth and not assume all test failures are prima facie evidence of a permit violation. We believe this is what the court intended in *Amoco Oil Co. v. EPA*, 501 F.2d 722 (D.C. Cir. 1974).

Table 2: Cumulative Probability of Type-I Errors in Monthly Test Regime

# of Type-1 Errors Observed in 80 Statistical Analyses	Probability of Observing EXACTLY as many Type-1 Errors	Probability of Observing AT LEAST as many Type-1 Errors
0 in 300	.00002%	---
1 in 300	.0003%	99.99998%
2 in 300	.003%	99.9997%
3 in 300	.01%	99.997%
4 in 300	0.1%	99.98%
5 in 300	0.2%	99.93%
6 in 300	0.4%	99.77%
7 in 300	0.9%	99.34%
8 in 300	1.8%	98.4%
9 in 300	3.1%	96.6%
10 in 300	4.7%	93.5%
11 in 300	6.6%	88.8%
12 in 300	8.3%	82.2%
13 in 300	9.7%	73.9%
14 in 300	10.5%	64.2%
15 in 300	10.5%	53.7%
16 in 300	9.9%	43.2%
17 in 300	8.7%	33.3%
18 in 300	7.1%	24.7%
19 in 300	5.6%	17.5%
20 in 300	4.1%	11.9%
21+ in 300	---	7.8%

Note: monthly testing will evaluate a total of 60 effluent samples over the course of 5 years. Each month five separate statistical analysis are performed: fish survival, fish growth, invertebrate survival, invertebrate reproduction and plant growth. Therefore a total of 300 separate statistical analysis will be conducted. And, each biological endpoint is statistically independent from all the others when a sample is stipulated to be non-toxic.

The plain English explanation for Table 2:

If 1,000 dischargers perform monthly chronic toxicity tests on three species, the probability of passing all 300 statistical analyses is 0.00002% (1 in 5 million) even if the effluent is chemically identical to the non-toxic control water used by the laboratory. Nearly 75% of the permittees will record 13 or more false failures during the 5-year monitoring period.

Most end-users of WET test results, including many state and federal regulators mistakenly expect non-toxic effluents to fail few if any WET tests because the risk of Type-1 error is very low for any individual test. Some regulatory authorities have misinterpreted previous EPA guidance to mean that there can be no instance where the "maximum" number of Type-1 errors ever exceeds 5% at any given location.⁶ The Implementation Guidance must correct this misunderstanding and describe how the absolute number of false positives varies as a function of the number of tests. EPA must inform those who rely on WET test results that 5% expected error rate is an average over a large number of tests. If 100 dischargers each perform 100 WET tests, it is very unlikely that all 100 will have exactly 5 Type-1 errors. Some will have more, some will have less - but the overall average will be about 5%. Table 1 illustrates just how variable individual discharger results can be over time.

Absent appropriate guidance on the proper interpretation of Type-1 errors, state and federal regulatory authorities will continue to misunderstand the meaning and import of a few WET test failures among a large number of tests performed. This, in turn, can (and has) led to mistakes in making reasonable potential determinations, compliance determinations and water quality assessments (305b and 303d reports).

It is essential to instruct state and federal authorities to account for the possibility analytical errors when drafting the NPDES permit. In addition, although EPA guidance now recommends that the nominal error rate be set to 5%, the new Implementation Guidance should inform end-users that states retain discretion to adopt a 1% threshold of statistical confidence as one means to reduce the risk of Type-1 error.

- 2) The Implementation Guidance should describe, in detail, how to use the Coefficients-of-Variation EPA published for each WET test method. This is particularly important because EPA was unable to develop a Method Detection Level (MDL) for WET. Consequently, although WET test methods exhibit variability that is comparable to chemical testing, there is presently no mechanism by which to evaluate the range of uncertainty associated with any given WET test result. The Coefficient-of-Variation should be used to remedy the problem.

⁶ USEPA. Understanding and Accounting for Method Variability in WET Applications Under the NPDES Program. EPA-833-R-00-003. June, 2000; See section 5.3.2.1 @ pg. 5-6.

EPA recently published the Coefficient-of-Variation (CV) for each WET test method based on results from a large-scale interlaboratory study. However, the Agency failed to tell state and federal regulatory authorities how to use this new information.

The Coefficient-of-Variation is a measure of test precision which can be used to estimate the range of uncertainty associated with any given WET test result. For example, according to EPA's calculations, the Ceriodaphnia dubia chronic reproduction test has a CV of 0.35. If a specific WET test finds that the IC-25 is 76%, then a more accurate way to report the result is as "IC25 = 76% ± 53%" or, alternatively, " IC25 = 76% (95% confidence range = 23% - >100%). "

Table 3 shows how range of precision varies in relation to the reported IC₂₅ value and the related Coefficient-of-Variation EPA identified for two popular chronic freshwater methods.

Table 3: 95% Confidence Range of Test Precision for IC₂₅ Values

Reported IC₂₅	Fathead minnow Growth (CV=21%)	C. dubia Reproduction (CV=35%)
5%	3% - 7%	2% - 9%
10%	6% - 14%	3% - 17%
15%	9% - 21%	5% - 26%
20%	12% - 28%	6% - 34%
25%	15% - 36%	8% - 43%
30%	17% - 43%	9% - 51%
35%	20% - 50%	11% - 60%
40%	23% - 57%	12% - 68%
45%	26% - 64%	14% - 77%
50%	29% - 71%	15% - 85%
55%	32% - 78%	17% - 94%
60%	35% - 85%	18% - >100%
65%	38% - 92%	20% - >100%
70%	41% - 99%	21% - >100%
75%	44% - >100%	23% - >100%
80%	46% - >100%	24% - >100%
85%	49% - >100%	26% - >100%
90%	52% - >100%	27% - >100%
95%	55% - >100%	29% - >100%
100%	58% - >100%	30% - >100%

Once again, in plain English, Table 3 shows that when a lab reports that the IC₂₅ is estimated at 25% for a given effluent sample, 95% of the time another lab analyzing an identical split sample using the same method will report an IC₂₅ value somewhere between 8% and 43% effluent for *Ceriodaphnia dubia* reproduction and between 15% and 36% for Fathead minnow growth. This is precisely the pattern of results EPA routinely observes in the annual DMR-QA studies when more than 100 labs perform chronic WET tests. It is also important to note that the range of precision is much narrower for the chronic survival because the level of natural background variability is much lower for those endpoints.

Similar results are also evident within labs as well as between labs. All laboratories that perform WET testing maintain control charts to track the health of their culture organisms. Reviewing these charts reveals that laboratories routinely observe a very wide variation in results from monthly tests conducted using reference toxicants (see, for example, Figure 3 in Appendix -A attached to this Comment Letter).

In practice, EPA deems all results that fall within the 95% confidence range as functionally-equivalent and "acceptable." However, all answers are not equally "acceptable" from the discharger's point of view. For example, if the maximum allowed instream waste concentration was 30%, then the reported value of 25% would cause the test to be considered a "failure." This occurs despite the fact that an identical split sample may have estimated the IC₂₅ at 32%; a non-toxic result. Both estimates, 25% and 32%, are well within the expected range of precision for the other. But, one results in a permit violation and the other does not. This is an improper and unfair use of WET test results. And, it is the most important issue that the Implementation Guidance must address.

In the interlaboratory WET variability study, EPA used several sophisticated tools, such as ASTM's h & h statistics, to identify and censor "outliers" from the dataset. Unfortunately, such tools require dischargers to perform a large number of WET tests on identical split samples. That is simply cost prohibitive. Moreover, it is unnecessary. The precision range can easily be calculated from the CV data EPA published without having to run parallel tests at different labs.

Calculating the range of certainty based on the CV is simply a short-cut to estimating how the results might vary if a very large number of identical split samples were analyzed simultaneously. The 95% "error band" is equivalent to saying that 95% of the identical split samples would report an IC₂₅ somewhere between 23% and 100% effluent.

The question of whether to report the 95% confidence range, the 99% confidence range, or some other confidence range is one that each state must answer for themselves. However, reporting WET test results without acknowledging the intrinsic level of precision published by EPA routinely leads to misuse and misapplication of the data by state and federal regulatory authorities. Even EPA's own expert peer reviewers concluded that:

*"It is very difficult to assess the quality of estimates of LC50 or IC25 without seeing the associated 95% confidence limits. It is difficult to assess variation within and among laboratories without this information being available. Although their inclusion will make the tables more difficult to read, reporting of these limits is essential."*⁷

As noted earlier, the federal courts have ruled that it is essential to know the error band for any given test result before that data can be used to make a compliance determination. Therefore, EPA must describe how to use the CV's for each method to calculate and report such error bands when stating WET test results on the monthly Discharge Monitoring Report.

- 3) When WET test results are reported as the No-Observed-Effect-Concentration (NOEC), EPA's recommended acceptance range should be applied before using the data to make regulatory decisions. While performing a large-scale interlaboratory study to validate WET test methods, EPA reported that normal measures of variability (such as the Coefficient-of-Variation) cannot be to NOEC-LOEC data. However, EPA established an "Acceptance Range" of the median value plus or minus one concentration interval in the dilution series.

In practice, this means that when multiple labs identify identical split samples of a single effluent, with a true NOEC of 25% the reported NOEC will range between 12.5% and 50%. All values within that range are deemed an "acceptable" estimate of the true NOEC. EPA relied on a similar range of precision to evaluate reported NOEC values in both the WET Interlaboratory Variability Study (2000) and in all of the DMR-QA studies performed since 1992.

The Implementation Guidance should make clear to state and federal regulatory authorities that any reported NOEC value must be evaluated in light of the published "normal" acceptance range. It would be unreasonable to accept that laboratory estimates of toxicity will vary by up to 400%⁸, but fail to acknowledge or account for this "error band" when using WET test results to evaluate compliance or assess reasonable potential. Doing otherwise will lead to significant errors in regulatory decision making.

The U.S. Court of Appeals, in *Edison Electric Institute, et al v. EPA*, warned of this in their recent decision regarding WET test methods:

⁷ USEPA. Summary Report for Peer Review of Preliminary Report: Interlaboratory Variability Study of EPA Short-Term Chronic and Acute Whole Effluent Toxicity Test Methods. March, 2001 (WET Docket IX-B.6) pg. 33 of 78

⁸ USEPA. Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms – Fourth Ed. EPA-821-R-02-013. See Section 9.3.1.1 @ pg. 39.

[Petitioner's] concern is that some discharge permits may specify an acceptable non-zero level of toxicity, which the effluent may not exceed, and that the WET tests have the potential to produce arbitrary permit violations. For example, if a permittee were subject to a toxicity limit of 3 TUC, and a WET test of its effluent would yield a 2 TUC result most of the time, but up to 4 TUC some of the time, the latter outcome would constitute a permit violation and potentially trigger an EPA enforcement action. This is certainly a problem for which EPA's system must account.⁹

In concurring with EPA that the level of intrinsic variability was not "excessive," the Court did not give regulators the license to ignore the acknowledged variability level altogether. Rather, is something that "must" be accounted for. Relying on WET test results that are reported without an appropriate confidence range fails to account for the inherent variability of the method as required by the court. And, just as the court noted, the variability can result in errors that either over- or under-estimate toxicity. In either case, the regulatory purpose is ill-served when such variability is not reported or considered.

EPA's recent recommendation to use the "TSD Method" to account for variability is inappropriate. That approach fails to distinguish between variations in effluent quality and analytical variability¹⁰.. as EPA recommended in Response to Public Comments:

The allowable frequency for criteria excursions should refer to true excursions of the criteria, not to spurious excursions caused by analytical variability or error.¹¹

The current TSD approach incorrectly treats all reported variation in WET test results as though it is, in fact, variations in effluent quality. Consequently, the equations used to calculate reasonable potential overestimate the probability of toxicity. In addition, those same equations will generate unnecessarily stringent permit limits to compensate for variations in effluent quality that are not actually occurring.¹²

⁹ Edison Electric Institute, et al v. Environmental Protection Agency; 2004 U.S. Lexis 25474; Dec. 10, 2004 pg. 8

¹⁰ USEPA. Understanding and Accounting for Method Variability in WET Applications Under the NPDES Program. EPA-833-R-00-003. June, 2000. See Section 6.1.1 @ pg. 6-2; also Section 6.6 @ pg. 6-10.

¹¹ USEPA. Technical Support Document for Water Quality-based Toxics Control; EPA-505/2-90-001; 1991. See Appendix entitled: "Technical Support Document for Water Quality Based Toxics Control - Responsiveness Summary;" May 9, 1991 @ pg. 11

¹² USEPA. Understanding and Accounting for Method Variability in WET Applications Under the NPDES Program. EPA-833-R-00-003. June, 2000. See Section 6.2 @ pg. 6-4.

In addition, EPA recently acknowledged for formal written submissions to the U.S. Court of Appeals – D.C. Circuit that converting NOEC or IC₂₅ values to "toxicity units" will significantly inflate the coefficient-of-variation.¹³ Consequently, EPA must inform the states that reasonable potential estimates and permit limit calculations should not be based on toxicity units unless the higher level of artifactual toxicity is first accounted for.

Finally, EPA should re-evaluate the propriety of using some WET test endpoints, particularly sublethal measures such as growth and reproduction, to assess compliance with pass/fail permit limits. This suggestion is consistent with similar recommendations made by EPA's own expert peer reviewers after evaluating the results of the Interlaboratory WET Variability Study (2000):

*...the results seem to show that some of these tests should not be used in the regulatory context because the successful completion rate is too low and Coefficient-of-Variation values are too high.*¹⁴

There is no form of advanced waste treatment that can diminish or eliminate disparities in WET test results that are caused by analytical variability. It is unreasonable to impose permit obligations related to factors that are beyond the discharger's control.

- 4) The Implementation Guidance should make clear that failure of a WET test, per se, does not necessarily constitute a violation of water quality standards. EPA has repeatedly written that it was not promulgating a water quality standard for toxicity when it adopted the WET test procedures into 40 CFR Part 136.¹⁵ This is consistent with the recent federal court ruling:

*States retain discretion, subject to EPA guidance and recommendations, to set their toxicity thresholds in order to compensate for local conditions at the permitting stage. In light of this discretionary, rather than mandatory, nature of state implementation of standards and thresholds...the WET program [is not] an illegal federal water quality standard...Before implementing a test method, EPA must establish that the measured characteristic bears a rational relationship to real-world conditions... Petitioners are worried that they might be subject to excessive restrictions; such limits, however, would be imposed by local authorities, and are not part of the rule-making under review in this case. The WET test methods offer only a means of measuring compliance with those limits, individual dischargers remain free to challenge their permits, on a case-by-case basis, if they believe the local authorities are regulating at a level that poses only a minimal risk to aquatic life.*¹⁶

¹³ USEPA. Declaration of John F. Fox on behalf of Respondent United States Environmental Protection Agency. Appendix to Brief submitted to U.S. Court of Appeals – D.C. Circuit entitled: "Respondent EPA's Opposition to WET Coalition and Westcas' Petition for Panel Re-Hearing" (submitted March 17, 2005).

¹⁴ Reviewer Z in Response to Comments: Peer Review of the "Preliminary Report-Interlaboratory Variability Study of EPA Short-term Chronic and Acute Whole Effluent Toxicity Test Methods;" September, 2001; p.19 (WET-IX Docket #B.6)

¹⁵ See, for example, 60 Fed. Reg. 199 @ 53529 (Oct. 16, 1995) and 67 Fed. Reg. 223 @ 69969 (Nov. 19, 2002)

¹⁶ Edison Electric Institute, et al v. Environmental Protection Agency; 2004 U.S. Lexis 25474; Dec. 10, 2004 pgs. 12-13

Before WET test results can be used in a reasonable potential determination or to assess permit compliance, state regulatory authorities must "translate" the narrative criteria using approved toxicity test methods. This is particularly important where the natural ionic chemistry of local water supplies may interfere with the conduct and interpretation of WET tests. Several examples serve to illustrate the problems encountered:

Example-A: In many cities, local water supplies are very soft. Extreme low hardness is known to adversely impact the normal rate of reproduction in *Ceriodaphnia dubia*. In this instance, it is the absence of essential elements rather than the presence of harmful pollutants that may lead to mistaken conclusions about effluent "toxicity."

Example-B: Storm water samples routinely fail chronic toxicity tests because rainfall (and most snowmelt from the Colorado Rockies), is naturally low in conductivity and pH. Generally speaking, rain water is ill-suited for optimum growth of the standard species used as indicator organisms in WET testing.

Example-C: Many states in the arid southwest rely extensively on groundwater to meet public demand for potable water supplies. The natural chemistry of groundwater is distinctly different from surface freshwater. The balance of ions in groundwater may be less than ideal to support maximum growth and reproduction in the standard test species.

In each of these real-world examples, the presence of toxic pollutants is often inferred prematurely from WET test failure. The issue of whether state narrative standards have been exceeded, or not, depends on a great many other factors: actual dilution available, ambient natural background chemistry, representativeness of standard test species, and the frequency, duration and magnitude of exposure.

The new Implementation Guidance must make clear that state permitting authorities are obligated to "translate" the narrative water quality criteria for toxicity into relevant and meaningful WET test endpoints in the permit.

EPA views the state implementation procedures to be an essential element of the state's narrative water quality standards. If EPA would not approve the state standards in the absence of such implementation procedures, then the Agency should so state. If the EPA did so, then the states would be required to enact WET requirements in the same manner as for any other chemical pollutant ... through formal notice and rule-making.

Finally, the Implementation Guidance should clarify that states are not required to use the sublethal endpoints when developing procedures to implement the narrative water quality criteria unless it is necessary to do so to meet state standards. Several of EPA's regional offices have improperly informed state regulators that the chronic sublethal endpoints are required because they are part of the method. This is occurring despite the fact that EPA published a summary review of field validation studies showing that, unlike the WET mortality endpoints, the sublethal test endpoints could not be correlated with actual instream conditions.¹⁷

As EPA previously acknowledged, promulgation of a toxicity test method does not constitute adoption of a water quality standard for WET. If EPA's Regional Offices persist in telling state authorities that sublethal endpoints "must" be used then, in fact, the method promulgation is being used to establish de facto water quality standards contrary to what EPA told the U.S. Court of Appeals.

CONCLUSION

Variability is a fact of life for all test methods- chemical and toxicological. EPA's previous guidance was an important and useful first step in minimizing analytical error in WET testing. However, there remains a level of variability that is simply intrinsic to the use of biological organisms as measurement "devices."

Long years of experience have taught permitting authorities that values below the chemical MDL are not sufficiently reliable for most regulatory purposes. The absence of an MDL for WET test methods has led to the mistaken belief that all reported estimates of toxicity are equally precise and reliable. This, in turn, leads directly to errors when assessing attainment with water quality standards, making reasonable potential determinations and when evaluating permit compliance.

The Colorado Wastewater Utility Council strongly recommends that EPA use the results of the WET Interlaboratory Variability Study (2000) and the DMR-QA data generated over the last 10 years to inform state and federal permitting authorities on how to interpret the precision WET test results correctly. This can be accomplished by coaching state regulators to work with a "precision range" just as EPA has been doing for years.

CWUC is not asking for any fundamental programmatic change; only that dischargers be given the same consideration EPA routinely gives when auditing laboratories – an acknowledgement that test results vary within a predictable range and that all values within the "error band" are functionally-equivalent and acceptable. Even occasional "spikes" in the reference toxicant control chart are deemed, by EPA, to be inevitable and unavoidable Type-1 test error when a larger number of tests are performed.¹⁸

¹⁷ USEPA. A Review of Single Species Toxicity Tests: Are the Tests Reliable Predictors of Aquatic Ecosystem Community Responses? EPA/600/R-97/114 (January, 1999)

¹⁸ USEPA. Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms – Fourth Ed. EPA-821-R-02-013. See Section 4.16.5 @ pg. 16.

EPA readily admits that the WET test isn't perfect. That the test has limitations does not diminish its utility unless those limitations are ignored during the regulatory process. Toxicity is a zero-tolerance pollutant. State authorities rightfully expect permittees to discharge non-toxic effluent. However, they also assume that the only way to demonstrate that has occurred is to pass each and every WET test that is performed. It is impossible to demonstrate treatment perfection with a test method that is imperfect.

Respectfully submitted on behalf of the Colorado Wastewater Utility Council,

A handwritten signature in black ink, appearing to read "T. Moore", with a long horizontal flourish extending to the right.

Timothy F. Moore
Risk Sciences
1417 Plymouth Dr.
Brentwood, TN 37027
Office: 615-370-1655
Fax: 615-370-5188
Email: tmoore@risk-sciences.com

Attachment

Appendix A

COMMENTS ON THE NATIONAL WHOLE EFFLUENT TOXICITY
IMPLEMENTATION GUIDANCE UNDER THE NPDES PROGRAM
FROM A WET TESTING LABORATORY'S PERSPECTIVE

Chadwick Ecological Consultants, Inc. (CEC) and Chadwick & Associates, Inc., and aquatic biological laboratory were requested by the Colorado Wastewater Utility Council (Utility Council) to review the Draft USEPA *National Whole Effluent Toxicity (WET) Implementation Guidance under the NPDES Program*. The Colorado Wastewater Utility Council is a nonprofit corporation with more than 43 members comprised of municipal and special district wastewater treatment entities seeking to promote Clean Water Act goals through compliance with the regulations implementing those goals.

We are interested in NPDES implementation of WET testing from a number of different perspectives. However, a key issue is related to when our clients do “split” WET testing – i.e., sending the same effluent sample to two different laboratories – and get different results. A natural question is why this would happen? And this question is directly related to the real world implications of the variability in the precision of WET tests, which the court has said EPA must account for. CEC and their accompanying laboratory, Chadwick & Associates, Inc., have long-term data on *Ceriodaphnia dubia* neonate reproduction in control treatments (i.e., synthetic moderately-hard dilution water), as well as their sublethal response to reference toxicants and from other WET tests. We looked at the potential sources of WET test variability based on real world data from our own testing experience.

Biological Variability

The biological variability associated with using live test organisms complicates the use of sublethal numerical or narrative WET limits in NPDES permits. Each WET test relies on the measured response of a new group of organisms to control water and a new effluent of unknown constituents. Even though experimental conditions are highly controlled, there is no way to control biological response. Using data from 330 tests conducted from January 1999 through April 2005, it is clear that *C. dubia* neonate production in control treatments exhibits a wide range of values - even under control conditions (Fig. 1A and B). Normalized reproduction values for control treatments during a three brood cycle for this dataset ranged from 1 to 57 neonates with a mean of 28.1 (Fig. 1A) and coefficient of variation (CV)

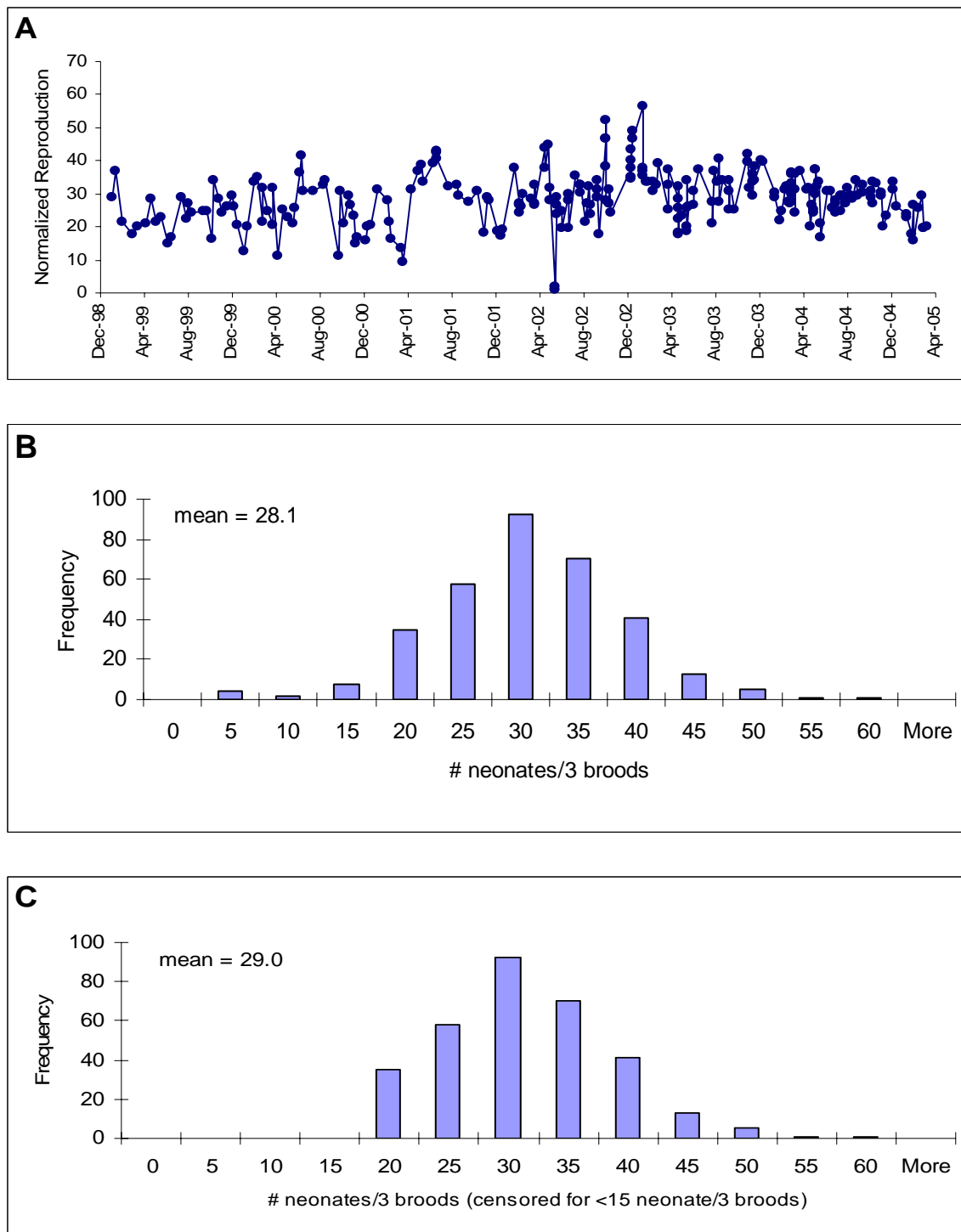


FIGURE 1: A) *Ceriodaphnia dubia* neonate production in reference toxicity and WET tests conducted by C&A since 1 January 1999. B) Frequency distribution of *C. dubia* neonate production during a three brood cycle using the same dataset. C) Frequency distribution of *C. dubia* neonate production during three brood cycle following imposition of EPA control performance criteria.

of 29.5%. Not surprisingly, the *C. dubia* reproduction data from control water are normally distributed, with 95% of the data within the range of 12 to 45 neonates (Fig 1B). Of course, EPA protocols require control reproduction to be 15 neonates or higher, invalidating 4.2% of the data. More importantly, in the absence of data under 15 neonates, the underlying distribution for control organisms shifts from a normal to a log-normal distribution (Fig 1C) and shifts the mean up to 29 neonates. Combining natural, normally distributed reproduction data with the one-tailed control performance criteria of EPA protocols results in two different distributions for WET test treatments (Fig. 1B) and WET test controls (Fig. 1C) over the population of tests. This “built in” difference could be one source of the variability observed in WET test results.

Within-test Variability

Another measure of variability in *C. dubia* performance is the percent minimum significant difference (PMSD) that quantifies the within-test variability. The PMSD is a measure of test precision based on the control performance and is calculated using the minimum significant difference (MSD) derived from the error mean square of ANOVA, and the critical value of Dunnett’s test statistic (USEPA 2000). The measure indicates the minimum detectable percent difference that is statistically significant when comparing a series of treatment means to a control mean, and provides insight into the sensitivity of the analytical method.

Over the past eight years, we have over 260 WET tests which show that PMSD data also vary considerably (Fig. 2A and B). The PMSD data are log-normally distributed (Fig. 2B). The C&A database includes values derived during both reference toxicant and NPDES WET tests, thus the range is indicative of the variability associated with sublethal *C. dubia* WET tests. The variability underlying experimental WET test controls, whether methodological or biological, indicates the low precision expected with the sublethal *C. dubia* WET methodology endpoints.

Results of the EPA method variability study using *C. dubia* (USEPA 2000) indicated that 18 of 33 laboratories (54%) had at least one sublethal reference toxicant test exceed the 90th percentile PMSD of 37%. These data indicate that variability within *C. dubia* sublethal WET tests is not limited to a few laboratories or studies.

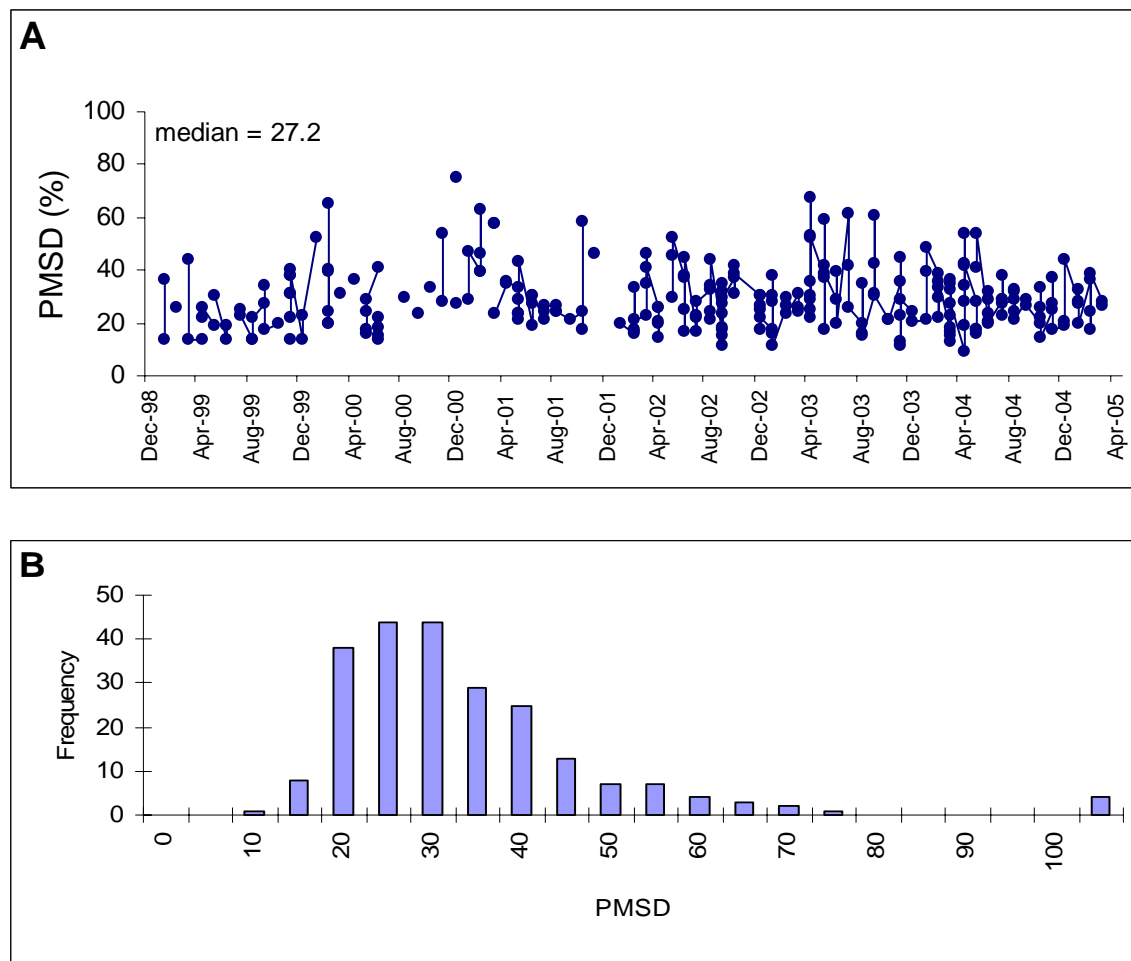


FIGURE 2: A) Percent minimum significant difference (PMSD) for *C. dubia* neonate production in reference toxicant and WET tests conducted by C&A since 1 January 1999. B) Frequency distribution of PMSD data from reference toxicant and WET tests.

Variable Response to Known Toxicant

Given the high degree of variability associated with *C. dubia* reproduction under standard laboratory control conditions (i.e., synthetic dilution water), it is reasonable to expect that this same variability will be expressed in regular WET testing statistics (i.e., IC25). Since 2000, C&A reference toxicant tests (NaCl) with *C. dubia* have used a standardized dilution ratio from zero to 1200 mg/l NaCl, with 50 tests during this period. Despite use of a standardized dilution series and standardized test

conditions required by USEPA methods, *C. dubia* IC25 values have exhibited a high degree of variability (Fig. 3) over time.

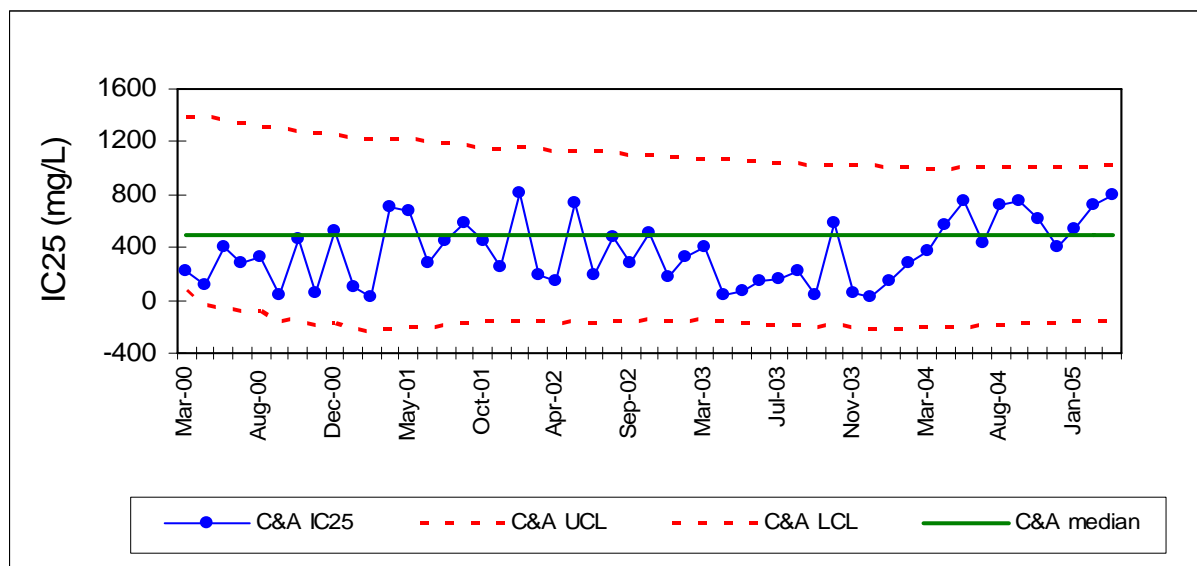


FIGURE 3: C&A reference toxicant (NaCl) inhibition concentrations and running laboratory coefficient of variation since June 2002.

These data indicate there is no predictable response by *C. dubia* to a known toxicant. This has considerable implications to real world applications. Furthermore, our data is consistent with USEPA's own experience with their interlaboratory variability study, where the variability of the *C. dubia* chronic reference toxicant tests was so great that EPA was forced to abandon the data outright (USEPA 2000). If *C. dubia* reproduction endpoints exhibit such random responses to a known toxicant, why should we expect a predictable response to effluent testing?

Laboratory Performance

We have investigated the variation observed in these reference toxicant tests and can only conclude that the variation is due to the physiological variability associated with using live organisms – not laboratory performance. Chadwick & Associates, Inc. (C&A) is an accredited WET testing facility with 15 years of experience performing acute and chronic toxicity tests for a variety of municipalities, mining, and industrial facilities in the western United States. Since 1992, C&A has received annual WET

test certification by the States of Washington and California, and has successfully passed WET testing audits by the USEPA in recent years. Furthermore, C&A participates annually in the acute and chronic portions of the EPA DMR-QA studies, meeting the test acceptability criteria annually. C&A performs both acute and chronic toxicity tests using *Ceriodaphnia dubia*, *Daphnia magna*, *Pimephales promelas*, *Selanastrum capricornutum*, and *Oncorhynchus mykiss*. In fact, C&A's DMR-QA results, using the same methodology as in the reference toxicant tests, have consistently been within USEPA's accepted range of values (Fig. 4). In addition, C&A was one of the laboratories selected for participation in USEPA's interlaboratory variability study, after successfully completing a rigorous review of our qualifications and historical performance.

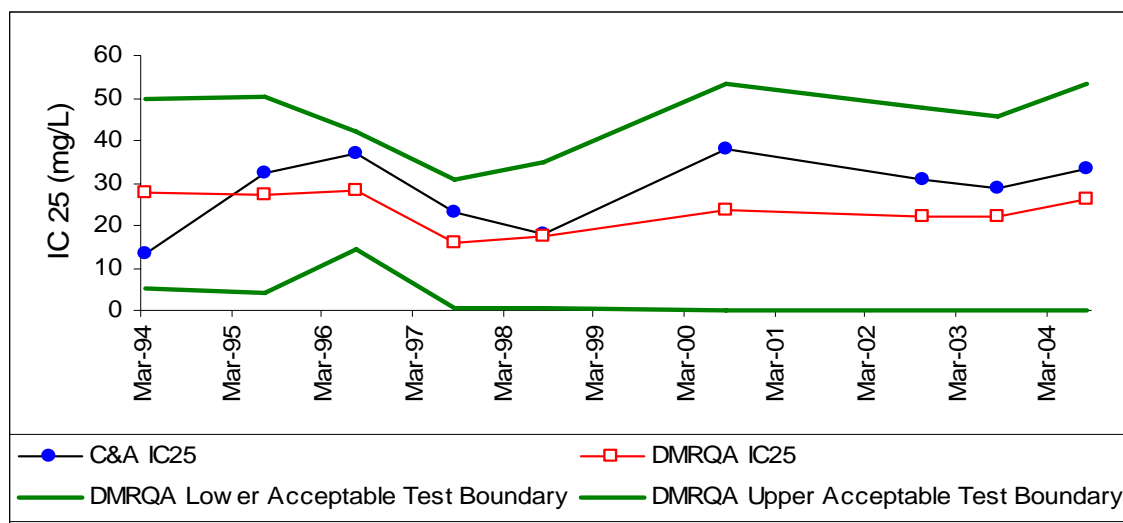


FIGURE 4: C&A performance in the DMR-QA program using *C. dubia* in chronic toxicity tests.

Recommendations

We are concerned that the variation observed in sublethal WET test endpoints using *C. dubia* is so great, that the results of many of the statistical tests is largely meaningless - and it would be difficult to distinguish whether these differences are due to effects of water quality or inherent biological variability. Of course, we recognize the importance of the use of WET tests in maintaining surface water quality, but we believe the use of *C. dubia* sublethal endpoints in the development of numerical or narrative triggers would be inappropriate. If chronic sublethal results indicated that an effluent would likely cause or

contribute to an excursion above water quality standards then the establishment of WET triggers should be based on acute or chronic-lethality toxicity limits.

As such, we believe that the NPDES implementation guidance should focus on acute and/or chronic lethal endpoints and allow individual States and tribes a greater flexibility in an approach for using sublethal endpoints.

References

USEPA. 2000. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program*. Eds. Denton, D.L., Fox, J., Fulk, F.A., Greenwald, K., Narvaez, M., Norberg-King, T.J., Phillips, L. EPA/833/R-00-003. Office of Water, Washington, DC.