OPTIONS IN DATA VALIDATION: PRINCIPLES FOR CHECKING ANALYTICAL DATA QUALITY

Ms. Shawna Kennedy, Staff Chemist
EcoChem, Inc., 801 Second Avenue, Suite 1401, Seattle, Washington 98104

ABSTRACT

US Environmental Protection Agency (EPA) Contract Laboratory Program National Functional Guidelines for Organic Data Review and EPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review (referred to as Functional Guidelines), along with regional modifications, provide guidance for validation of analytical data. However, these documents were written to accompany data analyzed under EPA Contract Laboratory Program Statement of Work methods (CLP SOW). Because analytical projects often use methods other than CLP SOW, data validation in these situations must rely on a combination of principles found in the applicable Functional Guidelines (with regional modifications, if any), the particular method, and professional judgment.

In addition, data validation can be performed under different levels of effort, from a limited review of reported results to full review of raw data, transcriptions, and calculations. The scrutiny applied to data depends on several factors including data quality objectives, familiarity with the laboratory's quality, project budget, and time constraints. A focused approach of applying limited and full review to subsets of data, as appropriate, can be an effective solution to meeting the requirements of the data, saving time and money as well as satisfying regulatory requirements.

INTRODUCTION

A review of data completeness, laboratory precision, data quality, and error checks can be performed using the principles found in Functional Guidelines, even if the data are not presented in the CLP SOW format. Using these principles, the data validation can be focused to meet the data user's needs. The validation level of effort depends on the data quality objectives, the intended use of the data, and an understanding of how each quality control (QC) element affects the final result. For example, false-negative and false-positive results are of special concern for data used in risk assessment; and compound identification issues are important in polychlorinated biphenyl (PCB) congener analyses, gasoline weathering studies, and other chemical 'fingerprinting' projects. As users of environmental data require analytical results that are more sophisticated and focused to a specific need, data users must also focus the accompanying QC evaluations to meet the specialized concerns.

What Is Data Validation?

Data validation is used to determine if the available project data satisfy the project's data quality objectives and data use requirements. It is the process of comparing laboratory chemistry data against criteria established for the data through an independent review, performed after the laboratory has completed its own in-house quality control checks. Validation determines if the data are acceptable by evaluating, at a minimum, the following categories.

Data package completeness: This step confirms that the laboratory has provided the deliverables required by the contract, method, and/or project plan. During data validation, receipt and completeness of deliverables is checked and documented against the project requirements.

Laboratory performance: Laboratory performance can be evaluated from QC summaries provided by the laboratory. Elements of laboratory performance common to most methods are:

- Holding times (did the laboratory analyze the samples within the required time frame?)
- Calibration (were instruments calibrated at the correct levels and frequencies?)
- Blanks (did the blanks contain target analyses that indicate samples may be contaminated from laboratory procedures?)
- Bias (do laboratory spiking tests show high or low recoveries that may bias associated sample results?)
- Precision (are results reproducible when duplicated?)

- Other quality control (QC) results (did method-specific items meet the QC goals?)

Error checks: Checking for quantitative and qualitative error is performed using supporting instrument and source data (raw data). Data transcriptions of both sample and QC data are reviewed; analyte identifications are evaluated; and quantitation of analyte concentrations are recalculated.

After the validation is completed, qualifiers are assigned to the data points that are affected by QC outliers. Qualifiers indicate to the data user that analyte concentrations may be affected by laboratory or field contamination (in the case of blank contamination), unusable because of QC deficiencies, and/or estimated due to possible bias or reduced confidence in the results.

Functional Guidelines provides guidance for the technical review of data generated using methods found in the EPA Contract Laboratory Program Statement of Work (CLP SOW). Historically, Functional Guidelines has been applied to other methods or protocols but project- or method-specific criteria (such as regional or state requirements) are not specifically covered in Functional Guidelines.

Data Validation Principles From Functional Guidelines

Some examples of laboratory performance principles found in Functional Guidelines that may be applied to methods other than CLP SOW methods are summarized in Table 1.

Table 1: Laboratory Performance Principles

Laboratory Performance Item	**Functional Guidelines General Principle**	**Notes**
Blank Contamination	Qualify data as undetected (U) if concentration in sample is less than five or ten times the blank concentration.	Criteria of five or ten times the blank concentration depends on whether analyte is known as a common laboratory contaminant or not.
Bias • Matrix Spike (pre- preparation) • System Monitoring Compound (surrogate) Spike (post-preparation) • Laboratory Control Sample (blank spike)	If recovery is low (low bias), qualify both positive and not detected results as estimated (J/UJ). If recovery is high (high bias), qualify only positive results as estimated (J). Results that are not detected are not jeopardized by high bias.	Organic and Inorganic Functional Guidelines have different guidance for spike results.
Precision • Matrix Spike/Matrix Spike Duplicate • Laboratory Duplicate • Field Duplicates	If precision is poor, qualify positive results as estimated (J).	Organic and Inorganic Functional Guidelines have different guidance for precision results. Also, data may or may not be qualified based on field duplicates.

Focus And Extent Of Data Validation

Different levels of data validation can be performed using scrutiny ranging from a limited review of reported results to full review of raw data, transcriptions' and calculations. The scrutiny applied to data depends on several factors including data usage, familiarity with the laboratory's quality, and budget and time constraints. A focused approach of applying limited or full review to appropriate subsets of data can be an effective solution for meeting the project data review requirements. The two general levels of validation contain the following QC items and effort levels.

Focused data validation can emphasize efforts in a full review on items above that have the most impact on the data, and apply limited review to remaining items.

Full validation may be used in the following situations:

- When the laboratory quality is unknown to the data user or has a history of errors
- When the data are to be used for litigation purposes
- When the data are to be used for a risk assessment and
- When the project specifies full data validation.

Limited (QC summary forms only)	**Full** (raw data reviewed)
Laboratory performance, including: • Completeness • Chain-of-Custody, • Holding times • Instrument tuning and system performance • Calibration results • QC results reported on summary forms • Detection limits • Other contractual items	Error checks on: • Laboratory performance • Preparation of standards and samples • Analyte identification and quantification from raw data

Limited validation may be used if the above situations do not apply (for example, if the data are from routine monitoring of a known site). Limited validation may also be used in conjunction with full validation to reduce the time and cost of validating large sets of data. If the entire data set receives limited review, a specified percentage of data, data from certain sensitive sampling areas, and/or data that revealed analytical problems during limited review can further receive a full review.

SUMMARY

The needs of data users must be considered when planning data validation for an environmental project. The plan depends on the data quality objectives, intended use of the data, and prioritizing the QC elements affecting the data. Using principles from Functional Guidelines, the extent of data review can be performed using various levels and focus.

REFERENCES

US Environmental Protection Agency. February 1994. *Contract Laboratory Program National Functional Guidelines for Organic Data Review.*

US Environmental Protection Agency. February 1994. *Contract Laboratory Program National Functional Guidelines for Inorganic Data Review.*