14 September 1998

Bill Telliard
Engineering and Analysis Division (4303)
OW/OST
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, D.C. 20460

RE: Comments on EPA's Proposed "Charge to Reviewers: Interlaboratory Study of WET Methods."

Mr. Telliard:

On behalf of the Western Coalition of Arid States, we thank EPA for providing us an opportunity to comment on your proposed design for the interlaboratory WET variability study. Many members of Westcas reviewed the document; I was asked to summarize all of the suggestions in this one letter.

1) The proposed methodology should provide more detail.

In order to assess the scientific validity of the proposed study, it is necessary to know more about many critical elements of the experimental design. For example:

(a) How will the specific test matrices (industrial, municipal and blank) be formulated? Will the municipal effluent matrix be formulated so as to represent the average water chemistry (hardness, alkalinity, ionic strength) of natural surface waters in the arid west?

(b) What reference toxicant(s) will be used with each matrix? We recommend that EPA consider using a metal salt rather than sodium chloride in order to better simulate the toxicants most likely to occur in effluent.

(c) What concentration of toxicant will be used? Will the concentration vary between matrices? Will different target concentrations be used for acute trials vs. chronic trials? Will the concentration specified for chronic trials be set at a level expected to induce lethal effects or only sub-lethal effects?

(d) What control water formulations will be allowed or disallowed (moderately hard reconstituted lab water vs. dilute mineral water)?

(e) What dilution series shall be used for each test trial?

2) The data analysis phase should be fully described.

(a) What data acceptance/rejection criteria will EPA apply to reported lab results?

(b) How will the coefficient-of-variation be calculated for each WET method? What measure of variability will be used for rank-order, semi-continuous variables such as NOEC?

(c) Will data derived from evaluating different effluent matrices be combined? If so, how?

(d) Will data derived from method blanks be combined with data from reference toxicants? If so, how?

(e) Will confidence intervals be calculated for the estimated coefficients-of-variation?

(f) If data is not normally-distributed, what transformations (if any) will EPA apply?

(g) How will EPA estimate the true rate of Type-I errors (false positives) for each method?

(h) How will EPA identify the various sources of variability? And, how will the relative contribution of each source be quantified?

(i) How will EPA test for the presence or absence of systemic bias within each method?

(j) How will EPA assess the presence or absence of a dose-response relationship in test results reported by the participating labs?

(k) Does EPA intend to state formal, testable research hypotheses to guide the study design and conduct? What critical value shall be used for alpha when testing each hypothesis statistically?

3) Quality assurance and control procedures should be more specific.

(a) Tests should be scheduled to ensure that sample identities remain "blind" to the participating laboratories. For example, if all municipal effluents are evaluated in one week, and all blanks in a another week, the labs are likely to be able to deduce whether the sample is expected to be toxic or not. It would be better to schedule the labs to conduct all fathead minnow chronic tests one week, all Ceriodaphnia dubia chronic tests another week, all selenastrum tests in a subsequent week, etc. In any given week, the lab would be conducting three identical tests on three unknown samples (some toxic, some not).

(b) The study design states that EPA will provide "specific instructions" to the participating labs. This does not appear to be consistent with paragraph 7 of Exhibit B in the settlement agreement which states that EPA will "provide each participating laboratory with specific instructions in accordance with their routine practices using acceptable test methods in the Final Rule." If such instructions are to be utilized, they should be made available for public review and comment. Does EPA intend to prohibit practices which are normally allowed and left to laboratory discretion under current agency guidance? If so, how? Will EPA revise the methods to permanently disallow such discretion?

(c) What is the role of the "referee lab?" How will the referee lab be selected? Are the results reported by the referee lab presumed to be more correct than results reported by any other lab participating in the study?

(d) Will EPA specify the statistical tools which must be used or may each lab rely on the stat software of their own choosing?

(e) What are "acceptable control charts," as opposed to unacceptable control charts, in the list of prequalification requirements? What are the rejection criteria?

(f) How will laboratory "proficiency in the application of appropriate statistical analyses for each test…" be evaluated? What are the rejection criteria?

(g) How will the agency assure QA/QC for water chemistry of all sample matrices?

4) Some of the laboratory pre-qualification requirements are unreasonable.

(a) It is appropriate to require labs to be experienced and to conform to existing guidance. We believe the best way to assess both is by limiting participation to labs that have conducted at least 50 tests in the previous 24 months. We also believe it is appropriate to limit participation to those labs which EPA deemed "acceptable" in the two most recent DMR-QA studies for each method. It is also reasonable to request each lab provide a copy of their existing control charts. Together, these requirements (along with other test acceptance criteria) provide adequate assurance of laboratory expertise and experience prior to participating in the round-robin study. However, we do not object if EPA wishes to select their nine laboratories using pre-qualification tests (see item 3-d, below).

(b) Special prequalification tests unreasonably restrict participation in the study contrary to Section 4 of Exhibit B in the Settlement Agreement. Sponsors were asked to commit to funding three WET test for each method. Prequalification tests increase overall cost by 33%. The added cost is likely to cause many sponsors to withdraw from participation. Since there are other means to assure expertise and experience (described above) there is no legitimate reason to forego the benefits of substantially larger sample sizes for each method evaluated.

(c) Prequalification tests may bias the study results. If the purpose of the study is to identify and quantify the sources of variability in the test methods, it is inappropriate to use those same methods to pre-judge data. Censoring the data, by eliminating the most extreme estimates will intentionally underestimate the true range of error experienced in the real world.

(d) We recommend that EPA accept data from all participating laboratories and, if necessary, add a field into the project database which records and tracks any variable which the agency believes might provide substantial explanation for observed variability. In this way, EPA could distinguish whether there was any difference in variability between those labs that passed a prequalification test, those labs that failed a prequalification test and those labs that did not run a prequalification test.

(e) Prequalification tests, per se, are not part of the method as promulgated. Therefore, it is inappropriate for EPA to mandate requirements which go beyond existing WET guidance. To do so, eliminates the generalizability of study results to the general population of labs approved for conducting WET tests and assessing NPDES permit compliance. Prequalification tests are not authorized by the settlement agreement.

(f) EPA failed to identify the specific criteria or thresholds by which laboratories would be accepted or rejected for participation in the study. What constitutes "acceptable" vs. "unacceptable" performance in the prequalification tests?

(g) How can labs which are otherwise qualified to perform WET tests for NPDES permitting purposes not be qualified to participate in the interlaboratory variability study?

(h) Does EPA intend to adopt the prequalification criteria into rule for the purpose of establishing national certification standards for biomonitoring labs? If EPA believes that high quality estimates can only be obtained by prequalifying laboratories, then does the agency expect permittees run monthly prequalification tests prior to selecting a lab to analyze effluent samples for whole effluent toxicity? Are WET results acquired without such prequalification invalid?

(i) When disqualifying a lab, based on prequalification testing, how will EPA distinguish between imprecision which results from lab incompetence vs. imprecision which derives from the inherent variability of the test method? To use the methods under investigation to prequalify labs would seem to predestine the outcomes and beg the very question we're trying to answer.

5) The purposes of the study should be stated more accurately.

(a) The draft design states that the purpose of the study is to "evaluate the precision of each test method." While this is true, there were also several other purposes identified in the settlement agreement. These should also be explicitly included in the study design and charge to peer reviewers:

> *"…among other things, shall quantify the interlaboratory variability, i.e., to determine an estimate of precision, including, at a minimum, a coefficient of variation for each endpoint as well as to determine the rate at which participating laboratories successfully completed the tests initiated and the rate at which the tests indicate toxicity is present when measuring reagent water, also known as 'blanks'." (Settlement Agreement, Exhibit B, Sections 1 & 2)*

Westcas believes that the primary reason blanks were included in the study design was to assess whether the test methods correctly identify the blank samples as non-toxic. This is a question of accuracy not precision.

(b) Some sections of the study design set forth inappropriate arguments. For example, stating that "accuracy" is an inappropriate data quality objective for WET tests is a matter for the peer review committee to evaluate. While it is not possible to know the exact response any organism will make to a reference toxicant, it is possible to evaluate the probability of arriving at an inaccurate conclusion regarding the presence or absence of toxicity (false positives and false negatives). Therefore, the study should be designed to assess the "accuracy" of such conclusions. This is particularly true given the stated purpose to use WET test conclusions to judge permit compliance status.

Some sections may mislead peer reviewers. For example, the second to last sentence in the first paragraph on page A-1 indicates that the current methods are the result of "many years of development and testing…" While there was considerable testing leading up to the methods promulgated, the final procedures themselves were not validated in any formal round-robin test program. In general, it would be better if EPA refrained from making arguments and confined themselves to describing the objectives and methods in the study plan.

We strongly recommend that, whenever possible, EPA use the words and phrases previously agreed to in the settlement documents when crafting a charge to peer reviewers.

(c) In various other official documents (see Public Docket DCN 4 & DCN 2), EPA established a series of data quality objectives (DQO) which should be used to evaluate the validity of proposed test methods. EPA should enumerate each recommended DQO and describe how the proposed study will address the requirement as required by EPA Order #5360.1.

Some DQOs may not be entirely appropriate when evaluating biological methods. Nevertheless, where that is the case, EPA should state that the particular DQO is excluded from the study purposes and provide the supporting scientific rationale for that determination. The settlement agreement specifically calls for the peer reviewers to assess the applicability of other method performance characteristics and data quality objectives. Similar language should be added to the official charge.

(d) In several places, EPA states its intent to "minimize variability" in the study design. Since the primary purpose of the interlab study is to quantify variability and to identify the sources of variability, the agency must be careful not to artificially reduce such variance. For example, it is appropriate for the agency to take all reasonable steps to insure that each lab receives substantially identical aliquots of the same sample matrix. It is not appropriate for EPA to mandate the use of moderately hard reconstituted lab water rather than dilute mineral water for controls. Agency guidance assumes that the choice of control water does not bias WET tests in either direction. Therefore, EPA should not pre-empt the legitimate use of discretion authorized by existing guidance. To do so would result in interlab variability studies which failed to evaluate the methods "as promulgated" and, therefore, failed to implement the terms of the settlement agreement.

6) Questions to peer reviewers should be revised and expanded.

(a) The questions regarding precision estimates are too vague. In order to evaluate scientific appropriateness of the study design, peer reviewers must know what level of precision EPA desires to establish. The agency may do so by postulating a quantitative objective and asking whether the study design is capable of meeting that objective. For example:

"Is the study design, as proposed, capable of ascertaining whether x% of the labs will be within plus or minus y% of the median estimate z% of the time?"

Alternatively, the agency may ask the peer reviewers to state what level of confidence, probability of agreement, and percent error the suggested study design is capable of projecting. In all cases, EPA should ask the peer reviewers to assess whether the number of labs provides adequate power and sensitivity to meet the specific quantitative precision objectives sought.

(b) EPA should ask the peer reviewers to consider each of the DQOs identified in the agency's 304(h) Report to Congress and ask whether the study design is adequate to make the necessary demonstrations. Where EPA believes that some DQOs are inappropriate for biological methods, the agency should invite the peer reviewers to evaluate the supporting rationale. The settlement agreement requires EPA to ask the peer reviewers whether there are other appropriate data quality objectives which should be considered when designing and conducting the study.

(c) EPA should ask the peer review panel to enumerate all potential sources of variability in the proposed study. The panel should also indicate which sources should be controlled and which should be allowed to vary in order to meet the stated objectives of the study.

(d) EPA should ask the peer review panel to consider the other relevant study objectives. For example, "what is the actual (vs. nominal) probability that a single WET test result will yield an incorrect conclusion regarding the presence or absence of toxicity."

(e) EPA should ask the peer review panel whether the proposed study design adequately tracks the relevant variables so that the relative contributions to overall variability can be partitioned.

(f) EPA should ask the peer review panel whether the process of pre-qualifying labs, so as to eliminate extreme values, over- or under-estimates the true level of test variability in the field.

(g) EPA should ask the peer review panel whether the proposed methods for preparing and distributing the sample materials adequate to ensure the objective and "blind" nature of the study.

(h) EPA should ask the peer review panel to validate the agency's conclusion that "toxicity is relative rather than absolute" when known non-toxic samples (blanks) are tested.

(i) EPA should ask the peer review panel to identify any limitations on the generalizability of conclusions drawn from the proposed sample population and study design.

(j) EPA should ask the peer review panel to evaluate any requirement to reject or censor data for its potential to bias results or conclusions from the study.

(k) The settlement documents should be part of the indexed documents provided to the peer review panel.

(l) Too often, the peer review panel is asked to make subjective assessments about the "appropriateness," "scientific credibility," "usefulness," "acceptability," and "adequacy" of various elements in the study design. EPA should provide the panel with more explicit and quantitative criteria by which to evaluate each of the aforementioned qualities.

(m) EPA should ask the peer review panel to review all of the data analysis and interpretation issues identified in comments 2a-2k above.

(n) On page 3, EPA states that the "review is primarily concerned with an evaluation of the study design in terms of its ability to assess test method precision, not the application or implementation of the methods." This instruction may confuse the peer review panel given earlier statements about how the agency may use results from a single test to assess permit compliance and initiate enforcement actions.

The peer review panel should be asked to evaluate the degree to which (im-)precision impacts the ability of the agency to rely on WET test results for the purposes stated. We assume this is what the agency meant when they said that the "primary function of the peer reviewer should be to judge whether the study design and interpretation of the data can provide meaningful and useful precision estimates." The meaning and utility of precision estimates is entirely related to how test results are applied and implemented. Therefore, this portion of the charge should be revised to more accurately reflect that intent.

7) Additional references should be cited and provided to the peer review panel.

(a) EPA's *Manual for Evaluation of Laboratories Performing Aquatic Toxicity Tests*, January, 1991. (EPA/600/4-90/031)

(b) The Westcas/UWAG White Paper on issues related to whole effluent toxicity testing submitted to EPA on May 1, 1996.

(c) American Society for Testing and Materials standard D6091-97: *Standard Practice for 99%/95% Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error.* (conformance with consensus industry standards is required of all federal agencies under the National Technology Transfer and Advancement Act of 1995.)

(d) The summary report of results from the Westcas WET Method Blank Study including the comprehensive appendices describing the experimental design and QA/QC procedures adopted for that study.

Westcas is genuinely interested in seeing this interlaboratory study implemented successfully. The large number of water and wastewater agencies that agreed to sponsor additional laboratory tests is strong evidence of that commitment.

If you have any questions regarding the suggestions made in this letter, please contact Mark Pifher or Tim Moore at the addresses and phone numbers given below.

Thank you for the opportunity to review and comment on your proposed study design.

Respectfully submitted,

Timothy F. Moore (on behalf of Westcas)
Risk Sciences
1417 Plymouth Dr.
Brentwood, TN 37027
(615) 370-1655
(615) 370-5188 fax
tmoore@risksciences.com

Mark T. Pifher
Anderson, Dude, Pifher & Lebel, P.C.
104 S. Cascade Ave., Suite 204
Colorado Springs, CO 80903
(719) 632-3545
(719) 632-5452 fax
adpl@worldnet.att.net

cc: Steven Sweeney, EPA Office of General Counsel
Mark Pifher, Attorney for Westcas
Steve Koorse, Attorney for Edison Electric Institute