8020 QUALITY ASSURANCE AND QUALITY CONTROL IN LABORATORY TOXICITY TESTS*

8020 A. General Discussion

Quality assurance and quality control (QA/QC) are essential elements of laboratory bioassay procedures. A good QA/QC program provides framework and criteria for assessing data quality, including a well-defined chain of responsibility, explicit data quality objectives, procedures and protocols for testing, and a mechanism for identifying and correcting potential problems. Elements to be included in a quality assurance plan (QAP) are outlined in Section 1020A; other resources for developing a comprehensive QAP for laboratory toxicity testing programs are available.[1–7] As a minimum, QAPs for laboratories performing aquatic toxicity testing should provide specific guidance on data quality objectives, test procedures, sample handling, data management, internal quality control, and corrective action.

* Reviewed by Standard Methods Committee, 1997. Editorial revisions, 2001, 2009.
Joint Task Group: 20th Edition—David W. Moore (chair), Rob Cuello, William L. Goodfellow, Robert E. Morcock, Russell H. Plumb.

References

1. American National Standards Institute & American Society for Quality. 2004. Quality Systems Requirements for Environmental Programs. ANSI/ASQ E4-2004, Milwaukee, Wisc.
2. Ratliff, T.A., Jr. 2003. The Laboratory Quality Assurance System—A Manual of Quality Procedures and Forms. 3rd ed. Wiley-VCH, New York, N.Y.
3. U.S. Environmental Protection Agency. 2001. EPA Requirements for Quality Management Plans. EPA QA/R-2, EPA/240/B-01/002. Off. Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.
4. U.S. Environmental Protection Agency. 2006. Guidance on Systematic Planning Using the Data Quality Objective Process. EPA QA/G-4, EPA/240/B-06/001. Off. Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.
5. U.S. Environmental Protection Agency. 2001. EPA Requirements for Quality Assurance Project Plans. EPA QA/R-5, EPA /240/B-01/003, Off. Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.
6. Environment Canada. 1990. Guidance document on control of toxicity test precision using reference toxicants. EPS 1/RM 12, Environment Canada, Ottawa, Ont.
7. Environment Canada.1995. Guidance document on measurement of toxicity test precision using control sediments spiked with a reference toxicant. EPS 1/RM 30, Environment Canada, Ottawa, Ont.

8020 B. Elements of QA/QC

1. Data Quality Objectives

Data quality objectives (DQO) are either qualitative or quantitative statements describing the overall acceptable uncertainty in results or decisions derived from environmental data. Such objectives for evaluating toxicity must ensure that the information obtained will provide an accurate and precise estimate of environmental effects. They identify the types of measurements to be made, the allowable bias, and desired precision of measurements.

Accuracy is the degree of agreement between an observed value and the true value or an accepted reference value. For water quality parameters, a measurement of accuracy might include calibration against a known standard. For toxicity testing, a reference toxicant test (i.e., exposing the test organism to a contaminated matrix of known toxicity) can be used as a measurement of organism-response accuracy. *Precision* is the degree of agreement among repeated measurements collected under identical conditions; it usually is described by a measure of variance (e.g., variance, standard deviation, coefficient of variation). For toxicity testing, precision of organism response can be described in a control chart of responses to a reference toxicant.

If the response (e.g., survival) of test organisms exposed to a sediment/water sample is significantly different from the response to a reference or control, then the organism has been affected by the sample. Traditionally, decisions of statistical significance are made at $\alpha = 0.05$. This means that the probability of a false positive result (detecting a difference when in fact none exists) must remain below 5%. Data quality objectives must control levels of bias (i.e., the difference between the measured value and true value) and precision to ensure that statistical significance is not affected by measurement error.

Minimum data quality objectives should be provided for: water quality in the test chamber (e.g., temperature, salinity, alkalinity, hardness, dissolved oxygen, pH, and ammonia); frequency and acceptable limits; minimum control survival; sensitivity of test organisms (e.g., reference toxicant testing); and frequency and number of observations.

Limits for bias and desired precision generally are not stipulated in standardized test protocols described herein, but should be specified in the laboratory's own manual of standard operat-

ing procedures (SOPs). Performance criteria (e.g., acceptable levels for control survival or water quality measurements) for most of these categories may be found in the test protocols for the organism of interest.

2. Test Procedures

Test procedures describe how to make all routine measurements associated with toxicity testing and related QA/QC activities. Follow these procedures to ensure integrity and quality of data.

Use SOPs and standardized data forms to ensure quality and consistency of toxicological testing and reporting. Write SOPs for all routine laboratory activities and periodically review and update them. Examples of quality control checklists, project schedule lists, procedural checklists, and test and reference toxicant procedures are available.[1–5]

Steps taken in the laboratory to reduce the potential for bias include blind testing, random assignment of organisms to test chambers, statistical designs (e.g., the randomized block), procedures to prevent cross-contamination, confirmation and witnessing of recorded observations, use of reference toxicant tests, and control charting.

Blind testing, in which the experimental treatment is unknown to the analyst, prevents the analyst from applying biases upon the treatments based on any preconceived expectations.

Use randomized designs to eliminate bias due to test chamber position in the test array. The completely randomized block, in which treatments are allocated to experimental units at random, is the simplest form of the design. Each unit has an equal chance of receiving a particular treatment. In addition, process the units in a random order at all subsequent stages of a test in which order could affect results. For example, position test containers maintained in a water bath under a light source randomly within the testing area. When replicates receiving a single treatment are placed together, observed differences cannot be attributed solely to treatment; differences also may have resulted from placement. Discussions of randomized block design, completely randomized block design, and other statistical aspects of experiment design are available.[6–10]

During setup and conduct of toxicity tests, prevent contamination from an external source and cross-contamination between treatments. Preventive measures include cleaning equipment between contact with treatments, properly conditioning laboratory test apparatus to minimize leaching, and covering test chambers to minimize loss of volatiles or extraneous contamination. Preferably, also analyze food, dilution water, and control water/sediment periodically for background contamination.

Periodic double checks of observations, data entry, and calculations and witnessing of all raw data sheets (i.e., having a coworker review and sign each raw data sheet) are good preventive steps for early identification and correction of errors. Important preventive procedures should include counting animals twice to ensure accuracy before adding them to the test chamber and periodically confirming calibration and measurements, particularly if environmental factors seem to be out of range.

Use reference toxicant tests to assess sensitivity of test organisms. Plot results from reference toxicant tests on control charts (Section 1020B.13) to determine whether the sensitivity of test organisms to a given reference toxicant is within a predetermined range of acceptability. Construct control charts by plotting successive values (e.g., LC_{50}s), for a reference toxicant, and evaluating temporal changes in sensitivity. Recalculate the mean and standard deviation with each plot until the statistics stabilize. Evaluate individual values in relation to the mean and standard deviation. Procedures for developing and using control charts are described in detail in Section 1020B.13 and elsewhere.[11]

3. Sample Handling, Storage, and Shipment

Consistency in sample handling and tracking is most important when testing samples with possible legal ramifications. To make technically sound decisions that withstand potential litigation, analysts must handle samples appropriately and be able to trace them to their source. Key components of this QA/QC element include established chain-of-custody procedures as well as procedures for sample sieving, subdividing, homogenization, compositing, storage, and monitoring.

Chain-of-custody procedures require an unbroken record of possession of a sample from its collection through analysis or testing, disposal, and possibly up to and during a court proceeding.[12] The goals of chain-of-custody are twofold: to ensure that the collected sample was the one tested and to ensure that the sample has not been tampered with or altered in any way. Chain-of-custody can be accomplished via custody seals and sample tracking forms. Examples of such forms are available.[1,12]

a. Water and wastewater: Guidance for handling effluent samples under the National Pollutant Discharge Elimination System (NPDES) program dictates that samples are to be stored at 4°C and that no more than 36 h should elapse between sample collection and initiation of testing.[2] However, holding times may be adjusted depending on study objectives and other specific logistical considerations (e.g., shipment of samples from remote areas). If water samples are to be stored, keep headspace to a minimum. Record condition of sample (e.g., pH, temperature, salinity, conductivity, dissolved oxygen) upon receipt and prior to use. Before storage, floating debris may be removed, if necessary, by pouring water samples through a 2- to 4-mm mesh sieve. If there is a possibility of interference due to the presence of indigenous organisms that show predation, competition, etc., pass samples through a 60-μm mesh sieve.[2] However, if volatile contaminants are of concern, take care to minimize aeration during collection, handling, storage, and testing.

b. Sediment: Sediment samples may require sieving before testing. Sieving decisions are driven by the presence of debris, such as twigs or leaves, that may affect recovery of test animals at test termination and/or the presence of indigenous species in the sample that may serve as food for, compete with, or actually prey on the test organism. In any case, test results can be biased. If sieving is required, press-sieve all sediments without adding water (including reference and control sediments) before testing. In most cases, a 0.5-mm screen size is sufficient for removing predators, and larger sieves may be used to remove debris. Recommendations about sieving test material usually are found in specific standardized test protocols.

Depending on test objectives, samples may be composited, homogenized, and/or subdivided before testing. Use clean, noncontaminating containers and implements to handle and store samples. Suggested materials are stainless steel, TFE, Lexan®, high-density polyethylene, and glass. Other appropriate materi-

als may be specified. Homogenize sediments to a consistent color and texture. Samples may be homogenized by hand with a spatula made of noncontaminating materials, or by mechanical mixing. Verify efficiency of homogenization by chemical analysis.

Sediments frequently are stored before testing. Current guidance for dredged material evaluations permits pre-test storage of sediment samples for up to 8 weeks from time of collection.[13] Preferably store sediment samples at 4°C with zero headspace or under an inert gas, such as argon. Rehomogenize samples just before testing.

Maximum time limits for sediment storage prior to testing are of concern; test samples as soon after collection as possible.

4. Data Recording, Reduction, Validation, and Reporting

Quality control of recording, reducing, validating, and reporting data is necessary to produce complete, scientifically defensible reports. Issues to be considered include maintenance of laboratory notebooks, data management, reporting and validation procedures, identification and handling of unacceptable data and outliers, measurements of completeness and comparability, and procedures for data archival.

Standardization of data recording facilitates electronic transfer and manipulation of data. At a minimum, standardize procedures for intralaboratory data entry. Identify no-data entries with a mark ("—") to indicate that data were not omitted. Use abbreviations for names of personnel and routine laboratory observations to reduce data recording and entry time; standardize these whenever possible. Attach a list of definitions and code descriptions to data sheets and project files. Record data in indelible ink; make corrections by drawing a single line through the mistake, correcting the mistake, dating and initialing the correction, and writing an initialed explanation for the lined-out data in a footnote at the bottom of the data sheet. More detailed guidance on maintaining laboratory notebooks can be found elsewhere.[14]

Validate all original data at each level of transcription (e.g., entering data from bound laboratory notebooks into computer databases). Arrange for an independent QA/QC review of at least 10% of the data. Review laboratory records daily for outliers or unusual observations so any necessary corrective action can be taken.

Criteria for establishing outlier values are program-specific. Toxicity endpoint outliers, such as survival, growth, or reproduction, may be more important than water quality outliers. Depending on program requirements, identify outliers and either accept them as "real" or reject and selectively remove. If outliers are removed from a data set, note this and clearly justify the reason. For example, an outlier for mortality in a given replicate might be reasonably excluded from a data set when it is clearly related to spurious low dissolved oxygen. If there is no rational explanation for the outlier, it must be assumed that the value is real and representative of the test system's variability.

Completeness and comparability are measures of data quality. *Completeness* is a measure of the amount of data obtained versus the amount of data originally intended for collection. Generally 80 to 90% is an acceptable level of completeness for water quality data. However, endpoint data, such as survival or reproduction, should be 100% complete; otherwise, the statistical power of the test may be compromised. If data are less than 80% complete, use professional judgment to assess the data's usefulness for decision-making. *Comparability* is the confidence with which one data set can be compared to another. Comparability and confidence can be enhanced through interlaboratory calibration, including use of reference toxicants and control charts.

5. Internal Quality Control Checks

Internal quality control checks are "in-house" procedures implemented by the laboratory to ensure high-quality data. Internal quality control checks include reviewing documentation to determine that all samples are tested, sample holding times are not exceeded, holding conditions are acceptable, test protocols followed, instruments calibrated and maintained, and control survival and water quality conditions are within acceptable ranges. Other important issues are verifying the taxonomy and viability of test organisms.

Document source and culture history of test organisms. If possible, preserve a subsample of the test organisms for future identification in the event of aberrant toxicity. The age, size, and/or maturity of the test organisms usually are specified in the test protocol; verify these. Specify appropriate holding time and acclimation procedures either in the test protocols or the laboratory's SOPs; ensure that resulting documentation is available for audit.

Two widely accepted ways to assess test organism viability are the use of test-validation controls and reference toxicant tests. A *test-validation control* is a group of organisms that, with the exception of the treatment factor, are handled in a manner identical to the other organisms in the test. Acceptable levels of mortality in the test-validation controls for most acute lethality tests are limited to $\leq$10% (i.e., survival $\geq$90%). If less than 90% survival is achieved in the test-validation control, the test is considered invalid and must be repeated. For chronic sublethal tests, the test-validation control also may include acceptable limits for other endpoint data, such as growth and reproduction. Reference toxicant tests are designed to assess sensitivity to a specific contaminant. In a *reference toxicant test*, organisms are exposed to a range of concentrations of a single contaminant or contaminant mixture in water-only exposures, and an LC_{50} (usually 48 or 96 h) is calculated. Evaluate results of reference toxicant tests in a laboratory control chart (see 8020B.2).

Before testing, develop guidance for defining deviations, deficiencies, and appropriate corrective action. Corrective action may be required when a deficiency or deviation from planning documents or procedures is discovered or when there are deviations from established data quality objectives.

Deviations are data outside the range specified in data quality objectives. Out-of-compliance data may be due to deviations in test protocols or deficiencies associated with toxicological tests. Examples of deviations from the DQO in toxicity tests include excessive control mortality, out-of-range water quality conditions, lack of randomization, lack of required reference, control, and/or out-of-range reference toxicant results.

Poor control survival, loss of control over exposure conditions, major mechanical errors, or mishandling of test organisms may result in a decision to retest. However, brief episodes of out-of-range water quality conditions or incomplete test monitoring information may require only that data be flagged and

TABLE 8020:I. SUMMARY OF TYPICAL TEST DEVIATIONS AND NEED FOR RETESTING

Deviation	Need for Retesting	
	Required	Possible*
Lack of test array randomization		✓
Testing not blind		✓
Required references, controls not tested	✓	
Test chambers not identical		✓
Test containers broken or misplaced		✓
Mean control mortality exceeds acceptable limits	✓	
Excessive control mortality in a single replicate		✓
Test organisms not randomly assigned to test chambers		✓
Test organisms not from the same population		✓
Test organisms not all the same species or species complex	✓	
Test organism holding time exceeded		✓
Water quality parameters consistently out of range	✓	
Brief episodes of out-of-range water quality parameters		✓
Test monitoring documentation incomplete		✓
Sample holding times exceeded	✓†	
Sample storage conditions outside acceptable ranges		✓†

* If not retested, data may have to be qualified depending on study objectives.
† Unless evidence provided to clearly show that sample quality (physico-chemistry and contaminant levels) has not been affected.

qualified. A number of typical test deviations and suggested corrective actions are summarized in Table 8020:I.

Corrective actions may include, but are not limited to, reviewing the data and calculations, identifying and qualifying suspicious data, and retesting. Review all "out-of-limit" events as soon as data are tabulated and validated.

6. References

1. MOORE, D.W., T.M. DILLON, J.Q. WORD & J.A. WARD. 1994. Quality Assurance/Quality Control (QA/QC) Guidance for Laboratory Dredged Material Bioassays—Results of QA/QC Workshop Held May 26–27, 1993, in Seattle, Washington, Misc. Paper D-94-3. U.S. Army Corps of Engineers Waterways Experiment Sta., Vicksburg, Miss.
2. U.S. ENVIRONMENTAL PROTECTION AGENCY. 2002. Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms, 5th ed., EPA-821-R-02-012. Off. Science & Technology, Washington, D.C.
3. U.S. ENVIRONMENTAL PROTECTION AGENCY. 2000. Methods for Measuring the Toxicity and Bioaccumulation of Sediment-Associated Contaminants with Freshwater Invertebrates, 2nd ed., EPA-600/R-99-004. Off. Science & Technology, Washington, D.C.
4. U.S. ENVIRONMENTAL PROTECTION AGENCY. 1994. Methods for Assessing the Toxicity of Sediment-Associated Contaminants with Estuarine and Marine Amphipods, EPA-600/R-94-025. Off. Research & Development, Washington, D.C.
5. U.S. ENVIRONMENTAL PROTECTION AGENCY. 1991. Manual for the Evaluation of Laboratories Performing Acute Toxicity Tests, EPA/600-4-90/031. U.S. Environmental Protection Agency, Cincinnati, Ohio.
6. SOKAL, R.R. & F.J. ROHLF. 1994. Biometry—The Principles and Practices of Statistics in Biological Research, 3rd ed. W.H. Freeman & Co., New York, N.Y.
7. COCHRAN, W.G. & G.M. COX. 1992. Experimental Designs, 2nd ed. John Wiley & Sons, New York, N.Y.
8. GAD, S.C. & C.S. WEIL. 1988. Statistics and Experimental Design for Toxicologists. Telford, Caldwell, N.J.
9. HICKS, C.R. 1999. Fundamental Concepts in the Design of Experiments, 5th ed. Oxford University Press, New York, N.Y.
10. HURLBERT, S.H. 1984. Pseudo replication and the design of ecological field experiments. *Ecol. Monogr.* 54(2):187.
11. AMERICAN SOCIETY FOR TESTING AND MATERIALS. 2002. Manual on Presentation of Data and Control Chart Analysis, 7th ed. ASTM Manual Ser. MNL 7, Revision of Spec. Tech. Publ. (STP) 15D. American Soc. Testing & Materials, Conshohocken, Pa.
12. RATLIFF, T.A., JR. 2003. The Laboratory Quality Assurance System—A Manual of Quality Procedures and Forms, 3rd ed. Wiley-VCH, New York, N.Y.
13. U.S. ENVIRONMENTAL PROTECTION AGENCY & U.S. ARMY CORPS OF ENGINEERS. 1998. Evaluation of Dredged Material Proposed for Discharge in Inland and Near Coastal Waters—Testing Manual, EPA-823-B-94-002. U.S. Environmental Protection Agency, Off. Water and Dep. of the Army, U.S. Army Corps of Engineers, Washington, D.C.
14. KANARE, H.M. 1985. Writing the Laboratory Notebook. American Chemical Soc., Washington, D.C.