DEFINING RELATIONSHIPS BETWEEN WHOLE EFFLUENT TOXICITY AND INSTREAM TOXICITY

Jerry Diamond, Christiana Daley, and Michael Barbour
Tetra Tech, Incorporated
10045 Red Run Blvd., Suite 110
Owings Mills, MD 21117

ABSTRACT

A database has been compiled that contains WET test, instream assessment, and supporting data for over 250 municipal and industrial wastewater facilities representing many of the major regions in the U.S. Sites included in the database: (a) used USEPA *Ceriodaphnia* and/or fathead minnow WET tests (acute, chronic, or both), (b) tested effluents that were representative of the actual discharge; (c) used a standardized benthic macroinvertebrate bioassessment method upstream as well as downstream of the discharge and some form of habitat assessment and (d) indicated that no intervening point or nonpoint sources (other than the discharge in question) were present between the upstream and downstream sampling sites. Supporting data included effluent instream waste concentration (IWC, based on average and/or design flows), pollutants present (categorized by general chemical class), and percent industrial waste. Bioassessment and WET data were standardized to allow inter-site and inter-region statistical comparisons. Stepwise multiple discriminant analyses indicated that survival endpoints (acute or chronic) were more related to instream conditions (impairment vs non-impairment) than sublethal endpoints. Multiple regression analyses, indicated that the number of tests performed at the site, type of chemicals discharged, the amount of effluent dilution (under low flow), and habitat quality all affected the relationship between WET and instream condition. We could not confirm a definite relationship between the degree of WET variability at a site and instream conditions partly due to the above confounding factors. However, we observed that facilities that failed < 25% of their tests had ≤ 15% chance of exhibiting instream measurement. Data available for arid west sites exhibited similar relationships as headwater sites in other parts of the U.S. Our results suggest that discharges associated with fair-poor habitat quality instream or that have ammonia or metals as the pollutant of concern, or that have considerable dilution (IWC < 20%), have a low probability of exhibiting impairment even if WET test failures were observed. Effluents meeting these conditions and that fail typical current regulatory WET criteria, may be subject to overly stringent WET limits.

KEYWORDS

effluent, toxicity testing, bioassessment, Ceriodaphnia, Pimephales, independent application

INTRODUCTION

With the increased use of whole effluent toxicity (WET) tests and regulatory-driven WET limits for evaluating wastewater effluent compliance, there are concerns about how WET should be implemented in discharge monitoring permits. One of the underlying issues is the accuracy of laboratory WET endpoints in predicting effluent effects on receiving stream biota. The recent SETAC Pellston Workshop on WET (Grothe et al. 1996) is evidence that this is a very important issue to many people.

Some past studies (as summarized in EPA 1991) reported good agreement between WET and instream biological condition. However, careful analysis of these “validation” studies has shown that relatively few independent sites were examined and that site selection was non-random. Furthermore, some studies involved highly acutely toxic effluents and/or multiple stresses, yet validation of chronic or sublethal WET endpoints for single discharges was claimed. Given the fact that most of these studies used very limited WET or ambient toxicity information (most effluents were tested only once or twice), it is remarkable that good correlations with instream effects were observed.

Re-analysis of the EPA studies by Dickson et al. (1992) indicated reasonably good correlations between *ambient* toxicity and instream biological conditions, however, the same correlation was not as consistently demonstrated for WET tests. Design and analysis were typically site-specific in most of these studies so they were unable to establish predictive relationships that could be applied elsewhere (Sprague 1995; Marcus and McDonald 1992). Except for a very few published studies (e.g., Dickson et al 1996), most sites were monitored during only a single year with no follow-up assessment.

Developing a predictive understanding of relationships between WET and instream biological condition is critical to the implementation of WET limits in discharge monitoring permits. Furthermore, this research could shed light on the debate concerning the appropriateness of EPA's policy of independent application (USEPA 1991). To help address these issues, the Water Environment Research Foundation supported our project that examined relationships between standard freshwater WET tests, typically required of permittees, and actual receiving stream effects as judged by commonly-used measures of biological integrity.

METHODOLOGY

We performed an extensive retrospective analysis using WET and bioassessment data from sites around the U.S. While we incorporated some data from EPA's Complex Effluent Toxicity Testing Program (CETTP), Eagleson et al. (1990), and other less recent studies that met our selection criteria (see below), most of our data were collected from post-1990 studies. Very few of EPA's CETTP sites met the data acceptability criteria for this research chiefly because of the presence of either multiple discharges or other known stressors in the area between the up and downstream sampling points. Recent data were preferred for this research because data quality control criteria were generally higher and test performance better (particularly in comparison with some of the CETTP WET data). Our research was limited to standard EPA acute and "chronic" *Ceriodaphnia* and *Pimephales promelas* WET tests and freshwater receiving streams. Effects upon stream biota were assessed using data from benthic macroinvertebrates. These organisms are sensitive, relatively nonmobile fauna of riverine systems that have been widely used as an ecological indicator assemblage (Plafkin et al. 1989; Eagleson et al. 1990; Cuffney et al. 1993; Gibson et al. 1996). All macroinvertebrate data were expressed as a standardized relative percent difference in community indices or metrics between sites up and downstream of each discharge and WET data were expressed as a function of effluent dilution in the receiving stream making it feasible to compare data across sites. A stream habitat has a profound effect on biota, we strived to obtain habitat data for all sites and categorized sites for later analysis on the basis of either poor, fair, or good-excellent habitat quality. Supporting data collected for each site included effluent dilution, pollutants discharged, and percent of industrial waste contribution (by flow).

Criteria for inclusion of studies in our database were: (1) WET tests used effluent representative of what was discharged (e.g., no lab effluent dechlorination or pH adjustment prior to WET testing); (2) WET tests met current EPA QA requirements; (3) WET tests were performed prior to but not more than 1 year before the bioassessment; (4) upstream and downstream bioassessment data were available; (5) stream habitat quality was similar up and downstream of the discharge; and (6) no point or nonpoint sources, other than the discharge of interest, were present between the up and downstream site. Supporting data were also collected for many sites including design effluent dilution (for the 10-year 7-day low flow condition - 7Q10), pollutants discharged and concentrations (through EPA's Permit Compliance System database and discharge reports), and type of treatment facility.

Several hypotheses were examined in this research including: 1) The relationship between WET and stream biota will be strongest for acute test endpoints and weakest for chronic test endpoints, particularly sublethal endpoints; 2) the relationship between WET and stream biota will be weak or non-existent when there is high variability among and within test endpoints for a particular effluent; 3) the relationship between WET and stream biota will be strongest for effluent-dominated systems and weakest for systems that greatly dilute the effluent; 4) the relationship between WET and stream biota will be strongest for effluents that contain persistent organic compounds or metals, and weakest for effluents that contain volatile or labile compounds such as ammonia; 5) the relationship between WET and stream biota will be strongest in those regions in which the test species naturally occur (eastern and midwest and U.S.) and weakest in regions in which test species do not naturally occur (arid west).

Table 1 summarizes WET endpoints and pass/fail criteria for WET and bioassessment data examined in this study. **(See Table 1. "Summary of effluent toxicity test (WET) endpoints and pass/fail criteria examined")**

RESULTS AND DISCUSSION

We analyzed data for 250 sites in 18 states comprising over 1,000 WET tests and 330 bioassessments. In addition, this study analyzed data for the Santa Ana River dischargers in California. This case study was analyzed separately because extensive chemical, toxicological (both WET and ambient toxicity tests) and biological data were available which allowed more in-depth analyses of WET than was feasible for other uses in our database. Nearly one-third of the sites had > 4 WET tests per bioassessment and approximately half the sites had both acute and chronic test data. Fifteen sites had multi-year assessments.

Database Characteristics

At more than half the sites, only one of the two species of interest were tested. *C. dubia* LC_{50} and acute percent mortality were the most prevalent WET endpoints in the database followed by *C. dubia* chronic (NOEC, LOEC) endpoints. This was because in some states this is the only test species required in permits. Chlorine and metals were the most prevalent pollutants of concern reported in discharges in this database followed by ammonia, cyanide, and non-volatile organics. The types of pollutants present in the database may be an artifact because only those chemicals having state water quality standards are required to be monitored.

A large percentage of the sites were located near or at the headwaters of the receiving stream and thus the effluent contributed > 80% of the flow under 7Q10 conditions. Despite efforts to obtain sites from all regions of the U.S., the majority were located east of the Mississippi River with very few sites representing the arid west. The higher frequency of usable sites in the east and midwest was due primarily to the longer history of WET testing in those regions as compared to most western U.S. states.

Agreement between WET results (using the pass/fail criteria in Table 1) and macroinvertebrate assessment results for all sites pooled together differed with the WET endpoint **(See Figure 1 "Percent Agreement Between Whole Effluent Toxicity (WET) Test Failure and Bioassessment Impairment by WET Endpoint")**. Certain endpoints (e.g., *C. dubia* reproduction NOEC) appeared to be relatively unrelated to stream biota while other endpoints (e.g., *C. dubia* chronic survival or *P. promelas* chronic survival) appeared to be more related to biological condition. Chi-square analysis indicated that the *Pimephales* percent acute mortality endpoint had the strongest relationship with biological condition ($p < .05$), while *C. dubia* acute lethality had an inverse relationship with stream biota (i.e., greater lethality in conjunction with good biological condition). Thus, acute WET test endpoints were not definitively better at predicting biological condition than chronic test endpoints (H_o #1). However, lethality endpoints were better prediction than sublethal endpoints as hypothesized.

Multiple discriminant analyses indicated that variability for each WET endpoint (independent of stream factors) was related to a different compliment of variables. However, number of WET tests performed by a discharger was consistently a significant factor across all endpoints (**See Figure 2 "Relationship between Whole Effluent Toxicity Test Failure Rate and the Number of Tests Conducted by a Facility").** This result could be due either to a higher probability of discerning toxic events with more frequent testing or to chance factors inherent in the tests themselves.

In general, the frequency and magnitude of WET "failures" at a site were unrelated to instream conditions regardless of the WET endpoint. Among those sites in which WET tests consistently failed, there was approximately a 50:50 chance of the stream being impaired. However, if WET consistently passed current state criteria, there appeared to be a higher likelihood of no instream impairment. Thus, we observed only partial confirmation of our hypothesis (H_o #2) that WET and stream biota were more related when WET test endpoint variability is low. One major factor affecting this relationship was habitat quality. Streams with good-excellent habitat were more likely to show clear relationships between WET and stream biota, even for chronic WET endpoints **(See Figure 3 "Degree of Instream Impairment Observed for Facilities as a Function of the Instream Habitat Quality")**.

A second factor was effluent dilution. Discharges that comprised > 60% of the stream flow under low flow conditions had a higher probability of failing WET criteria **(See Figure 4 "Percent of Facilities Failing their Whole Effluent Toxicity Requirements as a Function of Design Effluent Instream Waste Concentration (IWC)")**, probably due to more stringent WET limits as dilution decreases. Thus, as we hypothesized, (H_o #3), WET was more related to stream biota in effluent dominated systems.

A third factor was the type of pollutants discharged. Effluents with metals or ammonia exhibited poor relationships with stream biota. **(See Figure 5 "Example of Whole Effluent Toxicity Test Data and Bioassessment Data From a Facility With Ammonia as the Primary Pollutant of Concern")**. Effluents even with relatively high metal concentrations, and that failed WET criteria, tended to be no more associated with non-impairment than expected by chance alone. Thus, we were unable to confirm our hypothesis (H_o #4) that WET was more strongly related to stream biota for effluents containing metals as compared to effluents containing ammonia.

Analysis of the data gathered for the Santa Ana River indicated that certain WET test endpoints could reliably predict instream impairment. However, any single WET test could not provide sufficient predictive accuracy, precision or certainty. As demonstrated for the national data set as a whole, the Santa Ana River sites indicated that the fathead minnow survival endpoints (acute and chronic) were reliable predictors of real-world stress on the aquatic ecosystem. The fish growth endpoint did not add any additional predictive value. Results from a series of WET tests were statistically-correlated with chemical data to identify the most probable cause of toxicity. The fathead minnow tests were particularly well-suited to the Santa Ana River because they are resident to the Santa Ana River and many other arid west streams, and they are especially sensitive to chlorine and ammonia, the two most common pollutants discharged to the Santa Ana (at the time of these studies).

Results from WET tests using *Ceriodaphnia dubia* were not so clear. High hardness and ionic strength, common to many arid west streams, make it very difficult to run a properly controlled toxicity experiment with this species. As this species is not resident to the Santa Ana River, it is not known whether *C. dubia* serves as a reliable indicator species for the invertebrate fauna which do reside there. Finally, the poor habitat quality found in arid west streams may simply overwhelm the relative influence of water quality. Shifting sand stream bottoms provide little opportunity to develop rich and abundant invertebrate populations. Nevertheless, *Ceriodaphnia* acute mortality and chronic survival endpoints provided some indication of ecosystem stress. Chronic reproduction, however, provided no reliable indication of either impairment or attainment of uses. This was consistent with the findings from other sites in the national database. Results for the Santa Ana case study suggest that arid west systems may be fundamentally different from those in wetter regions of the U.S. but that relationships between WET and stream biota in arid systems share much in common with systems in other regions (H_o #5).

CONCLUSIONS AND RECOMMENDATIONS

We were unable to demonstrate significant agreement between WET and stream biological assessments in our analyses. Several factors were important in determining whether a given WET history resulted in biotic impairment or not. First, we observed that facilities with less effluent dilution instream had a greater probability of impairment downstream, particularly if they failed at least 25% of their tests and if the instream habitat quality was at least fair. Facilities that failed a smaller percentage of their tests, or that comprised $\leq$ 70% of the stream flow under design flow conditions, typically exhibited little or no relationship between WET and instream conditions. Pollutant type may play a role as well; facilities with either metals or ammonia may not exhibit significant relationships between WET and instream results presumably due to differences in bioavailability.

Ceriodaphnia acute endpoints were especially poor in terms of predicting stream biological conditions because effluents that had greatest toxicity using this test also tended to have greater dilution instream. Of all WET endpoints, the fathead minnow acute percent mortality and chronic NOEC or LOEC (based on survival) demonstrated the most significant relationships between WET and instream conditions. However, a weight of evidence based on the results of both species chronic tests appeared to be more predictive than either species endpoint alone. Throughout our analyses, we were less accurate at predicting stream impairment than non-impairment using WET and other facility information.

We examined several alternate WET compliance criteria in this research and determined that standard WET pass/fail criteria currently in use, could portray or predict instream effects if implemented in several specific ways. First, a specified test fail rate should be used (e.g., 25%), rather than a single test failure or an average endpoint value, to determine facility compliance. Using a test fail rate implies that several WET tests will be conducted over a one-year period. Based on our analyses of WET endpoint fail rates, 4 to 6 tests over a year should be considered a minimum test frequency unless consistent effluent compliance (test passes) has already been demonstrated. In a recent essay, Mount (1998) expressed the view that WET testing needs to be more frequent than typically required to more accurately determine toxicity potential. Our results support this contention, however, as discussed previously, the inherent variability in WET test methods needs to be defined first to avoid unacceptable Type I or II errors. Mount (1998) also suggested expressing WET limits as an average rather than a single value. While our results indicate that basing a WET limit on an average value is more appropriate than basing compliance on a single WET test, our analyses suggest that use of a percent fail rate (e.g., 25%) is a more effective predictor of effects on stream biota.

Second, unless sufficient WET information has already been demonstrated for a facility, initial testing should be conducted using both species rather than only one species. In addition, it may be appropriate to average WET results for a given type of test endpoint (acute or chronic) to determine effluent compliance, unless more detailed information is available.

Third, sublethal chronic endpoints (growth or reproduction) were rarely if ever strongly correlated with stream biological conditions. Chronic survival endpoints were more informative and more reliable predictors in this study.

Fourth, the use of an NOEC and design IWC for chronic test compliance appears to be overly stringent in most cases. Using instead the LOEC and design IWC, or better yet, the actual effluent concentration in the receiving stream during the time of testing, are likely to make WET results more predictive of effects on stream biota.

Fifth, for acute WET testing, fish survival (or mortality) in undiluted effluent may be a reasonable predictor of effects on stream biota and at least as useful as an LC_{50} value. This was not true for *C. dubia* in our dataset.

Sixth, test method variability among tests over time, independent of the effluent being tested, must be determined in order to accurately identify significant changes in effluent toxicity over time and, therefore, to accurately predict potential effects on biota.

Seventh, some form of *in-situ* or ambient toxicity monitoring, or biological assessment, should accompany standard WET testing to determine the degree to which laboratory tests accurately portray pollutant bioavailability instream. This additional form of monitoring is especially desirable if WET test results indicate unacceptable effluent toxicity. Those effluents that indicate consistent, non-toxic results using standard WET tests are less likely to cause undesirable effects on biota and therefore, may not require additional field monitoring.

ACKNOWLEDGMENTS

We wish to thank the Project Subcommittee Penny Firth (Chair), Rodney Cruz, Tom LaPoint, Ben Parkhurst, Irwin Polls, Jeff Moeller for their helpful comments and suggestions during this project. Tim Moore of Risk Sciences analyzed and interpreted the Santa Ana case study data and provided stimulating feedback for the project as a whole.

Funding for this project was provided by the Water Environment Research Foundation, grant number 95-HHE-1.

REFERENCES

Cuffney, T., M. Gurtz, and M. Meador. (1993). Methods for collecting benthic invertebrate samples as part of the National Water-Quality Assessment (NAWQA) Program. U.S. Geological Survey Open-File Report 93-406.

Dickson, K.L., W.T. Waller, J.H. Kennedy, and L.P. Ammann**.** (1992). Assessing the relationship between ambient toxicity and instream biological response. *Environ. Toxicol. Chem.* 11:1307-1322.

Dickson, K.L., W.T. Waller, J.H. Kennedy, L.P. Ammann, R. Guinn, and T.J. Norberg-King**.** (1996). Relationships between effluent toxicity, ambient toxicity, and receiving system impacts: Trinity River dechlorination case study. *In* Grothe, D.R., K.L. Dickson, and D.K. Reed-Judkins (eds.). *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving Stream Impacts*. SETAC Press, Pensacola, FL.

Eagleson, K., D. Lenat, L. Ausley, and F. Winborne**.** (1990). Comparison of measured instream biological responses with responses predicted by *Ceriodaphnia* chronic toxicity tests. *Environ. Toxicol. Chem.* 9:1019-1028.

Gibson, G., M. Barbour, J. Stribling, J. Gerritsen, and J. Karr. (1996). Biological criteria: technical guidance for streams and small rivers. EPA/822/B-96-001. Office of Science and Technology, Health and Ecological Criteria Division, U.S. EPA, Washington, DC.

Grothe, D.R., K.L. Dickson, and D.K. Reed-Judkins (eds.)**.** (1996). *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving Stream Impacts*. SETAC Press, Pensacola, FL.

Mount, D. 1998. Mid-course corrections in WET Testing Program. SETAC News, Vol. 18, Number 2, pp 19-20. Society of Environmental Toxicology and Chemistry, SETAC Press, Pensacola, FL.

Marcus, M. and L. McDonald**.** (1992). Evaluating the statistical bases for relating receiving water impacts to effluent and ambient toxicities. *Environ. Toxicol. Chem.* 11:1389-1402.

Plafkin, J., M. Barbour, K. Porter, S. Gross, and R. Hughes**.** (1989). Rapid bioassessment protocols for use in streams and rivers. Benthic macroinvertebrates and fish. EPA/440/4-89-001. Office of Water Regulations and Standards, USEPA, Washington, DC.

Sprague, J.B. (1995). Review of methods for sublethal aquatic toxicity tests relevant to the Canadian metal-mining industry, and design of field validation programs. Prepared for Aquatic Effects Technology Evaluation Program, Canadian Centre for Mineral and Energy Technology, Department of Natural Resources Canada, Ottawa.

U.S. EPA**.** (1991). Technical Support Document for Water Quality-Based Toxics Control. EPA/505/2-90-001. U.S. EPA Office of Water, Washington, DC.

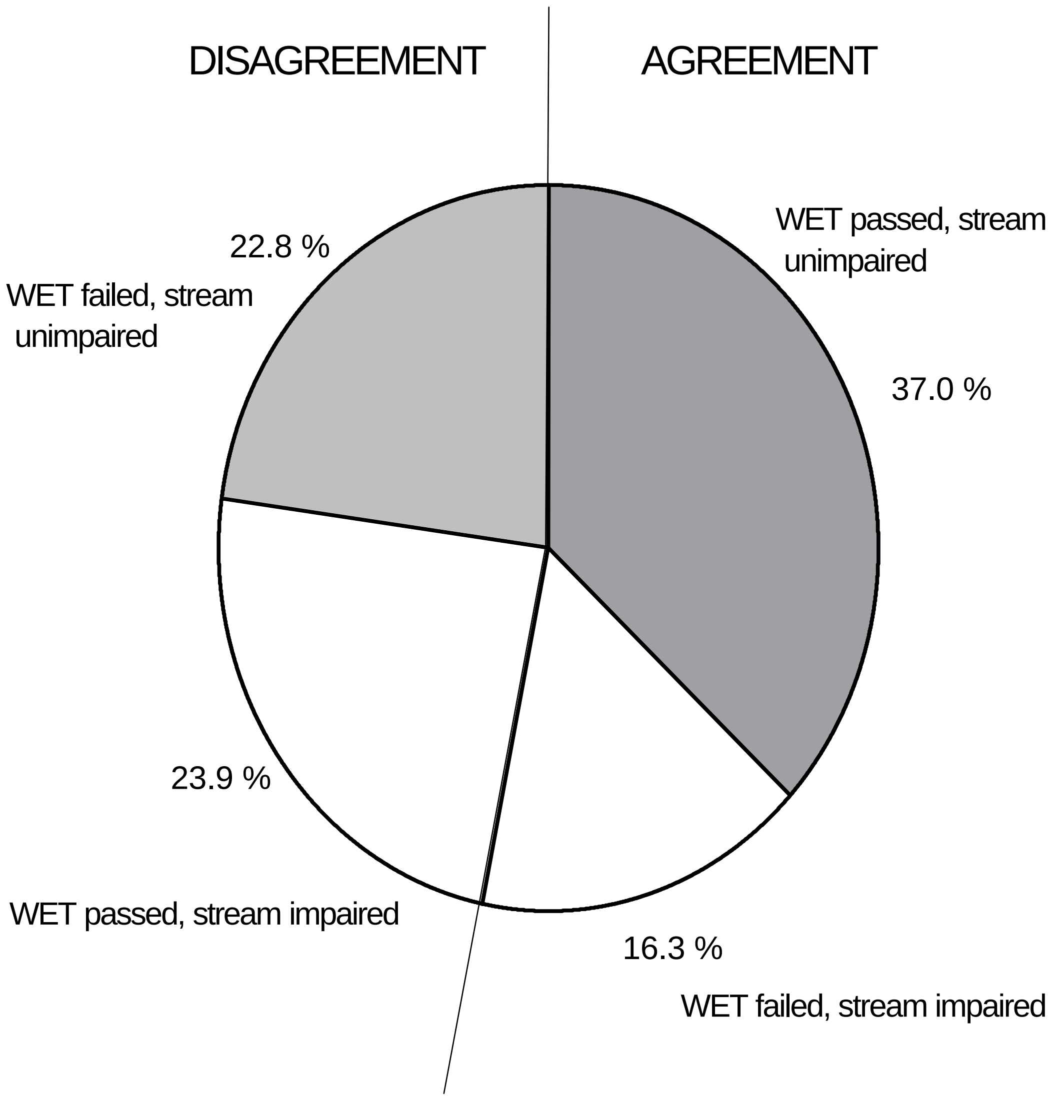


Figure 1. Percent Agreement Between WET Test and Instream Biological Condition Results

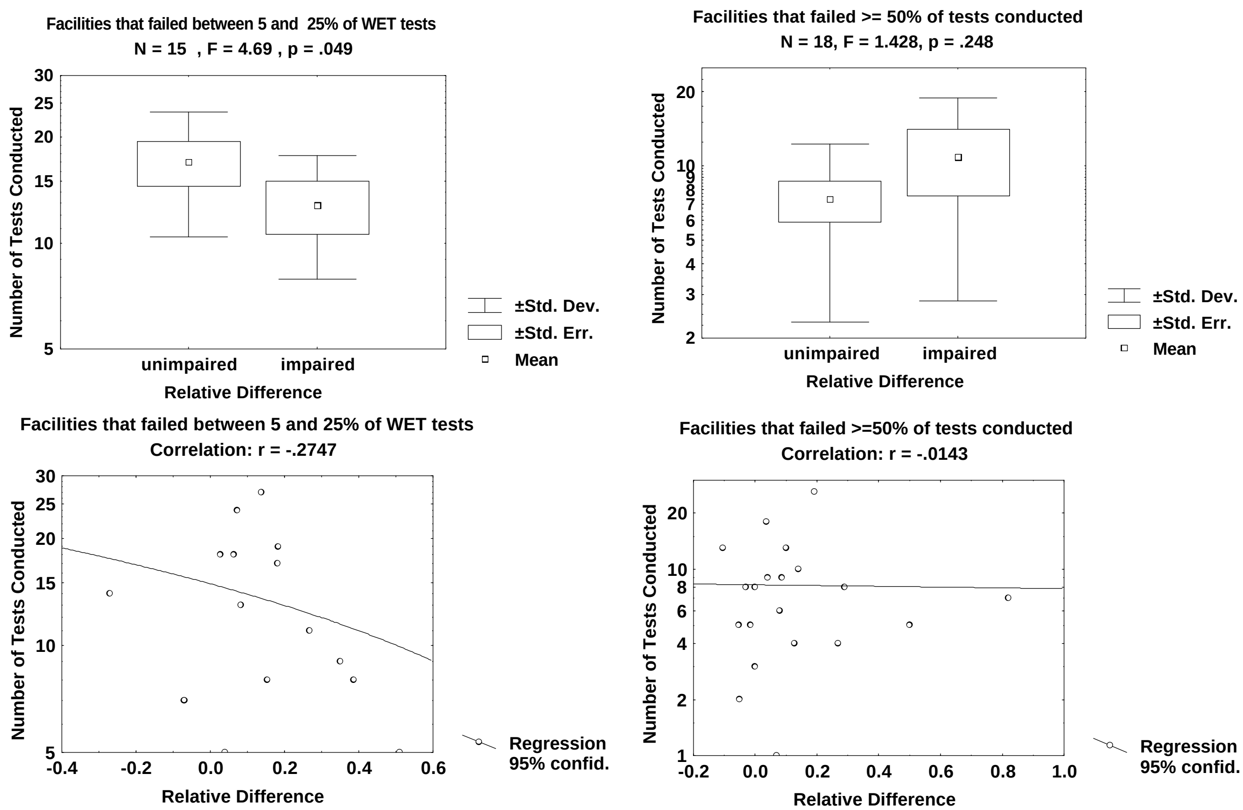


Figure 2. Comparison Between the Percent WET Tests Failed and Instream Biological Condition

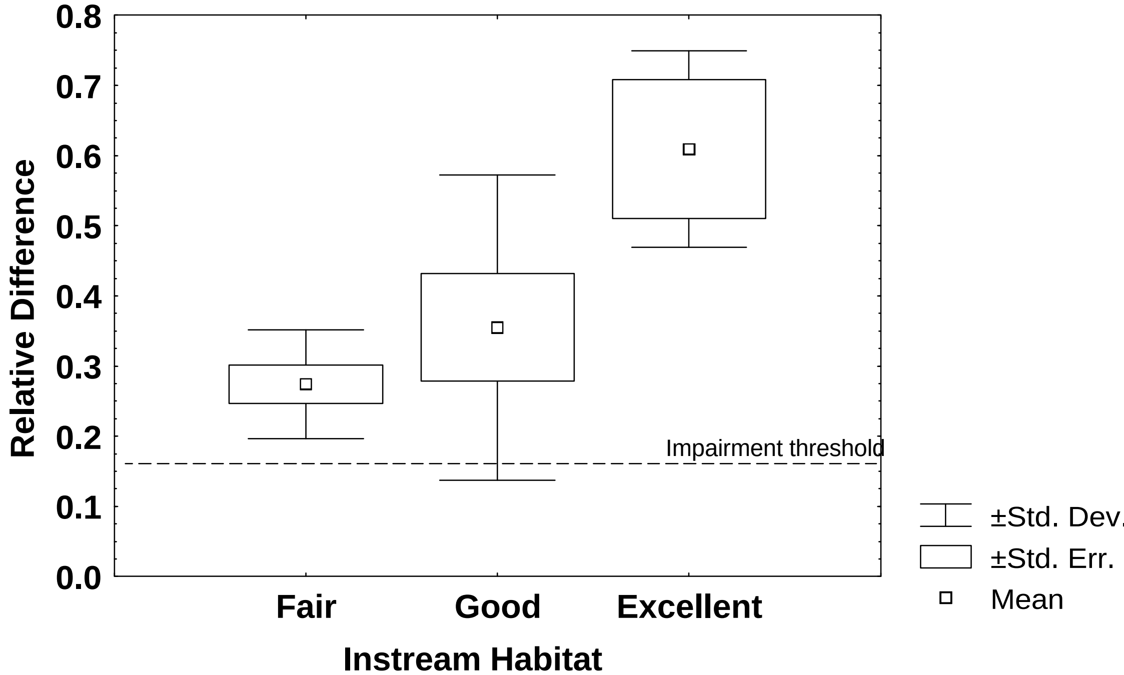


Figure 3. The Effect of Habitat Quality on the Observed Relative Difference Values for Facilities That Failed at Least 15% of Their WET Tests and for Which Downstream Impairment was Observed

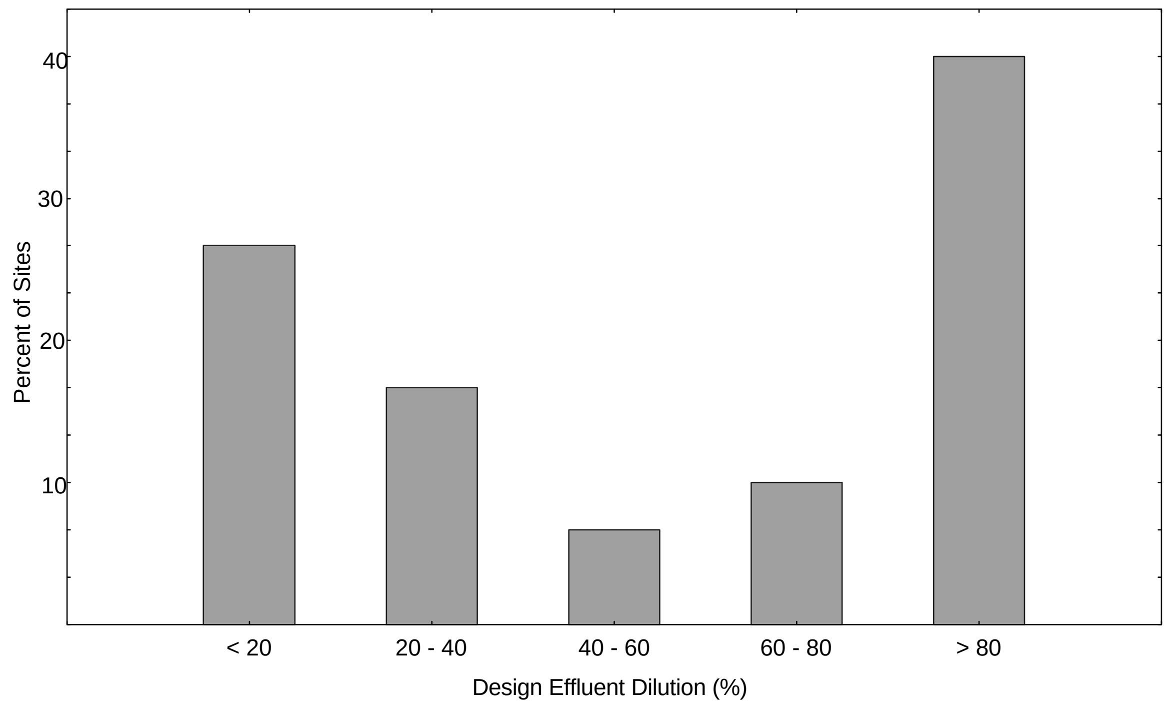
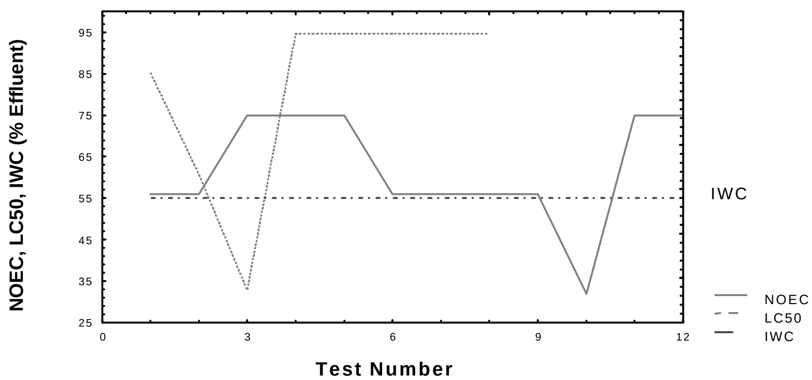


Figure 4. Percent Agreement Between WET Test and Instream Biological Condition as a Function of Instream Waste Concentration



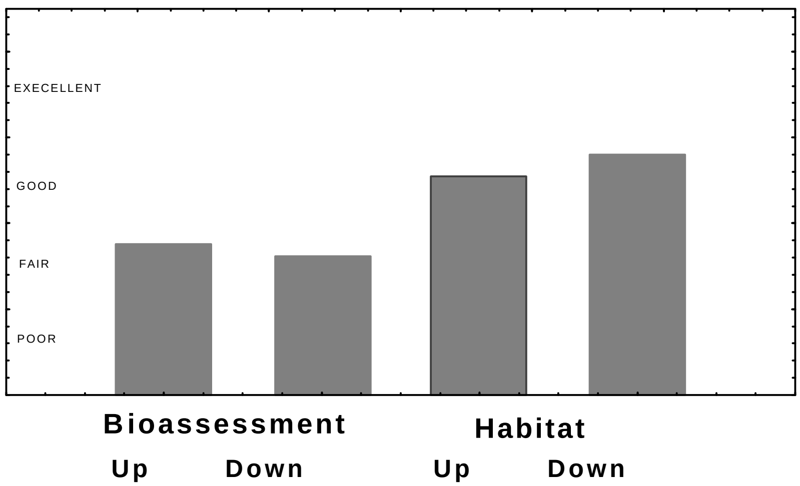


Figure 5. Example of Data for a Facility having Ammonia as the Chief Pollutant