8020

Quality Assurance and Quality Control in Laboratory Toxicity Tests

Reviewed by Standard Methods Committee, 2014. Editorial revisions, 2022. Joint Task Group: David W. Moore (chair), Jason M. Conder, Scott Hall.

8020 A. General Discussion

Quality assurance and quality control (QA/QC) are essential elements of laboratory toxicity test procedures. A good QA/QC program provides a framework and criteria for assessing data quality, including a well-defined chain of responsibility, explicit data-quality indicators (DQIs), test procedures, protocols, and a mechanism for identifying and correcting potential problems. Elements to be included in a quality assurance plan (QAP) are outlined in Section 1020 A and Section 9020 A; other resources for developing a comprehensive QAP for laboratory toxicity-testing programs are available.[1–7] At a minimum, QAPs for laboratories performing aquatic toxicity testing provide specific guidance on DQIs, test procedures, sample handling, data management, internal QC, and corrective actions.

References

1. American National Standards Institute and American Society for Quality Control. Quality systems requirements for environmental data and technology programs; ANSI/ASQ E4-2004. Milwaukee (WI): American Society for Quality Control; 2004.
2. Ratliff TA Jr. The laboratory quality assurance system—A manual of quality procedures and forms. 3rd ed. New York (NY): Wiley-VCH; 2003.
3. EPA requirements for quality management plans; EPA QA/R-2, EPA/240/B-01/002. Washington DC: Office of Environmental Information, U.S. Environmental Protection Agency; 2001.
4. Guidance on systematic planning using the data quality objective process; EPA QA/G-4, EPA/240/B-06/001. Washington DC: Office of Environmental Information, U.S. Environmental Protection Agency; 2006.
5. EPA requirements for quality assurance project plans; EPA QA/R-5, EPA/240/B-01/003. Washington DC: Office of Environmental Information, U.S. Environmental Protection Agency; 2001.
6. Guidance document on control of toxicity test precision using reference toxicants; EPS 1/RM/12. Ottawa (ON): Environment Canada; 1990.
7. Guidance document on measurement of toxicity test precision using control sediments spiked with a reference toxicant; EPS 1/RM/30E. Ottawa (ON): Environment Canada; 1995.

8020 B. Elements of QA/QC

1. Data Quality Indicators

Data quality indicators (DQIs) 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 provides an accurate and precise determination 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. *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, or coefficient of variation). For toxicity testing, precision of organism response can be described in a control chart of responses to a reference toxicant (i.e., repeatedly exposing test organisms to set concentrations of a known toxicant).

If the response (e.g., survival) of test organisms exposed to a sediment or water sample is significantly different from the response to a reference or control treatment, 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 none exists) must remain $<5\%$. Data quality indicators must set precision and accuracy limits to ensure that statistical significance is not affected by measurement error.

Minimum DQIs should be provided for water-quality measurements in the test chamber (e.g., temperature, salinity, alkalinity, hardness, dissolved oxygen, pH, and ammonia), including 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 standard operating 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 the integrity and quality of data.

Use SOPs and standardized data forms to ensure the quality and consistency of toxicological testing and reporting. Write SOPs for all routine laboratory activities and periodically review and update them. Examples of QC 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.

a. Blind testing, in which the experimental treatment is unknown to the analyst, prevents the analyst from potentially applying biases to treatments due to 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, the experimental units should be processed in a random order at all subsequent stages of a test in which the order could affect results. For example, random assignment of test containers to positions within a water bath under a light source helps ensure that potential variations in lighting or temperature in the water bath don't affect results. Discussions of randomized block design, completely randomized block design, and other statistical aspects of experiment design are available.[6–10]

b. Contamination: Any material in contact with the samples, solutions, control water, organisms, or food must be nontoxic. While setting up and conducting toxicity tests, it is critical to prevent contamination from any external source and cross-contamination between treatments. Preventive measures include cleaning equipment between contact with treatments, proper conditioning of laboratory test apparatus to minimize leaching, and covering test chambers to minimize loss of volatiles and evaporation and prevent introduction of extraneous contamination. It is also generally recommended to periodically (at least annually) analyze food, dilution water, and control water or sediment for contamination.

c. Preventing procedural error: 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 to identify and correct errors early. Important preventive procedures 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.

d. Sensitivity: Use reference toxicant tests to assess sensitivity of test organisms (for frequently used or in-house cultured organisms, monthly reference toxicant tests are recommended). Plot results from reference toxicant tests on control charts (Section 1020 B.13) to determine whether the test organisms' sensitivity 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 evaluate 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 1020 B.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 legal review, analysts must handle samples appropriately and be able to trace the sample to its point of origin. Key components of this QA/QC element include established chain-of-custody procedures, as well as procedures for sieving, subdividing, homogenizing, compositing, shipping and transporting, storing, and monitoring samples.

Chain-of-custody procedures require an unbroken record of sample possession from point of collection through subsequent handling, storage, shipment or transfer for 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 collected sample has not been tampered with, or altered in any way. Chain of custody can be accomplished via use of 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 ≤ 36 h should elapse between sample collection and test initiation.[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, it is generally recommended that headspace be eliminated or at least minimized to the extent practical depending on the nature of the sample and the study objectives. Record sample characteristics (e.g., pH, temperature, salinity, conductivity, dissolved oxygen, and residual chlorine) upon receipt and before use. Ideally, dechlorinate samples in the field before transport to the lab (except when residual chlorine is the potential toxicant of concern) and check for residual chlorine if collected and transported by outside personnel. If residual chlorine is detected and dechlorination is appropriate, dechlorinate samples before storage or test initiation. If samples are dechlorinated (e.g., using anhydrous sodium thiosulfate), the testing program must include appropriate dechlorination controls (e.g., thiosulfate). Before storage, debris may be removed by carefully 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 or competition, pass samples through a 60-µm mesh sieve.[2] 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 the presence of indigenous species in the sample that may serve as food for, compete with, or prey on the test organism. 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. Also, consider which toxicants are likely to be present in the sample and whether sieving would likely change sediment conditions, such as pH and redox potential, that may alter the toxicant. This is particularly of concern for metals whose toxicity is valence-state dependent.

Depending on test objectives, samples may be composited, homogenized, or subdivided before testing. Use clean, noncontaminating containers and implements to handle and store samples. Suggested materials are stainless steel, PTFE, Lexan, high-density polyethylene, and glass. Other appropriate materials may be specified, providing that they do not affect sample toxicity. 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 the efficiency of homogenization by chemical analysis.

Sediments frequently are stored before testing. Current guidance for dredged material evaluations permits pretest 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. Re-homogenize samples just before testing. Maximum time limits for sediment storage before testing are of concern; test samples as soon after collection as possible.

4. Data Recording, Reduction, Validation, and Reporting

Quality control of data recording, reduction, validation, and reporting 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.

The standardization of data recording facilitates electronic transfer and data manipulation. At a minimum, standardize procedures for intralaboratory data entry. Identify entries for which no data exist with a mark ("—") to indicate that data were not omitted. Use abbreviations in lieu of the 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 foot-note on 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 on a minimum of 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 (e.g., 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 before analysis. 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 levels. 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 originally specified for collection. Generally, 80% to 90% is an acceptable level of completeness for water-quality data. However, endpoint data (e.g., survival or reproduction) must be complete; otherwise the test's statistical power may be compromised. If the 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 can be enhanced through inter-laboratory calibration, including use of reference toxicants and control charts.

5. Internal Quality-Control Checks

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

Document the source and culture history of test organisms. Identify organisms to species, record, and include taxonomic references used. If possible, preserve a subsample of the test organisms for future identification in the event of aberrant toxicity. The age, size, and maturity of the test organisms usually are specified in the test protocol; verify these. Organisms must be healthy, exhibiting good survival (e.g., >80%) preceding testing. 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. If animals are provided by a supplier, request documentation from the supplier. For cultured organisms, documentation includes culture history [original source, conditions under which they are reared, reference toxicant data, analysis of food and water supplies, and (for bioaccumulation test species) tissue residue analysis]. For organisms collected in the wild, request collection-related data (e.g., date, location, and water quality at time of collection), acclimatization, and shipping procedures. Confirm species by taxonomic identification before use.

Two widely accepted ways to assess test-organism viability are the use of test-validation controls (or negative controls) and reference-toxicant tests. A *test-validation control* is a group of organisms that, except for the treatment factor, are handled identically to the other organisms in the test. Test-validation controls for most acute lethality tests limit acceptable mortality levels to ≤10% (i.e., survival ≥90%). If <90% of the test-validation control survive, the test is considered invalid and must be repeated. For chronic sublethal tests, the test-validation control may also 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 8020 B.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 DQIs.

Deviations are data outside the range specified in the DQIs. Out-of-compliance data may be due to deviations in test protocols or deficiencies associated with toxicological tests. Examples of deviations from DQIs in toxicity tests include excessive control mortality, out-of-range water-quality conditions, lack of randomization, lack of required reference, control, and 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 qualified. A number of typical test deviations and suggested corrective actions are summarized in Table 8020:1.

Table 8020:1. Summary of Typical Test Deviations and Need for Retesting

Deviation	Need for Retesting: Required	Need for Retesting: Possible[a]
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	•[b]	
Sample storage conditions outside acceptable ranges		•[b]
Reference toxicant test results outside of control limits		•

[a] If not retested, data may have to be qualified depending on study objectives.
[b] Unless evidence provided to clearly show that sample quality (physicochemistry and contaminant levels) has not been affected.

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

References

1. Moore DW, Dillon TM, Word JQ, Ward JA. 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; Miscellaneous Paper D-94-3. Vicksburg (MI): U.S. Army Corps of Engineers Waterways Experiment Station; 1994.
2. Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms. 5th ed.; EPA-821-R-02-012. Washington DC: Office of Science and Technology, U.S. Environmental Protection Agency; 2002.
3. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. 2nd ed.; EPA-600/R-99-004. Washington DC: Office of Science and Technology, U.S. Environmental Protection Agency; 2000.
4. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods; EPA-600/R-94-025. Washington DC: Office of Research and Development, U.S. Environmental Protection Agency; 1994.
5. Manual for the evaluation of laboratories performing aquatic toxicity tests; EPA/600-4-90/031. Cincinnati (OH); U.S. Environmental Protection Agency; 1991.
6. Sokal RR, Rohlf FJ. Biometry—The principles and practices of statistics in biological research. 4th ed. New York (NY): WH Freeman & Co.; 2012.

7. Cochran WG, Cox GM. Experimental designs. 2nd ed. Reprint of 1957 edition, New York (NY): John Wiley & Sons; 1992.
8. Gad SC, Weil CS. Statistics and experimental design for toxicologists. Caldwell (NJ): Telford; 1988.
9. Hicks CR, Turner RV Jr. Fundamental concepts in the design of experiments. 5th ed. New York (NY): Oxford University Press; 1999.
10. Hurlbert SH. Pseudo replication and the design of ecological field experiments. Ecol Monogr. 1984;54(2):187–211.
11. Manual on presentation of data and control chart analysis. 8th ed.; ASTM Manual Series. MNL 7-8th, Revision of Special Technical Publications (STP) 15D. West Conshohocken (PA): ASTM International. 2010.
12. Ratliff TA Jr. The laboratory quality assurance system—A manual of quality procedures and forms. 3rd ed. New York (NY): Wiley-VCH; 2003.
13. Evaluation of dredged material proposed for discharge in inland and near coastal waters—testing manual; EPA-823-B-94-002. Washington DC: Office of Water U.S. Environmental Protection Agency and U.S. Army Corps of Engineers, Department of the Army; 1998.
14. Kanare HM. Writing the laboratory notebook. Washington DC: American Chemical Society; 1985.

https://doi.org/10.2105/SMWW.2882.151