ENVIRONMENTAL QUALITY

GUIDANCE FOR EVALUATING PERFORMANCE-BASED CHEMICAL DATA

This manual is approved for public release; distribution is unlimited.

ENGINEER MANUAL

AVAILABILITY

Electronic copies of this and other U.S. Army Corps of Engineers (USACE) publications are available on the Internet at http://www.usace.army.mil/inet/usace-docs/. This site is the only repository for all official USACE engineer regulations, circulars, manuals, and other documents originating from HQUSACE. Publications are provided in portable document format (PDF).

DEPARTMENT OF THE ARMY
U.S. Army Corps of Engineers
Washington, DC 20314-1000

EM 200-1-10

CECW-E
CEMP-RT

Manual
No. 200-1-10

30 June 2005

Environmental Quality
GUIDANCE FOR EVALUATING PERFORMANCE-BASED CHEMICAL DATA

1. Purpose. This Engineer Manual (EM) provides specific guidance, procedures, criteria, and tools to improve the evaluation of chemical environmental data. In particular, the EM presents strategies to more effectively evaluate data in the context of its end use, which is referred to as "performance-based data evaluation." Performance-based data evaluation is recommended to help ensure that only scientifically defensible data are used to support the decision making process after project-specific data quality objectives have been establish. This EM is intended for use by USACE personnel as a critical companion document to ER 1110-1-263.

2. Applicability. This EM applies to HQUSACE elements, major subordinate commands, districts, laboratories, and field operating activities having responsibility for hazardous, toxic, and radioactive waste (HTRW) projects. This includes, but is not limited to, execution of the following programs: Defense Environmental Restoration Programs; Base Realignment and Closure; Superfund; Civil Works; Military Construction; installation environmental compliance; Defense Logistics Agency; Department of Energy; work for others; and any construction projects involving hazardous, toxic, and radioactive waste.

3. Distribution Statement. Approved for public release and unlimited distribution.

4. References:

a. ER 1110-1-263

b. ER 1180-1-6

c. ER 5-1-11

d. ER 1110-1-12

5. Discussion. This EM provides guidance for screening the usability of chemical data after sample collection and analysis and for documenting the evaluation, in support of satisfying the requirements of the USACE Environmental Quality Assurance (QA) Program for chemical data as prescribed in ER 1110-1-263 and other general USACE quality management policy.

FOR THE COMMANDER:

2 Appendices
(See Table of Contents)

JOHN R. McMAHON
Colonel, Corps of Engineers
Chief of Staff

DEPARTMENT OF THE ARMY
U.S. Army Corps of Engineers
Washington, DC 20314-1000

CECW-E
CEMP-RT

Manual
No. 200-1-10

30 June 2005

Environmental Quality
GUIDANCE FOR EVALUATING PERFORMANCE-BASED CHEMICAL DATA

Table of Contents

CHAPTER 1

Introduction

1-1. Opening Remarks.

a. The production of data of "known and acceptable quality" is a primary goal of every environmental restoration and compliance sampling effort. In general, some degree of data review should be performed for all data collection activities to help ensure that only scientifically defensible data are used to support project decisions. However, the extent of the review will be dependent upon the project's **data quality objectives (DQOs)** and will be limited by the physical contents of the data package.[1] For example, the reporting and evaluation requirements for **definitive data** and **screening data** will differ significantly.

b. This document provides guidance to the U.S. Army Corps of Engineers (USACE) and USACE contractors (e.g., to architect-engineering contractors and third-party data reviewers) for evaluating instrumental chemical data using a **performance-based approach**. A **performance-based method** is defined as an analytical procedure for which **data quality indicators** are documented and evaluated with respect to acceptance criteria that are established from project data quality objectives. In particular, the **PARCCS** parameters (i.e., **precision**, **accuracy**, **completeness**, **representativeness**, **comparability**, and **sensitivity**) are documented for the **target analytes** of concern at the levels of concern (i.e., at or below project action levels) in the environmental media of interest and are evaluated with respect to acceptance limits or **measurement quality objectives (MQOs)** that are designed to ensure that total measurement uncertainty is within the limits prescribed by project DQOs. This document assumes that DQOs and MQOs have been established and presents guidance for evaluating chemical data quality, as measured by PARCCS, as a *first step process* for data usability assessment. (Refer to Chapter 1.2.2 for additional discussion regarding data usability assessment.) [2] To more effectively assess data usability, it is recommended that existing data evaluation protocols and checklists be revised using at least some of the strategies presented in this document.

1-2. Scope and Limitations of Performance-Based Data Review.

a. In general, data packages must contain enough information to evaluate PARCCS.[3] Data packages must essentially contain summary results for instrument calibrations (initial and continuing), environmental samples, and associated **batch quality control (QC) samples** (e.g., **method blanks** and **laboratory control samples**), as well as select raw data. (Specific reporting requirements are addressed in Chapter 4 of this document.) Data packages that satisfy these reporting requirements will be referred to as **performance-based (PB) data packages** and the im-

[1]The definitions of a number of critical terms appear in bold print the first time they are used. These terms are defined in the body of the document or, more commonly, in the glossary at the end of the document. Italics are used to denote emphasis or special meaning.
[2]Since it is assumed that MQOs have been established and are consistent with DQOs, for simplicity, from hence forth, the distinction between the two terms will not be maintained; "DQOs" will also refer to the MQOs.
[3]Although this document constitutes guidance, terms such as "must," "shall," and "will" are used in the document when an item is viewed to be especially critical or when an activity is viewed to be typically appropriate.

plementation of the data evaluation activities described this document will be referred to as **performance-based (PB) data review.**

b. The data review protocols presented in this document should not be viewed as prescriptive algorithms but as strategies intended for the purposes of guidance. For example, quality control (QC) acceptance limits are specified in this document but these limits should be viewed as "baseline" limits that should be adjusted (i.e., increased or decreased) based upon the objectives of the project. Even if it were possible to specify a set of QC acceptance limits that would be applicable to all projects, the potential occurrence of multiple QC problems alone suggests that a prescriptive approach for data evaluation would be unfeasible (e.g., it would not be practical to propose an evaluation strategy for every combination of QC problems that could be encountered). Because of the complexities of environmental investigations and uniqueness of environmental samples, analytical data must ultimately be evaluated using **professional judgement** in the context of *project-specific data objectives*.

c. In order to successfully implement a PB review, the DQOs in project planning documents such as Work Plans, Sampling and Analysis (SAPs), and Quality Assurance Project Plans (QAPPs) must be well defined. The generation of DQOs is beyond the scope of this document. However, it should be noted that generic statements such as "definitive" or "Level IV data will be collected" will not suffice. A specific set of QC acceptance limits must be presented for the analytes of concern for the concentrations of interest for the environmental populations being sampled. In theory, project documents such as QAPPs contain comprehensive and appropriate QC specifications, but, in practice, this is not necessarily true (e.g., when the data reviewer is not adequately involved in the project planning process).

d. In order to perform a PB data review, project-specific data quality objectives must be scientifically defensible. The scientific defensibility of the data should take precedence over contract compliance issues or the QAPP when QAPP contains inappropriate specifications. For example, if the QAPP requires data to be evaluated solely upon the basis of method-specified QC criteria (e.g., such as those specified in SW-846 methods) or laboratory performance criteria, then the validity of assessing the data on this basis should be carefully evaluated before proceeding with the data review. In particular, sensitivity requirements should not be established solely on the basis of method-specified quantitation limits such as the Contract Laboratory Program (CLP) Contract Required Quantitation Limits (CRQLs). Method-specified quantitation limits may be inappropriately high for project-specific action levels (e.g., risk-based cleanup levels). Laboratory detection, quantitation, and reporting limits must be evaluated with respect to the project-specific action levels to demonstrate that the proposed analytical methods possess adequate sensitivity. Similarly, it would typically be inappropriate for the QAPP to establish method data quality objectives for precision and **bias** solely on the basis of a laboratory's statistical control limits (e.g., for **matrix spikes** and laboratory control samples). A laboratory's statistical limits may be indicative of the laboratory's routine performance but may be too wide to yield quantitatively reliable results.

e. Performance-based data review must not be performed as a "last-minute" activity that is initiated only after the completion of all sample collection and analysis. To the extent that is possible or practical, prior to performing a PB data review, the reviewer should possess a

complete understanding of the intended use of the data and the relationship of the QC results to the usability of the data. The reviewer must receive input from the end-data users regarding the objectives and expected results of the analyses (e.g., via the review of the Project Quality Assurance Plan and Sampling and Analysis Plan). For optimal results, the reviewer should be involved in the DQO process in the early planning stages of the project (e.g., should be involved in scoping meetings where project DQOs, scheduling, sampling techniques, analytical methodologies, and data evaluation criteria are established.) *When the data reviewer is not adequately involved in the DQO process, a PB data review may result in the rejection of a significant portion of the analytical data. Performance-based data evaluation strategies need to be specified during project planning*

1-2.1. Performance-Based Data Review Versus Data Validation.

a. This guidance is generally applicable to any instrumental performance-based method, regardless of the determinative or preparatory techniques used to process the environmental samples. The data review protocols will result in a relatively thorough evaluation of data quality and will be applicable to a variety of environmental projects. However, the data review strategies presented in this document may be insufficient for all data uses. Project-specific DQOs may require more comprehensive data evaluation activities than those performed during PB data review.

b. Performance-based data review does not constitute ***data validation****. Data validation* is a more in-depth evaluation of laboratory data quality and is beyond the scope of this document. As the term is used in this document, *data validation* refers to any independent systematic review of **comprehensive data packages** with respect to a predefined set of technical performance criteria for PARCCS. A *comprehensive data package* is defined as a data package that contains sufficient information to completely reconstruct the laboratory analyses that were performed and documents salient field sample collection and handling activities (e.g., contains the Chain-of-Custody and may contain field logs). Hence, comprehensive data packages contain summary data for environmental, **batch QC**, and **instrument QC sample** analyses as well as all the raw laboratory data (e.g., standard preparation logs and printouts of chromatograms). CLP data packages are examples of comprehensive data packages with distinct reporting requirements.

c. Data validation involves the evaluation of batch QC and calibration results, in addition to other instrument QC results using the raw data. Since all the raw laboratory data are not included in performance-based data packages (unlike for data validation), reported QC summary results (e.g., laboratory control sample and **surrogate** recoveries) are not verified to the level of the raw data (e.g., using chromatograms and other instrumental printouts). Furthermore, with the exception of calibration data, PB data packages do not contain instrument QC results (e.g., pesticide percent breakdown and tune checks). Hence, during PB data review, instrument performance (other than calibration) is assumed to be in control or out-of-control in a manner that is consistent with batch QC performance. This assumption is usually reasonable but is not always valid.

d. A more thorough data evaluation should be considered when significant QC problems are observed during PB data review or when data is being collected to support critical decisions. Since laboratories normally maintain files of all supporting data and documentation for the analyses performed (for the period of time that is normally specified in the contract for analytical services), the laboratory can be requested to provide copies of the raw data to perform a more comprehensive review when the need arises. However, it is recommended that requirements for archiving comprehensive data packages be explicitly addressed when contracting for analytical services.

e. During project planning, the objectives of the analyses, nature of the contamination, limitations of the analytical methodology, and historic information about the site should be evaluated to determine whether a more comprehensive review needs to be performed. In particular, if the analytical technique involves the use of a **2-D detector** rather than a **3-D detector**, then it is especially critical to take stability problems (e.g., photochemical and thermal degradation) and interferences into account when determining whether or not a more comprehensive evaluation is required. For example, a review of batch QC results alone would probably be inadequate to identify data quality problems when a high performance liquid chromatograph (HPLC) with a fluorescence detector is being used to measure low levels of polynuclear aromatic hydrocarbons (PAHs) such as benzo(a)pyrene at a site with high background fuel contamination. The evaluation of 4,4'-DDT and Endrin breakdown checks (e.g., as discussed in Method 8081A) may be required to determine whether or not **detections** of target analytes such as Endrin ketone and Endrin aldehyde are actually false positives arising from poor method implementation (e.g., the degradation of Endrin during instrumental analysis).

> Note: The evaluation strategies presented in this document may be less adequate for 2-D detector methods than 3-D detector methods. *However, this does not imply that the strategies are not appropriate or useful for 2-D detector methods.* The level of confidence for data will be a function of the nature of the analytical technique, regardless of the thoroughness of any data evaluation activity. It is merely being noted that, since 2-D methods inherently lack the specificity of 3-D methods, 2-D methods are more prone to data quality problems (e.g., false positives) that, under select circumstances, may only be identified via the evaluation of a full raw data package.

1-2.2. Performance-Based Data Review Versus Usability Assessment.

a. It is emphasized that the PB data review activities discussed in this document constitute only a first-step process for the assessment of data usability. A *full* assessment of data usability is a more complex and comprehensive activity than PB data review or validation; the former encompasses the latter and is potentially more subjective. The data user must ultimately assess the overall usability of data on the basis of *total measurement uncertainty* and the objectives of the investigation. .

b. Total measurement uncertainty consists of the sum of the *laboratory analytical uncertainty* and *field sampling uncertainty.* Unfortunately, *field sampling uncertainty* is often greater than *laboratory analytical uncertainty* and is not fully taken into account during data review or validation (i.e., data review and validation identify *laboratory analytical uncertainty* but do not

fully address *field sampling uncertainty)*. For example, data review and validation may identify incorrect preservation techniques but would not adequately characterize the representativeness of a sample collected from an environmental population with high spatial or temporal variability.

c. Little or no usability assessment is typically performed during data review or validation. Usability assessment is usually performed after data review or validation is completed. For example, when data validation is performed using the National Functional Guidelines, sensitivity is evaluated with respect to fixed CRQLs rather than project-specific action levels. However, meeting CRQLs does not ensure that the data will be usable (a problem which, unfortunately, many usability assessments also fail to identify). This document constitutes a more streamlined approach for data review. Data quality is evaluated during data review in the context of the end use of the data.

1-3. Overview of Performance-Based Data Review.

a. This section of the document presents a brief overview of the PB data review process. The reviewer initially receives input from the end-data users regarding the objectives and expected results of the analytical efforts (e.g., in the form of formal DQOs described in the Work Plan and QAPP). Prior to performing a PB data review, the reviewer performs a cursory evaluation of the data package to ensure that it contains all the required documentation. *This is critical since the evaluation of any data package will be limited by its physical content.* If the data package is essentially complete, the reviewer performs a more complete evaluation to determine if the data *potentially* meet the needs of the end user. The reviewer verifies that sample collection and handling activities were properly implemented in the field and subsequently evaluates the analytical quality of the laboratory data. A PB review includes the evaluation of the following QC elements:

(1) Completeness.

(2) Holding Time and Preservation.

(3) Initial Calibration.

(4) Initial Calibration Verification.

(5) Continuing Calibration Verification.

(6) Sensitivity (e.g., detection and quantitation limits).

(7) Blanks (e.g., field and method blanks).

(8) Laboratory Control Samples.

(9) Post-Digestion Spikes (for trace metal methods).

(10) Matrix Spikes.

(11) Matrix Spike Duplicates and Matrix Duplicates.

(12) Surrogates (for organic chromatographic methods).

b. Detailed definitions of these QC elements may be found in the USACE *Shell for Analytical Chemistry* and the glossary of this document.

c. Quality control samples are designed to evaluate the PARCCS parameters and identify quality problems in three specific areas: (i) Laboratory analytical performance, (ii) matrix effects, and (iii) field performance. For example, accuracy is assessed from calibration, laboratory control sample (LCS), matrix spike (MS), post-digestion spike (PDS), and surrogate data. Precision is evaluated from duplicate laboratory control and matrix spike samples. Sensitivity is evaluated using detection limits and quantitation limits. Representativeness is evaluated via the review of holding time and blank data. A laboratory's analytical performance is evaluated using calibration results (i.e., initial calibrations, initial calibration verifications, and continuing calibration verifications) and **batch QC samples** such as method blanks and laboratory control samples. Matrix effects are evaluated using matrix spike, surrogate spike, and post digestion spike recoveries**. Field duplicates**, **rinsate blanks**, and **trip blanks** are examples of QC samples that are used to assess QC problems associated with sample collection activities..

d. After (or during) the technical evaluation, the reviewer generates a **data review report** that summarizes the overall quality of the data and lists individual QC problems and any observations that may be relevant to the data's potential usability. Data review reports are discussed in Chapter 2.

CHAPTER 2

Data Review Reports

2-1. Introduction.

A data review report documents the PB data evaluation. At least one data review report must be generated for each data sample delivery group. The data review report may address the data packages for several analytical methods. The format of the document is not as important as its content. However, a uniform format is recommended to facilitate data evaluation activities. The elements presented below must be included in a data review report.

2-2. Cover Page.

The cover page specifies the following information:

a. Unique report ID number.

b. Name and address of data reviewer.

c. Contract number.

d. Client name and address.

e. Project name and site location.

f. Statement of data authenticity and official signature of release.

2-3. Cover Letter.

a. Project name (or brief description of the project).

b. Site name (location from which the samples were collected).

c. Parties responsible for evaluating the data and the date the evaluation was performed (including a point of contact for questions with phone and facsimile numbers).

d. Technical criteria used to evaluate the data (e.g., cited as a reference).

e. Laboratory that performed the analyses (name, location, and point of contact).

f. Description of the samples that were evaluated, including the following:

(1) Number of samples.

(2) Matrix.

(3) Environmental samples associated with the field QC samples.

(4) Field and laboratory ID numbers.

(5) Date samples were collected in the field.

(6) Preparatory and determinative methods of analysis (including method numbers).

(7) Target analytes or parameters.

(8) Date the laboratory analyses were performed.

(9) Date the data package from the laboratory was received.

2-4. Executive Summary.

a. The objective of the *Executive Summary* is to *concisely* describe the overall quality of the data package in a manner that is comprehensible to an individual *lacking an extensive background in analytical chemistry*. Major areas of concerns and any information which would aid the reader to better understand the quality or usability of the data must be discussed in general terms. For example, Executive Summary may state that a complete review of the data could not be performed because of missing information or it may state that no significant QC problems were observed.

b. When *major* QC problems are observed after the data review process, when possible, the Executive Summary must indicate whether these problems primarily resulted from unacceptable laboratory performance, **matrix interference**, or problems associated with the sample collection activities. If the QC problems resulted because the laboratory or field personnel failed to follow the requirements in the Work Plan or QAPP, this information should be highlighted. The Executive Summary should also recommend corrective actions to improve the quality of the data. The format and content of the Executive Summary are otherwise left to the discretion of the author.

2-5. Technical Summary.

a. The *Technical Summary* must discuss the quality of the data package in terms of specific QC elements and must be divided into subsections—one for each QC element in which problems were identified (e.g., “Holding Times,” “Laboratory Control Samples,” “Matrix Spikes,” and “Continuing Calibration Verifications”). The Technical Summary must discuss the effects of QC problems in the context of sensitivity (e.g., false negatives due to high detection limits), precision (e.g., high variability), accuracy (e.g., high or low bias), representativeness (e.g., blank contamination), completeness (e.g., missing information), and comparability (e.g., failure to use specified methodology).

b. The Technical Summary explain *why* each result was qualified (e.g., matrix interference and blank contamination). This is especially critical when project DQOs are not well defined for a particular parameter or when qualification is based upon a high degree of professional judgment (e.g., due to the complexity of the project's objectives or the environmental population being sampled). Any changes made to the laboratory's reported data (e.g., due to misidentification, transcription errors, or calculation errors) must be identified and the samples affected by each QC problem should be listed in a tabular format.

c. The Technical Summary must identify QC problems as having a *major* or *minor* impact on data quality or usability. The Technical Summary should also distinguish systematic errors from random errors. Systematic errors resulting from blunders (e.g., transcription errors) should also be distinguished form systematic effects that bias the results (e.g., from poor extraction efficiency). When possible, the Technical Summary should identify the *direction of bias (high or low).* In addition, problems giving rise to *qualitative* uncertainties must be distinguished from those giving rise to *quantitative* uncertainties. *Qualitative* uncertainty refers to uncertainty associated with the *identification* of an analyte in an environmental sample. *Quantitative* uncertainty refers to error associated with the determination of the amount of an identified analyte. (Refer to the definitions of the N and J qualifiers in Chapter 3.)

d. When possible, field sampling uncertainty must be distinguished from laboratory analytical uncertainty. Problems arising from missing data and QC failures resulting from substandard laboratory performance (e.g., out-of-control LCS recoveries) and substandard sample collection procedures (e.g., the lack of sample preservation) must be highlighted. When major QC problems are observed, corrective actions should be recommended. However, it should be noted that the major of objective of the evaluation is to determine the potential usability of the data and not contractual compliance (e.g., contractually noncompliant data may or may not be usable.)

e. The reviewer should *avoid* statements pertaining to the *ultimate* usability of the data. As defined in this document, PB data review results in "usability screening" rather than a full usability assessment. In particular, unless there is a high degree of confidence that a set of results must be rejected (e.g., the results are being qualified with the R flag), adjectives such as "unusable" and "unacceptable" should be avoided (e.g., the results should be described as "tentatively unusable.") The use of these terms in data review reports may be interpreted as contradictory by regulators in situations where the end users determine that the data are useful for project purposes in spite of the QC problems. Similarly, terms such as "usable," "acceptable," and "valid," should only be used when the report *explicitly* defines these terms to mean that the data are *potentially* usable or that the data review specifications have been satisfied. Examples of preferred terminology are presented below.

(1) "The results for the aqueous HVOCs (laboratory batch 50603, samples SL5-3031-1 to SL-3031-6) may possess a negative bias because the samples were analyzed one day beyond the **holding time limit**; detections are qualified with the J- flag (i.e., as estimated with suspected low bias) and **nondetections** are qualified with the UN flag to indicate the possibility of false negatives ...".

(2) "The **Chain-Of-Custody** was not signed in the field. This may adversely impact the legal defensibility of the data. However, no results were qualified based upon this observation ..."

(3) "The low LCS **recovery** (11%) for the semivolatile pentachlorophenol for (laboratory) Batch 49382 is indicative of a large negative bias for the associated samples (GW-2-21-972, GW-2-21-972-FD, GW-2-21-972-MS, and GW-2-21-973 to GW-2-21-980). Detected concentrations of the analyte are considered to be minimum values and nondetections are considered to be unreliable at the stated reporting limits. Reported concentrations (nondetections and detections) of the analyte below the project-specified action level are qualified with the X flag as tentatively unusable because they do not demonstrate that the analyte is present below the action level. Detections above the action level are qualified with the J- flag ..."

2-6. Data Summary Tables.

Present qualified results for the environmental samples in a tabular format and list the definitions of all data qualifiers. Use footnotes to briefly explain why the data were qualified. The summary tables should also list the detection and quantitation limits and any project-specific action levels or requirements for sensitivity. (It is recommended that this be done in a format that will enable the data to be readily exported to the project report.) The header information for each table typically includes the following information:

a. Project name and location.

b. Laboratory name and location (City and State).

c. Field and laboratory ID numbers.

d. Matrix type.

e. Preparatory and determinative method.

f. Date of sampling, analysis, and preparation.

g. Amount of sample processed and analyzed (including extract volume).

h. Dilution factors.

i. Percent moisture (for solid samples).

j. Concentration units.

2-7. Project Specific Communications.

This section contains pertinent communications (e.g., phone logs, E-mail, and letters) between the data reviewer and any agencies or parties that possess an interest in the quality of the data

(e.g., the analytical laboratory, the contractor that collected the samples, the USACE district field office, and regulators). Examples are listed below:

a. Requests to the analytical laboratory for the submittal of additional information.

b. Communications with the client concerning major data quality deficiencies.

c. Communications with the samplers to address QC problems associated with sample collection.

d. Inquiries from regulatory authorities.

e. Requests from the client for quick turnaround time.

f. Amendments of the data quality objectives from the client.

2-8. Project Specific Communications.

a. Include checklists that were used to review the data packages in an appendix of the data review report. Checklists demonstrate that the data packages were assessed for overall completeness prior to the technical evaluation and appropriate QC elements were assessed during the technical evaluation (e.g., holding times, initial calibration, and laboratory control samples).

b. Include worksheets that were used to verify the laboratory's reported results (e.g., any recalculations that were performed). When errors are observed, photocopies of the laboratory's original data and any relevant field documents should be used to *illustrate* the corrections performed. For example, if incorrect concentration units were reported for all the samples, it would only be necessary to illustrate the correction for one sample. Note that, depending upon the severity of the errors and the contractual requirements for the analyses, the laboratory may be required to correct the results and resubmit the data packages.

CHAPTER 3

Data Qualifiers

3-1. Introduction.

a. **Data qualification** is an integral component of data review and validation. During PB data review, **data qualifiers** or **flags** are applied to alert the end user to quality problems that may impact the usability of the data (e.g., QC acceptance limits that were not met). However, it is emphasized that data qualification essentially results in a *qualitative* evaluation of the data (e.g., measurement uncertainty is not evaluated in a quantitative manner). This is one of the major reasons why data review or validation is "only a first-step process for the assessment of data usability" (as stated in Chapter 1.2.2).

b. During PB data evaluation, results are either *accepted* or reported with *data qualifiers* or *flags*. Data that meet all QC acceptance limits are *potentially* usable and are *not* qualified. Data that fail one or more QC criteria are qualified as *estimated* (with the *J flag*), *tentatively rejected* (with the *X flag*), or *rejected* (with the *R flag*). The distinction between *estimated*, *tentatively rejected*, and *rejected* data resides in the degree of the QC failure and is *highly dependent upon the reviewer's understanding of the objectives of the project.*

c. In general, data that are believed to be *completely unusable* with a high degree of confidence (e.g., because of the gross failure of QC criteria) are qualified as *rejected* and would not normally be used to support decisions for an environmental project. Data associated with a *marginal* failure of QC criteria that are believed to be *tentatively usable* or "more usable than not" are qualified as *estimated.* Data that are "mostly unusable" or that fall into the "gray area" between *estimated* and *rejected* are qualified as *tentatively rejected.*

> Note: Typically, when data validation is performed, data are primarily qualified as either *estimated* (e.g., with the *J flag*) or *rejected* (e.g., with the *R flag*). However, since data are usually rejected only for the most severe or blatant QC problems, the *R flag* is rarely applied. When QC problems are observed, the data are frequently qualified as *estimated* and are subsequently used to support project decisions. Unfortunately, J-qualified data are often used to support project decisions without evaluating the impact of the QC problems on the usability of the data, resulting in an over estimation of data quality. To minimize the potential indiscriminate use of J-qualified data, an additional data qualifier, the *X flag*, has been defined. During a PB review, depending upon the severity of the QC problem, data are primarily qualified as either "estimated and tentatively accepted" (*J flag*), "estimated and tentatively rejected" (*X flag*) or "rejected" (*R flag*).

d. As stated earlier, full data usability assessment is a more complex and comprehensive activity than data review or validation and is usually performed by the end user (rather than by the data reviewer) because the data user typically possesses a greater understanding of the project's DQOs (e.g., because of a more extensive knowledge of the project's history). Therefore, the end user must ultimately determine the acceptability of the data. However, this does not im-

ply that the end user may apply qualified data in an indiscriminate fashion. This is particularly true of data that have been qualified as *tentatively rejected. Tentatively rejected data must not be used to support project decisions unless the data user presents (i.e., documents) some technical rationale for doing so.* In other words, *tentatively rejected* data must ultimately be *rejected* (e.g., using the R flag) in the absence of a scientifically defensible rationale to do otherwise. *This requirement should be explicitly specified in the QAPP.* Furthermore, when data qualified as *tentatively rejected* are used to support decisions for a project, the data reviewer should be consulted for a consensus unless it is clear that the reviewer did not possess a complete understanding of the objectives of the investigation (e.g., new DQOs were established after the data review was performed). It should be noted "Chemical Data Quality Assessment Reports" (as defined by EM 200-1-6, 10 October 1997) by USACE project chemists represent one possible mechanism to document a more comprehensive usability evaluation (e.g., X-flagged data may be converted to J-flagged or R-flagged data in Chemical Data Quality Assessment Reports).

e. Ideally, *estimated* (i.e., J-qualified) data, though presumed to be usable by the data reviewer, should be accepted by the end user only after the reasons for the data qualifications and their impact on the achievement of project DQOs have been examined. For example, when the direction of bias and the magnitude of the analytical uncertainty are well defined, a more thorough examination of the data may entail an evaluation similar to that presented in Paragraph 11-6.

3-2. Definitions of Data Qualifiers.

a. All data qualifiers or flags must be clearly defined. Project-specific requirements ultimately determine the types of qualifiers that are required (e.g., the EPA Functional Guidelines for validation require a distinct set of flags). However, in the absence of more appropriate conventions for data qualification, the flags defined below must be used. The definitions of the data qualifiers are summarized in Table 3-1.

(1) *R flag*.

(*a*) The datum is rejected. The qualifier typically indicates that a datum is completely unusable because it is of unknown quality (e.g., missing QC information) or because of gross QC deficiencies (e.g., extremely poor recoveries for the LCS).

(*b*) NFor gas chromatography/mass spectrometry (GC/MS) analyses, the R flag must be used to reject TIC tentatively identified compound results that are believed to be laboratory artifacts (i.e., common laboratory contaminants). Examples include reagent contaminants, solvent preservatives, siloxanes, and aldol condensation reaction products of acetone (e.g., 4-hydroxy-4-methyl-2-pentanone, 4-methyl-2-penten-2-one, and 5,5-dimethyl-2(5H)-furanone).

(2) *J flag*. The target analyte is *positively identified,* but the reported numerical result (e.g., analyte concentration) is an *estimated* value and the direction of bias is unknown. The flag indicates that a significant *quantitative* (as opposed to a qualitative) uncertainty exists. The J flag must *always* be used to report the following.

(*a*) Detections below the **method quantitation limit (MQL)**.

(*b*) Tentatively Identified Compounds (TICs).[1]

(3) *J-Flag.* The target analyte is present but the reported numerical result is an estimate that is believed to be *biased low* (e.g., the actual concentration in the environmental sample is believed to be *greater* than the reported concentration).

(4) *J+ Flag.* The analyte is present but the reported numerical result is an estimate that is believed to be *biased high* (e.g., the actual concentration in the environmental sample is believed to be *less* than the reported concentration).

(5) *N Flag.* The target analyte is reported as a *tentative detection* (e.g., because the identity of the analyte is in doubt). The N flag indicates a significant *qualitative* rather than quantitative uncertainty exits (i.e., the reported detection of the analyte may be a "false positive"). When used in combination with the U flag (i.e., the UN flag), the qualifier indicates that the absence of a target analyte at some stated reporting limit is in doubt (i.e., a false negative is possible at the stated reporting limit). Applications of the N flag include the following:

(*a*) Uncertain Aroclor identifications (e.g., weathered PCBs).

(*b*) Tentatively Identified Compounds.

(6) *U Flag*. The analyte was not detected relative to the **method reporting limit (MRL)**; that is, the result is less than the method reporting limit (MRL).

(7) *NJ Flag*. The presence or identity of the analyte is in doubt and the reported concentration is estimated. The estimation is both *qualitative and quantitative* in nature.

(8) *UN Flag.* The result is reported as a *tentative nondetection* (as opposed to tentative detection); there is uncertainty with whether or not the nondetection is valid at the stated method reporting limit (e.g., because of QC problems).

> Note: The UN flag is similar (but not identical) to the CLP "UJ flag," which is defined as follows: "The analyte was not detected above the reported sample quantitation limit. However, the reported quantitation limit is approximate and may or may not represent the actual limit of quantitation necessary to accurately and precisely measure the analyte in the sample." Both flags indicate that an uncertainty is associated with a nondetection. However, the UJ and UN flags differ in that the former is defined in terms of the CLP CRQLs, while the later is defined in terms of *project-specific reporting limits*. The UN and UJ qualifiers are essentially equivalent when the reporting limits equal the CRQLs. However, in general, setting the reporting limits equal to the quantitation limits is not recommended (e.g., unless the action levels for the project are high).

[1] TICs are typically qualified by the laboratory.

(9) *X Flag*. The datum is *tentatively rejected* because project-specific data quality objectives (e.g., for sensitivity, accuracy, or precision) were not met or were not demonstrated. When objectives for sensitivity are not met, the X flag typically indicates that a result (a detection or nondetection) is potentially unusable with respect to an action level (e.g., the result does not demonstrate that a target analyte is actually present in an environmental sample at a concentration above or below a risk-based decision limit).

> Note: When evaluating objectives for sensitivity, the R flag may be more appropriate than the X flag when action levels are fixed and statistical analyses are not being performed. The X flag may be appropriate when action levels are subject to change, a set of data is being evaluated with respect to different action levels, or when statistical analyses are being performed.

b. It may be desirable to use the X flag in combination with other flags as illustrated below:

(1) *X- Flag.* The detection is (quantitatively) grossly estimated with low bias and is tentatively rejected.

(2) *X+ Flag*. The detection is (quantitatively) grossly estimated with high bias and tentatively rejected.

(3) *XN Flag.* The detection is quantitatively and qualitatively highly estimated and is tentatively rejected.

(4) *XU Flag.* The nondetection is tentatively rejected.

c. When a datum is qualified, the reviewer must explain *why* the particular qualifier was applied. It is recommended that numerical subscripts be placed on flags to indicate why the flag was used (e.g., the flags J_1 and J_2 may indicate that a result is estimated because of poor target analyte **recovery** for the associated laboratory control sample and matrix spike, respectively). However, the use of subscripts may not be practical for projects that involve a large number of samples when the samples possess multiple QC problems. Similarly, it may be desirable to suppress the numerical subscripts when the reasons why the flags were applied are not important to a particular data user. However, the rationale for each qualification must be explained in the data review report.

d. When possible, the J flag must identify any suspected bias (high or low) in the data. If bias is known for an estimated result, use either the J- or J+ flag to qualify a result. However, if a datum is estimated because of multiple QC problems and the direction of bias is not well defined, it may be appropriate to qualify the datum as *tentatively rejected* rather than as *estimated*.

Table 3-1
Summary of Major Data Qualifiers

Qualifier	Definition
J	Estimated (quantitatively) and tentatively usable
J-	Estimated (quantitatively) with low bias
J+	Estimated (quantitatively) with high bias
U	Below reporting limit
N	Qualitatively estimated (tentative detection)
X	Tentatively rejected
R	Rejected
UN	Tentative nondetection
NJ	Quantitatively and qualitatively estimated

CHAPTER 4

Completeness [1]

4-1. Introduction.

The primary objective of this review is to ensure that the data package contains adequate documentation to perform the PB data review. The deliverables that constitute a PB data package are discussed below. Before performing a technical review of the data package, the package should be examined using a checklist to verify that all the required elements are present.

4-2. Minimum Reporting Requirements.

a. As discussed in more detail below, each data package must include a cover page, a table of contents, a Case Narrative, a **Chain-Of-Custody** form, a summary of the environmental sample results (e.g., method of analysis, date analyzed, and amount analyzed), and a summary of the batch QC results (e.g., method blank and LCS results). A summary of all instrument calibration results (i.e., initial calibrations, initial calibration verifications, and continuing calibration verifications) and copies of the sample preparation, standard preparation, and instrument run log sheets must also be included in the data package as described below.

b. The organization of the data package must be such that chemical data are reported on a per **batch** basis. All calibration, method, and batch QC results must be presented on summary forms using a tabular format. The use of CLP standard forms is not necessary. However, submission of standard instrument output alone is unacceptable to satisfy the reporting requirements for PB data packages. Batch QC samples must be clearly linked to their associated environmental samples. Instrument QC samples must be clearly linked to the associated environmental and batch QC samples.

c. The data package must contain sufficient information to determine how the final sample concentrations were calculated from the calibration curves, or, alternatively, how a calibration standard may be expressed as a final sample concentration. In particular, using the sample and calibration summary forms and the standard and sample preparation logs, it must be possible to express the low calibration standard as a sample concentration.

4-2.1. Cover Letter.

The cover sheet includes the following information:

a. Title of Test Report or "Test Certificate."

[1]The evaluation of "completeness" performed during PB data review should not be confused with the "completeness" evaluation that is performed during data usability assessment. The former relates to the data package while the latter is more global in nature and is performed to determine whether or not there is sufficient data of known and acceptable quality to support project decisions.

b. Name and location of laboratory.

c. Laboratory point of contact with phone and facsimile numbers.

d. Name and location of any subcontractor laboratories.

e. Contract number.

f. Client name and address.

g. Project name and site location (if provided by client).

h. Statement of data authenticity and official signature and title of person authorizing the release of the test report.

i. Amendments to previously released reports shall clearly identify previous reports and state the reason(s) for the report amendments..

4-2.2. Case Narrative.

A Case Narrative must be included in each report. The Case Narrative contains tables summarizing the samples received, providing a correlation between field sample numbers and laboratory sample numbers, and identifying which analytical test methods were performed. When analyses are subcontracted to other laboratories, the Case Narrative must clearly specify which laboratory performed each analysis. Samples that were received but not analyzed must also be identified. Extractions or analyses that are performed out of holding times must be appropriately noted. The Case Narrative must define all data qualifiers or flags used. Deviations of any calibration standards or QC sample results from appropriate acceptance limits must be noted and associated corrective actions taken by the laboratory must be discussed. Any other factors that could affect the sample results (e.g., air bubbles in VOC sample vials, excess headspace in soil VOC containers, the presence of multiple phases, sample temperature and sample pH excursions, container type or volume, etc.) must be noted.

4-2.3. Technical Summary.

Summary forms for each sample include the information specified below. Information need not be repeated if noted elsewhere in the data package.

a. Laboratory name and location (city and state).

b. Project name and unique ID number (if specified by client).

c. Field sample ID number as written on custody form.

d. Laboratory sample ID number.

e. Matrix (soil, water, oil, etc.).

f. Sample description.

g. Sample preservation or condition at receipt.

h. Date sample collected.

i. Date sample received.

j. Date sample extracted or prepared.

k. Date sample analyzed.

l. Analysis time when holding time limit is less than 48 hours.

m. Method (and SOP) numbers for all preparation, cleanup, and analysis procedures employed.

n. Preparation, analysis, and other batch numbers.

o. Analyte or parameter.

p. Method reporting limits adjusted for sample-specific factors (e.g., aliquot size, dilution or concentration factors, and moisture content).

q. Method quantitation limits.

r. Method detection limits**.**

s. Final analytical results (e.g., concentrations) *with the correct number of significant figures.*

t. Data qualifiers and definitions of data qualifiers.

u. Concentration units.

w. Dilution factors—All reported data shall reflect any dilutions or concentrations. The dilution factor must be noted on the analytical report. If analyses were performed for both the undiluted and diluted samples, both results must be reported.

x. Percent moisture or percent solids (e.g., soils, sediments, and sludges are typically reported on a dry weight basis).

y. Sample aliquot analyzed.

z. Final extract volume.

4-2.4. Sample Management Records.

These types of records include the documentation accompanying the samples (e.g., original chain-of-custody record, shipping documents, and laboratory notification sheets), records generated by the laboratory which describe the condition of the samples upon receipt at the laboratory (e.g., Sample Cooler Receipt forms, and any records of telephone conversations), and any records generated to document sample custody, transfer, analysis, and disposal.

4-2.5. Batch QC Summary Results.

a. The data package must include all batch QC sample results. This includes method blank (MB), laboratory control sample (LCS), **laboratory control sample duplicate** (LCSD), matrix spike (MS), matrix spike duplicate (MSD), matrix duplicate (MD), and post-digestion spike (PDS) results with the associated acceptance criteria. Summary forms for the MS, MSD, PDS, and LCS, must specify the spiking concentrations, the measured concentrations (e.g., the measured concentrations before and after spike addition for MS and PDS analyses), the **percent recovery,** and the percent recovery acceptance limits for each target analyte. The nature of the recovery acceptance ranges must also be specified (e.g., project-specified acceptance range or in-house laboratory statistical limits). The laboratory's statistical warning and control limits must be specified (in addition to any project required acceptance ranges) for each target analyte for the laboratory control samples (for each preparatory and determinative method).

b. Summary forms for replicate results (e.g., duplicate precision as measured by MS/MSD, LCS/LCSD, or sample/MD pairs), must specify the concentrations of the replicate results, the **relative percent differences (RPDs)** or **percent relative standard deviations (%RSDs)** for each set of replicates (e.g., each duplicate pair), and the acceptance limits (e.g., for the RPDs). The nature of the acceptance limits must also be specified (e.g., project-specified limits or in-house statistical limits for the RPDs).

4-2.6. Standard Preparation Logs.

a. Copies of all relevant *standard preparation log sheets* must be provided for all calibration standards and spiking standards associated with the environmental samples (e.g., the initial calibration, initial calibration verification, continuing calibration verification standards as well as the MS, PDS, and LCS spiking standards). At a minimum, the standard preparation logs must clearly specify the following for all standards:

(1) Source (e.g., manufacturer and lot number for commercial stock solutions).

(2) Composition (e.g., the concentration of all target analytes, surrogates, and internal standards).

(3) Date of preparation and expiration.

(4) Name of the analyst.

(5) ID number of the standard.

(6) Reagents and solvents added to standards (including source and lot numbers).

b. When a standard is prepared via the dilution of a stock solution, the following must be specified: The spiking volume and concentration of the stock solution, and the final volume and concentration of the diluted standard. Copies of manufacturer certificates for commercially purchased stock standards must be included in the standard preparation logs. When the laboratory prepares its own stock solutions, calculations and conversion factors must be shown in the standard preparation log (e.g., using equations or sample calculations).

4-2.7. Sample Preparation Logs.

Copies of the *sample preparation log sheets* must be included in the data package. The sample preparation logs must include the following information:

a. Sample and batch ID numbers.

b. Matrix.

c. Preparatory method (method and laboratory SOP ID number).

d. Date of sample preparation.

e. Initial volume or weight of the sample processed.

f. Final volume of sample processed (e.g., after digestion, extraction or cleanup).

g. Percent moisture (for solid samples).

h. Reagents and solvents added to the samples (including source and lot numbers).

i. Preservation and pH checks or adjustments.

j. Spiking standards (ID number of the LCS, PDS, and MS spiking solutions, volume added, and the final spike concentration).

k. Name of the analyst.

4-2.8. Instrument Run-Sequence Logs.

a. Copies of *instrument analysis log sheets* must be provided for each instrument for each day or analytical shift project samples or associated QC samples were analyzed. *Instrumental analysis logs are particularly important since they provide the basic link between*

the environmental sample analyses and QC data. The run sequence logs must include the following information:

(1) Date of analysis.

(2) Determinative method (including SOP ID number).

(3) Name of the analyst.

(4) Unique ID numbers for environmental and QC samples.

(5) Amount of the sample (instrumentally) analyzed (e.g., the injection volume for chromatographic methods).

(6) Instrument ID number and salient instrument features (column type for chromatographic methods).

(7) Reanalyses and dilution factors (when performed).

b. The run log must clearly identify all QC samples (e.g., continuing calibration verifications). It must also clearly indicate which environmental and batch QC samples are associated with each initial calibration, initial calibration verification (ICV), and continuing calibration verification (CCV). The order in which environmental and QC samples (e.g., ICVs and CCVs) are recorded in the run log must always be consistent with the temporal order in which the samples were actually analyzed. The time of analysis for CCVs and tunings (when performed) must be specified for chromatographic methods. When an autosampler is used (i.e., any device that enables the loading of multiple samples for sequential analysis) the position or sequence number should be recorded in the run log (e.g., the purge port number for purge-and-trap analyses). Lastly, any salient analytical problems (e.g., carry over) must be noted. In particular, *when a QC sample (e.g., a CCV) is reanalyzed, the run log must specify why the reanalysis was performed.*

4-2.9. Traceability.

The data package must clearly demonstrate complete ***traceability*** *of all standards.* For example, unique ID numbers must link all batch QC samples (e.g., MSs and LCSs) in the *instrument run log* to the spiking solutions listed in the *sample preparation log*. The spiking solutions in the *sample preparation log* must be traceable to the original (primary) stock standards via the *standard preparation log*. Similarly, unique ID numbers must link **instrument QC samples** (e.g., CCVs and ICVs) in the *run log* to the corresponding standards in the *standard preparation log*.

4-2.10. Calibration Summary Results.

a. The concentration and corresponding instrumental response (e.g., peak area or peak height) must be reported for each initial calibration standard. When the initial calibration (and quantitation) is performed using internal standards, instrumental response and the corresponding internal standard concentration (or amount and volume of internal standard analyzed) must be

reported for the initial calibration. When more than one internal standard is used, the data package must list the set of target analytes associated with each internal standard. A quantitative "goodness-of-fit" value (e.g., the correlation coefficient, the coefficient of determination, or the **%RSD**) must be reported for the calibration curve of each target analyte and surrogate (when surrogates are used).

b. Plots of the initial calibration curves and instrumental printouts of quantitation summary reports must also be included in the data package for all target analytes. Instrument response must be plotted on the y-axis as the dependent variable and concentration must be plotted on the x-axis as the independent variable. The equation of each calibration curve must also be specified. When calibrations are performed using response factors, the mean response factors must be included with the data package (e.g., in lieu plots and equations for the initial calibrations).

c. The "true" (i.e., reference) concentration (i.e., level spiked), measured concentration, the instrumental response corresponding to the measured concentration, and percent recovery must be reported for each CCV and ICV for each target analyte and surrogate (when surrogates are analyzed). When internal standards are used, the instrumental response for each internal standard and the corresponding internal standard concentration (or amount and volume of internal standard analyzed) must also be reported. When an ICV is not performed, the data package must clearly indicate (e.g., in the standard preparation logs) whether an independent-source standard was used to spike the CCVs or LCSs.

4-2.11. Chromatographic Methods for Organic Target Analytes.

The additional reporting requirements specified in this section of the document apply only to chromatographic methods for organic target analytes.

4-2.11.1. Initial Calibration.

When the initial calibration is performed using response factors, the **response factor** for each initial calibration standard and the mean response factor for each analyte and surrogate must be reported. The %RSD must also be reported for each set of initial calibration standards. A reporting format similar to the CLP "Initial Calibration Summary Form" for the GC/MS BNA and VOA analyses is recommended. Response factors must also be reported for CCVs and the initial calibration when minimum response factors are specified by the analytical method (e.g., GC/MS Methods 8260A and 8270B). (Note that this does not imply that regression analysis cannot be used to perform initial calibrations for these methods.)

4-2.11.2. Internal Standard Summary Information.

When internal standards are used, the internal standard areas and retention times must be summarized for each batch QC and environmental sample in a tabular format (e.g., using summary forms similar to the CLP "Internal Standard Area And RT Summary" forms). The internal standard retention time and retention time windows must also be specified for the most recent associated CCV. The area counts and area count acceptance ranges for the associated mid-level ini-

tial calibration standard must be specified on the internal standard retention time and peak area summary form.

4-2.11.3. Surrogate Results.

The expected concentrations of the surrogates, the surrogate recoveries, and surrogate recovery acceptance ranges must be reported for all the environmental samples, batch QC samples, continuing calibration verifications, and initial calibration verifications. The surrogate recoveries for the environmental and batch QC samples must be summarized in a tabular format (e.g., using forms similar to the CLP "Surrogate Recovery" forms). The nature of the surrogate recovery acceptance ranges must also be specified (e.g., project-specified versus in-house statistical control ranges). The laboratory's statistical warning and control limits must be specified (in addition to any project required acceptance range) for the surrogates in the laboratory control samples or method blanks (for each preparatory and determinative method).

4-2.11.4. Chromatographic Methods With 2-D Detectors.

a. Additional reporting requirements are required for chromatographic methods with 2-D detectors (e.g., FIDs, PIDs, and ECDs). When chromatographic analyses are performed using second column confirmation, the results for both analytical columns must be reported (e.g., surrogate recoveries, sample results, and initial and continuing calibration results).

b. The results for the environmental samples must be reported using summary forms similar to the CLP "Pesticide Identification Summary For Single Component Analytes" and the "Pesticide Identification Summary For Multicomponent Analytes." Retention times and retention time windows must be reported for all single component standards (e.g., initial calibration, continuing calibration verification, and internal standards), target analytes, and surrogates or both the primary and confirmatory columns. For multi component analytes, the retention times and retention time windows should be specified for at least three to five characteristic peaks. (Note that retention time and retention time windows must be reported for both the primary and confirmatory columns.) In addition, the results (e.g., concentrations) from the primary and secondary columns, as well as the corresponding RPDs must be reported.

4-3. Evaluation of Completeness.

a. Use professional judgement to determine the degree to which missing information can be tolerated. Distinguish sporadic occurrences of missing noncritical data from systematic noncompliances. For example, if the LCS and method blank results were not included, the data would typically be of unknown quality and would be rejected. However, the entire data package would not typically be rejected if the dilution factor were not specified for one of the environmental samples.

b. If the data package is not substantively complete, (i) missing information must be requested from the laboratory, (ii) the data package must be rejected, or (iii) a limited data review must be performed. When the data package is grossly deficient, the reviewer should consult with the Project Manager to determine which option is most appropriate for the data objectives of the

investigation. For example, if missing deliverables cannot be obtained from the laboratory or if data evaluation cannot be delayed because of scheduling constraints (e.g., because time-critical decisions must be made), a more limited data review may be appropriate. If the data are being used to support critical decisions (e.g., for litigation), it may be necessary to reject the data when standards are not traceable or the COC is missing (since the integrity the data has not been definitively demonstrated).

c. When QC sample results are not included in the data package, it may be possible to use the available results to evaluate or to make inferences concerning the quality of the data. *However, this approach must be used with caution since the missing QC samples may have been analyzed by the laboratory but not reported because of QC failure.* Some evaluation strategies for incomplete data are discussed below.

4-3.1. Missing Blanks.

a. When a blank is missing, a higher **hierarchy** blank may be used to qualify the corresponding field samples for contamination. For example, if the method blank is not available (e.g., because it was not processed with the batch of samples), then field samples may be qualified using another associated blank that was processed with the batch of samples. In particular, if a rinsate blank or trip blank were analyzed with the batch of samples, rather than rejecting the data, the data could be evaluated using the trip blank or rinsate blank data. The trip blank or rinsate blank would be indicative of the accumulative field and laboratory contamination.

> Note: This approach will not be appropriate when the field blank is not processed in the same batch (or in the same manner) as the environmental samples for which it is assumed to represent.

b. Similarly, an environmental sample in the preparation batch for which no target analytes were detected may serve as a *field blank* (since such a sample would demonstrate the lack of systematic field and laboratory contamination).

> Note: Since samples are processed through a variety of handling, preparatory, and analysis procedures, blanks are typically collected during various stages of these procedures in a manner which would establish the source of contamination and enable the implementation of corrective action. In theory, if a **comprehensive blank** is truly representative of all contamination that could have been introduced from the time of sample collection to analysis (especially if the blank is free of contamination), then missing lower **hierarchy** blanks should not affect the data qualification. However, it should be noted that, in practice, it is difficult to obtain a representative **comprehensive blank**. For example, common laboratory contaminants such as methylene chloride can appear in blanks in a sporadic manner (e.g., may be present in a method blank but may not be present in a higher hierarchy blank such as a field blank).

c. When the method blank is missing and a higher hierarchy blank is not available to evaluate the data, qualification may be required for the corresponding field sample results. (Contractual corrective actions may also be required for missing method blanks, because labo-

ratories are usually required to process a method blank with each batch of field samples.) *In general, nondetections must not be qualified (unless other QC problems are present).* Detections must be qualified with the R, X, or N flag. Project-specific DQOs and professional judgment must be used to determine which flag is more appropriate. However, *the R or X flag must be used to qualify detections in the absence of a technically defensible rationale for using the N flag.* Some situations for which the use of the N flag may be appropriate are presented below.

d. It may be appropriate to qualify an analyte detection with the N flag if the analyte was reliably detected in other site field samples that were processed in a separate batch with an acceptable method blank. Since the probability of external contamination as the source of a detection typically varies inversely with the magnitude of the detected concentration, it may be appropriate to qualify very high level detections with the N flag (e.g., detections ten times greater than the MQL). If an action level is available, it may be appropriate to qualify detections less than the action level (especially very low-level detections) with the N flag. However, in the absence of a technically defensible rationale to do otherwise (i.e., when information supporting the validity of detected concentrations is not available), detections greater than an action level must be qualified with the R or X flag when blank results are not available.

4-3.2. Missing Laboratory Control Samples.

In general, when LCS results are not available for a particular batch of field samples, the data are of unknown quality and the associated field sample results must be qualified with the R or X flag. Since a laboratory is typically required to process at least one LCS with each batch of samples, contractual corrective action for unacceptable performance may be appropriate. Possible exceptions are discussed below.

4-3.2.1. Matrix Spike Data.

a. If the LCS is missing but all target analytes are present in the MS and precision information is available from the MS/MSD or MD, then the associated field samples may be qualified (for bias and precision) using the MS/MSD or MS/MD results since this data gives bias and precision information for the overall method. In particular, if all the MS/MSD or MS/MD recoveries and RPDs are in control, then the sample results need not be qualified. When the MS/MSD or MS/MD is not in control, the associated field samples must be qualified but the QC failure may be due to poor laboratory method performance (rather than matrix interference).

b. If the LCS is missing and only a subset of the target analytes is present in a MS/MSD or MS/MD pair, then qualify the associated field samples as discussed above for the subset of spiked analytes. Reject or tentatively reject the associated field sample results for the unspiked target analytes. If the data are being used to support noncritical decisions, (for target analytes not present in the matrix spike), it might be appropriate to reject or tentatively reject nondetections but to qualify detections as estimated.

4-3.2.2. Surrogate Spike Data.

Surrogate recoveries may be used to make inferences about the performance of organic analyses. If acceptable surrogate results are available and the data will be used to support noncritical decisions, it is recommended that the data be qualified as estimated rather than as rejected. *However, this approach must be used with caution since the surrogates may not be representative of all the target analytes* (e.g., although ketones are often target analytes for SW-846 Method 8260B, none of the surrogates recommended in the method are ketones).

4-3.2.3. CCV Data.

When all environmental and QC samples undergo the same preparatory and determinative processes or when significant sample preparation is not performed (e.g., aqueous VOCs by purge-and-trap GC/MS and aqueous anions by ion chromatography), CCVs are essentially LCSs. Method performance may be evaluated using the CCV results. If the CCV results are acceptable, then the data would not be qualified on the basis of the missing LCS results.

4-3.3. Missing Detection Limits.

a. Detection limits must be included in the data package. Laboratory detection limits are typically the **method detection limits (MDLs)** defined in 40 CFR, Part 136, Appendix B. *Instrument detections limits (IDLs) must not be viewed as a substitute for MDLs.* For example, *MDLs* may be significantly higher than *IDLs* when an analytical method involves extensive sample preparatory procedures.

> Note: Laboratories frequently perform an MDL study for a single instrument and erroneously report the resulting MDLs as applicable to all similar instruments used for the same analytical method. Method detection limits are matrix, method, and *instrument-specific*. For example, if a laboratory performs Method 8260B using six GC/MS instruments, then six sets of MDL studies must be available or sensitivity must otherwise be demonstrated for all six instruments. A valid conservative approach may consist of performing a separate MDL study for each instrument but reporting only the highest MDLs for each instrument.

b. When detection limits are not available (i.e., not included in the data package), the laboratory should be contacted for this information or the method reporting limits (MRLs) for nondetections must be set equal to the MQLs (e.g., when the laboratory lists reporting limits that are less than the MQLs). However, the MQLs must be verified as discussed in Chapter 6.

(1) If an action level is unavailable, then qualify all detections less than the MQL with the J flag. Report all nondetections as "< MQL" or "MQL U," where "MQL" denotes the numerical value of the method quantitation limit.

(2) If an action level is available, then compare the MQL to the action level. If the MQL is less than the action level, then qualify detections and nondetections as described above. *How-*

ever, if the MQL is greater than the action level, adequate sensitivity has not been demonstrated. Qualify nondetections with the X or XU flag and qualify detections less than the action level with the X flag. A detection below the action level does not demonstrate that the target analyte is actually present in the environmental sample below the action level because analytical uncertainty is large or undefined below the quantitation limit. Detections greater than the action level but less than the MQL may be qualified with the X flag but, at a minimum, must be qualified with the J flag. When resampling or reanalysis cannot be performed, the use of the J flag normally constitutes the more conservative approach.

CHAPTER 5

Holding Times and Preservation

5-1. Introduction.

The primary objective of this review is to ascertain the representativeness of the analytical data in the context of **preservation** and **holding time limits**[1]. **Holding times** for environmental samples are calculated from the dates of sample collection to preparation and analysis. (Refer to the glossary).

5-2. Acceptance Criteria.

a. The data package must clearly indicate the dates of all sample handling processes and the method of sample preservation. Holding time and preservation acceptance criteria and corrective actions are determined by method requirements and project DQOs. If holding time or preservation requirements are not specified for the project, use the published holding times and preservation requirements that are summarized in Table B-1 (Appendix B).

> Note: The preservation requirements and holding time limits were primarily taken from SW-846 and water methods (40 CFR, Part 136.3). Other sample preservation and holding time criteria may be more applicable (e.g., depending on the methods selected and matrices being tested).

b. Published holding time limits are generally considered maximum times that samples may be held before analysis and still be considered compliant with method guidelines, and typically apply to preserved samples.

> Note: Published holding times listed in environmental methods and regulations are not necessarily scientifically valid. However, use of alternative holding times may have a profound impact on the legal defensibility of results. It is recommended that published holding times be extended or shortened, if the client and regulators have agreed to other holding times for a particular project (e.g., based upon the chemistry of the method and holding time studies).

5-3. Evaluation.

a. Holding times and preservation are evaluated using the COC (Chain Of Custody) form, the laboratory's Cooler Receipt form (e.g., refer to the format of the USACE Cooler Receipt form), Case Narrative, sample preparation logs, and instrument run logs.

[1]Sampling design (e.g., sampling locations) probably affects representativeness far more than any other factor. In the context of a full data usability assessment, the review described here should be viewed as a screening process to determine if the samples are *potentially* representative of the environmental matrices being sampled.

(1) Check sample holding times using the date of sample collection listed on the COC form with the dates of sample preparation and analysis reported on the laboratory's sample summary forms.

(2) Verify that digestion, extraction, or cleanup dates reported on the sample summary forms are identical to the dates listed on the sample preparation log sheets.

(3) Verify that the analysis dates reported on the sample summary forms are identical to those listed on the instrument run logs.

(4) Review the Case Narrative and Cooler Receipt form included in the data package to determine if all the samples were properly preserved and holding times were met. Note any problems that may have impacted the integrity of the samples (e.g., samples not maintained at $2^{o}C$ - $6^{o}C$, aqueous VOC samples with head space, custody seals that are broken, and holding times that are not met). Verify that the pH of chemically preserved samples was checked and appropriate pH values were obtained.

b. If samples are properly preserved, holding time limits are met, and no problems with the samples are indicated in the Case Narrative, the laboratory's Cooler Receipt form, and the COC form, then assume that the physical integrity of the samples is acceptable.

c. If there are holding time violations or preservation problems, then the integrity of the samples may have been compromised. In the absence of information to the contrary, assume that holding time and preservation problems give rise to a *low* bias.

> Note: This assumption typically constitutes a conservative approach but will not always be appropriate. For example, the assumption will not be valid when degradation products are also target analytes. In particular, holding time noncompliances would presumably give rise to a low bias for pesticides such as DDT and Endrin, but to a high bias for associated degradation products such as DDE and Endrin aldehyde. In addition, when samples are held for an extended period of time, target analyte may leach or permeate into the storage containers, giving rise to a high bias. To address this potential problem, the representativeness of the method blanks would need to be evaluated. For example, if the method blank and samples were stored and analyzed together after the holding time limit, the absence of blank contamination in the method blank would suggest that leaching and permeation did not give rise to a high bias.

d. Distinguish *gross* holding time and preservation noncompliances from *marginal* noncompliances. Using the guidance presented in the "USEPA Contract Laboratory Program National Functional Guidelines For Organic Data Review," any holding time that is greater than twice the holding time limit is considered to be a *gross* holding time noncompliance. However, if holding time studies were performed, a gross holding time noncompliance would be defined on the basis of these studies.

e. Professional judgment is critical for evaluating holding time noncompliances. The evaluation of preservation problems is highly dependent upon the nature of the target analyte, the matrix, the method of analysis, and the sample handling procedures. For example, if aqueous samples for alkalinity were chemically preserved using pH adjustment, the analyses would be considered unusable. However, BNA soil sample results would not be rejected if a cooler temperature of 7°C were reported. Because of the time required to reach thermal equilibrium (relative to that required to ship the samples), samples may not cool during shipment to the 2–6°C acceptance range even when adequate refrigerant is placed in the cooler.

5-4. Qualification.

The data qualification strategies presented in this section of the document are conservative in nature. A holding time noncompliance is assumed to give rise to a low bias.

5-4.1. Low Stability Target Analytes.

If there is a holding time noncompliance or a preservation problem for analytes that are known to readily volatilize or degrade in the matrix being tested (e.g., aqueous aromatic VOCs and hexavalent chromium), then qualify the results as follows:

a. Qualify all nondetections with the R flag. For example, if large bubbles are reported in vials of the aqueous VOC samples or unpreserved aromatic VOC samples were analyzed beyond the limit, then qualify all nondetections with the R flag.

b. At a minimum, qualify all detections with the J- flag. If an action level is specified and the analyte is detected at a concentration less than the action level, then qualify the detection with the X flag. Although the detection may be *qualitatively* reliable, it is not *quantitatively* reliable. Although the analyte was detected (despite the noncompliant holding time or preservation problem), the reported result potentially possesses a low bias and does not demonstrate that the analyte is actually present in the environmental sample at a concentration less than the action level.

5-4.2. High Stability Target Analytes

If there is a holding time or preservation noncompliance for analytes that are relatively stable in the matrix being tested (e.g., dioxins and trace metals such as lead in soils), then qualify the data as follows:

a. If the holding time limit is *marginally* exceeded or the preservation problem does not appear to be significant, then qualify nondetections with the UN flag and detections with the J- flag.

b. If there is a *gross* holding time or significant preservation problem, then qualify the data as discussed in Paragraph 5-4.1; namely, qualify nondetections with the R flag and detections with the J- flag or X flag.

Table 5-1
Data Qualification for Holding Time Noncompliance

Stability	Holding Time (*t*) [1]	Flag	Remarks [2]
Low	$t \leq$ HTL	None	Holding time limit is met.
		R	$y <$ MRL
		X	MRL $< y <$ AL
	$t >$ HTL	J-	$y >$ MRL and y > AL
High	$t \leq$ HTL	None	Holding time limit is met.
		UN	$y <$ MRL
	HTL $< t \leq$ 2 HTL	J-	$y >$ MRL
		R	$y <$ MRL
		X	MRL $< y <$ AL
	$t >$ 2 HTL	J-	$y >$ MRL and $y >$ AL

Notes: 1. The project-required holding time limit and the calculated holding time for the sample are denoted by HTL and *t*, respectively. 2. The concentration of the field sample, the action limit, and the method reporting limit are denoted by *y*, AL, and MRL, respectively.

CHAPTER 6

Data Review Reports

6-1. Introduction.

The primary objective of this review is to ensure that analytical sensitivity is adequate for project-specific action levels and to ensure that data are reported in a manner that is consistent with the laboratory's detection and quantitation limits. The evaluation of sensitivity will be a function of how the detection, quantitation, and reporting limits are defined and whether or not action levels are specified. Since it is impractical to discuss sensitivity in the context of multiple definitions for these limits, they will be defined as the **method detection limits (MDLs)**, **method quantitation limits (MQLs)**, and **method reporting limits (MRLs)** presented the glossary. In particular, it is assumed that the detection limit is the method detection limit of 40 CFR, Appendix B, Part 136. The method quantitation limit is defined (primarily) as the low level calibration standard adjusted for method specific factors. Lastly, the method reporting limit is defined as the threshold or censoring limit below which target analyte concentrations are reported as "ND" (i.e., not detected) or as "<" (i.e., "less than").

6-2. Method Reporting Limits.

6-2.1. Establishing Method Reporting Limits.

a. The definition of the MRLs must be declared in each data package or in project documents such as the Quality Assurance Project Plan (QAPP).

> Note: Merely listing numerical values for the MRLs will not satisfy this reporting requirement; the MRLs must be defined in terms of the laboratory's actual quantitation and detection limits.

b. In general, any analyte concentration greater than the detection limit may potentially be reported either as a "detection" or as a "nondetection" with respect to some censoring (reporting) limit greater than the detection limit. For example, if an action level, AL, is very large relative to the MQL, then it may be desirable to report all analyte concentrations less than 5% of the AL action level as "< MRL" or "MRL U," where MQL < MRL = 0.05 AL. In this context, "< MRL" indicates that (i) the analyte is present below the detection limit, or (ii) was detected at some concentration greater than the detection limit but less than 0.05 AL. Conversely, if low-level reporting is desirable, then it may be appropriate to establish a censoring limit (MRL) at some concentration greater than the method detection limit (MDL) but less than the MQL. Under these circumstances (MDL < MRL < MQL), analyte concentrations between the MRL and MQL would be reported as estimated and concentrations less than the MRL would be reported as "< MRL."

> Note: The term *reporting limit* is being defined in a more general manner than is conventionally used for environmental testing. For example, according to the CLP Statement of Work (SOW) for organic analyses, the reporting limit for nondetections is nec-

essarily the CRQL. If an analyte is "not detected," the reporting limit is the CRQL and detections below the CRQL are reported as estimated. *However, there is no a priori reason for setting the reporting limit equal to the quantitation for all data uses.* For example, presence-absence issues can typically be resolved at concentrations that are significantly less than the quantitation limits.

c. In the absence of project-specific guidance, assume that all reliable detections greater than the MDL or MRL must be reported. (Note that detections should be reported based upon the laboratory's detection limits as well as the analyst's judgement.) In addition, assume that the MRL for **a nondetection** must be no less than the **limit of identification (LOI)** or the **reliable detection limit (RDL)**. The RDL and LOI are approximately two times the MDL.

Note: Establishing *any* reporting limit constitutes a form of "data censoring." Censoring results (i.e., reporting nondetections to twice the MDL) will typically be appropriate when action levels have been established and the results (detections and nondetections) are being compared to the action levels on a point-by-point basis. *However, this approach will not be appropriate for all projects.* For statistical applications, it is usually desirable to report results *without any censoring* (e.g., to report results less than the MDL). The reviewer must refer to project-specific objectives prior to performing censoring or evaluating the data with respect to the reporting limits.

6-2.2. Qualification.

a. If the reporting limit is less than the LOI or the RDL (i.e., two times the MDL), then qualify nondetections (at the reporting limit) with the UN flag and discuss the potential high false negative probability at the reporting limit in the data review report. Alternatively, if the project action levels (ALs) are relatively high (e.g., at least 10 to 20 times greater than the MQL), increase the reporting limit to the quantitation limit (if the quantitation limit was established by the lowest calibration standard) and qualify nondetections with the U flag.

b. If an action level (AL) is available, compare the MRL to the AL. *If the MRL is greater than the AL, qualify nondetections with the X or XU flag (since false negatives have not been adequately addressed).*

c. It is recommended that the MRL be no higher than %5 to 10% of the AL. If the MRL is less than but near the AL, then use professional judgement to qualify nondetections, especially when the AL is less than the MQL or the LCS acceptance limits are wide. For example, if the MQL = 50 ppb, the MRL = 10 ppb, AL = 15 ppb, the LCS acceptance range is 50–150% (e.g., for a 100 ppb spike near the mid-calibration range), and the LCS recovery associated with a set of environmental samples is 55%, then nondetections reported as "< 10 ppb" do not demonstrate the 15-ppb action level was met. Under these circumstances, nondetections would be qualified with the X or XU flag

6-3. Method Quantitation Limits.

6-3.1. Establishing Method Quantitation Limits.

a. Project planning documents (e.g., the QAPP) must define what constitutes a quantitation limit. In general, project documents should specify tolerances for uncertainty at the quantitation limit and strategies for verifying the tolerances have been satisfied (e.g., a low-level LCS at the quantitation limit must be recovered to within 20% of its expected value). *Unfortunately, quantitation limits are often poorly defined.*

> Note: The laboratory's reported MQLs must not be evaluated solely upon the basis of "Practical Quantitation Limits" ("PQLs") or "Contract Required Quantitation Limits" (CRQLs) specified in published analytical methods or project documents unless these quantities are adequately defined (e.g., tolerances for uncertainty at the quantitation limits are specified).

b. The guidance presented below will typically be applicable.

(1) A low-level LCS or CCV (spiked with the target analytes at or near the MQL) may have been analyzed to verify the quantitation limit. Low-level CCVs would be appropriate for methods that do involve significant sample preparation or for methods in which the calibration standards are prepared with the environmental samples. Low-level CCVs can often be used to verify the quantitation limits for inorganic methods (e.g., when the sample preparatory process does not introduce too much uncertainty). However, this approach will not be valid for methods that involve significant sample preparation and the CCVs are not processed with the environmental samples. Under these circumstances, a low-level LCS (spiked with target analytes at or near the MQL) is required to verify the quantitation limit.

(2) If a low-level CCV (e.g., the lowest calibration standard) was used to check the MQL, verify that the CCV was recovered to within the tolerance for instrumental uncertainty (the acceptance limits must be equal to or slightly greater than the acceptance limits for mid-level CCVs). For example, for trace metals by ICP, the low-level CCV should be recovered to within 10% to 15% of its expected value. If a low-level LCS was used to check the MQL, then verify that the low-level LCS was acceptably recovered.

(3) If a low-level CCV or LCS spiked at the MQL or near the MQL (e.g., less than two times the MQL) was not analyzed, then compare the reported MDL to each corresponding MQL as discussed below (i.e., verify that each MQL is at least five to ten times greater than the MDL and was established using the lowest initial calibration standard).

(4) Use the calibration data to verify that the laboratory's reported quantitation limit for each analyte is established from the lowest calibration standard (or corresponds to a higher concentration that is within the calibration range).

Note: *This is not a sufficient condition to verify the project-required method quantitation limits*. It is often erroneously concluded that if the initial calibration curve is acceptable (e.g., as indicated by a high correlation coefficient), then the lowest calibration standard will be acceptable for establishing the MQL. However, an acceptable fit for the entire calibration curve does not necessarily imply that the uncertainty will be acceptable at concentrations near the lowest calibration standard. Conventional measures of fit are not adequately sensitive to high variability at the low concentration ranges. For example, when regression analysis is used to fit initial calibration results, a high correlation coefficient is possible when the lowest standard radically deviates from a linear fit (e.g., when instrumental response is inherently nonlinear at low concentrations).

(5) Compare the MDLs (if available) to the corresponding MQLs to ensure that the quantitation limits are sufficiently greater than the detections limits. *If the MQL is established from the lowest calibration standard but is not otherwise defined, ensure that the MQL is at least five to ten times greater than the method detection limit.*

Note: The quantitation limit will be dependent upon the magnitude of the analytical noise (whether chemical or electronic in nature) that constitutes the "background" signal or response for the analysis method, and the project-required tolerance for uncertainty for quantitation. Since the detection limit is measure of "background" response, the quantitation limit must typically be greater than the detection limit *by some multiplicative factor* in order to meet the project-required error tolerance. In general, when a low error tolerance is required, the quantitation limit must be significantly greater than the detection limit.

If it is assumed that the magnitude of the analytical uncertainty is approximately ± MDL, then the relative uncertainty will be about ± 20% at five times the MDL and ± 10% at ten times the MDL. (It is being assumed that the standard deviation determined from the MDL study is not strongly dependent upon concentration and there is no significant bias.) However, the actual relative uncertainty will often be higher than 10% to 20% at five to ten times the MDL (e.g., because the standard deviation is often an increasing function of concentration).

(6) If the laboratory's reported quantitation limit is less than the method quantitation limit calculated from the lowest initial calibration standard and the standard is at least five times greater than the MDL, then increase the quantitation limit using the lowest calibration standard.

(7) If the lowest calibration standard is not at least five times greater than the MDL and an acceptable low-level CCV or LCS was not analyzed to verify the MQL, then the initial calibration results must be evaluated. If the low-level calibration standard is less than five times the MDL, it may be appropriate to use the next highest calibration standard to establish the MQL. If possible, use the equation for the initial calibration curve to calculate the concentration of the lowest calibration standard (i.e., calculate the concentration of the lowest standard from the measured response) and ensure that the calculated value of the lowest standard is within the uncertainty tolerance for the CCV. If it is not possible or practical to determine the MQL from

the calibration data, then set the MQL to five to ten times the MDL, but indicate that the MQL is an estimate in the data evaluation report. Multiply the MDL by at least a factor of ten for ICP analyses.

6-3.2. Qualification.

a. Once the MQLs have been verified or established, *qualify all detections less than the MQLs as estimated using the J-flag* (e.g., unless the X or R flag is more appropriate because significant QC problems are observed).

b. If action levels are available, compare the MQLs to the action levels and ensure that the MQLs are *less* than the action levels. Although the MQLs should have been compared to the project's action levels during the planning stages of the project, sensitivity problems may still occur (e.g., because of dilutions). As a "rule of thumb" the MQL should not be greater than about one half of the AL for inorganic analyses and about one third of the AL for organic analyses.

c. *If the MQL is greater than a corresponding AL, adequate sensitivity has not been demonstrated; qualify detections less than the AL with the X flag.* Under these circumstances (MQL < AL), depending upon project DQOs, it may be appropriate to also qualify detections greater than the AL with the X flag (e.g., when a conservative estimate of contamination is not desirable).

Table 6-1
Data Qualification for Sensitivity When Action Levels Are Available

Sample Result (y) LOI ≤ MRLs [1]	Flag	Remarks
y < MRL < AL	U	Nondetections
y < MRL, MRL > AL	X, XU	
MRL ≤ y < AL < MQL	X	Detections
MRL ≤ AL < y < MQL	J or X [2]	
MRL ≤ y < MQL < AL	J	
MRL ≤ MQL < y	No flag	

Notes: 1. The action level, method reporting limit, and method quantitation limit are denoted as AL, MRL, and MQL, respectively. The concentration of the target analyte in a field sample is denoted as y. (It is assumed that the limit of identification is less than or equal to the MRL.) 2. A detection above the AL was obtained. However, because quantitative uncertainty is high, the target analyte may not actually be present in the sample at a concentration that exceeds the AL; the X flag may be appropriate. The use of the J flag constitutes a conservative interpretation of the data (namely, that the AL has been exceeded).

CHAPTER 7

Initial Calibration

7-1. Introduction.

The **initial calibration** is evaluated to ensure that the instrument was capable of producing acceptable quantitative data *prior* to the analysis of environmental samples. The concentration range and number of standards for the initial calibration will be dependent upon the instrument, method, and objectives of the project. The variation in instrumental response with concentration may define a straight line or some curve. Instrumental response may be expressed either as peak area (e.g., determined from a sum of detector signals) or as peak height (e.g., the maximum detector signal above background noise).

7-2. Acceptance Criteria.

The calibration criteria presented below are based upon the guidance presented in USACE Shell document and SW-846 (Method 8000B). However, the criteria are conservative and should be applicable to most instrumental methods. Calibration may be performed using linear or nonlinear fits. However, linear calibrations should be used in preference to nonlinear calibrations

7-2.1. Frequency.

a. An initial calibration must be performed prior to the analysis of samples and when a continuing calibration verification is unacceptable.

b. For inorganic analyses, the initial calibration is typically performed at the beginning of each analytical shift in which analyses are performed; that is, each time the instrument is set up to perform analyses (e.g., turned on and "warmed up"). When analyses are performed continually, the initial calibration is typically performed on a daily basis (i.e., every 24 hours).

7-2.2. Number of Calibration Standards.

a. The number of calibration standards (or points) is highly method dependent. The number of calibration standards will be proportional to the variability of instrumental response. For example, the higher the variability of response, the greater the number of calibration standards that will be required. As a general rule, at least three calibration standards should be used for linear calibrations. Additional calibration standards should be used if the linear calibration range is greater than one or two orders of magnitude or when nonlinear calibrations are performed.

b. When calibration curves are fitted with polynomials using regression analysis, the number of initial calibration standards must be sufficient for at least one statistical degree of freedom. The degrees of freedom for a regression curve, $df = n - 1 - k$, where $k > 0$. The variable k is the order of the polynomial and n is the number of initial calibration standards. For example, if a regression line is being used ($k = 1$), then the number of calibration standards (n) must be

greater than or equal to three. However, beyond the minimum number of standards required to perform the actual mathematical fit (e.g., three standards for linear regression lines), the number of initial calibration standards required is somewhat arbitrary: Any number standards may be used to generate a calibration curve that meets the specified tolerance for uncertainty. However, when regression analysis is not being performed, only a single calibration standard may be required. For example, if response has been demonstrated to be linear through the origin (e.g., as in ICP analyses), the initial calibration may be performed using only a single high-level standard.

c. It is recommended that the number of standards be equal to or greater than that specified in the most applicable promulgated USEPA method. For example, if SW-846 methods are specified in the Sampling and Analysis Plan (SAP), then the guidance presented in the individual SW-846 method should be followed *to the extent that is appropriate for the objectives of the project.* For the purposes of illustration, the guidance in Update III of SW-846 for linear calibrations is summarized below. However, additional criteria are also specified.

(1) At least five calibration standards are used for all single-component target analytes and surrogates. (Single-standard calibrations are not typically acceptable for surrogates).

(2) A blank and a minimum of three calibration standards are used for SW-846 trace metal analyses (e.g., GF-AA, FL-AA, and ICP analyses). However, a blank and a minimum of five calibration standards are used for mercury analyses.

(3) A blank and one calibration standard may be used for ICP analyses. However, though not specified in Method 6010B, *the initial calibration line should be initially verified using a low-level CCV in order to establish the method quantitation limit.*

(4) Calibration for multi component analytes such as Aroclors are performed using a single standard. Instrumental response is related to concentration using the peak areas or heights of several peaks (e.g., at least three **characteristic peaks** for the Aroclors). *However, contrary to the guidance presented in the method, a minimum of three calibration standards is recommended for chromatographic analyses for multi-component analytes.*

7-2.3. Linear Calibration.

Linear calibrations must be performed using regression analyses (with the possible exception of chromatographic analyses). Instrumental response, as measured by the slope of the calibration line, must be high relative to analytical uncertainty (e.g., calibration lines with very "shallow" slopes would not acceptable).

7-2.3.1. Organic Methods.

The coefficient of determination (i.e., the square of the correlation coefficient r) must be equal to or greater than *0.980* for each target analyte. Mean response factors may be used to perform linear calibrations through the origin for chromatographic analyses. The **percent relative standard deviation (%RSD)** for the mean response factors must be equal to or less than *15%* for each target analyte. However, if only three calibration standards are used (e.g., for multi compo-

nent analytes such as PCBs), the coefficient of determination should be *0.990* or greater or the %RSD should be *10%* or less. In addition, any method-specified minimum response factors must be met.

> Note: A number of analytical methods (e.g., SW-846 methods) specify a maximum %RSD of 20% for the initial calibration for chromatographic methods with 2-D detectors. An acceptance limit of 20% is not recommended. Initial calibration lines with %RSDs of 20% often exhibit poor linear fits (e.g., a calibration line may not adequately fit the data at the upper end of the calibration range or the regression coefficient may be less than 0.99).

7-2.3.2. Inorganic Methods.

The coefficient of determination (square of the correlation coefficient r) must be at least *0.990* for each target analyte. Note that when a single standard is used to perform the initial calibration for ICP analyses, a correlation coefficient cannot be calculated. Under these circumstances, the initial calibration must be evaluated using CCVs at multiple concentrations.

7.2.4. Nonlinear Calibration.

a. Nonlinear calibrations are appropriate when linear calibrations cannot be performed over a sufficiently wide working range (e.g., when detector response is inherently nonlinear over a calibration range that spans two orders of magnitude or less). Nonlinear calibrations are inappropriate to compensate for detector saturation at higher concentrations or to avoid proper instrument maintenance. A large reduction in instrumental response (e.g., "curve flattening" characteristic of detector saturation) must not occur near the upper portion of the curve corresponding to high concentrations. Instrumental response must be high relative to analytical uncertainty and must be similar in magnitude to that for linear calibration.

b. All nonlinear calibrations must be performed using regression analysis. Nonlinear calibration curves must be generated using polynomial fits of no higher than third order (e.g., $y = ax^3 + bx^2 + cx + d$) and must possess at least three statistical degrees of freedom. The coefficient of determination must be equal to or greater than *0.99*.

7-2.5. Intercept of Calibration Curve.

A calibration curve (whether linear or nonlinear) must *not* be forced through the origin unless it is demonstrated (e.g., during method development) that the intercept (i.e., $y[x = 0]$) is not statistically different from zero (e.g., by performing a t-test for the y-intercept or comparing it to the MDL.) Arbitrarily forcing a calibration curve through the origin may adversely impact low-level quantitative results. Similarly, when calibration curves are generated using regression analysis, the curves must not be artificially weighted toward the origin by including the point (0,0) one or more times.

7-3. Evaluation.

Review the standard preparation and run sequence log sheets to verify that the initial calibration was performed at the appropriate frequency using the appropriate number of standards. Review the calibration summary results (e.g., Chapter 4.2.10) to ensure that acceptable initial calibrations, CCVs, and ICVs were performed.

7-3.1. "Goodness of Fit."

a. Check the linearity or "goodness-of-fit" for the calibration curve for each target analyte and surrogate by performing a qualitative visual examination of each calibration plot. Any unusual problem observed during the visual examination of the calibration plots must be noted (e.g., shallow slopes, high curvature at the upper of the calibration curve indicative of detector "saturation," large negative or positive intercepts, and large scatter).

b. When regression analysis is performed, it is especially critical to evaluate the fit for the calculated curve near the origin. In particular, a high correlation coefficient can be obtained (e.g., $r > 0.995$) when the calibration points significantly deviate from the curve.

c. Determine whether each of the reported regression coefficients or %RSDs are acceptable. Using the calibration summary forms, recalculate the regression coefficient or %RSD (depending on how the initial calibrations were performed) for at least one target analyte, and compare the recalculated value to the reported value. If calibration was performed using the **internal standard technique,** recalculate the relative response factors, mean response factor and %RSD for one target analyte using the appropriate internal standard. The reported and calculated values must agree to within *two* significant figures.

7-3.2. Representativeness of Initial Calibration Curve.

a. Verify that the initial calibration is representative of the analysis. In particular, review the calibration summary forms to determine if any calibration points were erroneously "censored;" that is, whether any calibration standards were *inappropriately* omitted from the "goodness-of-fit" calculations (e.g., for the calculation of the %RSD or correlation coefficient). Since instrumental response will be linear over a sufficiently narrow concentration range, the standards at the extreme upper and lower ends of the calibration range (i.e., the highest and lowest calibration standards) may be dropped to obtain a better linear fit. *However, it is usually inappropriate to drop calibration standards in the middle of the concentration range!*

b. When any calibration standard between the highest and lowest initial calibration standards is omitted, verify that a legitimate reason is documented (e.g., operator blunder, instrument malfunction, and the inadvertent use of an expired calibration standard). Censoring calibration results (with the exception of the points corresponding to the extreme ends of the concentration range) for no other reason than to obtain a "better" curve fit is an *inappropriate* laboratory practice. An "errant" calibration result may actually be a valid result that reflects higher-than-expected random analytical error. Under these circumstances, the omission of the calibration result will under estimate the actual uncertainty associated with the environmental samples.

7-4. Qualification.

7-4.1. Frequency and Number of Standards.

If the instrument was not calibrated at the appropriate frequency with the appropriate number of standards, qualify the associated sample results using professional judgement. For example, if only four points were used to perform the initial calibration for the mercury analyses and a high correlation coefficient was obtained, the associated mercury results would not typically be rejected. The width of the calibration range and the goodness of fit must be taken into account. As a "rule of thumb," at least three calibration standards are required when the quantitation range is two orders of magnitude in width; additional standards are frequently required for larger calibration ranges.

7-4.2. Representativeness.

When mid-level standards are inappropriately dropped, if possible, recalculate the calibration curve and the "goodness of fit" (e.g., r^2 or %RSD). If the "goodness of fit" is unacceptable, then qualify the associated sample results as discussed in Chapter 7.4.3. If it is not possible or practical to recalcuate the curve and the "goodness of fit," then, at a minimum, qualify all associated detections with the J flag and all nondetections with the UN flag. The X flag may be appropriate if the data are being used to support critical decisions.

7-4.3. "Goodness of Fit."

a. Results must be qualified on the basis of quantitative acceptance limits for the calibration fits (e.g., the regression coefficients) and the *visual examination of the initial calibration plots*. In particular, it may be necessary to qualify results when an *acceptable* coefficient of determination or %RSD is obtained for the calibration. For example, if a nonlinear curve possesses an acceptable coefficient of determination, it would be appropriate to qualify high concentration samples if severe curve "flattening" from detector saturation were observed. Detections near the intercept may be unreliable when a large nonzero y-intercept is obtained.

b. A conservative approach is recommended when method-specified minimum response factors are not met. It is recommended that detections and nondetections be qualified with the X flag or the R flag. (Minimum relative response factors for GC/MS analyses are typically 0.05 or greater.) However, use professional judgement, and, at a minimum, qualify nondetections with the UN flag and detections with the J flag.

c. Table 7-1 lists (to three significant figures) the maximum linear correlation coefficient that is not significantly greater than zero as a function of the number of calibration points (n) and confidence level (CL). For example, when the initial calibration is performed using five calibration points (e.g., for organic analyses), a correlation coefficient of 0.93 or less does not demonstrate a positive linear correlation exists between instrument response and concentration at the 99% confidence level. Therefore (using the 99% confidence level), if the correlation coefficient is less than or equal to 0.93, then detections and nondetections are qualified with the X or R flag. Similarly, when the initial calibration is performed using four calibration points (e.g., three cali-

bration standards and a blank for the metal analyses), a positive linear correlation is not demonstrated at the 99% confidence level when the correlation coefficient is less than or equal to 0.98.

Table 7-1
Maximum *r* (Linear Correlation Coefficient) Not Significantly Greater Than Zero Versus Number of Calibration Points for Initial Calibration [1]

n	90% CL	95% CL	97.5% CL	99% CL
3	0.951	0.987	0.996	0.999
4	0.800	0.900	0.950	0.980
5	0.687	0.805	0.878	0.934
6	0.608	0.729	0.811	0.882
7	0.550	0.669	0.754	0.832

Notes: 1. CL denotes the confidence level and *n* denotes the number of points used to generate the initial calibration line. The above table was generated using a one-tail t-test for the null hypothesis that the correlation coefficient is not greater than zero

d. In the absence of project-specific requirements to do otherwise, qualify the results as discussed below. It is assumed that at least five points (i.e., calibration standards) are used to perform the initial calibration. When the initial calibration is performed using less than five points, the acceptance limits for the %RSD and the correlation coefficient should be more stringent.

e. When the initial calibration is performed using five or more calibration points, the acceptance criteria for the initial calibration are not met, and the correlation coefficient is *greater than 0.93 or* the %RSD is *less than or equal to 30%* (when mean response factors are used to perform the initial calibration), then, at a minimum, qualify nondetections with the UN flag and detections with the J flag (as illustrated in Table 7-2). If the correlation coefficient is less than or equal to 0.93 or the %RSD is greater than 30%, then qualify the results (i.e., the detections and nondetections associated with the noncompliant target analyte) with the R flag. However, professional judgement should be applied. For example, if the %RSD is grossly unacceptable, but a review of the response factors for the initial calibration indicates that the %RSD is unacceptable because of high variability at the high-end of the calibration line, it may be more appropriate to qualify nondetections with the UN flag rather than the R flag.

f. Qualification of results for unacceptable calibration fits may be avoided when it is possible to recalculate the calibration curve from the information enclosed in the data package. For example, if the coefficient of determination or %RSD is unacceptable for a linear calibration (e.g., because of curvature at the extreme low or high end of the calibration range), then dropping the high or low end calibration standard may give an acceptable calibration line (i.e., an acceptable fit). The new calibration line may then be used to recalculate all the reported sample results. However, at least three calibration standards are required to calculate a calibration line

using regression analyses. Furthermore, since this strategy narrows the quantitation range, results that fall outside of the new calibration range would need to be qualified (as described below).

> Note: The degree to which recalculations will be required for the initial calibrations to avoid data qualification will be highly situation-dependent. For example, it will be function of the completeness of the data package, the level of effort negotiated for the data review, and the degree to which QC requirements and corrective actions were specified in the contract for laboratory analytical services.

Table 7-2
Data Qualification for "Goodness of Fit" for Organic Analyses with Five-Point Linear Calibrations [1]

"Goodness of Fit"	Remarks	Sample (y)	*Flag*
$r > 0.99$ %RSD ≤ 15%	Acceptable Calibration	MRL < MQL < y	None
		MRL < y < MQL	J
		y < MRL	U
0.93 < r < 0.99 15% < %RSD ≤ 30%	Marginal Failure	y > MRL	J
		y < MRL	UN
$r \leq 0.93$ %RSD > 30%	Gross Failure	y > MRL	R
		y < MRL	R

Notes: 1. %RSD, r, MRL, MQL, and y denote the percent relative standard deviation of the response factors, the linear correlation coefficient, method quantitation limit, method reporting limit, and the concentration of analyte in the associated field sample, respectively. (The MRL is assumed to be greater than the MDL but less than the MQL.)

7-4.4. Calibration Range.

a. All detections that fall outside of the calibration range of the instrument must be qualified. For calibrations performed with multiple standards, the lowest and highest calibration standards determine the lower and upper limits of the calibration range, respectively.

b. If the initial calibration is performed using a single (high) calibration standard for ICP analyses, then the lower limit of the calibration range must be established by a low-level CCV standard that is analyzed immediately after the initial calibration. If the initial calibration was not verified via the analysis of a low-level CCV standard or if a low-level CCV was analyzed but possesses a recovery that does not fall within 85% to 115%, then multiply the MDL by a factor of ten (10) to establish the lower calibration limit (and the MQL). However, if the low-level CCV is at least ten times greater than the MDL and the recovery does not fall within 85% to

115%, then, at a minimum, qualify all detections between the low-level and mid-level CCVs as estimated and qualify any associated nondetections with the UN flag.

c. As stated in Chapter 6, detections that are less than the lower limit of the calibration range (e.g., the low calibration standard) must be qualified with the J flag. In addition, all detections that *marginally* exceed the high calibration standard must be qualified with the J flag; detections that *grossly* exceed the upper calibration range must be qualified with the X flag (or the R flag). The criteria for *marginal* versus *gross* failures must be determined using professional judgement. The determinative technique as well as the range, fit, and shape of the calibration curve must be taken into consideration. In general, if a sample result exceeds the upper calibration standard within the uncertainty tolerance for the CCV, it is recommended that the result be qualified as a marginal failure. For example, if the high calibration standard is 100 ppb and the CCV must be within 15% of its expected value, then detections greater than 100 ppb but less than 115 ppb should be qualified with a J flag.

CHAPTER 8

Initial Calibration Verification (ICV)

8-1. Introduction.

The **initial calibration verification (ICV)** is evaluated to assess the accuracy of the initial calibration standards[1].

8-2. Criteria.

The ICV must be performed after the initial calibration via the analysis of a mid-level standard. The working calibration standards and the ICV standard must be from *independent* sources (e.g., from two different manufacturers). The recovery of the ICV should be within 90% to 110% for inorganic analyses and within 80% to 120% for organic analyses.

8-3. Evaluation.

a. Review the standard preparation logs to verify that the ICV and initial calibration standards were prepared from independent NIST-traceable standards. Review the instrument printouts and run log sheets to verify that the ICV was analyzed after the initial calibration within its expiration date. Using the standard preparation log sheets and the ICV summary form, recalculate an ICV recovery and compare the calculated value with the reported value. If an ICV standard was not prepared, review the standard preparation log to determine whether or not any CCVs or LCSs were prepared from an independent-source standard

b. It should be noted that an ICV failure does not definitively demonstrate a source problem for the initial calibration standards. For example, failures may occur because of problems with the initial calibration curve (e.g., a poor fit) and analytical blunder. Prior to qualifying the data, it may be desirable to investigate the source of the failure (e.g., by requesting additional information from the laboratory).

[1]One could argue that an acceptable ICV does not definitively demonstrate the accuracy of the standards used for the initial calibration. For example, the spiking concentrations for both the initial calibration and ICV standards could be biased low (relative to the actual analyte concentrations in the standards). However, since both standards should be traceable to a reliable source (e.g., NIST), an acceptable ICV supports the conclusion that the standards are accurate. It is more likely than not that two different traceable standards that are in agreement are accurate.

8-4. Qualification.

8-4.1. Frequency.

a. When an ICV is not performed (i.e., when initial calibration standards are not verified with an independent-source standard) at a minimum, qualify all detections with the J flag and all nondetections with the UN flag. Alternatively, the data review report must state that all the results are potentially estimated. Rejection of the data may be appropriate when the data is being used to support critical decisions.

b. The CCVs or the LCS may have been prepared from an independent-source standard. If CCVs or LCSs are prepared from independent-source stock standards and the recoveries are acceptable, then the data must not be qualified. If an independent-source standard for the ICV is not commercially available, a standard from the same source material but a different preparation lot (e.g., different manufacturer's lot number) may be used for the ICV standard.

8-4.2. Percent Recovery.

8-4.2.1. Inorganics.

If the ICV recovery is unacceptable but falls within 80% to 120%, qualify detections with the J flag and nondetections with the UN flag. If the ICV recovery does not fall within 80% to 120%, then qualify the results with the X flag.

8-4.2.2. Organics.

If the ICV recovery is unacceptable but falls within 70% to 130%, qualify detections with the J flag and nondetections with the UN flag. If the ICV recovery does not fall within 70% to 130%, then qualify the results with the X flag.

8-4.3. Qualification for Bias.

When the ICV recovery is unacceptable or an independent-source standard is not used to verify the initial calibration standard (e.g., an ICV is not performed), *the direction of bias is unknown for the entire analytical process.* The recoveries of other QC samples (e.g., laboratory control samples and matrix spikes) must not be used to make inferences about the direction of bias (e.g., unless the uncertainty is much greater than that arising from the ICV noncompliance.)

Table 8-1
Data Qualification for ICV Results [1]

ICV %R	Method	Remarks	Sample (*y*)	Flag
90% ≤ %R ≤ 110% 80% ≤ %R ≤ 120%	Inorganics Organics	Acceptable %R	MRL < MQL < *y*	None
			MRL < *y* < MQL	J
			y < MRL	U
110% ≤ %R ≤ 120%, 80% ≤ %R ≤ 90% 120% ≤ %R ≤ 130%, 70% ≤ %R ≤ 80%	Inorganics Organics	Marginal Failure	*y* > MRL	J
			y < MRL	UN
%R > 120%, %R < 80% %R > 130%, %R < 70%	Inorganics Organics	Gross Failure	*y* > MRL	X
			y < MRL	X

Notes: 1. %R, MRL, MQL, and *y* denote the percent recovery of the target analyte in the ICV, method reporting limit, method quantitation limit, and concentration of the target analyte in an associated field sample, respectively. (The MRL is assumed to be greater than the MDL but less than the MQL.)

CHAPTER 9

Continuing Calibration Verification (CCV)

9-1. Introduction.

Continuing calibration verifications (CCVs) are evaluated to determine whether the instrument was within acceptable calibration throughout period in which samples were analyzed (i.e., to verify that the initial calibration was applicable during the sample analyses). In general, failure of the CCV indicates that the initial calibration is no longer valid and should trigger recalibration and the reanalysis of the associated samples in the analytical sequence.

9-2. Criteria.

9-2.1. Traceability and Reporting Requirements.

The initial calibration and the sample analyses associated with each CCV must be clearly indicated in the run log. The run log must also list the date each CCV standard was analyzed; the time of analysis must also be specified for chromatographic methods. In addition, the source, reference concentration (level spiked), measured concentration, and percent recovery must be reported for each target analyte and surrogate (when surrogates are analyzed). However, for chromatographic methods where the initial calibrations are performed using mean response factors, the **percent differences** for the CCV response factors may be reported instead of percent recoveries (e.g., when the instrument's software cannot readily report CCV recoveries).

9-2.2. Representativeness.

CCVs must be analyzed in the same fashion as other QC samples (e.g., LCSs) and environmental samples (i.e., must be analyzed in a manner that is representative of all other sample in the analytical sequence).

9-2.3. Frequency.

a. All environmental samples in an analytical sequence must be bracketed by (i) an initial calibration and a CCV or (ii) by two CCVs. Therefore, a CCV must be analyzed at the end of every analytical sequence.

b. If replicate CCVs are analyzed in succession before or after a set of samples, the CCVs analyzed *immediately* before and after the samples constitute the bracketing pair of CCVs. For example, "Sample-01" and "Sample-02" are qualified based upon the performance of "CCV-02 and "CCV-03" for the analytical sequence:

CCV-01, CCV-02, Sample-01, Sample-02, CCV-03, CCV-04 ...

c. However, it should be noted that a single reinjection of the CCV is typically performed when a CCV fails. Therefore, if CCV-03 were to fail, the bracketing CCVs would consist of CCV-02 and CCV-04.

9-2.3.1. Chromatographic Methods.

For chromatographic methods, a low-level or mid-level CCV standard must be analyzed at the following frequency: (i) At the *beginning* of the analytical shift/sequence (when an initial calibration is not being performed); (ii) every *12 hours* of analyses or every *10 to 20 samples*, whichever comes first; and (iii) at the *end* of the analytical sequence--*this includes GC/MS methods*.

> Note: The term "sample" refers to field samples and batch QC samples such as method blanks, laboratory control samples, matrix spikes, matrix spike duplicates, matrix duplicates.
>
> Note: *Laboratories do not typically analyze a CCV at the end of the run sequence for GC/MS analyses. In order to minimize the occurrence of estimated data, this requirement must be explicitly specified when contracting for laboratory analytical services. Alternatively, a rationale for not analyzing CCVs at the end of the run sequence for GC/MS analyses must be presented in project documents such as the QAPP.*

9-2.3.2. Inorganic Methods.

For inorganic methods, a low-level or mid-level CCV must be analyzed at the following frequency: (i) Every 10 to 15 samples and (ii) at the end of the analytical sequence.

9-2.4. Acceptance Criteria.

The acceptance limits for the CCVs will be highly dependent upon the analytical technique (as well as the end use of the data). Therefore, several assumptions were made to develop the data evaluation strategies that are presented in this section of the document. It was assumed that the acceptance limits for the CCV are more stringent than the acceptance limits for the LCS when the method of analysis involves significant sample preparation and the standards are not fully processed with the environmental samples. Similarly, it was assumed that the CCV and LCS limits will be similar when the method of analysis does not involve significant sample preparation. Lastly, a "gross" CCV failure was typically assumed to occur when a CCV exceeds twice its tolerance for uncertainty.

9-2.4.1. Inorganic Methods.

a. If the method involves significant sample preparation and the CCVs are mid-level standard solutions that are essentially instrument QC samples that are directly analyzed (e.g., CCVs for metals by GF-AA or ICP), then the recovery should be within 90–110%. If the calibration is verified using a CCV set at the low-level calibration standard, then the recovery should fall within 85–115%.

b. Wider acceptance ranges should be used when the CCV is processed in the same manner as the environmental samples (i.e., when the CCV is also an LCS) or when the CCV undergoes a significant preparatory process. For example, CCVs are typically LCSs for the Hg CV-AA analyses. When the CCV is processed in the same manner as the environmental samples, then the CCV should be evaluated using the LCS acceptance limits; an acceptance range of 80–120% should be used. An acceptance range of 85–15% is recommended when the CCV is not processed in an identical manner as the samples but nevertheless undergoes a significant preparatory process (e.g., cyanide CCVs that are distilled but that are not extracted with the environmental samples).

9-2.4.2. Organic Chromatographic Methods.

a. If the calibration is verified with a mid-level CCV, then the recovery should be within 85–115% for analyses of extractable organics (e.g., pesticides and Aroclors). If calibration is verified using a CCV set at the low-level calibration standard, then the recovery should be 80–120%. For purge-and-trap methods (where the environmental samples and CCVs are prepared and analyzed in the same manner), CCVs should be within 20% of their expected values. Wider acceptance ranges may be appropriate for other organic methods where the CCVs are processed in the same manner as the environmental samples.

b. Depending on the analytical method and the level of detail required for the evaluation, additional acceptance criteria may be applicable for chromatographic methods. In particular, for methods that require minimum response factors, the method-specified minimum response factor criteria must be met. Methods such as 8260B and 8270C specify acceptance limits for the responses and retention times of the internal standards in the CCVs. The evaluation of internal standards is discussed in Chapter 16. CCVs are often evaluated to determine if analyte identification criteria are being met. In particular, for chromatographic methods involving the use of two-dimensional detectors (e.g., FIDs and PIDs), CCV retention times are typically assessed to verify that they fall within established retention time windows.

9-3. Evaluation.

a. *Review the instrument run logs to verify that the CCVs were analyzed at an appropriate frequency.* Review the standard preparation log and note whether or not the CCVs and initial calibration standards were prepared from the same source.

b. Use a continuing calibration summary form (and any instrument printouts of quantitation reports) to recalculate a CCV recovery. For chromatographic methods where the initial calibration is performed using mean response factors and **percent differences** are calculated for response factors, calculate the percent difference for at least one response factor. Compare the calculated values with the reported values. The former must agree with the latter to within at least *two* significant figures.

c. "For each CCV, review the CCV summary form to verify that the reported percent recovery or percent difference for each target analyte is acceptable. For chromatographic methods for which minimum response factors are specified, note any response factor that is not compliant with method requirements.

9-4. Qualification.

9-4.1. Representativeness.

a. Qualify the associated sample results if the CCVs were not analyzed in a representative manner. In particular, the number of replicate analyses and system "clean-out" activities must not be applied to CCVs to a greater extent than to the environmental samples in the analytical sequence.

b. If the *run sequence log* indicates that multiple CCBs (continuing calibration blanks) are analyzed before the CCVs but not before any of the environmental samples, then the CCVs may not be representative. If the replicate CCB analyses were being performed to address "carry over", then qualify the associated sample results as estimated or rejected depending upon the severity of the blank contamination and the intended use of the data. Ideally, the laboratory should be required to provide the entire raw data package and the CCB with the highest level of "carry over" (typically the first CCB in the run sequence) should be used to qualify the associated sample results for blank contamination using the strategies in Chapter 10. However, when the CCB results are not available, at a minimum, qualify all detections in the associated environmental samples as estimated (with the J+ flag).

c. If multiple CCVs are being analyzed, the representativeness of the CCV results must be critically evaluated. For example, assume that the following run sequence is observed for a set of aqueous VOC analyses:

> CCV-01, CCV-02, CCV-03, CCV-04, Sample-01, Sample-02, MB, CCV-05, CCV-06, CCV-07 ...

d. Assume that CCV-04 and CCV-07 are acceptable (i.e., the CCV recoveries fall within the acceptance range), but the remaining CCVs are unacceptable. Although two acceptable CCVs bracket the samples, the run sequence suggests that the CCVs are not being analyzed in an appropriate (i.e., representative) manner. When analytical problems exist (especially when a method is only marginally out-of-control), if a sufficient number of QC samples (such as CCVs) are analyzed, then one of the QC samples will eventually fall within the acceptance limits by chance (i.e., because of random error)! Method performance appears to be acceptable but is actually substandard (most of the CCVs are not falling within the acceptance limits). Under these circumstances, qualify the associated sample results (e.g., Sample-01 and Sample-02) using the most noncompliant CCV recovery. If this information is not available (e.g., the recoveries for only CCV-04 and CCV-07 are reported), then qualify all the associated sample results for marginal CCV failure (refer to Chapter 9-4.4). If the data are being used to support critical decisions, it may be appropriate to qualify the sample results as tentatively unusable (using the X flag).

c. CCVs are occasionally used to "update" the instrument's calibration data (e.g., "resloping" for GF-AA analyses). This is not the objective of a CCV. A CCV is performed to verify (to within some tolerance for uncertainty) that the initial calibration remains valid and is not performed to alter the initial calibration curve. "Updating" the calibration using the CCV primarily amounts to replacing the original multiple-point calibration with a single-point calibration. When this occurs, recalculate the associated results using the original calibration curve or, at a minimum, qualify the results as estimated. However, professional judgement should be used. For example, when there is significant instrumental drift and a calibration line is updated using the CCV, results calculated from the CCV (particularly mid-range detections) may be more accurate than those calculated from the multiple-point calibration!

9-4.2. Frequency.

a. If a CCV is missing at the end of the analytical sequence, then, at a minimum, qualify all detections with the J flag and all nondetections with the UN flag *unless it can be otherwise demonstrated that the instrument remained in calibration for the entire analytical sequence.* For example, the laboratory may have analyzed extremely "dirty" environmental samples near the end of the run sequence and cleaned the instrument to eliminate "carry over" problems only for the next 12-hour CCV. Qualification of the associated sample results with the X flag may be more appropriate for some data uses (e.g., when the data is being used to support critical decisions).

b. If all samples are bracketed by two acceptable CCVs but the CCVs are not analyzed at the appropriate frequency (e.g., after every 10 to 20 samples), use professional judgement to determine whether data qualification is necessary. For significant nonconformances, qualify detections with a J flag and nondetections with the UN flag.

9-4.3. Tolerance for Uncertainty.

a. In general, if a CCV in an analytical sequence is not acceptable, then qualification is required for all samples following the last acceptable CCV and all samples preceding the next acceptable CCV. For example, consider the following run sequence:

> CCV-01, Sample-01, Sample-02, CCV-02, Sample-03, Sample-04, CCV-03, Sample-05, Sample-06, CCV-04 . . .

b. "Sample-01" to "Sample-04" would be qualified if CCV-02 were unacceptable. Qualification protocols for CCV failures are very similar to those for LCS failures. Marginal CCV failures are distinguished from gross failures as discussed below.

9-4.3.1. Inorganic Methods, CCVs Not Processed with Samples.

If the CCV does not undergo a significant preparatory process relative to the environmental samples, then evaluate the CCV results as follows: If the CCV recovery is unacceptable but falls within 80% to 120%, then qualify the data (i.e., the associated sample results) for *marginal* failure. If the CCV recovery is unacceptable and does not fall within 80% to 120%, then qualify the data for *gross* failure.

9-4.3.2. Inorganic Methods, CCVs Processed with Samples.

If the CCV is processed in the same manner as the environmental samples, then the CCV is essentially an LCS and must be evaluated using the LCS limits. The results should be qualified for *marginal* failure if the CCV is unacceptable but falls within 60% - 140% of the expected value. If the CCV undergoes a significant sample preparatory process but is not processed in an identical manner as the environmental samples, then it is recommended that results be qualified for marginal failure if the CCV is unacceptable but falls within 70% - 130% (e.g., cyanide CCVs that are distilled but not extracted with the environmental samples).

9-4.3.3. Organic Methods, CCVs Not Processed with Samples.

The following guidance applies to methods that require significant sample preparation (e.g., solvent extractions or cleanup procedures) and the CCV is not processed with the environmental samples. If the CCV is unacceptable but the percent recovery falls within 70% to 130% or the percent difference for the response factor is not greater than 30%, then qualify the data for *marginal* failure. If the CCV is unacceptable and the percent recovery does not fall within 70% to 130% or the difference for the response factor is greater than 30%, qualify the associated sample results for *gross* failure.

9.4.3.4. Organic Methods, CCVs Processed with Samples

If the method does not require significant sample preparation or the CCV is processed with the samples (e.g., aqueous purge-and-trap analyses), the CCV is unacceptable but the percent recovery falls within 40% to 160% or the percent difference for the response factor is not greater than 60%, then qualify the data for *marginal* failure. If the CCV is unacceptable and the percent recovery does not fall within 40% to 160% or the difference for the response factor is greater than 60%, then qualify the associated sample results for *gross* failure.

9.4.4. General Qualification Strategies.

a. Environmental sample results are qualified for CCV failure, based upon the (i) direction of bias, (ii) the magnitude of the failure, and (iii) the concentration of the target analyte relative to the AL. The direction of bias for a CCV failure is *well defined* when all other associated QC samples (e.g., ICVs and LCSs) are in control or exhibit bias in the same direction, i.e., if the CCV recovery is unacceptably high but the LCS recovery is unacceptably low, then the direction of bias is not well defined. Similarly, if the ICV is unacceptable or if a second source standard was used to prepare the CCV and the CCV is unacceptable, then the direction of bias cannot be inferred from the CCV recovery. Qualification strategies for CCV failures follow.

(1) If the CCV is *marginally* unacceptable and the direction of bias is *well defined*, then the data is qualified as follows: For *low* bias, qualify detections with the J- flag and nondetections with the UN flag. For *high* bias, qualify detections with the J+ flag and nondetections with the U flag.

(2) If the CCV is *marginally* unacceptable and the direction of bias is *not* well defined, then qualify detections with the J flag and nondetections with the UN flag.

(3) If the CCV is *grossly* unacceptable and the direction of bias is *well defined*, then qualify the associated sample results as follows:

(*a*) For *low* bias, qualify all nondetections with the R flag. When an AL is *not* specified, qualify detections with the J- flag. If an AL is specified, then qualify detections less than the AL with the X flag and qualify detections greater than the AL with the J- flag.

(*b*) For *high* bias, qualify all nondetections with the U flag. Qualify detections with the J+ flag. However, when an AL is specified, it may be appropriate to qualify detections greater than the AL with the X flag. Alternatively, it may be desirable to obtain additional information from the laboratory before completing the evaluation. For example, additional data could be requested to determine if the high CCV recovery resulted from "carry over" or improper integrations

(4) If the CCV is *grossly* unacceptable and the direction of bias is *not* well defined, then qualify nondetections with the R flag. When an AL is *not* specified, qualify detections with the J flag. If an AL is specified, qualify detections with the X flag. (However, if possible and practical, the magnitude of the uncertainty relative to the proximity of the detection to the AL should be taken into account.)

b. The qualification strategies discussed above are illustrated in Table 9-1 (where it is assumed that each CCV must be within 10% of its expected values). *However, CCV failures must be interpreted in the context of other instrumental and batch QC results using professional judgement.* In particular, a result may still be acceptable when an associated CCV does not fall within the CCV acceptance limits because the uncertainty tolerance for instrumental performance is typically more stringent than that for overall method performance. For example, if the CCV recovery must be within 10% of its expected value and the LCS must be within 20% of its expected value, but the CCV recovery is 85% and the LCS recovery is 80%, then overall accuracy of the associated sample results is still acceptable. In general, if the direction of bias is well defined and the LCS is in control, sample qualification is not required when the CCV recovery is marginally unacceptable. (However, under these circumstances contractual corrective action may be appropriate.)

Table 9-1
Data Qualification for CCV Results[1]

%R for CCV Bias	Remarks	Sample (*y*)	Flag
90% ≤ %R ≤ 110%	Acceptable %R	MRL < MQL < *y*	None
		MRL < *y* < MQL	J
		y < MRL	U
110% ≤ %R ≤ 120% or 80% ≤ %R ≤ 90% Undefined Bias	Marginal Failure	*y* > MRL	J
		y < MRL	UN
80% ≤ %R ≤ 90% Low Bias	Marginal Failure	*y* > MRL	J-
		y < MRL	UN
110% ≤ %R ≤ 120% High Bias	Marginal Failure	*y* > MRL	J+
		y < MRL	U
%R < 80% Low Bias	Gross Failure	*y* > MRL	X if y < AL J- otherwise
		y < MRL	R
%R > 120% High Bias	Gross Failure	*y* > MRL	J+ Possibly, X if *y* > AL
		y < MRL	U
%R > 120 or %R < 80% Undefined Bias	Gross Failure	*y* > MRL	J if AL not specified; X if AL specified
		y < MRL	R

Notes: 1. %R, MRL, MQL, AL, and *y* denote the percent recovery of the target analyte in the CCV, method reporting limit, method quantitation limit, AL, and concentration of the target analyte in an associated field sample, respectively. It is assumed that the MRL is greater than the MDL, less than the MQL, and less than the AL.

CHAPTER 10

Blanks

10-1. Introduction.

Blanks are assessed to determine the existence and magnitude of contamination problems and measure of the *representativeness* of the analytical process. Blanks reflect the amount of contamination introduced into the environmental samples during sample collection, transfer or analysis. In particular, **method blanks** reflect laboratory contamination from both the determinative and preparatory method. Field blanks (e.g., trip blanks and equipment or rinsate blanks) account for accumulative field and laboratory activities. In general, the samples associated with each blank (e.g., method and field blanks) must not be corrected for blank contamination (e.g., unless QAPP or the method of analysis describes a valid procedure for correcting for blank contamination).

> Note: Although blank contamination imparts a high bias to analytical results, blanks are not viewed to be a *measure* of bias or "positive" interference because a one-to-one correspondence between blank contamination and bias does not exits (e.g., high LCS recoveries can be obtained when contamination is not detected in any blanks). Blank contamination is indicative of an effect that is external to the native sample matrix and relates the "representativeness" of the sample.

10-2. Criteria.

10-2.1. Frequency.

a. At least one method blank must be reported for each preparation batch of samples.

> Note: Method blanks associated with a set of environmental samples must be analyzed with the environmental samples using the same instrument in the sample analytical run sequence. For example, if a batch of 20 samples is prepared with a method blank, some of the environmental samples are analyzed with the method blank on "day one," and the remaining environmental samples are analyzed on "day two," then the same method blank analyzed on "day one" should be analyzed on the second day of analysis. At a minimum, an instrument blank must be analyzed with the remaining environmental samples on "day two."

b. Trip blanks must be reported for each cooler containing VOC samples. Additional field blanks may be required for certain projects. The frequency of collection and types of field blanks must be evaluated against project-specific requirements.

10-2.2. Acceptance Limits.

The concentration of each target analyte in each blank must be less than the greater of the following: (i) the RDL for the target analyte; (ii) the MRL when the MRL is not greater than 5% of

the AL, (iii) 5 to 10% (depending on project DQOs) of the analyte concentration detected in each associated field samples; and (iv) 5 to 10% (depending on project DQOs) of the AL. *Environmental sample detections greater than the MRL but less than 10 times the corresponding blank detections must be qualified.* In instances where more than one blank is associated with a given sample (e.g., a rinsate blank and method blank), evaluate blank contamination using the associated blank containing the *highest* contaminant concentration

> Note: Laboratories commonly set the method blank acceptance criteria at the method reporting limit (MRL), which in turn is set equal to the method quantitation limit (MQL). *This is not appropriate when action level is near the MRL/MQL!* When blank acceptance criteria are established based upon the MRLs, blank contamination between the RDLs and MRLs must be reported when the MRL is greater than 5% of the AL.

10-3. Evaluation.

a. Review the Case Narrative and note any problems with method blank contamination. Review the summary forms for method blanks and any field blanks (e.g., trip blanks and rinsate blanks). Significant contamination in a blank may be an isolated occurrence. However, if the reviewer cannot reasonably demonstrate that a contamination problem is an isolated occurrence, a conservative approach must be used. Qualify the environmental sample results using the *highest* analyte concentration detected in the associated blanks (e.g., the method, field, and instrument blanks).

b. Although data qualification strategies for blank contamination are presented in Chapter 10.5, professional judgement is also required. Factors such as the magnitude and frequency of the blank contamination, the nature of the site contamination, the nature of the analysis, and historic data regarding the presence of blank contaminants should also be taken into account. For example, assume that methylene chloride has not been detected during prior sampling efforts (e.g., long-term groundwater monitoring) and methylene chloride has been historically detected in a sporadic manner in associated blanks at low-levels. Furthermore, assume that two batches of groundwater samples are reported for the most current sampling event, “Batch 1" and “Batch 2." Methylene chloride is detected at low levels in the environmental samples in “Batch 1" and “Batch 2,” but methylene chloride is detected only in the method blank for “Batch 1.” It would be reasonable to qualify the low-level methylene chloride detections for the samples of “Batch 2" on the basis of the method blank associated with “Batch 1," even though all the blanks associated with “Batch 2" are “clean.”

10-4. Contractual Considerations.

a. Since laboratories are normally required to reprocess (e.g., reextract and reanalyze) a batch of samples when the method blank is unacceptable, contractual corrective action for unsatisfactorily performance may be warranted when high levels of contamination are systematically observed in the method blanks or when a method blank is not processed. Similarly, contractual corrective action may be appropriate for unacceptable field blanks (e.g., rinsate and field blanks).

b. When high blank contamination is observed, the reviewer should consult with the Project Manager to determine whether the data package must be reviewed or rejected. For example, the laboratory may be required to reanalyze the environmental samples. Alternatively, it may be possible to adopt higher reporting limits (e.g., when the higher reporting limits are still much lower than the project's decision limits).

> Note: Meeting the method blank acceptance criteria on a routine basis may not be practical for common laboratory contaminants (e.g., methylene chloride, phthalates, and acetone); sporadic detections of contamination may occur and are difficult to control. Exercise professional judgment when evaluating contractual compliance for common laboratory contaminants.

10-5. Qualification for Blank Contamination.

a. When a target analyte is detected above the RDL in any blank, qualification for the associated environmental samples for blank contamination is *not* required when any of the following occur:

(1) The target analyte is not detected in the environmental samples.

(2) The target analyte is detected in the blank at a concentration less than 5% to 10% of the corresponding environmental sample concentration.

(3) The target analyte is detected in the blank at a concentration greater than the RDL and less than the MRL, where the MRL is less than 5% of the AL.

b. In general, qualification is required when a target analyte is detected in a blank at a concentration *greater than 5 or 10%* of the corresponding environmental sample concentration (e.g., even when the analyte is detected at *less than 5% of the AL*). Qualification for blank contamination is illustrated in Table 10-Samples are qualified for blank contamination using the following strategies:

(1) *J+ flag.* If the analyte concentration for an environmental sample is greater than five but less than ten to twenty times higher than the analyte concentration in the corresponding blank, qualify the reported sample result with a J+ flag. Under these circumstances, the J+ flag indicates that the analyte is present in the sample but the reported concentration of the analyte believed to be biased high because of blank contamination. When the analyte concentration for an environmental sample is *less than five times the analyte concentration in an associated blank*, data qualification will be highly dependent upon project-specific DQOs. In particular, qualification will be dependent upon whether or not action levels are available. Sample results are qualified with the U, UN, X, or N flag as discussed below.

(2) *UN flag.* If the analyte concentration for the environmental sample is less than five times the analyte concentration in the corresponding blank, then qualify the sample result with the *UN flag* if (i) an AL is *not* available or (ii) the sample result is *less* than the AL. The UN flag indicates that the analyte was not reliably detected because of blank contamination and the re-

ported result is viewed as a tentative nondetection at the reported concentration. Alternatively, multiply the blank by a factor of five and report (in place of the sample result) the resulting product with a U flag when (i) the product is significantly less than the AL (e.g., 5% or 10% of the AL) or (ii) an AL is not available.

(3) *X flag.* If the analyte concentration for the environmental sample is less than five times the analyte concentration in the corresponding blank but is *greater* than the AL, then qualify the sample result with the *X flag*. Under these circumstances, the X flag indicates that the analyte was not reliably detected (above the AL) because of blank contamination and should be rejected. In effect, blank contamination has increased the reporting limit for the analyte to a concentration that is greater than the AL. A nondetection reported at the elevated limit does not demonstrate the target analyte is present in the environmental sample above or below the AL. *The sample result must not be qualified with the UN (or U flag) unless a defensible technical rationale for the use of the flag UN is presented*.

(4) *N flag.* If the analyte concentration for the environmental sample is less than five times the analyte concentration in the corresponding blank and the analyte concentration is greater than the AL, then qualify the sample result with an *N flag* only when it can be demonstrated that the UN flag or X flag is *not* appropriate. For example, the N flag may be appropriate when it is desirable to establish an upper limit for site-related contamination. When used in this manner, the N flag indicates that a target analyte result is being reported as a detection but the detection may not be reliable because of contamination problems.

Table 10-1
Data Qualification for Blank Contamination [1]

Blank (BLK) (ppb)	Reported Result (*y*) (ppb)	Qualified Result (ppb)	Remarks AL = 100 ppb MRL = 1 ppb MQL = 5 ppb	Blank acceptance criteria in Chapter 10.2.2 met?
1 U	6	6	No significant contamination detected.	Yes BLK < RL < 5% of AL
23	< 1	1 U	Contamination detected, but no action required.	No [3] BLK > 5% of AL
2 J	4	4 UN or 10 U	*y* < 5 BLK and *y* < AL [2]	Yes BLK < 5% of AL
2 J	11	11 J+	5 BLK < *y* < 20 BLK	Yes BLK < 5% of AL
2 J	60	60	*y* > 20 BLK	Yes BLK < 5% of AL
80	150	150 X	*y* < 5 BLK and *y* > AL	No [4] BLK > 5% of y

Notes: 1. The concentration of analyte detected in the field sample and blank are denoted by *y* and BLK, respectively. For the purposes of illustration, it is assumed that MRL = 1 ppb, MQL = 5 ppb, and AL = 100 ppb, where RDL ≤ MRL. Note that the MRL is less than 5% of the AL. 2. The same flags would be applied to the sample result if an AL were not available. 3. Although acceptance criteria for blank contamination in Chapter 10.2.2 were not met, the result is still usable. The laboratory would not typically be required to reprocess the sample for method blank contamination but should be expected to investigate the source of the contamination. 4. The laboratory would typically be required to reprocess the sample for method blank contamination

CHAPTER 11

Laboratory Control Samples (LCSs)

11-1. Introduction.

Laboratory control samples are evaluated to assess overall method performance and are the primary indicators of laboratory performance. In general, laboratory control samples are similar in composition as the environmental samples, contain known concentrations of *all* the analytes of interest, and undergo the same preparatory and determinative procedures as the environmental samples. LCS recoveries are used to measure accuracy. The relative percent difference (RPD) for duplicate LCS recoveries is normally used as a measure of precision. When both a laboratory control sample (LCS) and laboratory control sample duplicate (LCSD) are processed for a batch of samples, there is no significant physical distinction between the LCS and LCSD. *Both the LCS and LCSD must satisfy the same recovery acceptance criteria.* Therefore, for simplicity, the term *LCS* will refer to one or more laboratory control samples (e.g., the term "LCS acceptance criteria" will refer to the acceptance criteria for both the LCS and LCSD).

11-2. Criteria.

11-2.1. Frequency.

At least one LCS must be reported with each batch of samples. A laboratory control sample and a laboratory control sample duplicate (LCSD) may be analyzed to provide information on the precision of the analytical method. The generation of control chart limits for precision via the analysis of LCS/LCSD pairs is an effective means to measure method precision. Multiple LCSs may be required to evaluate method precision and accuracy at different spiking concentrations.

11-2.2. Acceptance Limits.

a. Project documents such as the QAPP should specify the acceptance limits for LCS recoveries. To the extent possible, LCS acceptance limits should be established based upon project DQOs rather than upon contractual specifications, the limitations of the laboratory, or the limitations of the analytical method. Laboratory statistical control limits should be evaluated during the planning stages of the DQO process to assure that project-required acceptance limits will be met.

b. Laboratory statistical control limits must not be the sole basis upon which project-required acceptance limits are established. Statistical control limits generated by the laboratory may be representative of routine method performance but may be too wide to satisfy project-specific DQOs. Furthermore, statistical control limits for laboratory control samples tend to adversely impact laboratory-to-laboratory comparisons (e.g., when USACE QA split sample analyses are being performed, an LCS recovery that falls within the wide acceptance range of one laboratory will not necessarily fall within the tighter acceptance range of the referee laboratory or vice versa).

c. Acceptance limits for bias and precision are presented in various analytical methods (e.g., SW-846 and CLP methods) but *many of these limits may be inappropriately wide.* Acceptance limits for accuracy and precision are presented in the USACE Shell. Although these limits were established to ensure a moderate to high level of data quality, they are ultimately contractual in nature (e.g., permit poor performance for select target analytes because of inherent limitations of the analytical methodology). It may not be practical or possible (even after method modification and development) for a method to routinely meet the acceptance limits for every target analyte. Under these circumstances, the reviewer must distinguish contractual compliance and laboratory performance issues from data usability issues.

d. Inappropriately wide LCS acceptance ranges may be specified for a method in project-documents such QAPPs, SAPs, and Work Plans. These acceptance ranges are often based upon contractual, method-specified, or laboratory control chart limits. For example, erroneously wide LCS acceptance ranges may be specified when ALs are equal to or near the MQLs. *The specification of an acceptance limit in a project document per se does not imply that limit is scientifically sound with respect to project objectives. When, in the reviewer's professional judgment, project-specified LCS acceptance limits are not consistent with project DQOs, evaluate the data package with respect to scientifically defensible limits.*

e. In the absence of reasonable LCS recovery limits, the following limits are recommended: The recovery for each target analyte should fall within 80 to 120% for *inorganic* analyses and within 60 to 140% for *organic* analyses. For *purge-and-trap* GC and GC/MS analyses, recoveries should fall within 80% to 120% when the CCV is being used as the LCS. If the LCS is an independent source standard, the LCS should fall with 70 to 130% for purge-and-trap analyses.

f. In the absence of project-specific limits for precision, it is recommended that the acceptance limit for the RPD be equal to one half of the width of the corresponding LCS recovery acceptance range or to the laboratory's RPD acceptance limit, whichever is less. For example, the laboratory may have established statistical RPD acceptance limits by processing an LCS/LCSD pair for each batch or from interbatch LCSs (i.e., LCSs from consecutive batches).

11-3. Evaluation.

Evaluate the LCS results using the following strategies:

a. Using the standard preparation logs verify that *all* target analytes were spiked into the LCS and note whether or not an independent-source standard was used to prepare the LCS.

> Note: A number of published analytical methods do not require the LCS to contain all the target analytes. Unless a scientifically defensible rationale for not spiking all the target analytes is presented in the analytical method or in project documents such as the QAPP, assume that all "single-component" target analytes must be spiked into the LCS. However, when several multi component target analytes are being simultaneously analyzed (e.g., the set of Aroclors in Method 8082), it may not be possible (or desirable) to spike all the analytes into a single LCS. Depending on the nature of the analysis and

the data quality objectives for the project, a set of laboratory control samples (e.g., one LCS for each multi component target analyte) may be required or only a single LCS containing "representative" components may be appropriate (e.g., an LCS containing Aroclors 1016 and 1260 is typically assumed to be representative of the other Aroclors analyzed by Method 8082).

b. Using the sample preparation log and the instrument run log verify that the LCS was processed with the samples through the entire analytical method.

c. Using the LCS summary form, calculate the LCS recovery for at least one target analyte and compare the calculated value to the reported value. Similarly, recalculate the RPD for an LCS/LCSD pair for one target analyte and compare the calculated value to the reported value. The calculated LCS recoveries and RPDs must agree with the reported values to within two significant figures.

d. For each target analyte, compare the LCS recoveries and RPDs reported on the laboratory's summary forms to the corresponding LCS acceptance limits for bias and precision. In the absence of appropriate acceptance limits, establish a set of limits to properly evaluate the LCS results. A batch of samples is acceptable only for those target analytes that satisfy the LCS criteria for bias and precision. All failures must be noted. Data qualification is required when the LCS acceptance criteria are not met.

e. Review the Case Narrative and note any problems discussed for the LCS. When an LCS recovery is unacceptable, examine the Case Narrative and note why the batch was not reprocessed (e.g., reextracted and reanalyzed) for the failed analyte. However, it should be noted that even when method implementation is optimal, a small percentage of sporadic failures should be expected for the LCS (especially when a large number of target analytes are being simultaneously analyzed).

11-4. Contractual Considerations.

a. Contractual considerations may impact the data review. Since laboratories are normally required to reprocess (e.g., reextract and reanalyze) a batch of samples when the LCS is unacceptable, contractual corrective action for unsatisfactorily performance is typically required for gross systematic LCS failures. When gross systematic failures occur, the reviewer should consult with the Project Manager to determine whether or not to proceed with the review or to reject the data package as a whole (e.g., the laboratory may be required to reanalyze the environmental samples). However, the reviewer should exercise professional judgment when determining whether contractual compliance will impact the data review. In particular, for methods containing large lists of target analytes (e.g., Method 8270C) or "poor performers" (e.g., the ketones of Method 8260B or other analytes which cannot meet QC limits because of inherent method limitations), it is highly probable that the recoveries of several target analytes will be unacceptable.

b. *Sporadic marginal LCS failures should be expected and should not trigger a consultation with the Project Manager or the rejection of a batch of samples.* For example, a "marginal

sporadic failure" may be said to exist if an LCS recovery falls between the three- and four-sigma control limits for no apparent reason for a particular batch of samples but the laboratory control samples for prior and subsequent batches are acceptably recovered. The table below lists the maximum number analytes expected to fall outside of the three-sigma control limits for an LCS when the LCS contains a large set of target analytes..

c. For example, according to Table 11-1, if there are 20 target analytes, as many as two analytes in the LCS may fall outside of the three-sigma acceptance limits because of random error. Typically, these types of sporadic failures should not trigger reanalyses of the batch but the associated environmental sample results should be qualified.

Table 11-1
Number of Target Analytes Versus Number of Expected LCS Failures

n[1]	f[2]
10–15	1
16–45	2
46–85	3
86–130	4

Notes: 1. n = Total number of target analytes being simultaneously analyzed. 2. f = Maximum number of analytes expected to fall outside of the three-sigma control limits with 99% confidence if the probability of a random failure is less than or equal to 1%.

Note: Review project documents (e.g., the Quality Assurance Plan) to ensure that the noncompliant analyte is not a critical analyte (e.g., a human or ecological "risk driver"). For example, if 60 VOCs are being analyzed by 8260B, but vinyl chloride is the primary contaminant of concern, then reanalyses for vinyl chloride should be expected when the LCS recovery is not acceptable.

d. If precision is unacceptable for a particular analyte (e.g., the RPD is higher than the acceptance limit), then the associated field sample detections above the MQL (or the MRL if it is greater than the MQL) must be qualified as estimated data. To satisfy project-specific requirements, the laboratory may be required to reprocess a batch of samples when the LCS does not satisfy precision acceptance criteria. Under these circumstance, verify that this was done. However, it should be noted that laboratories do not typically reprocess environmental samples for unacceptable RPDs when the LCS recoveries are acceptable.

11-5. Qualification.

a. The qualification strategies presented in this section of the document will generally be applicable.

(1) When multiple laboratory control samples (e.g., an LCS and LCSD) are processed for a single batch of samples, and one or more LCS recoveries are unacceptable for a particular target, then the associated samples must be qualified on the basis of the most noncompliant target analyte recovery. However, it should be noted that replicate laboratory control samples may not be required or reported. For example, if the RPD for an LCS/LCSD pair is calculated using interbatch laboratory control data (i.e., the LCSD is not extracted with the LCS but is the control sample for a consecutive batch of samples), the LCSD recovery may not have been reported..

(2) Data qualification must be a function of both the magnitude and direction of the QC failure. *Gross* QC failures must be distinguished from *marginal* failures and the direction of bias must be taken into account. When the LCS recovery is unacceptable, the direction of bias will be said to be *well defined* if the direction of bias for other batch and **instrument QC samples** (e.g., ICVs, surrogates, and replicate LCSs) is consistent with the noncompliant LCS recovery. For example, if both an LCS and LCSD are extracted with a batch of samples and the LCS recovery is less than the lower control limit but the LCSD recovery is greater than the upper control limit, then the direction of bias is not well defined. Similarly, the direction of bias is not well defined when the RPD for an LCS/LCSD pair is used to evaluate duplicate precision and the RPD is unacceptable, but the LCSD recovery is not reported

b. Specific qualification protocols for laboratory control samples are presented below and are illustrated in Table 11-2 (where it is assumed that all QC samples other than the LCSs are in control).

(1) If the LCS recovery is *marginally* unacceptable and the direction of bias is *not* well defined, then qualify detections of the target analyte with the J flag and nondetections with the UN flag.

(2) If an LCS recovery is *marginally* unacceptable and the direction of bias is *well defined*, then qualify the data as follows: For *low* bias, qualify detections with the J- flag and nondetections with the UN flag. For *high* bias, qualify detections with the J+ flag and nondetections with the U flag.

(3) If an LCS recovery is *grossly* unacceptable and the direction of bias is *well defined*, then qualify the associated sample results as follows:

(*a*) For *low* bias, qualify all nondetections with the R flag. If an AL is *not* specified, qualify detections with the J- flag. If an AL is specified, then qualify detections less than the AL with the X flag and detections greater than the AL with the J- flag.

(*b*) For *high* bias, qualify all nondetections with the U flag. Qualify detections with the J+ flag. However, when an AL is specified, it may be appropriate to qualify detections greater than the AL with the X flag (e.g., when a conservative estimate is not being sought).

(4) If the LCS recovery is *grossly* unacceptable and the direction of bias is *not* well defined, then qualify nondetections with the R flag. If an AL is *not* specified, then, at a minimum,

qualify detections with the J flag (the X flag may be more appropriate). If an AL is specified, qualify detections less than the AL with the X flag. Depending on project DQOs, qualify detections greater than the AL with the J or X flag.

c. In addition to the qualification strategies discussed above, use the following protocols when duplicate laboratory control samples are processed with each batch of samples:

(1) If the LCS/LCSD recoveries are acceptable, the RPD is *marginally* unacceptable, and the direction of bias is *not* well defined, then qualify detections with the J flag and nondetections with the UN flag.

(2) If the RPD is *grossly* unacceptable and the direction of bias is *not* well defined, then qualify nondetections with the R flag. (The X flag may be appropriate if additional information to determine the direction of bias will be obtained). Qualify detections with the J flag when an AL is *not* specified. If an AL is specified, then qualify detections less than the AL with the X flag and qualify detections greater than the AL with the J flag or the X flag.

d. In the absence of valid project-specific limits for bias and precision, a *gross* failure is defined to occur when one of the following conditions is satisfied:

(1) For *inorganic* analyses, a gross failure occurs for a target analyte when the percent recovery does not fall within 60 to 140%. For *organic* analyses involving *significant sample preparation* (e.g., solvent extraction), a gross failure occurs when the LCS recovery does not fall within 20 to 180%. However, for *purge-and-trap* analyses, a gross failure occurs when the LCS recovery does not fall within 40 to 160%.

(2) A *gross failure* occurs when the RPD for the LCS/LCSD is greater than 40% for *inorganic* analyses, *60%* for *purge-and-trap* analyses, and *80%* for *extractable organic* analyses.

11-6. Qualification Strategies Using Estimates of the Uncertainty.

a. This section of the document describes some optional data qualification strategies that may be used when analytical uncertainty can be estimated from laboratory control samples. These strategies will be applicable when matrix interference and sample heterogeneity are not significant components of the analytical uncertainty or when it is desirable to establish a *lower bound* for the total uncertainty. Laboratory uncertainty is estimated from the laboratory's in-house statistical warning and control limits for LCS recoveries. If representative matrix spike warning and control limits are available, it is recommended that these limits be used instead of the LCS limits. The use of matrix spike warning and control will result in better estimates of the uncertainty (e.g., since LCS limits do not account for the uncertainty associated with matrix effects). However, it should be noted that representative matrix spike recovery limits are not typically available from environmental production laboratories and must be generated on a project-specific basis. (Refer to Chapter 12 for additional information.)

b. When an analytical result is being compared to a decision limit, it may be useful (e.g., for the purposes of data qualification) to estimate an upper or lower confidence limit for the re-

sult. If there is significant analytical bias (i.e., the percent recovery for the LCS is statistically different from 100%), the result can be corrected for bias prior to estimating confidence limits. Since low bias is more common than high bias for environmental analyses (e.g., for extractable organic compounds) and is more likely to adversely impact data quality than high bias, only low bias will be addressed. Upper confidence limits (UCLs) will be approximated by correcting for low bias and taking random error into account. The upper confidence limits will then be compared to project action levels to qualify results. This strategy will constitute a relatively conservative approach for risk-based applications.

c. If the percent recovery of a target analyte in the associated LCS is not too close to zero (e.g., the percent recovery is least 20–30%), precision is in control, then an upper confidence limit for a laboratory result may be approximated using the following equation[1]:

$$UCL(C,\%R,\alpha) = u(C,\%R,\alpha)\,[\,C / (\%R / 100)\,] \qquad (11\text{-}1)$$

d. The measured concentration of the sample and percent recovery for the associated laboratory control sample are denoted by C and %R, respectively. The second term in Equation 11-1 (enclosed in brackets) is the "biased corrected concentration." The first term, $u(C,\%R,\alpha)$, will be referred to as the "uncertainty factor" because it accounts for the random error associated with the measured result C and the calculated percent recovery %R. The factor is primarily a function of C, %R, and the desired level of statistical confidence, α. The factor will be some positive value greater than one. The use of a high value for the uncertainty factor will result in a conservative estimate for the UCL (e.g., will minimize false negatives when comparing results to an AL).

e. If normality is assumed and the relative uncertainty (i.e., the relative standard deviation) is assumed to be constant within the quantitation range of the method, then the "uncertainty factor" for the 95% UCL may be estimated using the following equation: [2]

$$u(95\%) \approx 1 + (2)^{1/2}\,(L_{95\%} / \%R) \qquad (11\text{-}2)$$

where $L_{95\%}$ is half the width of the warning range for the LCS percent recoveries (e.g., from the laboratory's control charts). The half width of the control range, $L_{99\%}$, gives an upper 99% upper confidence limit.

$$u(99\%) \approx 1 + (2)^{1/2}\,(L_{99\%} / \%R) \qquad (11\text{-}3)$$

Note that the uncertainty increases as the width of the warning or control ranges increases and the percent recovery decreases.

---

[1]For a rigorous treatment of propagation of analytical measurement uncertainty, refer to the following reference: "Draft EURACHEM/CITAC Guide Quantifying Uncertainty in Analytical Measurement," Second Edition, June 1999, EURACHEM Measurement Uncertainty Working Group.

[2]Georgian, T. Estimation of laboratory uncertainty using laboratory control samples. "Environmental Testing and Analysis," Vol. 9, No. 6, p. 20. November/December 2000.

f. The assumption that the relative standard deviation is constant will be valid for sample concentrations sufficiently near the spiking concentration for the LCS (typically the mid-calibration range) and will be appropriate when the standard deviation is approximately a linear (increasing) function of concentration. Uncertainty is often proportional to analyte concentration when the measurements are well above the detection limits. The above equations will probably result in reasonable estimates when there is no appreciable matrix interference or sample heterogeneity, measurements are within the calibration range of the method, and the analyte levels are near the LCS spiking concentration. Note that the variability associated with the heterogeneity of the sample matrix is not taken into account because the total uncertainty is estimated from the LCS, which is typically a "clean" matrix such as reagent water or purified sand.

g. The use of the mean LCS recovery (%R), rather than the use of a single LCS recovery, %R, associated with a batch of samples, will generally result in a more reliable estimate of the UCL. This is especially true when extreme low bias (e.g., %R < 20% or 30%) or high method variability exists. Under these circumstances, bias correction should be performed using the mean percent recovery. If the mean LCS recovery is available (e.g., at least 20 or 30 data points were used to establish the laboratory's in-house statistical warning and control limits) *and the method is in statistical control*, then substitute (%R) for %R in Equation 11-1 and use the following uncertainty factors:

$$u(95\%) \approx 1 + (L_{95\%} / \langle \%R \rangle) \tag{11-4}$$

$$u(99\%) \approx 1 + (L_{99\%} / \langle \%R \rangle) \tag{11-5}$$

h. Note that (when bias correction is performed) the use of the mean recovery decreases the uncertainty (and the UCL) because the mean recovery is a more confident representation of "true" bias than any single recovery value.

i. If there is no significant bias (i.e., %R = 100%), the relative uncertainty is approximately constant within the quantitative range of the method and the associated LCS recovery is in control for the sample batch, then Equation 11-1 and either Equation 11-4 or Equation 11.5 may be used to estimate an upper confidence limit, by setting $\langle \%R \rangle = 100$:

$$UCL(95\%) = u(95\%)\, C \approx (1 + L_{95\%} / 100)\, C \tag{11-6}$$

$$UCL(99\%) = u(99\%)\, C \approx (1 + L_{95\%} / 100)\, C \tag{11-7}$$

j. Note that the total uncertainty is larger when a bias correction is performed. This occurs because Equation 11-1 contains two sources of uncertainty (the uncertainty associated with %R and C) while Equations 11-6 and 11-7 contain only one source of uncertainty (uncertainty associated with C).

k. To illustrate the use of the above equations, assume that %R = 40% and C = 2 ppb. If the LCS warning range is 60–140%, then $L_{95\%} = 40\%$. It follows from Equations 11-1 and 11-2 that the upper confidence limit for the measured result C is:

$$UCL(95\%) = (1 + 1.4)\,[\,2\text{ ppb} / (40\% / 100)\,] \approx 12\text{ ppb}$$

l. If there is no significant method bias and the LCS recovery is in control, then the upper confidence limit can be estimated using Equation 11-1 and Equation 11-6:

$$UCL(95\%) = 1.4\ (2\text{ ppb}) \approx 3\text{ ppb}$$

m. Once an upper confidence limit is calculated, the upper confidence limits can be compared to the project decision limits and this information can be used to qualify the data. To illustrate, let %R = 40%, C = 2, and $L_{95\%}$ = 40% (the first example presented above). Assume that the project-required acceptance range for the LCS is 80–120% and the project action level (AL) is 50 ppb. Since the LCS recovery is 40%, the result C = 2 must be qualified (e.g., as estimated or rejected). Since UCL(95%) = 12 ppb < AL = 50 ppb, despite the low bias, it is not likely that the analyte is actually present in the sample at a concentration that exceeds the AL. Hence, it would be appropriate to qualify the 2-ppb result with the J-, flag. However, if AL = 5 ppb, since the UCL > AL, it may be more appropriate to qualify the result with the X flag (e.g., when statistical analyses are not being performed and each reported sample concentration is being directly compared to the AL). The low-biased result of 2 ppb does not demonstrate that the analyte is present at a level that is less than the 5-ppb action level

n. It should be noted that the uncertainty factor does not typically exhibit a large amount of variability in the context of the tolerances normally applied to laboratory environmental analyses. The uncertainty factor will typically assume values between two to four, and, at worst, will probably be less than ten. For example, if %R = 20% and the LCS control range is 20% to 180%, conditions that are indicative of rather poor method performance for a target analyte, then an uncertainty factor of less than seven would be calculated from Equation 11-3. Therefore, if the LCS recovery is unacceptably low but the recovery is not less than about 20%, then it may be more convenient to calculate an UCL for a measured sample concentration by correcting the measured concentration for bias and then simply multiplying the bias-corrected result by a factor of five or ten. The UCL could then be compared to the AL to qualify a sample result associated with the noncompliant LCS recovery. For example, if the UCL were less than the action, then the result would be qualified as estimated (e.g., using the J- flag). If the UCL were greater than the AL, then the sample result would be qualified potentially rejected (using the X flag).

Table 11-2
Data Qualification for LCS Results [1]

Acceptance Criteria: 80% ≤ %R ≤ 120%, RPD ≤ 20%			
%R [RPD]	**Remarks [Bias]**	**Sample (*y*)**	**Sample Flag**
90% [18%]	%R and RPD in control	MRL < MQL < *y*	Flag not required.
		MRL < *y* < MQL	J
		y < MRL	U
90% [30%]	%R acceptable RPD OFC [Unknown]	*y* > MRL	J
		y < MRL	UN
70% [15%]	%R < LCL [Low]	*y* > MRL	J-
		y < MRL	UN
140% [10%]	%R > UCL [High]	*y* > MRL	J+
		y < MRL	U
10% [15%]	%R << LCL [Low]	*y* > MRL	J- X if *y* < AL
		y < MRL	R
250% [20%]	%R >> UCL [High]	*y* > MRL	J+ Possibly X if *y* > AL
		y < MRL	U
250% [200%] or 10% [200%]	%R >> UCL or %R << LCL RPD grossly OFC [Unknown]	*y* > MRL	J X if *y* < AL Possibly X if *y* > AL
		y < MRL	R

Notes: 1. %R and RPD denote the percent recovery for the LCS and the relative percent difference for the LCS/LCSD, respectively. The concentration of the field sample is denoted by *y* and the action level by AL. (It is assumed that MRL < AL.) The terms "out of control," "upper control limit," and "lower control limit" are abbreviated as OFC, UCL, and LCL, respectively. The inferred direction of bias is enclosed in brackets. The symbols "<<" and ">>" denote "much less than" and "much greater than," respectively.

CHAPTER 12

Matrix Spikes, Matrix Spike Duplicates, and Matrix Duplicates

12-1. Introduction.

a. **Matrix spike (MS), matrix spike duplicate (MSD), and matrix duplicate (MD)** results are examined to evaluate the impact of matrix effects on overall analytical performance and the potential usability of the data. A matrix spike is a *representative* environmental sample that is spiked with target analytes of interest *prior* to being taken through the entire analytical process in order to evaluate analytical bias for an actual matrix. A matrix duplicate is a collocated (e.g., a VOC soil sample) or a homogenized sample that is processed through entire analytical procedure in order to evaluate overall precision for an actual matrix. Duplicate or replicate matrix spikes are also used to evaluate overall precision.

b. Matrix spike recovery failure and poor precision may arise because of (i) poor sampling technique, (ii) inadequate homogenization, or (iii) from matrix effects associated with the preparatory or determinative portion of an analytical method. For example, inappropriate sample collection and handling procedures for VOC soil samples (e.g., as described in Method 5030) may result in variable losses of VOCs, giving rise to poor precision and low bias. Sludges, clayey soils or sediments, multi phasic samples, and samples with macroscopic particles of analytes such as explosives and metals, may defy homogenization attempts during sample preparation or compositing procedures used for sample collection, giving to unacceptable duplicate precision or matrix spike recoveries.

> Note: In this document, sample heterogeneity arising from the spatial or temporal distribution of the analytes in a study area is viewed as a characteristic of the environmental population being sampled and not as an "interference" that the method of analyses must be optimized to address.

12-2. Interpretation of Matrix Spike and Duplicate Results.

a. In general, when evaluating accuracy using matrix spike recoveries, a matrix effect is inferred when (i) all instrument and method QC samples (the LCSs and CCVs) are acceptable, (ii) the spiking concentration for the matrix spike is high relative to the **native analyte** concentration, and (iii) the recovery of the matrix spike does not fall within the laboratory's corresponding statistical range for *laboratory control samples*. Similar reasoning applies to the evaluation of precision using RPDs for MS/MSDs and MDs results. Namely, an interference is inferred when (i) instrument and method QC is in control, (ii) the native analyte concentrations are sufficiently high (e.g., above the quantitation limits), and (iii) some measure of precision (such as the RPD) exceeds the corresponding statistical LCS limits.

b. Laboratory and project documents (e.g., laboratory standard operating procedures and QAPPs) often state that the presence or absence of matrix effects is determined by establishing statistical control ranges using *MS* rather than *LCS spike* recovery data. Once the MS control limits are established, a matrix effect is subsequently inferred for a batch of environmental sam-

ples if an associated matrix spike recovery falls outside of the statistical MS control range (rather than outside of the LCS control range). *This approach will typically be inappropriate!* In order for this strategy to be viable, the matrix used to establish the MS control range must be relatively uniform, similar in composition to the environmental matrix of interest, and known to lack significant interferences.

c. Because of the variety and complexity of environmental matrices, it is usually impractical for environmental production laboratories to establish matrix-specific control limits. Most (if not virtually all) environmental laboratories that maintain statistical MS control ranges, establish MS limits by method rather than by matrix. For example, groundwater, surface water, rain water, and waste water are often erroneously considered to be the sample "matrix" for the purpose of calculating statistical MS control limits because the samples are processed using the same aqueous preparatory and determinative methods. Furthermore, the MS control ranges are frequently calculated using MS recoveries that have been impacted by matrix effects. *These problems frequently result in very wide MS control limits that are difficult to interpret and frequently do not satisfy project objectives.* Furthermore, since the MS control ranges are often calculated using spiked samples affected by significant matrix inferences, the absence of a matrix effect is not demonstrated when a MS recovery for a batch of environmental samples falls within the MS recovery range. At best, the result may demonstrate that a matrix effect (if present) is *no larger than is typically observed* for a variety of matrices analyzed by the same preparatory and determinative method.

d. In general, matrix spike control limits are not available from environmental production laboratories as "off-the-shelf" commodities but must be established on a project-specific basis. In order to obtain representative matrix spike control limits, a relatively large number of matrix spike samples (e.g., 20 to 30 samples) must be taken from each environmental medium in each project study area. *When a project's matrix spike acceptance ranges are established solely upon the basis of a laboratory's statistical MS control limits and these limits were developed using MS recoveries from non-project related media or dissimilar matrices that have been impacted by interferences, then the matrix spike control limits will probably be inappropriate.* Before proceeding with the data evaluation, assess the validity of the matrix spike acceptance limits (e.g., determine whether the acceptance ranges are unrepresentative or too wide to satisfy project's data objectives). A strategy for approximating statistical matrix spike control ranges using LCS recovery data is presented in Paragraph 12-3

e. Lastly, it should be noted that matrix spike recoveries are evaluated, at least potentially, to fulfill two separate objectives: *(i) To determine whether or not matrix effects exist and (ii) to determine whether or not project-specific objectives for accuracy were satisfied for the analytes in the matrices of interest.* The distinction between the two objectives is somewhat subtle but important to recognize when qualifying data because data are frequently qualified (e.g., as estimated) on the basis of the second objective rather than the first.

f. To illustrate the evaluation of matrix spike and LCS results, assume that a laboratory's statistical control range for LCS recoveries for aqueous lead analyses is 80–120%, the project-required acceptance range for MS recoveries is 50–150%, and three separate sets (batches) of samples were analyzed with associated MS recoveries of 90, 65, and 40%. Assume that the

spiking concentrations for all three MS samples are high relative to the native analyte concentrations and QC is otherwise acceptable. Since the 90% MS recovery lies within the statistical LCS acceptance limits, this recovery suggests the absence of any matrix effects. Since the MS recovery of 65% falls well outside of the LCS statistical acceptance range, the recovery is indicative of a matrix effect that is within the project-required tolerance for accuracy (50–150%). Although the recovery is indicative of matrix interference, data qualification would not necessarily be required. The recovery of 40% is indicative of a matrix effect that is greater than the project-required tolerance for matrix effects. At a minimum, data qualification would typically be required.

12-3. Estimating Statistical Matrix Spike Recovery Ranges.

a. If the spiking concentration for the MS is at least twice as large as the native analyte concentration, the laboratory's in-house statistical control or warning limits for LCS recoveries can be used to establish acceptance limits for MS recoveries:

$$\langle \%R \rangle \pm L_{95\%} \, (100 / \langle \%R \rangle) \, (\langle \%R \rangle / 100 + C_B / C_S) \tag{12-1}$$

$$\langle \%R \rangle \pm L_{99\%} \, (100 / \langle \%R \rangle) \, (\langle \%R \rangle / 100 + C_B / C_S) \tag{12-2}$$

b. As defined in Chapter 11-6, [%R] is the mean LCS recovery, $L_{95\%}$ is the half width of the LCS warning range and $L_{99\%}$ is half the width of the control range. The variable C_B denotes the native analyte concentration (i.e., the measured pre-spike sample concentration) and C_S denotes the calculated spike concentration in the sample matrix (i.e., the analyte concentration added to the sample matrix). If method bias is not significant (i.e., [%R] is near 100%), then the following equations may be used to estimate the MS acceptance ranges:

$$\langle \%R \rangle \pm L_{95\%} \, (1 + C_B / C_S) \tag{12-3}$$

$$\langle \%R \rangle \pm L_{99\%} \, (1 + C_B / C_S) \tag{12-4}$$

c. For example, if the LCS acceptance range is 80–120% (i.e., 100% ± 20%) and the spike concentration is twice the native analyte concentration, then the acceptance range for the MS recovery is as follows:

$$100 \pm 20\% \, (1 + ½) = 100 \pm 30\% = 70–130\%$$

d. Therefore (in this example), if the LCS recovery for a batch of environmental samples falls within 80–120% but the recovery of the associated matrix spike does not fall within 70–130%, then a matrix effect would be demonstrated.

e. The acceptance range for MS recoveries may be set equal to the acceptance range for LCS recoveries when the MS spike concentration is much higher than the native analyte concentration (e.g., by a factor of five to ten) or when it is desirable to establish a conservative (i.e., a more narrow) MS acceptance range.

Note: Since two measurements are required to calculate a MS recovery (the "pre-spike" and "post-spike" sample concentrations) but only one measurement is required to calculate the LCS recovery (the "post-spike" sample concentration), in order to establish MS acceptance limits from the statistical LCS acceptance limits, the random error associated with the additional MS measurement must be taken into account. (A "pre-spike" sample concentration is not measured for the LCS; since the LCS is a spiked blank, the "pre-spike" sample concentration is assumed to be zero.) The correction factors enclosed in parentheses in Equations 12.1 to 12.4 account for the additional measurement uncertainty associated with MS recovery determinations. The correction factors were calculated by assuming that the standard deviation is a linear function of concentration and give first-order approximations for the MS acceptance limits.

12-4. Criteria.

12.4.1. Representativeness.

a. Before evaluating matrix spike results, review the SAP, QAPP and similar planning documents. These documents should describe how representative matrix spikes will be selected for the environmental matrices of interest, particularly for heterogeneous matrices such as soils.

b. The composition of a matrix spike sample must be similar to that of the associated environmental samples. For example, when soil sampling is performed, the SAP should describe how the on-site geologist will select representative matrix spikes. This typically entails classification of soil type. For example, a matrix spike should be collected for a set of samples high in sand and a separate matrix spike should be collected for a set of samples high in clay. However, this does not imply that matrix spikes should be collected solely on the basis of grain size classification (e.g., sand, silt, and clay). For example, the origin of the geological formation (fill, glacial deposits, stream deposits, etc.) should also be taken into account. Therefore, unless all soil samples are being collected in a single geological formation of relatively uniform composition or matrix interference has been well characterized during prior investigations, a batch of samples should typically contain several matrix spikes (each representing a different soil type and general origin). Similarly, if only one matrix spike were collected for a set of groundwater samples but the groundwater samples were collected from two hydraulically isolated aquifers being investigated at the site (e.g., a "shallow" and a "deep" aquifer), then, in general, one should not assume that the matrix spike would be representative of the groundwater in both aquifers.

c. If the matrix spike sample for the preparation batch originates from a different project site or is suspected to be of dissimilar composition from the other samples in batch, it must not be used to qualify the other field samples. In order to consolidate small numbers of samples from different project sites, the laboratory may analyze samples from different projects together in the same preparation batch for the same parameters. However, the MS results would not be applicable to the samples collected from the other sites. Allowing the laboratory to choose the samples to be spiked often results in the selection of unrepresentative matrix samples. Similarly, matrix spikes must not be selected by field personnel in a manner that is solely designed to satisfy frequency requirements. For example, the collection of all matrix spikes on the last day of

sampling activities to satisfy a 5% frequency requirement for the collection of matrix spikes will typically result in unrepresentative matrix spike samples.

d. In general, a matrix spike sample must contain all the target analytes of interest. A subset may be used when it can be demonstrated that the subset of target analytes characterizes (i.e., represents) method performance for the remaining (unspiked) target analytes.

> Note: When only a subset of the target analytes is included in the matrix spikes, project documents such as the QAPP must present a scientifically defensible rationale for not spiking the entire set of target analytes. A number of promulgated analytical methods recommend specific target analytes for the matrix spikes. Merely referencing a subset of analytes recommended in a published analytical method (e.g., the six MS compounds listed in SW-846 Method 8260B) does not constitute a scientifically defensible rationale for not spiking all the target analytes (e.g., unless the method explains why the subset of spiked analytes is representative of the remaining target analytes).

12-4.2. Frequency.

Review the appropriate project documents (e.g., the QAPP) to determine the required frequency of MSs, MSDs, and MDs. A MS and MSD or MS and MD (representative of each type of matrix analyzed) are usually required for every batch of samples processed. MD pairs are typically used for inorganics (especially metals) and MS/MSDs for organics. Matrix spikes and matrix duplicates are usually collected at a frequency of at least 5% if the matrix is relatively uniform in physical composition.

12-4.3. Acceptance Limits.

Bias and precision specifications for matrix spikes and matrix duplicates are dependent upon the DQOs of the investigation. Acceptance limits for matrix spikes and duplicates should be specified in project documents such as the QAPP. Guidance for establishing "default" acceptance limits for matrix spikes and matrix duplicates (e.g., when acceptance limits are not specified) is presented below.

12-4.3.1. Project Specific Communications.

a. The laboratory's statistical LCS acceptance limits should not be greater than the project-required acceptance limits for matrix spikes and matrix-dependent duplicates. When this criterion is not satisfied (i.e., project-required acceptance limits are more stringent than the statistical LCS acceptance limits) and matrix spikes or matrix-dependent duplicates fail to meet the project-required acceptance limits, it is not generally valid to assume that the failures resulted from matrix effects. For example, assume that the statistical LCS recovery range is 60–140%, the project-required MS recovery range is 80–120%, and a MS recovery for a batch of environmental samples is 65%. The associated environmental samples must be qualified (e.g., using the J flag) for not meeting the project-required tolerance for accuracy. *However, the associated sample results must not be qualified for matrix interference.* (In this example, the MS recovery of 65% falls well within the statistical LCS acceptance range).

b. If the acceptance limits for matrix spikes are *not specified or are inappropriate* (e.g., refer to Paragraph 12-2) and the laboratory's *statistical* LCS acceptance ranges are *comparable to or more stringent* than the *project-required* LCS acceptance ranges (e.g., the warning or control ranges for the LCS recoveries fall approximately within the corresponding project-required acceptance ranges for the LCS recoveries), then approximate the statistical matrix spike recovery ranges as discussed in Chapter 12-3. Compare the calculated MS acceptance ranges and the project-required LCS acceptance ranges. Qualify the environmental data using the most extreme limits from the two sets of acceptance ranges. However, it is emphasized that this approach is applicable only if the project-required LCS ranges are greater than or equal to the laboratory's statistical control ranges.

> Note: It is recommended that the ranges be rounded (e.g., to the nearest 5% or 10%) to more readily compare the laboratory's statistical acceptance range to a project-required acceptance range. It is also recommended that the laboratory's statistical limits be viewed to be comparable to the project limits, when the LCS warning range falls approximately within the project-required LCS acceptance range. Alternatively, the width of the control range should be no greater than about 1.5 times the project's acceptance range. For example, if the project-required recovery range for the LCS is 90% - 110%, then a warning range of 90% - 110% or a control range of 85% - 115% would be considered to be acceptable.

c. To illustrate the above approach, assume that a matrix spike acceptance range is not specified, the laboratory's statistical control range for the LCS is 67–113% (i.e., 90% ± 23) and the project-required acceptance range for the LCS is 70–130%. The laboratory's statistical control range approximately falls within project-required LCS acceptance range. If it is assumed that the spiking concentration for the MS is at least twice as large as the native analyte concentration (e.g., which will typically result in a conservative estimate for the MS acceptance range), then, using Equation 12.4 in Chapter 12.3. In this example, the acceptance range for the MS is 90% ± 23% (1.5) = 55–125%. (Note that if Equation 12-2 were used, the acceptance range would be only be slightly wider: 90% ± 23% (1.6) = 53–127%.) Since the calculated MS acceptance range is 55–125% and the project-required LCS acceptance range is 70–130%, set the MS acceptance range for the project using the most extreme limits; use 55–130% as the MS acceptance range. Therefore, a MS recovery that does not fall within 65–130% is indicative of a significant matrix effect *and* the associated environmental samples would be qualified (e.g., as estimated or potentially rejected).

d. When acceptance limits for matrix spikes recoveries are *not specified or are inappropriate* and the laboratory's *statistical* LCS control ranges are significantly *wider* than the *project-required* LCS acceptance ranges, then a conservative approach is recommended. *Evaluate the matrix spike recoveries using the project-required LCS acceptance limits.* For example, if the LCS acceptance range is 80% - 120%, then the matrix spike acceptance range should be set to 80% - 120%. If LCS acceptance limits are not specified, then use the guidance presented in Chapter 11 of this document to establish a set of "default" LCS/MS acceptance limits. In general, if the MS recovery falls outside of the LCS acceptance range, then qualify the associated results as estimated or rejected. *However, it is inappropriate to attribute the unacceptable MS*

recovery solely to matrix interference. The evaluation strategies for the matrix spike and matrix duplicates are essentially the same as those for laboratory control samples described in Chapter 11.

12-4.3.2. Establishing Acceptance Limits for Matrix-Dependent Duplicates.

If acceptance limits are not specified for **matrix-dependent duplicates** (i.e., MDs and MS/MSD pairs), if appropriate, then calculate the matrix spike limits using the procedure Chapter 12.3 and set the maximum RPD equal to one half the calculated MS acceptance range. Alternatively, evaluate the RPD results using the project-required RPD acceptance limits for *laboratory control samples*. If RPD limits are not specified for laboratory control samples, set each RPD acceptance limit for matrix-dependent duplicates equal to one half of the width of the project-required recovery range for the corresponding LCS, or to the laboratory's statistical RPD acceptance limit when derived from LCS data, whichever is less. For example, if the project-required LCS recovery range is 80% - 120% and the laboratory does not maintain statistical limits for duplicate precision using LCS data, set the RPD acceptance limit for matrix-dependent duplicates to 20%.

12-5. Evaluation.

Review the standard preparation logs to verify that all target analytes were included in the matrix spike. Using the laboratory summary forms for the matrix spike and matrix duplicate results, recalculate the recovery and the RPD for at least one target analyte. Compare the calculated values to the values reported on the laboratory's summary form. The result must agree to within two significant figures. Review the Case Narrative and all of the recovery and precision results on the laboratory summary forms and note any failures.

12-6. Contractual Considerations.

a. Contractual issues may impact the review of MS, MSD, and MD data. However, contractual considerations for matrix spikes and matrix duplicates are more complex than those for blanks and laboratory control samples because the results are dependent upon matrix effects as well as sample preparation and analysis errors. For example, the heterogeneity of soil grab samples and sequentially collected groundwater samples complicates the evaluation of MS/MSD results because uniform concentrations are assumed for the native analytes. Therefore, laboratories do not typically base batch control on the results of MS, MD, or MSD samples unless a general method failure is indicated.

b. When matrix spikes or matrix duplicates grossly fail QC acceptance limits in a systematic manner, examine the Case Narrative and any laboratory communications (e.g., phone logs) included in the data package to determine if the Project Manger was notified and corrective actions other than data qualification were performed. Refer to project planning documents such as the Scope of Work for laboratory analytical services and the QAPP to determine whether corrective actions other than data qualification are required.

c. When gross failures occur and expected laboratory corrective actions are not performed, the reviewer should consult with the Project Manger to determine whether to proceed

with the PB review or to reject the data package as a whole (e.g., the laboratory may be required to reanalyze the environmental samples). Some probable corrective actions for matrix interferences are listed below:

(1) If a matrix spike recovery is unacceptable and matrix interference is suspected, then the laboratory should be expected to make a reasonable attempt to remedy the problem. Corrective action for matrix interference may include the implementation of cleanup procedures or other method modifications. For example, cleanup methods should be performed to address matrix interferences for extractable organic analyses such as the BNA, pesticides, and PCB analyses (e.g., as described in SW-846 Method 3600). The method of standard additions may be required for metal analyses. Under these circumstances, verify that appropriate method modifications were performed to minimize the matrix interference.

(2) When a MS recovery is unacceptable but matrix interference is not otherwise apparent, the MS sample would normally be reprocessed (e.g., reextracted and reanalyzed) by the laboratory to verify the effect. However, the MS sample would not be reprocessed if the failure is consistent with historical data. The matrix effect is confirmed if the second result is similar to the original result (in magnitude and direction of bias). It should be noted that some methods specify other verification procedures. For example, if low matrix spike recoveries are obtained for hexavalent chromium in soil, Method 3060A indicates that additional analyses should be performed (e.g., pH and oxidation-reduction potential) to determine whether or not the low matrix spike recovery results from reducing conditions within the environmental sample. When unacceptable matrix spike recoveries are obtained, examine the data package to determine if appropriate confirmatory procedures were implemented.

12-7. Qualification.

Data that fail quality objectives because of matrix effects may be unusable to support decisions and must be qualified. Data quality may also be adversely impacted if the matrix spike sample is not representative of the other environmental samples in the batch. Data are qualified for matrix effects primarily using the same qualification strategies for laboratory control samples. In particular, data qualification must take both magnitude and direction of bias into account. When both a MS and MSD are processed for a batch of samples, use the most noncompliant matrix spike recovery to evaluate and qualify the data. Additional guidance is presented below.

12-7.1. Matrix Spikes and Matrix Spike Duplicates.

a. For both the MS and MSD, compare the spiking levels to the concentrations of the native analytes in the sample selected for spiking. If the native concentration of a target analyte is high relative to the spiking concentration, then this may contribute a significant uncertainty to the recovery calculations; the MS recovery may not be representative of actual method performance for the matrix. In the absence of other guidance, *evaluate the MS recovery when the spiking concentration is at least two times greater than the native analyte concentration*. If environmental samples were qualified by the laboratory for matrix interference but the spiking levels are low relative to the native analyte concentrations, then the flags must be omitted. However, professional judgment is important when evaluating the native analyte concentration relative to

the spiking concentration. For example, if the spiking concentration is near but less than two times the native analyte concentration, a gross MS recovery failure (e.g., a MS recovery of 5%) is probably indicative of a matrix effect (rather than a low-spiking concentration) and the associated results must be qualified for matrix interference. In general, if the MS spiking concentration is between one and two the native analyte concentration, then data qualification is recommended only when gross MS recovery failures occur.

b. If the LCS results are acceptable, the spiking levels for the MS are high relative to the native analyte concentrations (i.e., at least two times the native analyte concentration), the matrix spike sample is representative of the other environmental samples, and the MS recovery falls outside of the acceptance limits, then significant matrix interference may exist. Qualify the *associated* sample results (e.g., environmental samples of a similar matrix collected from the same site) as follows:

(1) If all target analytes are present in the matrix spike, and the recovery of a particular analyte is unacceptable, then qualify all detections of the analyte in the associated environmental samples using the strategies discussed in Chapter 11. For example, if the MS recovery for a target analyte falls grossly below the lower recovery acceptance limit, then qualify all detections less than the AL with the X flag. Note that in those instances where it can be determined that the MS or MSD results affect only the sample spiked, qualification must be limited to this sample alone.

(2) If all the target analytes are not present in the matrix spike, then use professional judgment to determine the extent to which qualification of the non-spiked target analytes is required. In general, each spiked analyte must be clearly linked to each of the unspiked target analytes. If one of the spiked analytes clearly represents some subset of the target analytes, then qualify only the target analytes of the subset on the basis of the MS recovery. For example, if analyte “A” in the matrix spike sample is representative of the subset of target analytes {A, B, C} in the environmental samples, then qualify analytes “A,” “B,” and “C” for the environmental samples using the MS recovery of analyte “A.” However, if a clear association does not exist (e.g., and the lack of matrix interference was not demonstrated during a prior sampling event), then a conservative approach is recommended. At a minimum, qualify detections and nondetections for the unspiked analytes in the environmental samples as estimated (i.e., qualify detections with the J flag and nondetections with the UN flag). However, if the recovery of one or more of the spiked analytes is unacceptable, then qualify all of the unspiked analytes using the most noncompliant MS recovery.

c. If a MS sample is not available or is not representative of the other samples in the batch, then the performance of the method in the matrix of concern has not been well characterized. At a minimum, qualify the environmental sample results as estimated. If the data are being used to support critical decisions and method performance in the matrix of concern is not otherwise known (e.g., the environmental population of interest has not been previously sampled and the surrogate recoveries are not available or representative of the target analyte), then it may be appropriate to qualify the sample results as tentatively rejected.

12-7.2. Matrix-Dependent Duplicates.

a. Precision is typically measured using the RPDs for MS/MSD or MD pairs. MS/MSD pairs would normally be used to evaluate duplicate precision when low-level contamination is anticipated (i.e., analyte concentrations less than the MQLs) and MDs would normally be used to evaluate duplicate precision when high levels of contamination are expected. Compare the RPDs reported for all target analytes to the corresponding RPD acceptance limits.

b. Evaluate target analyte RPDs for MS/MSD pairs when the spike concentration is at least *two times* the native analyte concentration. Evaluate target analyte RPDs for MD pairs for analytes detected at or above the MQL. (The RPD is evaluated when a target analyte detection is greater than or equal to the MQL for at least one sample of the MD pair.) RPD results that do not satisfy these criteria (e.g., RPDs calculated from detections at concentrations less than the MQLs) must not be used to evaluate duplicate precision.

> Note: Sometimes an acceptance criterion for duplicate precision is specified for the MQL and a different acceptance criterion is specified for concentrations that are greater than the MQL by some multiplicative factor. Evaluate the appropriateness of the duplicate precision acceptance criterion that is nearest to the decision limit prior to performing data qualification. For example, assume that the QAPP requires the maximum RPD to be 40% for results equal to or greater than five times the MQL and requires results to agree to within ± MQL for concentrations between the MQL and 5 x MQL. Also assume that AL = 32 ppb, MQL = 20 ppb, and the following duplicate results are obtained: 20 ppb and 40 ppb. Since the duplicate results are less than 5 x MQL (100 ppb) and agree within ± MQL (i.e., ± 20 ppb), according to the QAPP, the results should not be qualified. However, since the MQL is near the AL and the RPD for the duplicate pair is high (RPD = 67%), the duplicate results do not demonstrate that contamination is above or below the AL. Contrary, to the criteria specified in the QAPP, qualified the associated sample results as estimated (e.g., unless quantitative statistical methods are being used to quantify the uncertainty and to compare the results to the AL).

c. If (i) the LCS results are acceptable, (ii) the spiking levels for the MS/MSD are high relative to the native analyte concentrations (i.e., at least two times the native analyte concentration) or the native analyte concentrations for the sample/MD are at least as high as the MQL, and (iii) the RPD is unacceptable, then a significant matrix effect may exist.

d. If precision is evaluated using MS/MSD pairs containing only a subset of the target analytes of interest and the analytes are representative of the set of unspiked target analytes, then qualify the sample results using the subset of target analytes in the MS/MSD. If it is unknown whether or not the subset of target analytes adequately represents the unspiked target analytes, then a conservative approach is recommended. Evaluate the unspiked target analytes using the most noncompliant RPD for the MS/MSD. However, even when duplicate precision is acceptable for the subset of target analytes in the MS/MSD, it may be appropriate to qualify all detections and nondetections of the unspiked target analytes as estimated (e.g., when statistical

analyses are not being performed to characterize the variability of these analytes in the matrix of concern).

e. When the RPD is unacceptable, qualify the associated sample results using the same strategies presented in Chapter 11 (e.g., Table 11-2). For example, when precision is evaluated using MD pairs or MS/MSD pairs and the direction of bias is unknown, then qualify all detections of the analyte in the associated environmental samples with the J flag and nondetections with the UN flag when marginal failures occur. However, when the RPD is marginally unacceptable and the direction of bias can be determined from other QC information, then qualify the detections using J+ or J- flag (instead of the J flag). For example, assume that the acceptance range for matrix spike recoveries is 80–120%, the acceptance limit for the RPD is 20%, and an RPD of 33% was calculated from matrix spike recoveries of 70 and 50%. Since the RPD is marginally unacceptable and bias is low, the associated detections would be qualified with the J- flag. However, in those instances where it can be determined that the results affect only the MD or MS/MSD pairs (and not the other samples in the preparation batch), then qualification must be limited to those samples alone.

f. It may not be possible to collect representative duplicates. For example, if duplicates are collocated samples (e.g., a pair of VOC soil samples) or cannot be homogenized because of the nature of material being sampled (e.g., multi phase wastes), then high RPDs are probably the result of sample heterogeneity rather than method performance problems in the matrix being investigated (e.g., digestates with high concentrations of dissolved salts, being analyzed for trace metals by Method 6010B, are not intermittently clogging the ICP nebulizer, giving rise to erratic results). If precision failures occur (gross or marginal) sample heterogeneity, then it is recommended that detections be qualified with the J flag and nondetections be qualified with the UN flag. The data review report must state that representative duplicates were not collected and the data user should determine whether or not the environmental sample and matrix-dependent duplicate results can be used to support project decisions.

CHAPTER 13

Surrogates

13-1. Introduction.

Surrogates are organic compounds that are similar in chemical composition to the analytes of interest and spiked into environmental and batch QC samples *prior* to sample preparation and analysis. Surrogate recoveries for environmental samples are used to evaluate matrix interference on a sample-specific basis. However, in order for this approach to be viable, the surrogates must behave in the same manner as the corresponding target analytes that are native to the matrices of interest (e.g., must partition between various phases in the same manner as the native target analytes). Unfortunately, in practice, this equivalency is typically difficult to demonstrate and is often more assumed than empirically derived. The most representative surrogate will typically be an isotopically-modified version of the target analyte. Therefore, when evaluating surrogate results, the representativeness of the surrogates should always be taken into account.

13-2. Criteria.

a. The acceptance for surrogate recoveries must take the end use of the data into account and must not be based solely upon contractual or method-specified limits. Method-specified surrogate acceptance limits (e.g., for SW-846 and CLP methods) are often inappropriately wide. Statistically-based acceptance limits generated by the laboratory may be representative of routine method performance but may also be too wide (i.e., may not satisfy project-specific DQOs).

b. The acceptance ranges for surrogate and target analyte spike recoveries must be similar (particularly for laboratory control samples and blanks), since, by definition, surrogates and target analytes are chemically similar compounds.

> Note: It is common for statistical control limits for surrogates to be significantly wider than the control limits for target analytes. This often occurs when surrogate control limits are calculated by inappropriately grouping surrogate recoveries from LCSs, MSs, and environmental samples into a single data set.

c. When the surrogate acceptance ranges are significantly wider than the acceptance ranges for the target analyte, then the appropriateness of the surrogate acceptance ranges must be carefully evaluated prior to performing data review or validation. When the surrogate acceptance limits are inappropriately wide, establish "default" acceptance limits using the target analyte acceptance ranges if these ranges appear to be reasonable. For example, if the acceptance ranges for the target analytes are approximately 70–130% (e.g., for the LCS) and the surrogate acceptance limits are 20–150%, set the acceptance range for the surrogates to 70–130%. Otherwise (i.e., in the absence of more appropriate acceptance limits), surrogate recoveries for organic methods should be evaluated using the acceptance ranges of 80–120% for purge-and-trap methods and 60–140% for extractable organic methods. However, if the LCS is prepared from an independent-source standard, then an acceptance range of 70–130% may be used for purge-and-trap methods.

d. If an analytical method requires *no more than two surrogates,* then surrogate results are acceptable only if *all* of the surrogate recoveries are in control. If three or more surrogates are associated with a set of target analytes, then one surrogate may be marginally (but not grossly) out of control. However, the marginal failure must not be systematic in nature (i.e., must occur in a sporadic or random manner). In particular, if several consecutive failures are observed for the same surrogate, then the data must be qualified.

13-3. Evaluation.

Review the laboratory Case Narrative and the summary forms and note any surrogate failures that are reported. A significant amount of professional judgement is required to evaluate surrogate results. However, the following strategies are generally applicable:

a. Prior to reviewing the surrogate data, examine the Case Narrative to determine whether any of the surrogate results should *not* be used to qualify the environmental sample results.

(1) Do not qualify environmental samples for matrix interference when surrogate recoveries are unacceptable because of localized chromatographic problems. For example, if several surrogates are associated with a group of target analytes and some (but not all) of the surrogate recoveries are unacceptable because of coeluting interferences, then qualification is not required.

(2) Do not qualify environmental samples for matrix interference when surrogate recoveries are unacceptable because of dilutions. For example, if all of the surrogate recoveries for an environmental sample are unacceptable because the surrogates were "diluted out," but the surrogate recoveries for the LCS and associated blanks are acceptable, then no further action is typically required.

(3) It is recommended that the raw data be requested for review when zero-percent surrogate recoveries are reported and these recoveries are not attributed to dilution. Zero-percent recoveries may arise from retention time shifts rather than from losses (e.g., during extraction).

b. If an unacceptable surrogate recovery is associated with only a subset of the target analytes (e.g., the surrogate is representative of the performance for only the acid fraction of the BNAs analyzed by Method 8270B), then qualify the results for only the subset of analytes.

c. Surrogate recoveries for laboratory control samples and method blanks characterize overall laboratory method performance in the absence of matrix interference are evaluated in much the same manner as target analyte recoveries. Distinguish unacceptable surrogate recoveries arising from matrix effects beyond the control of the laboratory from failures arising from poor laboratory analytical technique. When a surrogate recovery is out-of-control for an environmental sample but is also out-of-control for the LCS or an associated blank (e.g., the method blank), a laboratory performance problem rather than a matrix effect must be assumed.

d. Check for transcription and calculation errors for a representative number of samples. Using the laboratory summary form for the surrogate results, recalculate the recovery of at least one surrogate and compare the calculated value to the reported value. The two results must agree to within two significant figures.

13-4. Contractual Considerations.

a. Contractual considerations may impact the data review when surrogate failures are observed for laboratory control samples and blanks. A laboratory would normally be expected to reprocess a batch of field samples when a surrogate recovery is unacceptable for an LCS or blank. When surrogate recoveries for laboratory control samples or blanks are unacceptable and the batch of samples is not reprocessed, examine the Case Narrative and note why the corrective action was not performed. When surrogate recoveries for laboratory control samples and blanks grossly and systematically fail QC acceptance criteria, qualify the affected data accordingly and notify the Project Manager to determine whether to continue the PB data evaluation. (If the review were discontinued under these circumstances, the entire data package would be rejected.)

b. When surrogate failures are noted for environmental samples, refer to project documents such as the QAPP and the Scope of Work for analytical services to determine what corrective actions need to be documented in the laboratory's data package. Corrective actions typically performed for surrogate failures are discussed below:

(1) If matrix interference is not apparent in the chromatogram, an unacceptable surrogate recovery for an environmental sample is normally confirmed by reextracting and reanalyzing the sample. (The extract would be reanalyzed for confirmation if there were insufficient sample for reextraction.) The matrix effect is confirmed when the repeated result is within the same order of magnitude and exhibits bias in the same direction as the original result. Under these circumstances, examine the data package to determine if confirmatory analyses were performed. However, it should be noted that the laboratory may not routinely reprocess environmental samples with unacceptable surrogate recoveries unless surrogate failures in method blanks or laboratory control samples are indicative of a general method failure.

(2) When surrogate recoveries are unacceptable because of matrix interference, the laboratory may be required to perform method modifications or cleanup procedures (e.g., as described in Method 3600 of SW-846 for the SVOC analyses). Under these circumstances, examine the data package to determine if cleanups were performed. Note that when there are unacceptable surrogate recoveries followed by successful reanalyses, the laboratory is typically required to report only the successful run. When there are unacceptable surrogate recoveries followed by unsuccessful reanalyses, the laboratory is typically required to report both runs.

13-5. Qualification.

a. The qualification protocols for surrogate recoveries are similar to those for LCS recoveries. Qualification is generally required when the surrogate acceptance criteria of Chapter 13.2 are not met. If two surrogates are associated with (i.e., are representative of the performance of) a set of target analytes and both surrogate recoveries are unacceptable, qualify the sample result using the most noncompliant surrogate recovery. Similarly, if three or more surrogates are used and one or more surrogates are grossly out of control, then data qualification must be based upon the most noncompliant surrogate recovery. However, no action is required if three or more surrogates are used and one surrogate is marginally out of control in a sporadic manner.

b. Data qualification for noncompliant surrogate recoveries is dependent upon the direction and magnitude of the failure. Distinguish *gross* surrogate recovery failures from *marginal* failures. In the absence of more appropriate guidance, a *gross failure* is defined to occur when any surrogate recovery does not fall within 20–180% for *extractable organic* analyses and 60–140% for *purge-and-trap* analyses.

c. When a surrogate recovery for an environmental sample falls outside of the acceptance limits, the direction of bias will be said to be "well defined" when the remaining surrogates and all associated QC samples are in control or exhibit bias in the same direction. For example, if the recovery of a surrogate exceeds the upper control limit but the recoveries of other surrogates are below the lower control limit, then the direction of bias is not well defined (i.e., has not been adequately demonstrated). When there are several surrogates, a high or low recovery for a single surrogate is not necessarily indicative of the direction of bias or method extraction efficiency.

d. A direction of bias must not be inferred from the surrogate recoveries of volatiles analyzed by purge-and-trap (e.g., when the recoveries of all the surrogates are unacceptably low or high) unless the responses of the internal standards are available for review. Similar compounds are used for internal standards and surrogates for purge-and-trap analyses. The direction of bias will not be well defined when the surrogate and internal standard recoveries are not consistent with one another. For example, a high surrogate recovery can be obtained when the internal standard response (e.g., peak area) is extremely low (since the concentration of the surrogate is determined from the ratio of the surrogate response to the internal standard response).

e. In general, when the criteria of Chapter 13.2 are not met, qualify the target analytes (associated with the surrogate) as discussed below. The qualification strategies below apply (i) when two surrogates are used, (ii) and when *three or more* surrogates are used *and gross or systematic surrogate failures are observed*. These qualification strategies are illustrated in Table 13-1.

(1) "If any surrogate recovery is *marginally* unacceptable, bias *is* well defined, and there are *no gross recovery failures* for other associated surrogates, then the data must be qualified as

follows: For *low* bias, qualify detections with the J- flag and nondetections with the UN flag. For *high* bias, qualify detections with the J+ flag and nondetections with the U flag. (Note that qualification is not required when three or more surrogates are used and one sporadic marginal failure is observed.)

(2) "If any surrogate recovery is *marginally* unacceptable, bias is *not* well defined, and there are *no gross recovery failures* for other associated surrogates, then the data must be qualified as follows: Qualify detections with the J flag and nondetections with the UN flag. (Note that qualification would not be required if three or more surrogates were used and one sporadic marginal failure were observed.)

(3) If any surrogate recovery is *grossly* out of control and the direction of bias is *well defined* (i.e., the recoveries of the remaining surrogates are in control or exhibit bias in the same direction), then qualify the data as follows:

(*a*) For *low* bias, qualify all nondetections with the R flag. If an AL is *not* specified, then qualify detections with the J- flag. If an AL is specified, then qualify detections less than the AL with the X flag.

(*b*) For *high* bias, qualify all nondetections with the U flag. Qualify detections with the J+ flag. However, when an AL is specified, it may be appropriate to qualify detections greater than the AL with the X flag (e.g., when a conservative estimate is not being sought).

(4) If any surrogate recovery is *grossly* out of control and the direction of bias is *not* well defined, then qualify all nondetections with the R flag. If an AL is *not* specified, qualify all detections with the J flag. If an AL is specified, then qualify detections less than the AL with the X flag. Depending on project DQOs, qualify detections greater than the AL with the J or X flag.

Table 13-1
Data Qualification for Surrogate Recoveries [1]

Sample Surrogate Recoveries	Field Sample Result (y)	Flag
$\%R_1$ and $\%R_2$ in control: $LCL_1 \le \%R_1 \le UCL_1$ $LCL_2 \le \%R_2 \le UCL_2$	MRL < MQL < y	Flag not required
	MRL < y < MQL	J
	y < MRL	U
$\%R_1$ or $\%R_2$ marginally OFC with low bias: $\%R_1 < LCL_1$ or $\%R_2 < LCL_2$	y > MRL	J-
	y < MRL	UN
$\%R_1$ or $\%R_2$ marginally OFC with high bias: $\%R_1 > UCL_1$ or $\%R_2 > UCL_2$	y > MRL	J+
	y < MRL	U
$\%R_1$ or $\%R_2$ marginally OFC with inconsistent bias: $\%R_1 < LCL_1$, $\%R_2 > UCL_2$ or $\%R_1 > UCL_1$, $\%R_2 < LCL_2$	y > MRL	J
	y < MRL	UN
$\%R_1$ or $\%R_2$ grossly OFC with low bias: $\%R_1 << LCL_1$ or $\%R_2 << LCL_2$	y > MRL	J- X if y < AL
	y < MRL	R
$\%R_1$ or $\%R_2$ grossly OFC with high bias: $\%R_1 >> UCL_1$ or $\%R_2 >> UCL_2$	y > MRL	J+ Possibly, X if y > AL
	y < MRL	U
$\%R_1$ or $\%R_2$ grossly OFC with inconsistent bias: $\%R_1 << LCL_1$, $\%R_2 > UCL_2$ or $\%R_1 >> UCL_1$, $\%R_2 < LCL_2$ or $\%R_1 < LCL_1$, $\%R_2 >> UCL_2$ or $\%R_1 > UCL_1$, $\%R_2 << LCL_2$	y > MRL	J X if y < AL Possibly X if y > AL
	y < MRL	R

Notes: 1. It is assumed that the LOI ≤ MRL < AL. For the purposes of illustration a field sample result is evaluated using the recoveries of two surrogates. The subscripts indicate which surrogate is being referenced. For example, $\%R_1$ denotes the percent recovery of the first surrogate. The following abbreviations are used: %R = Recovery of surrogate spiked into field sample; y = Concentration of a target analyte in the field sample; AL = Action Level; LCL = Lower control limit for surrogate recovery; UCL = Upper control limit for surrogate recovery; OFC = Out of control

CHAPTER 14

Data Review Reports

14-1. Introduction.

Post digestion spikes are typically evaluated for trace metal analyses to assess the ability of a method to successfully recover target metals from an actual sample matrix *after* the digestion process has been performed. The PDS results are used with MS results to evaluate matrix interferences.

14-2. Criteria.

14-2.1. Frequency.

a. Like matrix spikes and matrix duplicates, the frequency of post digestion spikes is ultimately established from the project's data objectives. No PDSs may be required or a PDS may be required for every sample in the batch. In general, a PDS should *not* be required for a set of environmental samples when a *representative* MS sample is processed and the MS recovery is acceptable. Ideally, when a PDS is required, the matrix spike and PDS should be prepared from the same environmental sample.

> Note: Project documents (e.g., QAPPs) often require PDSs to be analyzed at the frequency specified in standard analytical methods (e.g., the CLP SOW requires a PDS for each sample). Unfortunately, the frequency for PDSs may be poorly or inappropriately defined some methods. For example, Method 6010B of SW-846 states that a PDS should be analyzed "whenever a new or unusual sample matrix is encountered." However, the term "new or unusual" is not well defined. Furthermore, even if "new or unusual matrix" were defined, PDSs would not be required to demonstrate performance if representative matrix spikes were processed and acceptable MS recoveries were obtained.

b. When project documents do not specify the PDS frequency or the PDS frequency is deemed to be inappropriate by the reviewer and an unacceptable matrix spike recovery is observed, use professional judgement to determine whether or not a PDS analysis should have been performed. For example, a PDS would not be required to confirm the presence of matrix interference if a serial dilution analysis (SDA) were performed and confirms the matrix effect. However, in the absence of a technically defensible rationale to do otherwise, assume that a post digestion spike (PDS) must be analyzed when the MS is unacceptable.

14-2.2. Acceptance Limits.

The acceptance range for each PDS recovery must be no wider than the corresponding acceptance range for the matrix spike recovery. When project-specific limits are not specified, an acceptance range of 85% to 115% is recommended when the concentration of the PDS is at least

two times the native sample concentration. An acceptance range of 80% to 120% is recommended when the spiking concentration is one to two times the native analyte concentration.

> Note: The acceptance range for the PDS specified in Method 6010B (75–125%) is wider than acceptance range for the MS (80–120%). When metal analyses are performed using Method 6010B and the spiking concentration is high relative to the native analyte concentration, the acceptance range for the PDS should be no wider than 80–120%.

14-3. Evaluation.

Examine the standard preparation logs to verify that the PDS contains all the target metals. Examine the sample preparation log to determine whether the PDS was prepared from the same sample used to prepare the MS. Review the Case Narrative and the PDS summary forms and note any PDS failures. Using the laboratory's PDS summary form, recalculate the PDS recovery for at least one target analyte and compare it to the reported value. The reported and calculated result must agree to within *two* significant figures.

14-4. Qualification.

a. PDS results are qualified using the same strategies for matrix spikes. In particular, compare the PDS spiking levels to the concentrations of the native analytes in the sample selected for spiking. If the native concentration of a target analyte in the sample (digestate) is high relative to the spiking concentration, then the PDS recovery may not be representative of actual method performance. Evaluate the PDS recovery when the spiking concentration is at least *two* times greater than the native analyte concentration (e.g., unless the spiking concentration is slightly less than two times the native analyte concentration and a gross failure occurs). If environmental samples were qualified (e.g., by the laboratory) for matrix interference but the spiking level for the PDS is low relative to the native analyte concentration, remove the data qualifiers.

b. If a single field sample is used to prepare the PDS and MS and the spike concentrations of both batch QC samples are at least two times greater than the native analyte concentrations, then evaluate the data as discussed below.

14-4.1. MS Recovery Acceptable and PDS Recovery Unacceptable.

If the MS (and LCS) recovery for a metal falls within the QC acceptance range but a PDS was analyzed and the PDS recovery is unacceptable, a matrix effect should *not* be suspected. The laboratory would normally be expected to reanalyze the PDS sample (digestate) to confirm the result. Contractual corrective action for unacceptable laboratory performance may be appropriate when a number of failures of this nature are observed and confirmatory reanalyses are not performed. When a problem of this nature occurs, it is recommended that the Project Manager be notified. It may be appropriate to request the raw data to perform a more comprehensive review. If there is a gross discrepancy between the PDS and MS recoveries for a particular metal (e.g., the MS recovery is within 80% - 120% but the PDS recovery is not within 50–150%), rejection of the data would constitute the most conservative approach.

14-4.2. MS Recovery Unacceptable and PDS Recovery Acceptable.

In general, if the MS recovery for a metal does not fall within the QC acceptance range but the PDS recovery is acceptable, then a matrix effect (associated with the preparatory process) should be suspected and the field sample results must be qualified on the basis of the matrix spike recovery. However, when historical data for the effect does not exist, the laboratory would normally be expected to perform a second digestion and reanalysis of the MS to confirm the result. The result would be confirmed if the MS recoveries and PDS recoveries for both sets of analyses were similar in magnitude and bias.

14-4.3. MS Recovery Unacceptable and PDS Recovery Unacceptable.

a. When both the MS recovery and PDS recovery for a particular metal fall outside of QC acceptance range in the same manner (i.e., the PDS and MS failures are of similar magnitude and the direction of bias is the same), confirmatory analyses are unnecessary. Assume that a matrix interference exists and use the most noncompliant recovery (the MS or PDS recovery) to qualify the data.

b. When both the MS and PDS are unacceptable, the laboratory should be expected to make a reasonable effort to correct for matrix interference before qualifying the field samples for matrix interference. Review the Case Narrative to determine what corrective actions were performed. Corrective actions for matrix interference may include the use of a different matrix modifier, different instrument operating conditions, the method of standard additions, internal standards, a different digestion or analytical procedure, and serial dilutions (if action levels can be met).

> Note: If project documents do not clearly demonstrate that the matrix spike sample is representative of the samples in the batch (which often occurs in environmental investigations), then the benefits of extensive corrective actions by the laboratory to minimize a matrix effect should be considered to be minimal. In other words, if the laboratory did not make a “reasonable” attempt to correct for the matrix interference, but the matrix spike sample is not representative of the samples in the batch, the lack of “representativeness” should be considered to be much more significant than the lack of corrective actions.

c. When the PDS and MS are prepared from two different environmental samples and the spike concentration is at least two times the native metal concentration, then evaluate the PDS as follows:

(1) If both the PDS and MS recoveries for a target metal fall outside of the QC acceptance range in the same manner (i.e., the PDS and MS failures are of similar magnitude and the direction of bias is the same), then assume matrix interference exists and qualify the data using the strategies discussed in Chapter 12.

Table 14-1
Evaluation of PDS and MS Data

%R [MS/PDS] [1]	Summary of Evaluation	References
PASS / PASS	Results not qualified.	
PASS / FAIL	1. Check for confirmatory analyses for the digestate 2. Request additional information from the laboratory and use professional judgement to either reject the data, qualify the data using the unacceptable PDS recoveries, or qualify the data using the acceptable MS recoveries.	Chapter 14-4.1
FAIL / PASS	1. Check for confirmatory analyses for the matrix spike. 2. Qualify for matrix interference based on the MS %R.	Chapter 14-4.2
FAIL / FAIL	1. Check if corrective action was taken to address the matrix interference. 2. Qualify field samples for matrix interference based upon the most noncompliant of the MS and PDS recoveries.	Chapter 14-4.3

Notes:1. It is assumed that the PDS and MS were prepared from the same environmental sample. A PDS or MS recovery is denoted to be in *FAIL* status when percent recovery, %R, does not fall within the recovery acceptance range.

(2) When the PDS recovery is acceptable but the MS recovery is not, use the MS recovery to qualify the associated field sample results. The laboratory may be required to analyze additional PDSs for the batch of samples (e.g., one PDS for every sample in the batch) when the matrix spike recovery is unacceptable. Under these circumstances, verify that the additional analyses were performed.

(3) When inconsistent PDS and MS recoveries are observed for two different samples in the preparation batch (e.g., the PDS recovery is biased high and the MS recovery is biased low), then the representativeness of the PDS and MS results for the remaining samples of batch must be carefully assessed. In particular, if the MS recovery for a soil sample is acceptable but the PDS recovery for a second soil sample is not, then the MS and PDS samples may not be representative of the remaining samples of the preparation batch and qualification of these samples may not be appropriate.

CHAPTER 15

Data Review Reports

15-1. Introduction.

a. In general, second column confirmation is required for chromatographic methods with 2-D detectors when the analytes of concern have not been well characterized. For single-component (i.e., single-response) analytes, a "tentative" identification of a target analyte occurs when the peak associated with the analyte falls within the retention window for the "primary" column. A "confirmed" identification occurs when the analyte peak also falls within the retention time window for "secondary" or "confirmatory" column. The confirmatory and primary columns must be dissimilar columns (i.e., must possess different stationary phases) so that the elution order for the target analytes reported from the primary and secondary columns differ. Target analyte identification for multi component (i.e., multi response) analytes (e.g., Aroclors by GC) are primary performed using pattern recognition. Hence, second column confirmation would typically be performed only if the identity of the analyte were in doubt (e.g., would be performed for weathered Aroclors by GC).

b. A quantitative result from the "primary" column and a confirmed identification from the "confirmatory" column are minimally required for second-column confirmation. In other words, if the "primary" column possesses quantitative capability, only detection capability is minimally required for the "confirmatory" column. However, it is usually desirable to apply the same QC criteria to both the "primary" and "confirmatory" columns and to report quantitative results from both analytical columns. (Note, that under these circumstances, the column designations "primary" and "confirmatory" are arbitrary; results reported from either column are equally reliable.) This strategy is advantageous because it provides a measure of instrument duplicate precision. In addition, when a chromatographic interference occurs for the primary column but does not prevent confirmation, a quantitatively reliable detection may still be reported from the confirmatory column.

15-2. Criteria.

15-2.1. Frequency.

Unless the analytes of concern have been well characterized or confirmation will be performed using an instrument with a 3-D detector, second column confirmation must be performed for all detections (i.e., all results above the reporting limits).

15-2.2. Duplicate Precision.

Unless otherwise specified, assume that quantitative results must be reported from both the primary and confirmatory columns. If a target analyte is detected with both the primary and confirmatory column and the result reported from one (or both) of the columns is greater than the quantitation limit, then the RPD calculated for the pair of results must be less than or equal to the absolute value of twice the uncertainty tolerance for the CCVs. In particular, if the error toler-

ance for the CCV is ± 20% (e.g., 80% - 120%), then the RPD must be within 40% (2 × 20%). The RPD for each pair of results is calculated from the equation:

$$\text{RPD} = 100 \, | y_1 - y_2 | \, / \, [(y_1 + y_2) / 2]$$

where y_1 and y_2 denote the results from the primary and secondary columns.

15-3. Evaluation.

Verify that all single component analyte detections were confirmed. Confirmation for multi component analytes will be dependent upon the nature of the contamination and the objectives of the investigation. All results above the reporting limit must fall within the retention time windows for both analytical columns. If possible, verify that dissimilar chromatographic columns were used for the primary and confirmatory columns. Calculate the RPD for a pair of results and ensure that the calculated and reported values agree to within two significant figures.

15-4. Qualification.

a. The qualification strategies must distinguish *quantitative* reliability from *qualitative* reliability. *If second-column confirmation is required for the project but was not performed, then, at a minimum, qualify all single-component analyte detections with the N flag (the results are not qualitatively reliable).* Based upon the objectives of the project, the X or R flag may be more appropriate. *If the nature of the site contamination has not been well characterized, then qualify all the detections with the X or XN flag.* For example, if PAH analyses, by liquid chromatography with a UV detector, are being performed for a new study area for a risk assessment and some valid confirmation procedure was not performed, then it would probably be appropriate to qualify detections (especially low-level detections) as tentatively rejected. The X flag should be used when detections are greater than project-specific action levels and a conservative estimate is inappropriate (for the particular phase of the project). However, note that, if a sample is analyzed using second-column confirmation, but an analyte is not confirmed in the sense that the analyte peak is detected with the primary column but not with the confirmatory column, then the analyte result is reported as “not detected” (e.g., using the U flag).

b. At a minimum, qualify all detections with the N flag, if, in the reviewer’s professional judgment, the two analytical columns are not sufficiently dissimilar (e.g., a C-18 column is used with a C-8 column instead of a CN column for explosives by HPLC).

c. Chromatographic interferences from coelutions can affect the quantitative as well as the qualitative reliability of the data. A high (i.e., noncompliant) RPD may result because one or more non-analyte peaks that elute in the retention time window for the analyte of interest. Qualify the results with high RPDs as follows:

(1) If the RPD is unacceptable high, at least one of the results is above the method quantitation limit, and the chromatograms are not available for review, or a coelution cannot otherwise be definitively identified, then, at a minimum, qualify the results from both the primary and confirmatory columns as qualitatively and quantitatively estimated using the NJ flag. Qual-

ify the results with the X flag when a gross failure occurs and the reason for the unacceptable RPD is not apparent. However, if a decision level is available and both results are less than the decision level, then NJ flag may be more appropriate. It is recommended that a gross failure be defined to occur when the calculated RPD is greater than two times the RPD acceptance limit (e.g., when the RPD > 80%).

(2) If the result from the primary column and the corresponding result from the secondary column are both less than the quantitation limit and a high RPD is obtained, then qualify both results with the J flag (rather than with the N flag).

(3) If it can be determined that a high RPD value arises from a coelution problem but confirmation is unaffected, then only qualify the result from the column with the coelution problem as quantitatively estimated or rejected. For example, assume that detections greater than the quantitation limit for a particular target analyte are reported from both the primary and confirmatory columns, but the result from the primary column is not quantitatively reliable because a non-target analyte gives rise to a very large broad shoulder on the target analyte peak. Since confirmation is unaffected and a quantitative result is available from the confirmatory column, the result from the confirmatory column would be reported as unqualified (assuming that all other QC criteria are met), but the result from the primary column may be rejected for quantitative reliability. However, it should be noted that a comprehensive data package would typically be required to perform this type of evaluation.

Table 15-1
Qualification for Second-Column Confirmation [1]

RPD	Result	Reported (Qualified) Result
RPD < 40%	MRL < MQL < y_1 MRL < MQL < y_2	y_1
	MRL < y_1 < MQL MRL < y_2 < MQL	y_1 J
RPD not calculated because y_2 < MRL	y_1 > MRL	MRL U
40% < RPD < 80%	MRL < MQL < y_1 MRL < MQL < y_2 $y_2 > y_1$	y_1 NJ and y_2 NJ [2] or y_2 NJ
	MRL < y_1 < MQL MRL < y_2 < MQL	y_1 J
RPD > 80%	MRL < MQL < y_1 MRL < MQL < y_2 $y_2 > y_1$	y_1 X and y_2 X If y_1, y_2 < AL, then y_1 NJ and y_2 NJ
	MRL < y_1 < MQL MRL < y_2 < MQL	y_1 J
RPD not calculated because confirmation was not performed	MRL < MQL < y_1	y_1 N or y_1 X
	MRL < y_1 < MQL	y_1 JN or y_1 X
	y_1 < MRL	MRL U

Notes: 1. Assume both columns are acceptably calibrated and all QC samples are in control (with the possible exception of the RPD). The result, y_1, is being reported from the primary column and result from the confirmatory column is denoted by y_2. The acceptance limit for the RPD is assumed to be 40%. 2. When the RPD > 40 and the reason for the high RPD is unknown, then the preferred approach is to report the results from both columns. As per the USEPA OSW memorandum "Clarification Regarding Use of SW-846 Methods" of 7 August 1998, "an approach that is conservative relative to environmental protection is to report the higher of the two values when the relative percent difference is greater than 40% *and* no interferences or chromatographic anomalies are evident." However, if it can be determined that the high RPD is from a chromatographic interference for one of the columns, then report the result from the remaining column (unqualified).

CHAPTER 16

Internal Standards for Organic Chromatographic Methods

16-1. Introduction.

For organic analyses, internal standards are compounds that are similar in chemical composition to the analytes of interest. However, unlike surrogates, internal standards are spiked into all instrument QC, batch QC, and environmental samples immediately prior to instrumental analysis. (Surrogates are spiked into batch QC and environmental samples prior to sample preparation and analysis.) For environmental applications, the internal calibration technique is typically used for mass spectrometry methods but may also be used for chromatographic methods with 2-D detectors. Internal standard response should be monitored throughout instrumental analysis to help evaluate instrument performance (e.g., sensitivity and stability) and matrix effects.

16-2. Criteria.

16-2.1. Frequency.

When quantitation is performed using internal standards, known quantities of internal standards must be added to all instrument QC, batch QC, and environmental samples immediately prior to instrumental analysis.

16-2.2. Acceptance Criteria.

a. The compounds used for internal standards, the concentrations, and the acceptance criteria will be highly dependent upon the analytical technique and the set of analytes of interest. In general, instrumental response for an internal standard must fall well within the calibration range. Internal standards in all samples must fall within the retention time windows for the most recent CCV (especially, for 2-D chromatographic methods). Ideally, internal standards should also elute at retention times that are near the retention times of the associated target analytes.

b. Unless a more appropriate criterion is available, the peak area for each internal standard in all instrument QC, batch QC, and environmental samples should be within -50% to +100% of the corresponding peak area for the mid-level initial calibration standard. The mean internal standard peak area for the set of initial calibration standards may be used in lieu of the internal standard peak area of the mid-level initial calibration standard.

> Note: This is contrary to the guidance presented in the "USEPA Contract Laboratory Program National Functional Guidelines for Organic Data Review." Internal standard peak areas for samples are evaluated using the internal standard peak area for the most recent CCV rather than the internal standard peak areas for the initial calibration. For example, for the "Volatile Data Review," Section X ("Internal Standards") states: "Internal standard area counts must not vary more than a factor of two (-50% to +100%) from the associated 12hr calibration standard." This practice is not recommended since it is based on the assumption that sensitivity, as measured by the internal standard areas

of the CCVs, will not progressively degrade (i.e., decrease) during the analysis. In other words, it is assumed that the internal standard peak area of each CCV will not systematically be significantly smaller than that for the preceding CCV.

16-3. Evaluation.

Review the internal standards summary form(s) and ensure that all the internal standards fall within the appropriate retention time windows and the internal standard areas fall within appropriate acceptance limits. Verify, using at least one of the CCVs, that the internal standard peak area acceptance limits were correctly calculated.

16-4. Qualification.

a. In general, environmental samples must be qualified for an unacceptable internal area count even when the samples are bracketed by acceptable CCVs.

Note: The composition of internal standards and the target analytes are similar but, in general, it should not be assumed that internal standards will behave in an identical manner as all the target analytes during all environmental conditions. For example, a matrix interference for a particular environmental sample may affect the internal standards more than the associated target analytes (e.g., the preferential absorption of an internal standard relative to the associated target analytes). Because of differences in volatility during purge-and-trap analyses, an internal standard may be lost more readily than its associated target analytes. For example, because of a poor seal for one of the purge vessels, the internal standard bromochloromethane can be lost more readily than the associate target analyte acetone.

b. The concentration of a target analyte is inversely proportional to the internal standard area. Hence, an unacceptably low area count tends to give rise to a high bias and an unacceptably high area count tends to give rise to a low bias. However, the analyte concentration is ultimately dependent upon the response for the internal standard as well as that for the target analyte. Therefore, the evaluation of bias on the basis of internal standard response is often problematic when a comprehensive data package is not available. If the internal standard area count of a sample does not fall within the acceptance range for internal standard area of the mid-level initial calibration, then qualify the results as summarized in the table below. It is assumed that a high internal standard response gives rise to a low bias. However, in order to obtain a more conservative evaluation of the data, a direction of bias is not assumed for low internal standard response.

c. An internal standard peak area that does not meet the threshold criteria for a detection may give rise to a false negative for the associated target analyte and surrogate results. If an extremely low area count is reported and the chromatograms are not available for review (e.g., to assess signal to noise ratios), a conservative approach must be used; qualify nondetections with the R flag. However, if a more detailed review of the data is planned, then the X flag would be more appropriate.

d. A more detailed review of the data is recommended when a sample's internal standard does not fall within the retention time acceptance windows. For example, the data evaluator should request a comprehensive data package and the raw data (e.g., chromatograms and quantitation reports) should be examined to determine if any false positives or false negatives exist.

Table 16-1
Data Qualification of Internal Standard Areas [1]

Internal Standard Area of Sample (A)	Sample Result (*y*)	Flag
½ $A_0 \leq A \leq 2\ A_0$	MRL < MQL < *y*	None
	MRL < *y* < MQL	J
	y < MRL	U
2 $A_0 < A < 5\ A_0$	*y* > MRL	J-
	y < MRL	UN
$(A_0/5) < A <$ ½ A_0	*y* > MRL	J
	y < MRL	U
A > 5 A_0	*y* > MRL	J- X if *y* < AL
	y < MRL	R
A < $(A_0/5)$	*y* > MRL	X
	y < MRL	R

Notes: 1.I t is assumed that the LOI ≤ MRL < AL. The following abbreviations are used: A = Internal standard area count sample; A_0 = Internal standard area count sample; y = Concentration of a target analyte for the field sample; AL = Action Level; MRL = Method Reporting Limit; MQL = Method Quantitation Limit.

APPENDIX A

Glossary

A-1. Accuracy.

Accuracy refers to *closeness* to the *true value*. The *true value* typically refers to the expected or prepared spike concentration. For statistical applications, the *true value* typically refers to the population mean. Accuracy is a conceptualization and cannot be known with complete certainty. Accuracy is evaluated in a variety of ways. For laboratory analyses, accuracy is commonly inferred from the percent recoveries of spike samples (matrix spikes, laboratory control samples, surrogate spikes, performance evaluation samples, etc.).

The terms "accuracy" and "bias" are commonly used interchangeably, but, strictly speaking, these terms are not synonymous. Accuracy is generally a function of both "random error" and "systematic error." "Random error" is characterized by unpredictable variations for the measured results of a parameter of interest, cannot be corrected directly, but can be reduced by increasing the number of measurements. (It should be noted that "random error" is distinguished from "spurious error," which is also unpredictable, but arises from factors such as human blunder and gross instrument malfunction.) The term "bias" refers to systematic directional error from the "true value." Unlike random error, "bias" or "systematic error" remains constant or varies in a predictable manner, but is independent of the number of measurements. The mean spike recovery for the LCS is a measure of method bias and the standard deviation for the LCS recoveries is a measure of method precision. The mean LCS recovery with the standard deviation essentially constitutes a measure of accuracy.

A-2. Analyte.

See "Target Analyte."

A-3. Batch.

See "Preparation Batch."

A-4. Batch Quality Control (QC) Sample.

See "Quality Control (QC) Sample."

A-5. Bias.

Bias refers "systematic error." Bias is a directional error that arises from a constant or predictable distortion of the measurement process. A measurement or estimate is said to be *unbiased* if the mean approaches the true value as the number of replicates increases. An estimate is said to possess a *low bias* if it is consistently less than the true value and is said to possess a *high bias* if it is consistently higher than the true value. The adjectives "high" and "low" are used to refer to the *direction* rather than the magnitude of the deviation from the true

value. Adjectives such as “small,” “slight,” “marginal,” “large,” and “gross” will be used to refer to the *magnitude* of the deviation. For chemical analyses, *consistently* low or high recoveries for batch QC samples (e.g., laboratory control samples and matrix spikes) are indicative of bias.

A-6. Chain-Of-Custody (COC).

Chain-Of-Custody (COC) procedures and forms primarily document the possession of the samples from collection to storage, analysis, reporting, and, ultimately, disposal. Each cooler sent from the field to a laboratory is accompanied by a unique COC record. (The COC form is typically sealed in a Ziploc-type bag and is taped to the inside of the cooler lid.) COC forms must become part of the permanent record of all sample handling and shipment. The COC form lists the samples in a cooler, and includes the following information: project identification, unique project-specific sample identifications, dates and times of sample collection, number of containers, general testing procedures, and any special remarks. Couriers’ shipping documents should also be included.

A-7. Characteristic Peaks.

For multi component target analytes (e.g., Aroclors), *characteristic peaks* are those peaks that are at least 25% of the height of the largest peak in chromatogram for the pure multi-component standard.

A-8. Comparability.

Comparability refers to the equivalency of two sets of data. This goal is achieved through the use of standard or similar techniques to collect and analyze representative samples. Comparable data sets must contain the same variables of interest and must possess values that can be converted to a common unit of measurement. *Comparability* is normally a qualitative parameter that is dependent upon the other data quality elements. For example, if the detections limits for a target analyte were significantly different for two different methods, the two methods would not be comparable.

A-9. Completeness.

Completeness refers to the percentage of data that is *valid* or usable; that is, which satisfies project-specific DQOs. The highest degree of completeness that can be achieved is normally desired. Completeness acceptance criteria would normally be defined for both field and laboratory activities. A typical acceptance criterion for completeness is 80% to 90%. A higher completeness acceptance criterion may be required for critical samples. In general, when calculating percent completeness, R-qualified and X-qualified data must not be included in the set of valid data.

A-10. Comprehensive Blank.

See “Hierarchy of Blanks.”

A-11. Comprehensive Data Package.

A *comprehensive data package* is defined as a package of "definitive" or "effective" chemical data that satisfies the minimum data reporting requirements of this document and contains sufficient information to completely reconstruct the chemical analyses that were performed. Comprehensive data packages include all **batch, method, and instrument QC** results as well as raw data (e.g., run logs, sample preparation logs, standard preparation logs, and printed instrumental output such as chromatograms).

A-12. Continuing Calibration Verification (CCV).

A continuing calibration verification (CCV) refers to the use of a mid- to low-level instrumental standard to *check* rather than to alter instrument calibration. CCVs are typically analyzed on a continuing basis (e.g., at the beginning, middle, and end of an analytical sequence) and are evaluated to determine whether the instrument was within acceptable calibration throughout the period of time samples were instrumentally analyzed. The CCV is usually (but necessarily) prepared from a standard that is from the same source as the initial calibration standards.

A-13. 2-D and 3-D Detectors.

A "3-D" detector differs from a "2-D" detector in that the former furnishes quantitative and comprehensive qualitative information for definitive compound identification, while the latter primarily furnishes only quantitative information. Detectors such as PIDs, ECDs, and FIDs are referred to as "2-D" or "two-dimensional" detectors since they essentially yield a two-dimensional plot of gross instrumental response versus time (i.e., "single-channel" time-versus-response data). Two-dimensional detectors cannot provide sufficient qualitative information for analyte identification. Detectors such as mass selective and infrared (IR) detectors are examples of "3-D" or "three dimensional" detectors since they provide time-versus-response data for multiple mass ions and wavelengths, respectively.

A-14. Data Quality Indicators.

See "PARCCS."

A-15. Data Quality Objectives (DQOs).

Data quality objectives (DQOs) are perhaps more appropriately referred to as data objectives. DQOs refer to the quantitative and qualitative statements that identify the goals, decision strategies, and boundaries for a particular study (e.g., acceptable levels of uncertainty); in addition, DQOs define the type, quality, and quantity of data required to support project decisions by the data users. The DQOs are developed during the planning stages of a project based upon the scientific method of inquiry. With respect to the chemical testing, DQOs are developed prior to sample collection and analysis, in order to determine appropriate analytical methodology, quality control acceptance limits (i.e., specifications for *data quality indicators*), and corrective actions.

A-16. Detection.

A detection typically refers to a reported measured concentration of a target analyte that is greater than the detection limit or the reporting limit.

A-17. Definitive Data.

The distinction between *definitive* and *screening data* is rather subjective. Definitive data are typically produced using "rigorous" analytical methods, such as EPA reference methods. The analytical results are frequently evaluated with respect to relatively stringent quality control specifications and PARCCS criteria are well defined. Recently, the term "effective" data (versus "definitive" data) has been used to describe data of sufficient quality to support project decisions. Screening data are essentially data that are not "fully effective"--data that cannot be used to support project decisions without higher quality data.

For example, since screening methods often lack specificity, they tend to give rise to false positives. Therefore, screening data are usually confirmed by testing a percentage of the environmental samples (e.g., 10%) with *definitive* methods or "more effective" methods of analysis. Quantitative data from screening methods also tend to be less precise and accurate than that from definitive methods. Screening data are typically generated by methods of analysis that are relatively rapid (typically involving minimal sample preparation) and performed in the field (as opposed to an off-site laboratory). However, real-time data generated in the field is not necessarily of inferior quality to fixed laboratory data.

A-18. External Calibration.

The *external calibration* technique is primary used for organic chromatographic analyses involving detectors other than MS detectors (e.g., FID, PID, ECLD, ECD and NPD). A **calibration factor** is calculated for each analyte and surrogate in each initial calibration standard using the equation:

$$CF = \frac{\text{Peak Area or Height of analyte in calibration standard}}{\text{Amount of the Compound Injected (e.g., mass in nanograms)}}$$

For multi component analytes, the numerator is the sum of the area or heights of several peaks. In other words, the calibration factor is the ratio of detector response to the amount of analyte in the calibration standard. The amount of analyte in an environmental sample is calculated by dividing the instrumental response for the analyte by the mean calibration factor for all the initial calibration standards.

A-19. Field Duplicates.

Field Duplicates are similar to "matrix duplicates." They differ in that the former are prepared in the field while the latter are prepared in the laboratory. A field duplicate is an environmental sample that is homogenized and split into two separate aliquots in the field rather than at the

laboratory. This document distinguishes between field duplicates and collocated samples. Collocated samples are field samples that are collected "near" each another during a single sampling event but which are not homogenized. However, for simplicity, field duplicates and colocated samples are not distinguished from laboratory duplicates when homogenization cannot be performed because of the nature of the analyte or the methodology. For example, samples collected for low-level VOC analysis by closed-system purge-and-trap cannot be homogenized. Hence, for these type of analyses, the term "matrix duplicate" refers to collocated samples.

A-20. Field QC Samples.

Field QC samples are *QC samples* that are prepared in the field or that are impacted by field activities. Examples of field QC samples include trip blanks, rinsate (equipment) blanks, and field duplicates. Matrix spike samples may or may not be field QC samples. For example, if an environmental sample were homogenized in the field and subsequently split into three aliquots for MS and MSD analyses, then the MS and MSD samples would be considered to be field QC samples.

A-21. Hierarchy of Blanks.

When environmental (field) samples are subjected to multiple handling, preparatory and analytical procedures, blanks may be introduced in a sequential manner to measure the level of contamination arising from each procedure or from select sets of procedures. For example, assume that a sample is sequentially processed using two distinct preparatory techniques, which will be referred to as "technique 1" and "technique 2." The sample and a blank, BLK(1,2), are processed using technique 1. The sample, the blank BLK(1,2), and a second blank, BLK(2), are then processed using technique 2. When blanks are processed in this manner, they can be used to evaluate the contamination associated with each stage of the preparatory process. For example, assume that contaminants are detected in BLK(1,2), but none are detected in BLK(2). It would be assumed that the contamination resulted from the first preparatory technique. Furthermore, since BLK(1,2) accounts for the contamination introduced from the entire preparatory process, only BLK(1,2) would minimally be required to evaluate environmental samples for contamination. For example, if BLK(2) were not processed, then the samples would be evaluated using BLK(1,2) alone.

A blank that measures contamination for a set of handling, preparatory, or analytical procedures is said to possess a *higher hierarchy* than a blank that measures contamination for only a subset of the procedures. In the example cited above, the blank BLK(1,2) possesses the highest hierarchy because it measures contamination from all of the preparatory techniques. The highest hierarchy blank will be referred to as a **comprehensive blank** if it accounts for contamination from all sample handling, preparatory, and analytical procedures. In general, a blank with a higher hierarchy is more critical than one with a lower hierarchy. In particular, when a blank is missing for a set of environmental samples, the samples may be qualified for contamination using a blank that possesses a higher hierarchy than the missing blank. Blanks for environmental analyses are listed in order of increasing hierarchy below:

Calibration/Instrument Blanks < Storage/Holding Blanks < Method Blanks

< Trip Blanks < Rinsate/Equipment Blanks

Note that trip blanks are collected only for VOC analyses. Furthermore, a rinsate blank may be substituted for a trip blank only when the rinsate blank is stored and shipped in the same cooler as the field samples. Under these circumstances, the rinsate blank would account for contamination arising from cleaning procedures, cross contamination in the sample coolers, and laboratory contamination.

A-22. Holding Time.

The **preparation holding time** (e.g., the **extraction** or **digestion holding time**) is defined as the period of time from the date an environmental sample is collected in the field to the date the sample is processed with the preparatory method (e.g., the date the sample is first exposed to the extraction or digestion solvent). The **analysis holding time** is defined as the period of time from the date of sample preparation (e.g., extraction or digestion) to the date of sample analysis using some determinative (i.e., instrumental) method.

A-23. Holding Time Limit.

The *holding time limit* is defined as the maximum acceptable holding time for sample preparation or analysis.

A-24. Initial Calibration.

Initial calibration refers to the establishment of a quantitative relationship between instrumental response and analyte concentration (or amount) prior to the analysis of samples. The correlation between instrumental response and analyte concentration is established via the analysis of a set of standards of known concentration and is demonstrated using quantitative performance specifications (e.g., linear correlation coefficients). The initial calibration must demonstrate that, over some concentration range of interest, a change in analyte concentration is associated with a predictable change in instrumental response and vice versa (i.e., there is a continuous functional and inverse functional relationship between instrumental response and concentration).

A-25. Initial Calibration Verification (ICV).

An *initial calibration verification (ICV)* refers to the use of a mid-level, second-source, instrumental standard to verify the accuracy of the standards used to perform the initial calibration. The ICV is typically performed immediately after the initial calibration. The acceptance limits for ICV recoveries should be similar to the acceptance limits for other instrumental QC samples such as CCVs.

A-26. Instrument Quality Control (QC) Sample.

See “Quality Control (QC) Sample.”

A-27. Internal Calibration.

The *internal calibration* technique is primarily used for organic chromatographic analyses involving MS detectors. An **internal standard** is added to each sample and calibration standard immediately prior to analyses. An internal standard is a substance that is similar to the target analytes in chemical behavior, which is not normally found in the environmental samples, and which is added at a fixed, known concentration to all samples and calibration standards. A **relative response factor** is calculated for each analyte and surrogate in each initial calibration standard according to the equation:

$$RRF = \frac{A_s C_{is}}{A_{is} C_s}$$

where

A_s = Peak area or height of the analyte or surrogate
A_{is} = Peak area or height of the internal standard
C_s = Concentration of the analyte or surrogate (e.g., µg/L)
C_{is} = Concentration of the internal standard (e.g., µg/L)

For multi component analytes such as dioxins, the terms A_s and A_{is} represent the sum of the integrated ion abundance of multiple quantitation ions. The calculation of the amount of analyte in an environmental sample involves dividing instrumental response for the analyte by the instrumental response for the internal standard and the mean relative response factor for the set of initial calibration standards. The internal standard technique is superior to the external standard technique because target analyte loss is taken into account for the portion of the analytical process that takes place after the internal standard is spiked into the sample (e.g., loss during sample injection). The internal standard procedure is used primarily with MS detectors because the signal intensities used for quantitation would not otherwise be adequately stable and the masses of the internal standards can be resolved from those of the target compounds even when chromatographic resolution cannot be achieved.

A-28. Laboratory Control Sample (LCS).

The *laboratory control sample* (LCS) is used to assess laboratory method performance. Laboratory control samples are analyzed to assess the ability of the laboratory to successfully recover the method's target analytes from a sample of known composition. Precision may be evaluated by analyzing an LCS and an LCSD (laboratory control sample duplicate) for each preparation batch of samples. A laboratory control sample must be chemically and physically similar to the environmental samples and must contain a known amount of each target analyte at an appropriate concentration. A laboratory control sample typically consists of a clean matrix (e.g., reagent water or purified sand) that has been spiked with the target analytes of interest. In general, an LCS must contain all single-component target analytes of interest and must be processed through the entire sample preparatory and analytical methods. The LCS usually contains only a subset of the target analytes when multi component analytes such as Aroclors are

being analyzed. Ideally, the concentrations of the target analytes in the LCS should be determined by project-specific DQOs (e.g., should be near the regulatory or risk-based decision limits), but the LCS is typically spiked between the low-level and mid-level calibration standards.

Internal LCS acceptance limits for accuracy and precision are produced by the laboratory by performing statistical calculations (using at least 30 data points). However, since duplicate precision is not as critical as accuracy, many laboratories do not generate statistical acceptance limits for duplicate precision but use "default" (e.g., method specified or arbitrary) RPD acceptance limits. Control charts and tables are maintained to establish the bias and precision of the method, and are updated periodically (typically, on a quarterly basis). A representative subset of the target analytes for each method is normally graphed to observe method trends. Unfortunately, LCS acceptance limits for environmental sampling and analysis activities are often based upon the laboratory's internally generated control chart limits or method-specified limits rather than project-specific DQOs. Ideally, project-specific acceptance limits should be equal to or greater than the laboratory's in-house statistical control limits.

When an LCS result falls outside of the laboratory's internal acceptance limits, the laboratory must implement some form of corrective action. In general, the preparation batch must be reprocessed when the associated LCS recovery falls outside of the acceptance range. When an LCS RPD is out-of-control but the LCS recovery is acceptable, the laboratory must implement corrective action but the associated environmental samples would not typically be reprocessed.

A-29. Laboratory Control Sample Duplicate (LCSD).

See "Laboratory Control Sample (LCS)".

A-30. Limit of Identification (LOI).

The *limit of identification* (LOI) is the lowest concentration of analyte that can be detected with *99%* confidence; that is, the LOI is the concentration at which the probability of a false negative (Type II decision error) is *1%*. The LOI is adjusted for method specific factors (e.g., sample size) and may be approximated as twice the detection limit. The LOI may be set equal to about two times the MDL (e.g., if it is assumed that the standard deviation is not strongly dependent upon concentration).

A-31. Matrix-Dependent Duplicate.

See "Matrix Spike Duplicate (MSD) and Matrix Duplicate (MD)."

A-32. Matrix Spike (MS).

The *matrix spike* (MS) is used to assess the performance of the method as applied to a particular matrix; they are analyzed to assess the ability of the method to successfully recover target analytes in the environmental population being sampled. An MS is an environmental sample to which known concentrations of all of the method target analytes have been added before it is

carried through all sample preparation, cleanup, and analytical procedures. MS results are evaluated in conjunction with other QC information (e.g., surrogate and LCS recoveries) to determine the effect of the matrix on the accuracy of the analysis.

The target analytes added to a matrix spike sample would typically be identical to those added to a laboratory control sample. The environmental sample selected for MS analysis must be representative of the environmental population being sampled and would normally be specified in the field. Control charts may be maintained for MS recoveries, but, in general, laboratories do not base batch control on the results of MS samples unless a general method failure is indicated. Matrix spikes are typically analyzed at a frequency of at least 5% but frequency requirements are project-specific.

A-33. Matrix Duplicate (MD) and Matrix Spike Duplicate (MSD).

The *matrix duplicate* (MD) and *matrix spike duplicate* (MSD) are used to assess the precision of a method in an actual matrix. A MSD is a duplicate of an MS. An MSD is also used to evaluate the accuracy of a method in a particular matrix and is evaluated using the same criteria for the MS. An MD is an environmental sample that is divided into two separate aliquots. (Care must be taken to ensure that the sample is properly divided into homogeneous fractions.) Both the MD and MSD are carried through the complete sample preparation, cleanup, and analytical procedures. For brevity, MS/MSD and matrix duplicate pairs are referred to as **matrix-dependent duplicates**.

Frequency requirements for MDs and MSDs are normally established on a project-specific basis. An MD is normally processed with each preparation batch when target analytes are expected to be present. A MSD is normally processed with each preparation batch when method target analytes are not expected to be present. As a "rule of thumb," a MSD is used for organic methods and a MD is used for inorganic methods. The results of the MD or MSD are evaluated, in conjunction with other QC information, to determine the effect of the matrix on the precision of the analysis. Control charts, or tables, may be maintained for these samples to monitor the precision of the method for each particular matrix and may be required by certain projects.

A-34. Matrix Interference.

As used in this document, the term "matrix interference" typically refers to an effect that arises from the native physical or chemical composition of an environmental sample that produces a negative or positive bias in the results.

For example, high concentrations of non-target analytes that coelute with the analytes of interest in the instrumental portion of a chromatographic method may give rise to a positive interference (i.e., high bias). Substances such as peat and clay may bind the target of interest and prevent complete extraction of the target analytes in the preparatory portion of an analytical procedure (especially when analyte concentrations are low), may give rise to a "negative" interference (i.e., low bias). However, sample heterogeneity is viewed as a characteristic of the matrix (e.g., the spatial variability of the environmental population being sampled) rather than as an "interference" for which the method of analysis must be optimized to reduce.

A-35. Measurement Quality Objectives (MQOs).

Measurement quality objectives are acceptance criteria for PARCCS for the various phases of the measurement process (e.g., sampling and analysis) that are established to ensure that total measurement uncertainty is within the range prescribed by project DQOs.

A-36. Method Blank (MB).

Method blanks are used to assess laboratory contamination. A method blank is defined as an interference-free matrix which is similar to the field sample matrix, lacks the target analytes of interest, and is processed with the environmental samples using the same preparatory and determinative methods. Hence, all reagents added to samples during extraction, cleanup, and analysis are also added to method blanks in the same volumes or proportions. Analyte-free reagent water is frequently used to prepare method blanks for aqueous analyses and a purified solid matrix (e.g., sand) is frequently used for solids.

Contamination may result in false positives or elevated reporting levels for target analytes. Method blanks are analyzed to assess contamination for the entire analytical process. Therefore, when a batch of samples is analyzed on separate instruments or separate analytical shifts, the method blank associated with the batch (e.g., extracted with the samples) must also be analyzed with the samples for each instrument and analytical shift.

A-37. Method Detection Limit (MDL).

The *method detection limit* (MDL) is the minimum concentration of a substance that is significantly greater than zero (an analytical blank) at the *99%* limit of confidence and is determined using the procedure described in 40 CFR, Part 136, Appendix B. The standard deviation, *s*, is calculated for *n* replicate aliquots (where $n > 7$) that are spiked near (e.g., one to five times) the estimated MDL and processed (as environmental samples) through the full analytical procedure. The standard deviation for the set of replicate analyses is subsequently multiplied by the Student t value corresponding to the 99% percentile of the t-distribution with n-1 degrees of freedom.

Since it is not practical to establish an MDL for each specific matrix received at any given laboratory, MDLs are usually estimated in interference-free matrices (typically reagent water for aqueous analyses and a purified solid matrix such as sand for the analysis of solid matrices). However, certain projects may require the determination of method detection limits in site-specific matrices.

As defined in 40 CFR 136 Appendix B, MDLs are method, matrix and instrument specific. MDL samples must be processed using the sample determinative and preparatory methods as the environmental samples (e.g., using the same extraction and cleanup procedures) and must be adjusted for method-specific procedures such as dilutions. When multiple instruments are used to perform the same method, MDLs may be *demonstrated* on individual instruments (including individual chromatographic columns and detectors) via the analysis of *MDL check samples*.

An *MDL check sample* is prepared by spiking an interference free matrix with all target analytes of interest at about two times the estimated MDL and subsequently processing it through the entire analytical procedure. If a target analyte is not recovered in the MDL check sample, then the MDL study should be repeated for that target analyte. It is recommended that a laboratory's MDLs be verified quarterly by analyzing detection limit check samples. MDL studies should be performed at least annually and whenever the basic chemistry or instrumentation for method is changed.

It should be noted that the statistical approach described in 40 CFR 136 Appendix B does not take calibration uncertainty into account. It is implicitly assumed that the calibration curve is known with certainty. Since the variability associated with the estimate of the calibration curve is not taken into account, when samples to not undergo a significant preparatory process, it may be desirable to establish detection limits using the procedure described by Andre Hubaux and Gilbert Vos *(Decision and Detection Limits for Linear Calibration Curves*, Analytical Chemistry, Volume 42, No. 8, July 1970).

A-38. Method Quantitation Limits (MQL).

The *method quantitation Limit* (MQL) is the concentration of an analyte in a sample that is equivalent to the concentration of the lowest initial calibration standard adjusted for method-specified sample weights and volumes (e.g., extraction volumes and dilutions). Typically, MQLs are equal to or greater than the lowest initial calibration standard and are at least five times greater than the MDL. MQLs must also be less than project-specific action levels. It is usually desirable for the MQL to be equal to some fraction of the project's action levels (e.g., one half or one third of the action levels).

A-39. Method Reporting Limit (MRL).

The *method reporting limit* (MRL) is the threshold or censoring limit below which target analyte concentrations are reported as "< MRL" or " MRL U," where "MRL" is the numerical value of the method reporting limit. The method reporting limit is usually established based on the laboratory's LOIs, MQLs, or project-specific action levels. The MRL for *undetected* analytes should not be less than the LOI or RDL and must not be greater than the AL.

A-40. Native Analyte.

In the context of environmental testing, the term "native analyte" refers to the analyte incorporated into the test material by natural processes or from past waste handling activities (e.g., as opposed to spike addition).

A-41. Nondetection.

A *nondetection* typically refers to a target analyte concentration that is less than the detection limit or the method reporting limit.

A-42. PARCCS.

The term "PARRCS" is an acronym for the primary elements of data quality: *Precision, accuracy, representativeness, completeness, comparability,* and *sensitivity*. It should be noted that *sensitivity* is often omitted and the acronym PARCC is commonly used; these five data quality elements (PARCC) are often referred to as *Data Quality Indicators* (DQIs).

A-43. Percent Difference (%D).

The percent difference of a measurement, X_o, *of a variable* X is defined by the equation:

$$\%D(X) = (| X_o - \langle X \rangle | / \langle X \rangle) \, 100$$

where

$$\langle X \rangle = \sum X_i / n$$

is the mean of a set of *n* replicate measurements of X (that excludes X_o). For brevity, unless otherwise specified, the term *percent difference* (%D) will refer to the percent difference for the **response factor** of a continuing calibration standard for an organic chromatographic method.

A-44. Percent Recovery (%R).

The *percent recovery* for a matrix or post-digestion spike is defined by the equation:

$$\%R = 100 \, (X_F - X_O) / S$$

where

X_F = Measured concentration of environmental sample after spike addition
X_O = Measure concentration of environmental sample prior to spike addition
S = Spike (reference) concentration

For CCVs, ICVs, and LCSs, the percent recovery is defined as:

$$\%R = 100 \, (X / S)$$

where

X = Measured concentration of QC sample

For brevity, the *percent recovery* is referred to as the **recovery**.

A-45. Percent Relative Standard Deviation (%RSD).

The *percent relative standard deviation* for n replicate measurements of a variable *X* is defined by the equation:

$$\%RSD(X) = (SD/\langle X\rangle)\,100$$

where

$$\langle X\rangle = \sum X_i / n$$

is the mean of the variable. For brevity, the term *relative standard deviation* (%RSD) will refer to the percent relative standard deviation of the **response factors** for the initial calibration standards for a chromatographic method.

A-46. Performance-Based Method/Approach.

This term does not appear to be well-defined in the literature. As applied to chemical testing, the term *performance-based* implies that the methodology used to produce an analytical result is secondary to the quality of the result itself. When a performance-based approach is implemented, specifications are primarily imposed upon the data (the "end product" of the analytical process) rather than upon the process by which the data are produced. Chemical data are generated by any analytical method which can demonstrate project-specific PARCCS requirements are met. Method QC elements such as detection limits, method blanks, laboratory control samples, and matrix spikes are minimally required to demonstrate method performance.

A-47. Post Digestion Spike (PDS).

A *post digestion spike (PDS)* is typically analyzed for metals to assess the ability of a method to successfully recover target metals from an actual matrix after the digestion process. A PDS is an environmental sample to which known concentrations of target metals are added after the digestion process. The spiking concentration for the PDS should not be less than about two times the native analyte concentration. The same target analytes should be spiked into the LCS, MS, and PDS. A PDS should be analyzed when the MS is unacceptable. When the MS is unacceptable, an aliquot of the same environmental sample should be selected for the PDS. Alternatively, a PDS should be routinely processed with each MS so that every batch of samples contains at least one sample that is spiked *before and after* the digestion process.

A-48. Precision.

Precision refers to the repeatability of measurements. For statistical applications, *precision* refers to the spread or distribution of values about the population mean and is frequently measured by the standard deviation. For the chemical analyses of environmental samples, *precision* is commonly determined from duplicate samples (e.g., matrix spike duplicates, matrix

duplicate and laboratory control sample duplicates) and is commonly measured using either the relative percent difference (RPD) or the percent relative standard deviation (%RSD).

A-49. Preparation Batch.

A *preparation batch* is defined as a set of samples that are prepared together by the same person or group of people; using the same equipment, glassware, and lots of reagents; by performing manipulations common to each sample in the same sequence and within the same time period (usually not to exceed one analytical shift). Ideally, the samples in a preparation batch must be from the same study area and must be of similar composition. Samples taken from the same study area would normally be grouped together for batching purposes within the constraints imposed by the method holding times. However, laboratories may find it necessary to group samples from different clients into a single batch.

Environmental and QC samples must be prepared, analyzed, and reported in a manner that is traceable to individual batches. Hence, each preparation batch must be uniquely identified within the laboratory. A preparation batch is normally limited to twenty field environmental samples of a similar matrix and also contains the appropriate QC samples (e.g., a laboratory control sample and a method blank). The QC samples undergo the same preparatory procedures (e.g., using the same extraction and cleanup methods) as the environmental samples. Samples in the same preparation batch would normally be analyzed together using the same instrument.

A-50. Preservation.

The term "preservation" refers to any technique (frequently involving the addition of laboratory-grade reagents) that retards biological, chemical, or physical processes that would alter the "representativeness" of the sample relative to the environmental population of interest (e.g., alter the analyte concentration in the sample matrix being tested). The most common preservation methods include pH adjustment, dechlorination, and temperature adjustment (i.e., cooling or freezing).

A-51. Professional Judgment.

As per ISO/IEC Guide 25 (August 1996 draft), the term *professional judgment* refers to "the ability of a single person or a team to draw conclusions, give opinions and make interpretations based on measurement results, knowledge, experience, literature and other sources of information." A "professional judgment" must be supported by appropriate documentation. The information or factors taken into account during the decision making process must be discussed.

A-52. Quality Control (QC) Sample.

This document distinguishes between *preparatory methods* (e.g., Method 3010A) and *determinative methods* (e.g., Method 6010A) of analyses. A QC (quality control) sample that is *independent of matrix effects* and analyzed *only in a determinative method* is referred to as an **instrument QC sample** (e.g., a CCV and CCB). A non-instrument QC sample that is *processed* with the same *preparatory and determinative* methods as the environmental samples (e.g., matrix

spikes and laboratory control samples) is referred to as a **method QC sample**. Note that method QC samples (e.g., MDL study samples) are not necessarily analyzed on a per batch basis. A non-instrument QC sample that is analyzed on a per batch basis is referred to as a **batch QC sample**. Hence, a batch QC sample is a method QC sample that is analyzed on a per batch basis, *or* is a QC sample which is analyzed in only the determinative method but which is dependent upon matrix effects (e.g., post-digestion spikes).

A-53. Recovery.

See **Percent Recovery**.

A-54. Relative Percent Difference.

The *relative percent difference* for a set of duplicate measurements of the variable X, RPD(X), is defined by the equation:

$$RPD(X) = (|X_1 - X_2| / \langle X \rangle)\, 100$$

where

$$\langle X \rangle = (X_1 + X_2)/2$$

is the mean of the pair of variables. The RPD is a measure of precision. For brevity, unless otherwise specified, the term *relative percent difference* refers to the relative percent difference of duplicate spike recoveries.

A-55. Reliable Detection Limit (RDL).

The *reliable detection limit (RDL)* is the upper 95% upper confidence limit of the MDL defined in 40 CFR, Part 136, Appendix B. RDLs for the upper 100ξ% upper confidence limit are calculated as follows:

$$RDL = [\, (n-1) / \chi_{n-1,(1-\xi)/2} \,]^{1/2}\, MDL$$

where the MDL is calculated by multiplying the standard deviation, *s,* by the 99th percentile point of the t-distribution with n–1 degrees of freedom, $t_{n-1, 0.99}$:

$$MDL = t_{n-1, 0.99}\ s$$

The number of replicate analyses used to compute the MDL is denoted by *n*. The $100\,(1-\xi)/2$ percentile of the Chi-Square distribution with *n* - 1 degrees of freedom is denoted by $\chi_{n-1,(1-\xi)/2}$. For n = 7 and $\xi = 0.95$

$$\chi_{n-1,(1-\xi)/2} = \chi_{6, 0.025} = 1.24$$

and

$$RDL = (6 / 1.24)^{½} MDL \approx 7\,s \approx 2\,MDL$$

A-56. Representativeness.

Representativeness refers to the degree to which a sample or set of samples estimates the characteristics of a target population. For the chemical analysis of environmental samples, representativeness is a usually a qualitative parameter that is dependent upon the design of the field sampling program and laboratory methods (e.g., subsampling techniques). An evaluation of *representativeness* would include an assessment of laboratory holding time and method blank data. For example, samples that are not properly preserved or that are analyzed beyond acceptable holding times may not provide representative data.

A-57. Response Factor.

The term "response factor" refers to the *calibration factor* or *relative response factor*. Refer to **internal calibration** and **external calibration**.

A-58. Rinsate Blank.

Equipment or *rinsate blanks* consist of reagent water passed through or over sampling equipment following sample collection and sample equipment decontamination. Contaminated equipment blanks indicate inadequate decontamination between samples and a likelihood of cross-contamination between samples.

A-59. Sample.

The term "sample" refers to non-instrument QC samples (i.e., batch QC and method QC samples) and environmental (field) samples.

A-60. Sensitivity.

Sensitivity refers to the amount of analyte necessary to produce a detector response that can be reliably detected or quantified. Detection limits (e.g., instrument and method) and quantitation limits are commonly used to measure sensitivity.

A-61. Surrogate.

In the context of environmental testing, a surrogate is a relatively pure organic compound which is added to samples prior to preparation and analysis and which is similar to the analytes of interest (in physical and chemical behavior), but which is not normally found in environmental samples. *Surrogates* are typically spiked into environmental samples as well as batch QC and instrument QC samples for chromatographic methods. Surrogate recoveries in environmental samples are primarily used to assess overall performance on a sample-specific basis. Surrogate recoveries for environmental samples measure matrix effects (e.g., and extraction efficiency for

organic analysis involving solvent extractions), are evaluated in a similar manner as matrix spikes, but are evaluated on a sample-specific rather than batch-specific basis.

Surrogate recoveries for instrument QC samples (such as continuing calibration standards) are dependent upon instrument performance. Surrogate recoveries for the LCS and MB are used to evaluate the performance of the preparatory and analytical procedure. Laboratories should maintain surrogate control charts using LCSs or MBs results to monitor method performance and to evaluate surrogate recoveries in actual environmental matrices.

A-62. Target Analyte.

A *target analyte* is an environmental compound or element that is being measured or identified in a chemical test to satisfy project-specific data objectives. Target analytes are distinguished from compounds or elements analyzed solely for the purposes of quality control (e.g., surrogates and internal standards). For brevity, target analytes are often referred to as *analytes*.

A-63. Traceability.

Traceability is formally defined as follows: "The property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties" ("International Vocabulary of basic and general standard terms in Metrology," ISO, Geneva, Switzerland 1993, ISBN 92-67-10175-1).

A-64. Trip Blank.

Trip blanks are prepared from reagent water and accompany each shipment of aqueous samples to be analyzed for volatile organic compounds. Analysis of the trip blanks indicates whether sample cross-contamination occurred during shipment and/or storage.

A-65. Validation.

A number of definitions for the term *validation* are currently being used in the environmental testing industry. In this document, *data validation* or *validation* refers to a systematic review of **comprehensive data packages,** *performed external to the data generator,* with respect to a predefined set of technical performance criteria for precision, accuracy, representativeness, completeness, comparability, and sensitivity. Validation is an objective sample and analyte-specific evaluation process that involves the application of scientific rather than contractual criteria to determine whether requirements for a specific intended use are potentially fulfilled. Validation results in a higher level of confidence when determining whether an analyte is actually present in an environmental sample at a particular level of interest, but usually results in a qualitative evaluation of the data. Data validation occurs prior to determining whether the overall project-specific objectives have been satisfied (i.e., prior to drawing conclusions from the body of the data).

APPENDIX B

Holding Times and Preservation

Table B-1
Qualification for Holding Times

Parameter	Preservative [4]		Holding Time		Flagging for Holding Time Noncompliance	
	Liquid	Solid	Liquid	Solid	Liquid	Solid
VOLATILE ORGANICS	Cool 4°C [12] No head space HCl to pH <2[8]	Cool 4°C, $NaHSO_4$ (aq) to pH < 2	14 d [9]	14 d [10]	Note 1	Note 1
		Cool 4°C Methanol				Note 2
Purgeable Aromatic Hydrocarbons	Cool 4°C No head space HCl to pH <2[8, 12]	Cool 4°C, $NaHSO_4$ (aq) to pH <2	14 d [9]	14 d [10]	Note 1	Note 1
		Cool 4°C Methanol				Note 2
Purgeable Halocarbons	Cool 4°C No head space[8,12]	Cool 4°C, $NaHSO_4$ (aq) to pH <2	14 d	14 d [10]	Note 2	Note 1
		Cool 4°C Methanol				Note 2
SEMIVOLATILE ORGANICS	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Benzidines	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [9]	Note 1	Note 1
Chlorinated Herbicides	Cool 4°C	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Chlorinated Hydrocarbons	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Chlorinated Pesticides	Cool 4°C	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Polychlorinated Dibenzo Dioxins & Furans	Cool 4°C [12]	Cool 4°C	30 d /45 d [6]	30 d/45 d [6]	Note 2	Note 2
Explosives	Cool 4°C	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Haloethers	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Nitroaromatics & cyclic ketones	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Nitrosamines	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
PCBs	Cool 4°C	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2

Parameter	Preservative [4]		Holding Time		Flagging for Holding Time Noncompliance	
	Liquid	Solid	Liquid	Solid	Liquid	Solid
Phenolics	Cool 4°C, H_2SO_4 to pH < 2	Cool 4°C	28 d	28 d	Note 2	Note 2
Phenols	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Phthalate Esters	Cool 4°C	Cool 4°C	7 d/40 d [3]	14 d/40 d [5]	Note 2	Note 2
Polynuclear Aromatic Hydrocarbons	Cool 4°C [12]	Cool 4°C	7 d/40 d [3]	14 d /40 d [5]	Note 2	Note 2
METALS Metals except Cr(IV) and Hg	Cool 4°C, HNO_3 to pH < 2	Cool 4°C	6 months	6 months	Note 2	Note 2
Chromium (VI)	Cool 4°C	Cool 4°C	24 hours	24 hours	Note 1	Note 1
Mercury	Cool 4°C, HNO_3 to pH < 2	Cool 4°C	28 d	28 d	Note 1	Note 1
MISCELLANEOUS Cyanide (Total & Amenable)	Cool 4°C, NaOH to pH >12 [11]	Cool 4°C	14 d	14 d	Note 1	Note 1
TRPH	Cool 4°C, H_2SO_4 to pH < 2	Cool 4°C	28 d	28 d	Note 2	Note 2
Acidity	Cool 4°C	NA	48 hours	NA	Note 1	NA
Alkalinity	Cool 4°C	NA	48 hours	NA	Note 1	NA
Ammonia	Cool 4°C, H_2SO_4 to pH < 2	NA	28 d	NA	Note 1	NA
Biochemical Oxygen Demand (BOD)	Cool 4°C, H_2SO_4 to pH < 2	NA	48 hours	NA	Note 1	NA
Biochemical Oxygen Demand (Carbonaceous)	Cool 4°C	NA	48 hours	NA	Note 1	NA
Bromide	None	NA	28 d	NA	Note 2	NA
Chemical Oxygen Demand (COD)[1]	Cool 4°C, H_2SO_4 to pH< 2	NA	28 d	NA	Note 1	NA
Chloride	None	None	28 d	28 d	Note 2	Note 2
Chlorine, Total Residual	None	NA	A.S.A.P.	NA	NA	NA
Coliform, Fecal & Total	Cool 4°C [12]	Cool 4°C	6 hours	6 hours	Note 1	Note 1
Color	Cool 4°C	NA	48 hours	NA	Note 1	NA
Common Ions	Cool 4°C	Cool 4°C	28 d	28 d	Note 2	Note 2
Dissolved Oxygen, Probe	None Required	NA	A.S.A.P	NA	Note 1	NA
Dissolved Oxygen, Winkler Method	Fix on Site /Store in dark	NA	8 hours	NA	Note 1	NA
Fecal Streptococci	Cool 4°C [12]	NA	6 hours	NA	Note 1	NA
Fluoride	None Required	NA	28 d	NA	Note 2	NA

Parameter	Preservative [4]		Holding Time		Flagging for Holding Time Noncompliance	
	Liquid	Solid	Liquid	Solid	Liquid	Solid
Hardness	HNO_3 or H_2SO_4 to pH < 2	NA	6 months	NA	Note 2	NA
Hydrogen Ion (pH)	None	Cool 4°C	A.S.A.P.	A.S.A.P.	NA	
Kjeldahl & Organic Nitrogen	Cool 4°C, H_2SO_4 to pH < 2	NA	28 d	NA	Note 2	NA
Nitrate	Cool 4°C	Cool 4°C	48 hours	48 hours	Note 1	Note 1
Nitrate-Nitrite	Cool 4°C, H_2SO_4 to pH < 2	Cool 4°C	28 d	28 d	Note 1	Note 1
Nitrite	Cool 4°C	NA	48 hours	NA	Note 1	NA
Oil & Grease	Cool 4°C, H_2SO_4 to pH < 2	Cool 4°C	28 d	28 d	Note 2	Note 2
Organic Carbon	Cool 4°C, HCl or H_2SO_4 to pH < 2	Cool 4°C	28 d	28 d	Note 2	Note 2
Orthophosphate	Filter Immediately Cool 4°C	NA	48 hours	NA	Note 1	NA
Phosphorus (Elemental)	Cool 4°C	NA	48 hours	NA	Note 1	NA
Phosphorus (Total)	Cool 4°C, H_2SO_4 to pH < 2	NA	28 d	NA	Note 1	NA
Gross Alpha	HNO_3 to pH < 2	Cool 4°C	6 months	6 months	Note 2	Note 2
Gross Beta	HNO_3 to pH < 2	Cool 4°C	6 months	6 months	Note 2	Note 2
Radium (Total)	HNO_3 to pH < 2	Cool 4°C	6 months	6 months	Note 2	Note 2
Residue, Filterable	Cool 4°C	NA	7 d	NA	Note 1	NA
Residue, Non-Filterable	Cool 4°C	NA	7 d	NA	Note 1	NA
Residue, Settleable	Cool 4°C	NA	48 hours	NA	Note 1	NA
Residue, Total	Cool 4°C	NA	7 d	NA	Note 1	NA
Residue, Volatile	Cool 4°C	NA	7 d	NA	Note 1	NA
Silica	Cool 4°C	NA	28 d	NA	Note 2	NA
Specific Conductance	Cool 4°C	NA	28 d	NA	Note 2	NA
Sulfate	Cool 4°C	Cool 4°C	28 d	28 d	Note 2	Note 2
Sulfide	Cool 4°C, 4 mL ZnAc plus NaOH to pH > 9	Cool 4°C	7 d	7 d	Note 1	Note 1
Sulfite	Cool 4°C	NA	A.S.A.P.	NA	NA	NA
Surfactant	Cool 4°C	NA	48 hours	NA	Note 1	NA
TCLP Volatile Organics	Cool 4°C	Cool 4°C	14 d/NA/ 14 d [7]	14 d/NA/ 14 d [7]	Note 1	Note 1

Parameter	Preservative [4]		Holding Time		Flagging for Holding Time Noncompliance	
	Liquid	**Solid**	**Liquid**	**Solid**	**Liquid**	**Solid**
TCLP Extractable Organics	Cool 4°C No headspace	Cool 4°C, No Headspace	14 d/7d/ 40 d [7]	14 d/7d/ 40 d [7]	Note 1	Note 1
TCLP Inorganics, Hg	Cool 4°C	Cool 4°C	28 d/NA/ 28 d [7]	28 d/NA /28 d [7]	Note 1	Note 1
TCLP Inorganics, all other metals	Cool 4°C	Cool 4°C	180 d/NA/ 180 d [7]	180 d/NA/ 180 d [7]		
Temperature	None Required	NA	A.S.A.P.	NA	NA	NA
Total Organic Halogens (TOX)	Cool 4°C, H_2SO_4 to pH < 2	28 d	28 d	28 d	Note 2	Note 2
Turbidity	Cool 4°C	NA	48 hours	NA	Note 1	NA

Notes:

1. Qualify all nondetections with the R flag and all detections with the J- flag or X flag when the holding time limit is exceeded.
2. When the holding time limit is exceeded but a gross holding violation does not occur, qualify all nondetections with the UN flag and all detections with the J- flag. When the holding time is grossly exceeded, qualify all nondetections with the R flag and all detections with the J- flag or X flag.
3. 7 days until extraction; analyzed within 40 days of extraction.
4. High concentration samples only require cooling to 4°C.
5. 14 days until extraction; analyzed within 40 days of extraction.
6. 30 days from collection to extraction; 45 days from collection to analysis.
7. Holding times shown are as follows: From collection to TCLP extraction; from TCLP extraction to preparative procedure; from preparative procedure to analysis.
8. All aqueous VOC samples must be protected from light. Project-specific preservatives (e.g., mercuric chloride, copper sulfate, and sodium azide) may be used when matrix interference (e.g., high carbonates) precludes the use of acid preservation.
9. For aqueous aromatic VOCs, the holding time is 7 days if the samples have not been preserved with HCl or a project-specific preservative (but have otherwise been properly preserved).
10. If the samples are stored in an air-tight vessel (e.g., the EnCore sampler) at 4°C without taking additional preservation measures (e.g., methanol immersion and acidification), then the holding time is only 48 hours. Sample results should be qualified for holding time exceedences as described in Note 2.
11. If oxidizing agents are present (e.g.,residual chlorine), add 0.06 g of ascorbic acid or 5 mL of 0.1 N $NaAsO_2$ per liter of sample.
12. Free chlorine must be removed by the addition of 0.008% $Na_2S_2O_3$.

d = day
NA = Not Applicable
A.S.A.P = As soon as possible, immediately