Articles Fail To Note Variability of WET Testing

Three articles in the October 1998 ***Water Environment & Technology*** dealt with the subject of analytical variability in [whole effluent toxicity] WET testing ["Agreement on WET Test for Clean Water Permits Reached," p. 8; "Survey Provides 'Nationwide' Picture of WET Testing Program," pp. 18-19; and "WET Test Methods Settlement Gives EPA 3 Years to Make Changes," p. 80]. Like all other laboratory measurements, some degree of variability in WET testing can be expected, which in turn can occasionally produce false positives. Although not as widely recognized, analytical chemistry data are often as, or more, variable (for example, see Grothe ***et al.*** 1990. ***Water Environ. Technol.*** 2:62–68).

However, the WESTCAS [Western Coalition of Arid States] study, cited by Mark T. Pifher in October's Legal Perspective article, overestimates the true rate of WET false positives. This is because of several serious flaws in the design and conduct of the study. The "nontoxic" synthetic fresh water (SFW) distributed to laboratories for testing with the three-brood ***Ceriodaphnia*** test may have actually been slightly toxic, abrogating the purpose of the study. The primary reason for this is that the investigators did not follow the methods described in 40 *CFR* 136 for preparation of moderately hard SFW. For example, rather than using ASTM [American Society for Testing and Materials] Type 1 deionized water to make up the SFW, as specified in EPA/600/4-91/002, a lower quality Type II-III hybrid water was used and incorrect' quantities of salts were added.

The investigators attempted to certify that the water was nontoxic through chemical analyses. This is conceptually difficult because the reason for WET testing is that toxic chemicals or mixtures may be present at toxic concentrations that are not detected by conventional analytical methods designed primarily for priority pollutants. For example, surfactants would not be identified by the methods used in the study yet can be quite toxic, with chronic effects at concentrations as low as 100 μg/L.

The detection limits for the analyses of the deionized water were not sufficiently low enough to detect some of the pesticides, which can be toxic in the low nanogram-per-liter concentrations. Similarly, based on the metal concentrations detected and the detection limits for those not detected (for example, 5 μg/L for [nickel]), additive metal toxicity cannot be ruled out. The limited chemical analyses (volatile organics and metals) performed on the "nontoxic" SFW identified bis(2-ethylhexylphthalate) (30 [to] 70 μg/L) as not a problem because it "is not toxic," even though the daphnid chronic values for the [carbon 6 to carbon 13] diakyl phthalates fall in the range of 42 [to] 150 μg/L (Group 1986. ***Environ. Health Perspect.*** 65:337–340).

Problems also existed in the way in which samples were stored and distributed. Samples were stored up to 2 weeks in soft polyethylene cubitainers; effluent samples are never stored for more than 72 [hours] in these same containers. The participating permittees were instructed to label the containers "reference toxicant samples," that is, the laboratories anticipated toxicity and thus were not testing "in the blind." The investigators did not include the (passing) results of the test conducted on the same sample in their own lab because they were not "tested in the blind" (they anticipated nontoxicity), further inflating the ratio of "toxic" test results. Samples should have been labeled differently to avoid subconscious bias by technicians.

Finally, if the study rate of false positives (50%) were correct, then one-half of all permittees required to conduct the threebrood ***Ceriodaphnia*** test would be in [toxicity reduction evaluation]! This obviously is not the case. The WESTCAS study is best interpreted not as a [method detection limit] study, but a biased interlaboratory variability study conducted with a marginally toxic sample. Only reproduction, not survival, was affected, and all but [two] out of the 22 tests (9%) reported NOEC [no observed effect concentration] values equal to either 100% "nontoxic" SFW or the next lowest dilution (50%), as would be expected for a slightly toxic sample. When the data are pooled for all valid tests, a significant ($p=0.003$) depression in reproduction is observed in 100% "nontoxic" SFW Test results were consistent within all labs conducting multiple analyses. It is not possible to discern from the data, but if the [two] tests with the low NOEC values were conducted by the same lab, then the out-of-control lab rate is only [0.07] or 7%, essentially 5% for sample size of 14.

The authors propose the use of 99%, rather than 95%, confidence levels to avoid "false positives" and suggest that each discharger submit several blanks, duplicates, and reference spikes annually to their labs. However, as the WESTCAS study shows, it is difficult to prepare good blanks and most permittees do not have the equipment, supplies, or expertise to do so. This would only unduly increase the total cost of the tests and, most likely, lead to a lot of frustration and anger for all involved.

The solution to the whole issue is actually rather clear-cut. Interlaboratory variability (and hence false positives) can be best reduced by increasing standardization of test methods for all analyses (chemical, biological, and physical) and ensuring that laboratories have the correct equip ment and facilities, as well as adequate training and experience, to conduct the analyses and that they strictly adhere to test standards.

Peter ***F.*** *De Lisle*
President
Coastal Bioanalysis Inc.
Gloucester, Va.

Timothy F. Moore, president of Risk Sciences in Brentwood, N.J.; Steven P. Canton, vice president of Chadwick Ecological Consultants in Littleton, Colo.; and Max M. Grimes, vice president of Regulatory Management Inc. in Colorado Springs, Colo., respond: Dr. De Lisle raises many important questions in his review of WESTCAS' WET Method Blank Study. Our results are so startling (even to us) that it is easy to understand why others assume there must be "serious flaws" in the research. But, the scientific evidence strong ly refutes Dr. De Lisle's suppositions.

By way of review, 17 labs analyzed 25 nontoxic sample blanks using ***Ceriodaphnia dubia.*** There were no test failures reported based on survival. Nine of 23 valid tests (39%) reported toxicity based on reduced reproduction. Six of the nine failures reported IC-25 [25% impairment concentration] estimates greater

than 2 TU_c [chronic toxicity units]. The NOEC endpoint did somewhat better: seven failures in 23 valid tests, [and] three of the seven were greater than 2 TU_c. Legally, all nine WET test failures may constitute potential permit violations regardless of the magnitude of failure. Copies of the study results and methodology can be downloaded at http://www.toxicity.com.

The most critical question is whether the SFW used in the study might be toxic. Two of three labs testing the sample said it was not. If the sample water was "marginally toxic," as Dr. De Lisle claims, then the most sensitive tests would identify it first. Quite the opposite occurred. There was a statistically significant positive correlation between test sensitivity and the probability of passing the test (p=.03, r^2=60%, n=23).

Dr. De Lisle suggests that our sample water was "slightly toxic" because it did not conform to ASTM Type-I standards. EPA guidance recommends Milli-Q "or equivalent" water. Our production process did produce equivalent (or better) water. We used activated carbon, deionization, and microfiltration to produce presumptively nontoxic water. Subsequent testing by the chemistry lab, using more sensitive instrumentation, confirmed that its treatment system produces 18 megaohm water that meets ASTM Type-I standards.

EPA guidance strongly recommends relying on dilution water that is "routinely used with success in the laboratory" (EPA/600/4-91/002 @ p. 11). We chose the sample water based on our long history of using it to culture *Ceriodaphnia* and meet EPA's test acceptance criteria for control performance. Our lab has consistently passed EPA's DMR-QA [Discharge Monitoring Report Quality Assurance] tests each of the last 7 years. If our dilution water was "marginally toxic," there would be a tendency to underestimate the true toxicity of EPA's check samples. That hasn't happened.

Dr. De Lisle theorizes that trace levels of pesticides or heavy metals may have caused the observed "toxicity." In prestudy testing, we identified these potential problems and revised our procedures to reduce the risk by pre-rinsing sample bottles. There was no evidence of "toxics in toxic amounts" in the water that was shipped to the bioassay labs. There is a statistically significant concentration-response relationship when all of the valid data is pooled (p=0.008; r^2<0.5%; n=1361). This means 99.5% of all the observed differences in reproduction were caused by factors unrelated to our sample even if we assume it was "slightly toxic." That is not nearly a strong enough relationship to explain the large number of false positives reported.

Finally, Dr. De Lisle suggests that bioassay technicians may be subconsciously biased toward finding toxicity when sample bottles are labeled "reference toxicant." How does the technician's opinion influence *Ceriodaphnia* reproduction or the mathematics of statistical analysis? The possibility of "subconscious bias" makes it essential to evaluate laboratory performance in the blind.

Scientific objectivity depends on the premise that the labs are not aware that they are being evaluated. In our study, we also did as Dr. De Lisle suggested. The samples were labeled with a wide variety of different identifiers, including: "reference toxicant," "effluent," "process control sample," and other less meaningful alphanumeric descriptors. Systematic bias is unlikely.

Dr. De Lisle was right about two things. He stated that if our study results were true, then half the dischargers around the country would be performing TREs [toxicity reduction evaluations]. They are. Random low level chronic test failures have triggered many costly and unproductive investigations. Two new Water Environment Research Foundation studies also document an unexpectedly high rate of inconclusive TREs and statistical anomalies in WET testing throughout the nation.

Second, Dr. De Lisle stated that it is "difficult to prepare good blanks." Yet, dischargers whom he claims lack the "equipment, supplies, and expertise to do so" are legally required to produce hundreds of millions of gallons of nontoxic wastewater every day. Dr. De Lisle is correct in concluding that this leads "to a lot of frustration and anger for all involved." If water which has been deionized, treated with activated carbon, and microfiltered cannot be expected to pass EPA's toxicity tests, it is no wonder angry and frustrated [publicly owned treatment works] filed suit challenging the standard methods for WET.

To settle the lawsuit, EPA agreed to conduct a large-scale study of interlaboratory variability in WET testing. This will provide an excellent opportunity to replicate our results. In addition, we have asked the Society of Environmental Toxicology and Chemistry (SETAC) to initiate an independent peer-review of the WESTCAS WET Method Blank Study. Their recommendations, along with Dr. De Lisle's, will provide invaluable assistance as EPA prepares to conduct its own study.