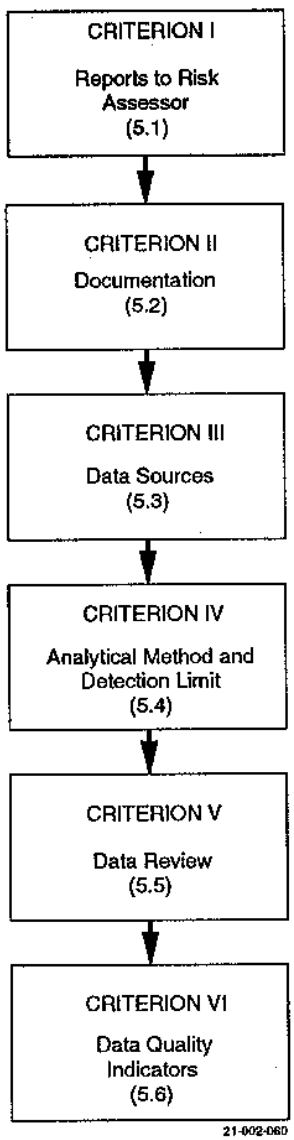
Chapter 5
Assessment of Environmental Data for Useability in Baseline Risk Assessments

This chapter provides guidance for the assessment and interpretation of environmental data for use in baseline human health risk assessments. Ecological risk assessments follow a similar logic but may differ in some details of sampling and analytical methodologies and minimum data requirements. The discussion of data assessment is presented as six steps that define the assessment process for each data useability criterion. Exhibit 60 lists the six criteria in the order that a risk assessor would evaluate them. It also gives references to the sections in this chapter where they are further discussed.

EXHIBIT 60. DATA USEABILITY ASSESSMENT OF CRITERIA



The four basic decisions to be made from data collected in the RI are:

- What contamination is present and at what levels?
- Are site concentrations sufficiently different from background?
- Are all exposure pathways and exposure areas identified and examined?
- Are all exposure areas fully characterized?

The uncertainty associated with each data useability criterion affects the level of confidence associated with each of these decisions.

How to conduct the data assessment. The risk assessor or RPM examines the data, documentation, and reports for each assessment criterion (I - VI) to determine if performance is within the limits specified in the planning objectives. The data assessment process for each criterion should be conducted according to the step-by-step procedures discussed in this chapter. Minimum requirements are listed for each criterion. Potential effects of not meeting the minimum requirements are also discussed and corrective action options are presented. Exhibit 61 summarizes the major impact on assessment if the minimum requirements associated with each data useability criterion have not been met.

Acronyms

CLP	Contract Laboratory Program
CV	coefficient of variation
CRDL	contract required detection limit
CRQL	contract required quantitation limit
DQO	data quality objective
GC	gas chromatography
ICP	inductively coupled plasma
MDL	method detection limit
MS	mass spectrometry
QA	quality assurance
QC	quality control
RAGS	Risk Assessment Guidance for Superfund
RI	remedial investigation
RME	reasonable maximum exposure
RPD	relative percent difference
RPM	remedial project manager
SAP	sampling and analysis plan
SOP	standard operating procedure
SQL	sample quantitation limit

EXHIBIT 61. MINIMUM REQUIREMENTS, IMPACT IF NOT MET, AND CORRECTIVE ACTIONS FOR DATA USEABILITY CRITERIA

Data Useability Criterion	Minimum Requirement	Impact on Risk Assessment if Criterion Not Met	Corrective Action
5.1 Reports to Risk Assessor	• Site description • Sampling design with sample locations • Analytical method and detection limit • Results on per-sample basis, qualified for analytical limitations • Sample quantitation limits and detection limits for non-detects • Field conditions for media and environment • Preliminary reports • Meteorological data • Field reports	• Unable to perform quantitative risk assessment	• Request missing information • Perform qualitative risk assessment
5.2 Documentation	• Sample results related to geographic location (chain-of-custody records, SOPs, field and analytical records)	• Unable to assess exposure pathways • Unable to identify appropriate concentration for exposure areas	• Request locations identified • Resampling
5.3 Data Sources	• Analytical data results for one sample per medium per exposure pathway • Broad spectrum analysis for one sample per medium per exposure pathway • Field measurements data for media and environment	• Potential for false negatives or false positives • Increased variability in exposure modeling	• Resampling or reanalysis for critical samples
5.4 Analytical Method and Detection Limit	• Routine (federally documented) methods used to analyze chemicals of potential concern in critical samples	• Unquantified precision and accuracy • False negatives	• Reanalysis • Resampling or reanalysis for critical samples • Documented statements of limitation for non-critical samples
5.5 Data Review	• Defined level of data review for all data	• Potential for false negatives or false positives • Increased variability and bias due to analytical process, calculation errors or transcription errors	• Perform data review
5.6 Data Quality Indicators	• Sampling variability quantified for each analyte • QC samples to identify and quantify precision and accuracy • Sampling and analytical precision and accuracy quantified	• Unable to quantify confidence levels for uncertainty • Potential for false negatives or false positives	• Resampling for critical samples • Perform qualitative risk assessment • Perform quantitative risk assessment for non-critical samples with documented discussion of potential limitations

21-002-061

The following activities should be performed for each assessment criterion:

- Identify or determine performance objectives and minimum data requirements.

  Quantitative or qualitative performance objectives should be specified in the sampling and analysis plan for all components of the acquisition of environmental data (as discussed in Chapter 4). The first step in assessing each criterion is to assemble these performance objectives and note any changes. Performance objectives should also be compared with the minimum acceptable requirements for data useability presented in this chapter. These minimum requirements can be adopted as performance objectives if objectives were not specified. For example, the requirement that there must be a broad spectrum analysis for at least one sample in each medium for each exposure area would be a performance objective, if performance were not specified during planning.

- Determine actual performance compared to performance objectives.

  The next step in the assessment of each criterion is to examine results to determine the performance that was achieved for each data useability criterion. This performance should then be compared with the objectives established during planning. Take particular note of performance for samples or analyses that are critical to the baseline risk assessment. All deviations from the objectives should be noted. In those cases where performance was better than that required in the objective, it may be useful for assessment of future activities to determine if this is due to unanticipated characteristics of the site or to superior performance in some stage of the data acquisition. Corrective action is the next step where performance does not meet performance objectives for data critical to the risk assessment.

- Determine and execute any corrective action required.

  ☛ *Focus corrective action on maximizing the useability of data from critical samples.*

Corrective action should be taken to improve data useability when performance fails to meet objectives for data critical to the risk assessment. Corrective action options are described in Exhibit 62. These options require communication among the risk assessor, the RPM, and the technical team. Sensitivity analysis may be performed by the risk assessor to estimate the effects of not meeting performance requirements given the certainty of the risk assessment. Corrective actions may improve data quality and reduce uncertainty, and may eliminate the need to qualify or reject data.

EXHIBIT 62. CORRECTIVE ACTION OPTIONS WHEN DATA DO NOT MEET PERFORMANCE OBJECTIVES

- Retrieve missing information.
- Resolve technical or procedural problems by requesting additional explanation or clarification from the technical team.
- Request reanalysis of sample(s) from extract.
- Request construction and re-interpretation of analytical results from the laboratory or the project chemist.
- Request additional sample collection and analysis for site or background characterization.
- Model potential impact on risk assessment uncertainty using sensitivity analysis to determine range of effect.
- Adjust or impute data based on approved default options and imputation routines.
- Qualify or reject data for use in risk assessment.

21-002-062

Using a worksheet to organize the data assessment. The level of certainty associated with the data component of risk assessment depends on the amount of data that meet performance objectives. The risk assessor determines whether the data for each performance measure are satisfactory (data accepted), questionable (data qualified) or unsatisfactory (data rejected). The worksheet provided in this chapter may be used as a guide or organizational tool.

Use the Data Useability Worksheet, Exhibit 63, to document data assessment decisions. Record the decision as accepted, accepted with qualification, or rejected for use in the risk assessment for each data

EXHIBIT 63. DATA USEABILITY WORKSHEET

Data Useability Criterion		Decision*	Comments
I	Reports to Risk Assessor		
II	Documentation A. Work Plan/SAP/QAPjP		
	B. SOPs		
	C. Field and Analytical Records		
III	Data Sources A. Analytical		
	B. Non-analytical		
IV	Analytical Methods		
V	Data Review		

* Decision: Accept, Qualified Accept, Reject

21-002-063

EXHIBIT 63. DATA USEABILITY WORKSHEET (Cont'd)

Data Useability Criterion		Decision*		Comments
VI	Data Quality Indicators A. Completeness	Sampling		
		Analytical		
		Combined		
	B. Comparability	Sampling		
		Analytical		
		Combined		
	C. Representativeness	Sampling		
		Analytical		
		Combined		
	D. Precision	Sampling		
		Analytical		
		Combined		
	E. Accuracy	Sampling		
		Analytical		
		Combined		

* Decision: Accept, Qualified Accept, Reject

21-002-063-01

useability criterion. Outline the justification for each decision in the comments section.

The remainder of this chapter explains how to assess data using the data useability criteria. Assessment of Criterion I involves identifying the data and documentation required for risk assessment (Section 5.1). Assessment of Criteria II through V examines available data and results in terms of the assessment of data useability criteria for documentation (Section 5.2), data sources (Section 5.3), analytical method and detection limit (Section 5.4), and data review (Section 5.5). Criterion VI includes the assessment of sampling and analytical performance (Section 5.6) according to five data quality indicators: completeness, comparability, representativeness, precision, and accuracy.

5.1 ASSESSMENT OF CRITERION I: REPORTS TO RISK ASSESSOR

Minimum Requirements

- Site description.
- Sampling design with sample locations, related to site-specific data needs and data quality objectives.
- Analytical method and detection limit.
- Results on per-sample basis qualified for analytical limitations.
- Sample quantitation limits and detection limits for non-detects.
- Field conditions for media and environment.
- Preliminary reports.
- Meteorological data.
- Field reports.

Data and documentation supplied to the risk assessor must be evaluated for completeness and appropriateness, and to determine if any changes were made to the work plan or the sampling and analysis plan (SAP) during the course of the work. The SAP discusses the sampling and analytical design and contains the quality assurance project plan and data quality objectives (DQOs), if they have been developed. The risk assessor should receive preliminary and final data reports, as described in the following sections.

5.1.1 Preliminary Reports

☛ *Use preliminary data as a basis for identifying sampling or analysis deficiencies and taking corrective action.*

Preliminary analytical data reports allow the risk assessor to begin assessment as soon as the sampling and analysis effort has begun. These initial reports have three functions:

- The risk assessor can begin to characterize the baseline risk assessment on the basis of actual data. Chemicals of interest will be identified and the variability in concentration can be estimated.
- Potential problems in sampling or analysis can be identified and the need for corrective action can be assessed. For example, additional samples may be required, or the method may need to be modified because of matrix interferences.
- RI schedules are more likely to be met if the risk assessment process can begin before the final data reports are produced.

The major advantage of preliminary review of data by the risk assessor is the potential for feedback and corrective action while the RI is still in process. This can improve the quality of data for risk assessment.

5.1.2 Final Report

☛ *Problems in data useability due to sampling usually can affect all chemicals involved in the risk assessment; problems due to analysis may only affect specific chemicals.*

The minimum data reports and documentation needed to prepare the risk assessment are:

- A description of the site, including a detailed map showing the location of each sample, surrounding structures, terrain features, receptor populations, indications of air and water flow, and a description of the operative industrial process (if any),
- A description and rationale for the sampling design and sampling procedures,
- A description of the analytical methods used,
- Results for each analyte and each sample, qualified for analytical limitations, and a full description of all deviations from SOPs, SAPs, and QA plans,
- Sample quantitation limits (SQLs) and detection limits for undetected analytes, with an explanation of the detection limits reported and any qualifications,
- A narrative explanation of the level of data review used and the resulting data qualifiers. The narrative should indicate the direction of bias, based on the assessment of the results from QC samples (e.g., blanks and field and laboratory spikes), and
- A description of field conditions and physical parameter data as appropriate for the media involved in the exposure assessment.

It may not be possible to perform a quantitative baseline risk assessment if any of these materials are not available and cannot be obtained. The RPM or risk assessor should attempt to retrieve missing deliverables from the source.

Additional reports and data that are useful to the risk assessor, such as data results on Contract Laboratory Program (CLP) diskettes, are listed in Exhibit 19. Access

to this information can improve the efficiency and quality of the risk assessment. However, not having access does not necessarily require the data to be qualified or rejected. Minimum requirements for reports to the risk assessor are listed in Exhibit 61.

5.2 ASSESSMENT OF CRITERION II: DOCUMENTATION

Minimum Requirements

- Sample results related to geographic location (chain-of-custody records, SOPs, field and analytical records).

Three types of documentation must be assessed: chain-of-custody records, SOPs, and field and analytical records. Chain-of-custody records for risk assessment must document the sample locations and the date of sampling so that sample results can be related to geographic location and specific sample containers. If a sample result cannot be related to a sampling date and the point of sample collection, the results are unuseable for quantitative risk assessment. Full scale chain-of-custody procedures (from sample collection through analysis) are required for enforcement or cost recovery.

SOPs describe and specify the procedures to be followed during sampling and analysis. They are QA procedures that increase the probability that a data collection design will be properly implemented. SOPs also increase consistency in performing tasks and, as a result, determine the level of systematic error and reduce the random error associated with sampling and analysis. Knowledge that SOPs were developed and followed increases confidence that the quality of data can be determined, and the level of certainty in risk assessment can be established. The existence of SOPs for each process or activity involved in data collection is not a minimum requirement, but SOPs can be useful if data problems occur, particularly in assessing the comparability of data sets.

Field and analytical records document the procedures followed and the conditions of the procedures. Field and analytical records, such as field logs and raw instrument output, may be useful to the risk assessor as back-up documentation, but they are not minimum requirements. QC data from blanks, spikes, duplicates, replicates, and standards should also be accessible, in either raw or summary formats, to support qualitative or quantitative assessments of the analytical results. Like SOPs, such records are critical to resolving problems in interpretation, but they may not directly affect the level of certainty of the risk assessment. Minimum requirements for documentation are listed in Exhibit 61.

5.3 ASSESSMENT OF CRITERION III: DATA SOURCES

Minimum Requirements

- Analytical sample data results for each medium within an exposure area.
- Broad spectrum analysis for one sample per medium per exposure area.
- Field measurements data for media and environment.

Data source assessment involves the evaluation and use of historical and current analytical data. Historical analytical data should be evaluated according to data quality indicators and not source (e.g., analytical protocols may have changed significantly over time).

The minimum analytical data requirement for risk assessment is that results are produced for each medium within an exposure area using a broad spectrum analytical technique, such as GC-MS methods for organic analytes or ICP for inorganic analytes. The useability of data will almost always increase as more broad spectrum analyses are performed for each exposure area. The absence of a broad spectrum analysis from a fixed laboratory results in an increased probability of false negatives; all chemicals of potential concern at the site may not be identified. In the absence of a broad spectrum analysis, the best corrective action is to take additional samples. If additional samples cannot be obtained, the probability of false negatives and false positives should be considered high, and the level of certainty of the risk assessment is decreased.

The broad spectrum analysis, and any other analytical data, are subject to the basic documentation and data review requirements discussed in this chapter. The location of the sample data point must be known, as well as the method and SQL achieved for analytical results. Guidance for the assessment of analytical data to determine false positives and false negatives and the precision and accuracy of concentration results is provided in Section 5.6.1.

Field measurements of physical characteristics of the site, medium, or contamination source are a critical data source, whose omission can significantly affect the ability of the risk assessor to perform a quantitative assessment. Physical site information is also required to perform exposure fate and transport modeling. Examples

of such data are particle size, pH, clay content and porosity of soils, wind direction and speed, topography, and percent vegetation. RAGS, Part A, Exhibit 4-2, "Examples of Modeling Parameters for Which Information May Need to be Obtained During a Site Sampling Investigation," (EPA 1989a) provides a list of data elements according to medium modeling category. These measurements must be collected during sampling. The use of default options and routines to estimate missing values allows the use of the model but increases the uncertainty associated with the exposure assessments.

5.4 ASSESSMENT OF CRITERION IV: ANALYTICAL METHOD AND DETECTION LIMIT

Minimum Requirements

- Routine (federally documented) methods used to analyze chemicals of potential concern in critical samples.

The risk assessor compares SQLs or method detection limits (MDLs) with analyte-specific results to determine their consequence given the concentration of concern. Assessment of preliminary data reports provides an opportunity to review the detection limits early and resolve any problems. When a chemical of potential concern is reported as not detected, the result can only be used with confidence if the quantitation limits reported are lower than the corresponding concentration of concern. The minimum recommended requirement is that the MDL be no more than 20% of the concentration of concern, so that the SQL will also be below the concentration of concern. Chemicals identified above this ratio of detection limit to concentration of concern can be used with good confidence. For example, if the concentration of concern for arsenic in groundwater is 70 ug/L for an average daily consumption of 2 L of water by a 70 kg adult, the detection limit of a suitable method for examination of groundwater samples from such a site should be no greater than 14 ug/L. Minimum requirements for analytical methods and detection limits are listed in Exhibit 61.

If the concentration of concern is less than or equal to the detection limit, and the chemical of concern is not detected, do not use zero in the calculation of the concentration term. When the MDL reported for an analyte is near to the concentration of concern, the confidence in both identification and quantitation may be low. This is illustrated in Exhibit 64. Information concerning non-detects or detections at or near detection limits should be qualified according to the degree of acceptable uncertainty, as described in Section 5.6.1.

The concentration of concern for ecological risk may be different than the concentration of concern for human health risk. In addition, aquatic life criteria should be examined to determine if they are based on ecological or human health risk.

5.5 ASSESSMENT OF CRITERION V: DATA REVIEW

Minimum Requirements

- Defined level of data review for all data.

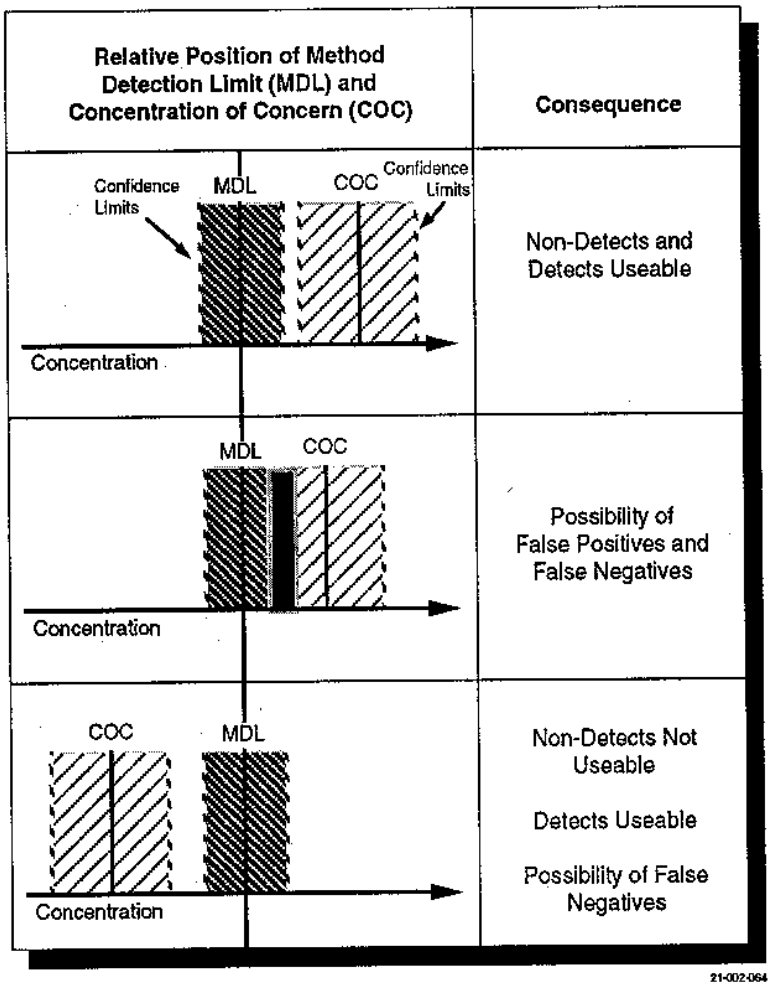
Data review assesses the quality of analytical results and is performed by a professional with a knowledge of the analytical procedures. The requirement for risk assessment is that only data that have been reviewed according to a specified level or plan will be used in the quantitative risk assessment. Any analytical errors, or limitations in data that are identified by the review, must be noted in the risk assessment if the data are used. An explanation for qualifiers used must be included with the review report.

All data should receive some level of review. The risk assessor may receive data prior to the quantitative baseline risk assessment that were not reviewed. Data that have not been reviewed must be identified because the lack of review increases the uncertainty for the risk assessment. These data may lead to false positive or false negative assessments and quantitation errors. Unreviewed data may also contain transcription errors and calculation errors. Data may be used in the preliminary assessment before review, but must be reviewed at a predetermined level before use in the final risk assessment.

Depending upon data user requirements, the level and depth of the data review are variable. The level and depth of the data review may be determined during the planning process and must include an examination of laboratory and method performance for the samples and analytes involved. This examination includes:

- Evaluation of data completeness,
- Verification of instrument calibration,
- Measurement of laboratory precision using duplicates; measurement of laboratory accuracy using spikes,
- Examination of blanks for contamination,

EXHIBIT 64. RELATIVE IMPORTANCE OF DETECTION LIMIT AND CONCENTRATION OF CONCERN: DATA ASSESSMENT



- Assessment of adherence to method specifications and QC limits, and
- Evaluation of method performance in the sample matrix.

Specific data review procedures are dependent upon the method and data user requirements. Section 5.6.1 details procedures for evaluating QC samples for laboratory and method performance. CLP data review procedures are performed according to criteria outlined in *National Functional Guidelines for Organic Data Review* (EPA 1991e) and *Laboratory Data Validation: Functional Guidelines for Evaluating Inorganics Analyses* (EPA 1988e). Minimum requirements for data review are listed in Exhibit 61.

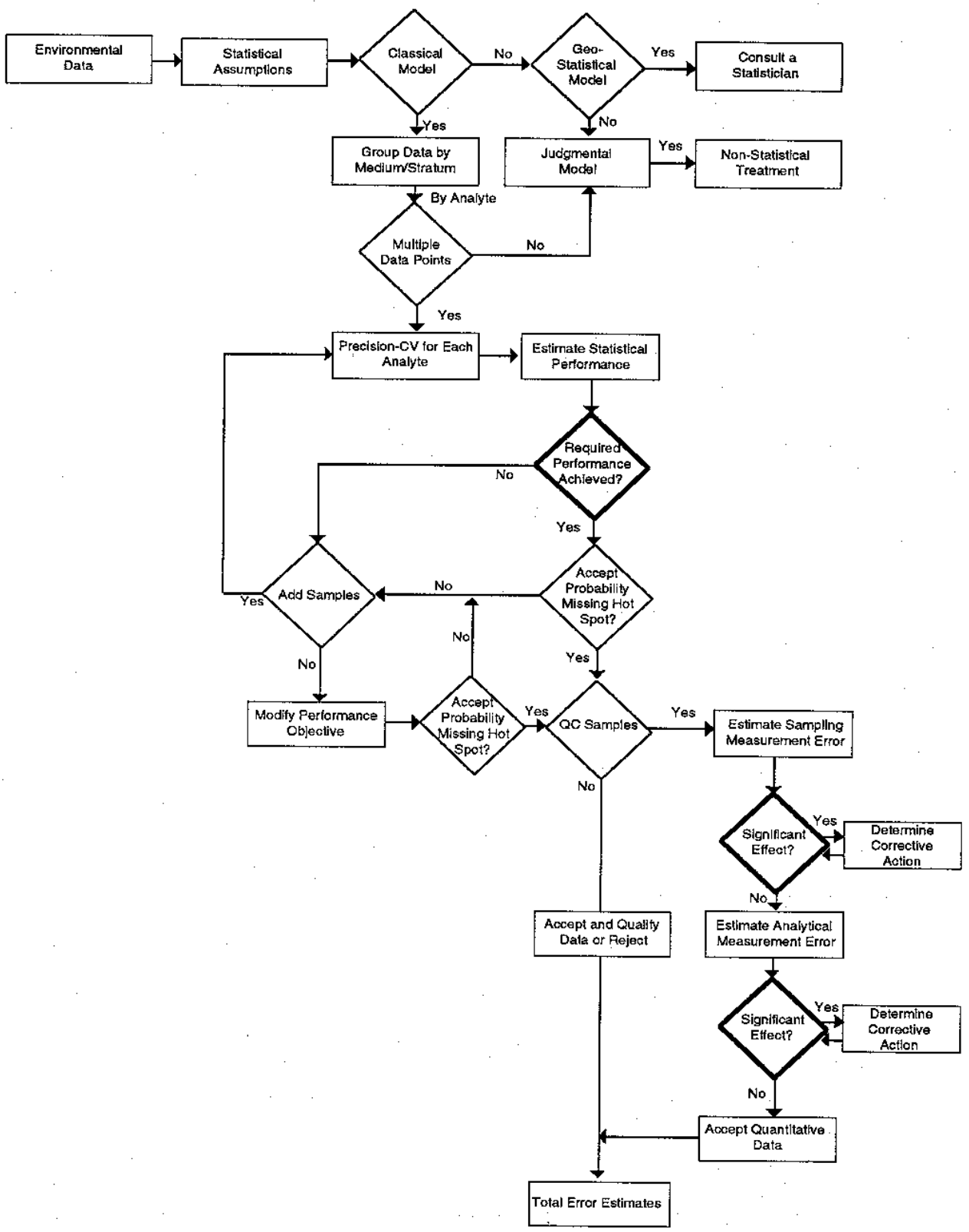
5.6 ASSESSMENT OF CRITERION VI: DATA QUALITY INDICATORS

Minimum Requirements

- Sampling variability quantitated for each analyte.
- QC samples required to identify and quantitate precision and accuracy.
- Sampling and analytical precision and accuracy quantitated.

The assessment of data quality indicators presented in this chapter is significant to determine data useability.

EXHIBIT 65. CONSEQUENCES OF ALTERNATIVE SAMPLING STRATEGIES ON TOTAL ERROR ESTIMATE



☛ *Qualified data can usually be used for quantitative risk assessments.*

The assessment of data quality indicators for either sampling or analysis involves the evaluation of five indicators: completeness, comparability, representativeness, precision, and accuracy. Uncertainties in completeness, comparability, and representativeness increase the probability of false negatives and false positives when the data are used to test particular hypotheses as part of the site evaluation. This increase in uncertainty can affect the confidence of chemical identification. Variation in completeness, comparability, representativeness, precision, and accuracy affects the uncertainty of estimates of average concentration and reasonable maximum exposure (RME). Once the indicator is examined or a numerical value is determined, the results can be compared to the performance objectives established during RI planning. This comparison determines the useability of the data and any required corrective actions.

A summary of the minimum requirements for data quality indicators is presented in Exhibit 61, and the evaluation process is illustrated in Exhibit 65. Specific requirements for each indicator are presented in the following sections.

5.6.1 Assessment of Sampling and Analytical Data Quality Indicators

The major activity in determining the useability of data based on sampling is assessing the effectiveness of the sampling operations performed. Samples provided for analysis must answer the four basic decisions to be made with RI data in risk assessment (cited at the beginning of this chapter) that are translated into site-specific objectives based on scoping and planning decisions.

Independent data review evaluates laboratory results, not sampling. Determining the useability of analytical results begins with the review of QC samples and qualifiers to assess analytical performance of the laboratory and the method. It is more important to evaluate the effect on the data than to determine the source of the error. The data package is reviewed as a whole for some criteria; data are reviewed at the sample level for other criteria, such as holding time. Factors affecting the accuracy of identification and the precision and accuracy of quantitation of individual chemicals, such as calibration and recoveries, must be examined analyte-by-analyte. The qualifiers used in the review of CLP data are presented and their effect on data quality is discussed in this section. Exhibit 66 presents a

EXHIBIT 66. USE OF QUALITY CONTROL DATA FOR RISK ASSESSMENT

Quality Control Criterion	Effect on Identification When Criterion is not Met	Quantitative Bias	Use
Spikes (High Recovery)	--	High	Use data as upper limit.
Spikes (Low Recovery)	False Negative[1]	Low	Use data as lower limit.
Duplicates	None, unless analyte found in one duplicate and not the other. Then either false positive or false negative.	High or Low[2]	Use data as estimate--poor precision.
Blanks	False Positive	High	Set confidence level 5x blank. Use data above confidence level. Use data below confidence level as estimate.
Calibration	--	High or Low[2]	Use data as estimate unless problem is extreme.
Tune	False Negative	--	Reject data or examine raw data and use professional judgment.
Internal Standards (Reproducibility)[3]	--	--	Use data as estimate--poor precision.
Internal Standards (High Recovery)	--	Low	Use data as lower limit.
Internal Standards (Low Recovery)	False Negative[1]	High	Use data as upper limit.

[1] False negative only likely if recovery is near zero.
[2] Effect on bias determined by examination of data for each individual analyte.
[3] Includes surrogates and system monitoring compounds.

21-002-066

summary of the QC samples and the data use implications of qualified data. Corrective action options are shown in Exhibit 62.

Sample media can be more complex than expected in environmental analysis. For example, sludge or oily wastes may contain interfering chemicals whose presence cannot be predicted in precision and accuracy measurements. The risk assessor must examine the reported precision [relative percent difference (RPD)] and accuracy [percent recovery (%R)] data to determine useability. Ranges used for rejection and qualification of CLP data have been determined based on the analysis of target compounds in environmental media. These ranges, documented in the Functional Guidelines (EPA 1991e, EPA 1988e) can be used in the absence of specifications in the planning documents.

Completeness. Completeness for sampling is calculated by the following formula:

$$\text{Percent Completeness} = \frac{\text{(Number of Acceptable Data Points)} \times 100}{\text{Total Number of Samples Collected}}$$

This measure of completeness is useful for data collection and analysis management but misses the key risk assessment issue, which is the total number of data points available and acceptable for each chemical of potential concern. Incompleteness should be assessed to determine if an acceptable level of data useability can still be obtained or whether the level of completeness must be increased, either by further sampling or by other corrective action. Any decrease in the number of samples from that specified in the sampling design will affect the final results. In this case, the option of obtaining more samples should be reviewed.

Minimum Requirements for Completeness	Impact When Minimum Requirements Are Not Met	Corrective Action
• Percentage of sample completeness determined during planning to meet specified performance measures. • 100% of all data for analytes in critical samples (at least one sample per medium per exposure area). • All data from critical samples considered crucial. Background samples and broad spectrum analyses are usually critical.	• Higher probability of false negatives. • Reduction in confidence level and power. • A reduction in the number of samples reduces site coverage and may affect representativeness. Data for critical samples have significantly more impact than incomplete data for non-critical samples. • Useability of data is decreased for critical samples. • Useability of data is potentially decreased for non-critical samples. • Reduced ability to differentiate site levels from background. • Impact of incompleteness generally decreases as the number of samples increases.	• Resampling or reanalysis to fill data gaps. • Additional analysis of samples already at laboratory. • Determine whether the missing data are crucial to the risk assessment (i.e., data from critical samples).

21-002-081

Typical causes for sample attrition include site conditions preventing sample collection (e.g., a well runs dry), sample breakage, and invalid or unuseable analytical results. Incompleteness can increase the uncertainty involved in risk assessments by reducing the available number of samples on which identification and estimates of concentration of chemicals at the site are based. The reduction in the number of samples from the original design further affects representativeness by reducing site coverage and increases the variability in concentration estimates. Only the collection of additional samples will resolve the problem, unless the samples involved were duplicates or splits. In this case, or if the cause was laboratory performance, the extracts may be considered for reanalysis.

Completeness for analytical data is calculated by the following formula:

$$\text{Percent Completeness} = \frac{\text{(Number of Acceptable Samples)} \times 100}{\text{Total Number of Samples Analyzed}}$$

The completeness for analytical data required for risk assessment is defined as the number of chemical-specific data results for an exposure area in an operable unit that are determined acceptable after data review.

An analysis is considered complete if all data generated are determined to be acceptable measurements as defined in the SAP. Results for each analyte should be present for each sample. In addition, data from QC samples necessary to determine precision and accuracy should be present. QC samples and the effects of problems associated with these samples are discussed later in this section.

Comparability. Comparability is not compromised provided that the sampling design is unbiased, and the sampling design or analytical methods have not changed over time. If any of these factors change, the risk assessor may experience difficulties in combining data sets to estimate the RME. The determination of the RME is based on the principal of estimating risk over time for the exposure area. The ideal situation occurs when samples can be added within the basic design, decreasing the level of uncertainty.

> ☛ *Anticipate the need to combine data from different sampling events and/or different analytical methods.*

Comparability is a very important qualitative data indicator for analytical assessment and is a critical

Minimum Requirements for Comparability	Impact When Minimum Requirements Are Not Met	Corrective Action
• Unbiased sampling design or documented reasons for selecting another sampling design. • The analytical methods used must have common analytical parameters. • Same units of measure used in reporting. • Similar detection limits. • Equivalent sample preparation techniques.	• Non-additivity of sample results. • Reduced confidence, power, and ability to detect differences, given the number of samples available. • Increased overall error.	For Sampling: • Statistical analysis of effects of bias. For Analytical Data: • Preferentially use those data that provide the most definitive identification and quantitation of the chemicals of potential concern. For organic chemical identification, GC-MS data are preferred over GC data generated with other detectors. For quantitation, examine the precision and accuracy data along with the reported detection limits. • Reanalysis using comparable methods.

21-002-082

parameter when considering the combination of data sets from different analyses for the same chemicals of potential concern. The assessment of data quality indicators determines if analytical results being reported are equivalent to data obtained from similar analyses. Only comparable data sets can readily be combined for the purpose of generating a single risk assessment calculation.

The use of routine methods simplifies the determination of comparability because all laboratories use the same standardized procedures and reporting parameters. In other cases, the risk assessor may have to consult with an analytical chemist to evaluate whether different methods are sufficiently comparable to combine data sets. The RPM should request complete descriptions of non-routine methods. A preliminary assessment can be made by comparing the analytes, useful range, and detection limit of the methods. If different units of measure have been reported, all measurements must be converted to a common set of units before comparison.

Representativeness. Representativeness of data is critical to risk assessments. The results of the risk assessment will be biased to the degree that the data do not reflect the chemicals and concentrations present in the exposure area or unit of interest. Non-representative chemical identification may result in false negatives. Non-representative estimates of concentration levels may be higher or lower than the true concentration. Non-representative sampling can usually only be resolved by additional sampling, unless the potential limitations of the risk assessment are acceptable.

It is important to determine whether any changes have occurred in the actual sample collection that convert an originally unbiased sampling plan into a biased sampling episode. Bias in unbiased designs is difficult to assess because no measure of the true value is known. Bias is assumed in non-statistical designs.

Representativeness is primarily a planning concern. The solution is in the design of a sampling plan that is representative. Once the design is implemented, only the sampling variability is evaluated during the assessment process, unless contamination occurs in the QC samples or blanks, or problems exist during sample preparation that affect sample results. Incompleteness of data potentially decreases representativeness and increases the potential for false negatives and the bias in estimations of concentration.

Representativeness is determined by examining the sampling plan, as discussed in Section 3.2. In determining the representativeness of the data, the evaluator examines the degree to which the data meet the performance standards of the method and to which the analysis represents the sample submitted to the laboratory. Analytical data quality affects representativeness since data of low quality may be rejected for use in risk assessments. Holding time, sample preservation, extraction procedures, and results

Minimum Requirements for Representativeness	Impact When Minimum Requirements Are Not Met	Corrective Action
• Sample data representative of exposure area and operable units. • Documented sample preparation procedures. Filtering, compositing, and sample preservation may affect representativeness. • Documented analytical data as specified in the SAP.	• Bias high or low in estimate of RME. • Increased likelihood of false negatives. • Inaccurate identification or estimate of concentration that leads to inaccurate calculation of risk. • Remaining data may no longer sufficiently represent the site if a large portion of the data are rejected, or if all data from analyses of samples at a specific location are rejected.	• Additional sampling. • Examination of effects of sample preparation procedures. • For critical samples, reanalyses of samples or resampling of the affected site areas. For non-critical samples, reanalyses or resampling should be decided by the RPM in consultation with the technical team. • If the resampling or reanalyses cannot be performed, document in the site assessment report what areas of the site are not represented due to poor quality of analytical data.

21-002-083

from analyses of blanks affect the representativeness of analytical data (see Appendix V).

Precision. The two basic activities performed in the assessment of precision are estimating sampling variability from the observed spatial variation and estimating the measurement error attributable to the data collection process. Assumptions concerning the sampling design and data distributions must be examined prior to interpreting the results. This examination will provide the basis for selecting calculation formulas and knowing when statistical consultation is required.

The type of sampling design selected is critical to the estimation of sampling variability as discussed in Sections 3.2 and 4.1. If the sampling design is judgmental, the nature of the sampling error cannot be determined and estimates of the average concentrations of analytes may not be representative of the site.

☛ *Determine the distribution of the data before applying statistical measures.*

The nature of the observed chemical data distribution affects estimation procedures. The estimation of variability and confidence intervals will become complex if the distribution cannot be assumed normal or to approximate normal when transformed to log normal. Estimates of the 95% upper confidence limit of the average concentration for the RME should be based on an analysis of the frequency distribution of the data whenever the database is sufficient to support such analysis. Statistical tests may be used to compare the distribution of the observed data with the normal or log normal distribution (Gilbert 1987). Graphs of data without statistical test results may also be acceptable for some data sets. Statistical computer software can assist in the analyses of data distribution.

Sampling variability. Exhibit 67 summarizes the assessment procedures for the evaluation of variability from different sampling procedures. The estimation of confidence levels, power, and minimum detectable relative differences requires assumptions about the coefficients of variation from sampling variability for

Minimum Requirements for Precision	Impact When Minimum Requirements Are Not Met	Corrective Action
• Confidence level of 80% (or as specified in DQOs). • Power of 90% (or as specified in DQOs). • Minimum detectable relative differences specified in SAP and modified after analysis of background samples if necessary. • One set of field duplicates or more as specified in the SAP. • Analytical duplicates and splits as specified in the SAP. • Measurement error specified.	• Errors in decisions to act or not act based on analytical data. • Unacceptable level of uncertainty. • Increased variability of quantitative results. • False negatives for measurements near the detection limits.	For Sampling: • Add samples based on information from available data that are known to be representative. • Adjust performance objectives. For Analysis: • Analysis of new duplicate samples. • Review laboratory protocols to ensure comparability. • Use precision measurements to determine confidence limits for the effects on the data. • The risk assessor can use the maximum sample results to set an upper bound on the uncertainty in the risk assessment if there is too much variability in the analyses.

21-002-084

EXHIBIT 67. STEPS TO ASSESS SAMPLING PERFORMANCE

1. Confirm statistical assumptions.

2. Summarize analyte detection data by strata: media within site or site subgroups and strata within media.

3. Transform analyte concentration data so distribution is approximately normal.

4. Calculate the coefficient of variation for each analyte detected.

5. Using Exhibit 47 "Relationships Between Measures of Statistical Performance and Number of Samples Required," look up the range of power, confidence level and minimal detectable relative differences for the calculated coefficient of variation.

6. Compare the statistical performance measures required to those achievable given the coefficient of variation and sample size.

7. If the performance objectives are achieved, go to Step 9.

   If the required statistical performance levels are not met, then additional samples must be taken or one or more of the performance parameters must be changed.

   If samples are to be added, Exhibit 47 and the calculation formulas in Appendix IV can be used to determine the number needed.

8. If the performance parameters are to be changed, the parameter to be changed should be the one which will increase the probability of taking unnecessary action as opposed to unnecessary risk.

9. Examine the results of the QC samples. Sample results must be considered to be qualitative if no results are available for QC samples.

10. If the QC sample results indicate possible bias through contamination, take appropriate corrective action.

21-002-067

each chemical of potential concern. The RPM or risk assessor should discuss the implications of these assumptions with a statistician to determine their potential impacts on data useability.

> ☛ *Determine the statistical measures of performance most applicable to site conditions before assessing data useability.*

Once the statistical assumptions and observed analyte variability are known, selected statistical performance measures can be assessed to determine the data quality achieved. Additional samples may be needed, or modified DQOs required, as a result of evaluating sampling variability. Three issues are involved in the assessment of required statistical performance:

- Level of certainty or confidence,
- Power, and
- Minimum detectable relative difference.

The required level for each of these performance measures should be included in the SAP as DQOs. The user's data quality requirements defined by these statistical measures determine the number of samples that are taken during data collection. Recommended minimum statistical performance parameters for

discriminating contaminant concentrations from background levels in risk assessment are provided in Exhibit 68.

EXHIBIT 68. RECOMMENDED MINIMUM STATISTICAL PERFORMANCE PARAMETERS FOR RISK ASSESSMENT

Null Hypothesis: On-site Contaminant Concentrations are not Higher than the Background
• Confidence level: [1] 80% minimum, reject null when true (take unnecessary action).
• Power: [2] 90% minimum, accept null when false (fail to take action when action is required).
• Minimum detectable relative difference: 10% - 20%, usually depends on concentration of concern.
1 (1-false positive estimate) or (1 -α).
2 (1-false negative estimate) or (1 -β).
Source: EPA 1989f.

21-002-068

First, summarize the sample results at the analyte level by stratum and strata within media to determine whether the performance objectives have been met. Sampling error is not relevant if a particular combination of stratum and analyte yields only a single data point. In that case, assessment proceeds to that of analytical error for that stratum and analyte combination.

The distribution for stratum and analyte combinations with multiple data points should usually be examined for normality and transformed to log normal. The coefficient of variation is calculated for each stratum and analyte combination. If the distribution resulting from the transformation is not normal, a new distributional model will need to be identified and validated in consultation with a statistician. Non-parametric procedures which require no distributional assumptions may also be used.

Conversely, the statistical performance achieved can be determined, given the coefficient of variation. This performance should be compared to the requirements stated in planning. If the performance objectives are achieved, the risk assessor can proceed to the assessment of measurement error.

If the required statistical performance objectives are not met, additional samples must be taken, or one (or more) of the performance parameters must be changed. If samples are added, the tables and formulas provided in Chapter 4 and Appendix IV can be used to calculate the number of samples required. If a performance parameter is changed, it should be the one that will increase the probability of taking unnecessary action as opposed to an increased probability of unnecessary risk. The uncertainty level will then be reduced first, the minimum detectable relative difference will be increased second, and the level of power will be reduced last. Minimum recommended levels for performance parameters in risk assessment in the absence of site-specific DQOs are 80% confidence levels, 90% power, and 10-20% minimum detectable relative differences (EPA 1989f). Exhibit 68 summarizes the recommended DQOs for statistical performance parameters.

Measurement error. Measurement error is estimated using the results of field duplicate samples. Field duplicates determine total within-batch measurement error, including analytical error if the samples are also analyzed as laboratory duplicates. The estimate is of the difference between analytical values reported for duplicates. This type of variation has four basic sources: sample collection procedures, sample handling and storage procedures, analytical procedures, and data processing procedures.

The formula for computing the relative percent difference between duplicates is:

$$RPD = \frac{|R_1 - R_2|}{(R_1 + R_2)/2} \times 100$$

where R_1 and R_2 are the results from the first and second duplicate samples, respectively. Precision is a measure of the repeatability of a single measurement and is evaluated from the results of duplicate samples and splits.

Low precision can be caused by poor instrument performance, inconsistent application of method protocols, or by a difficult, heterogeneous sample matrix. The last effect can be distinguished from the others by evaluation of laboratory QC data.

If split samples have been analyzed by different methods or different laboratories, then data users have a measure of the quality of individual techniques. Splits are particularly effective when one laboratory is a reference laboratory. If both sets of data exhibit the same problems, then laboratory performance can usually be ruled out as a source of error. Splits are also useful when using non-routine methods or comparing results from different analytical methods.

Accuracy. Accuracy is a measure of overestimation or underestimation of reported concentrations and is evaluated from the results of spiked samples. The procedure for determining accuracy will vary according to differences in the number of measurements and the precision of the estimates. Data that are not reported with confidence limits cannot be assigned weights based on precision and should not be combined for use (Taylor 1987).

Spiked samples are particularly useful in the analysis of complex sample types because they help the reviewer determine the extent of bias on the sample measurement. A set of standards at known concentrations is mixed into a portion of the sample or into distilled water prior to sample preparation and analysis. The analytical results are compared to the amount spiked to determine the level of recovery. It is important to note that unless every sample is spiked, spike recoveries indicate only a trend rather than a specific quantitative measure.

Accuracy is controlled primarily by the analytical process and is reported as bias. The absolute bias of a sampling design cannot be determined unambiguously because the true value of the chemicals of concern in the exposure area can never be known. However, statistically based sampling designs described in Chapter 4 are structured to produce unbiased results.

Bias can be estimated using field spikes on field evaluation or audit samples to assess the accuracy and comparability of results. These estimates will reflect the effects of sample collection, handling, holding time, and the analytical process on the result for the sample collected.

Bias is estimated for the measurement process by computing the percent recovery (%R) for the spiked or reference compound as follows:

$$\%R = \frac{\text{(Measured Amount - Amount in Unspiked Sample)} \times 100}{\text{Amount Spiked}}$$

Because of the inherent problems associated with the spiking procedure and the interpretation of recovery, spikes are considered minimum requirements only if specified in the SAP. Field matrix spikes are currently not recommended for use in soils (EPA 1989f).

Field blanks are evaluated to estimate the potential bias caused by contamination from sample collection, preparation, shipping and/or storage. Results for the analysis of field blanks indicate whether contamination resulted in bias, but they are not estimates of accuracy. Bias pertaining to analytical recoveries is computed as follows:

$$\text{Percent Bias} = \frac{\text{(Measured Amount-Amount in Unspiked Sample)} \times 100}{\text{Amount Spiked}}$$

Minimum Requirements for Accuracy	Impact When Minimum Requirements Are Not Met	Corrective Action
• Field spikes to assess accuracy of non-detects and positive sample results if specified in the SAP. • Analytical spikes as specified in the SAP. • Use analytical methods (routine methods whenever possible) that specify expected or required recovery ranges using spikes or other QC measures. • No chemicals of potential concern detected in the blanks.	• Increased potential for false negatives. If spike recovery is low, it is probable that the method or analysis is biased low for that analyte and values of all related samples may underestimate the actual concentration. • Increased potential for false positives. If spike recovery exceeds 100%, interferences may be present, and it is probable that the method or analysis is biased high. Analytical results overestimate the true concentration of the spiked analyte.	• Consider resampling at affected locations. • No correction factor is applied to CLP data on the basis of the percent recovery in calculating the analyte concentration. • If recoveries are extremely low or extremely high, the risk assessor should consult with an analytical chemist to identify a more appropriate method for reanalysis of the samples.

21-002-085

Blanks are of primary concern for the analysis of bias involved in sampling because of the difficulty in performing field spikes and the availability of appropriate reference standards and matrix for evaluation samples.

Results from blanks can be used to estimate the extent of high bias in the event of contamination. The following procedures should be implemented to prevent the assignment of false positive values due to blank contamination:

- If the field blanks are contaminated and the laboratory blanks are not, the RPM or risk assessor can conclude that contamination occurred prior to receipt of the samples by the laboratory. If the contamination is significant (i.e., it will interfere with the determination of risk), consider resampling at affected locations.
- If it is not possible to resample, the RPM or risk assessor must assess the effect of the contamination on the potential for false positives. Often, this determination can be made by examining data from samples located nearby. If all samples and blanks show the same level of a particular chemical, the presence of the chemical in the samples is most likely due to contamination.
- If the laboratory blanks are contaminated, the laboratory should be required to rerun the associated analyses. This is especially important in the case of critical analytes or samples. Before reanalyses, the laboratory must demonstrate freedom from contamination by providing results of a clean laboratory blank. Note: If laboratory blanks are contaminated, field blanks will generally also be contaminated.
- If reanalysis is not possible, then the sample data must be qualified. The Functional Guidelines provide examples of blank qualification. Chemicals detected in the associated samples below the action level defined in the Functional Guidelines are considered undetected.

Data qualifiers. All data generated by the routine analytical services of the CLP are reviewed and qualified by Regional representatives according to the guidelines found in the Functional Guidelines as modified to fit the requirements of the individual Regions.

> ☛ *Use data qualified as U or J for risk assessment purposes.*

Analytes qualified with a U are considered "not detected." If precision and accuracy are acceptable (as determined by the QC samples), data are entered in the data summary tables in the data validation report as the SQL or corrected quantitation limit (MDL corrected for dilution and percent moisture), and qualified with a U. Note that the same chemical can be reported undetected in a series of samples at different concentrations because of sample differences.

Data qualified with an R are rejected because performance requirements in the sample or in associated QC analyses were not met. For example, if a mass spectrometer "tune" is not within specifications, neither the identification nor quantitation of chemicals can be accepted with confidence. Extremely low recoveries of a chemical in a spiked sample might also warrant an R designation for that chemical in associated samples because of the risk of false negatives (see Appendix VI).

Data qualified with a J present a more complex issue. J-qualified data are considered "estimated" because quantitation in the sample or in associated QC samples did not meet specifications. The justification for qualifying the data should be explained in the validation report. Draft revisions of the Functional Guidelines propose that the justification be included on a qualifier summary table submitted with the validation report.

Data can be biased high or low when qualified as estimated. The bias can often be determined by examining the results of the QC samples. For example, if interfering levels of aluminum are found in inorganic analysis of the interference check sample, the sample results are probably biased high because the signal overlap is added to the signal being reported. When volatile organic compounds are qualified J for holding time violations, the results are usually biased low because some of the volatile compounds may have volatilized during storage.

Data associated with contaminated blanks are not considered estimated and are not flagged J. The presence of the blank contaminant chemical in the analytical samples is questionable at levels up to 5 to 10 times those found in the blank, depending on the nature of the analyte. An action level is determined for each chemical based on the quantity found in the blank. Data above the action level are accepted without qualification and data between the contract required quantitation limit (CRQL) and the action level are qualified U (undetected).

Estimated organics and inorganics data that are below the CRQL or contract required detection limit (CRDL) are qualified as UJ. This qualifier signifies that the quantitation limit is estimated because the QC results did not meet criteria specified in the SAP.

Other qualifiers may be added to the analytical data by the laboratory. A set of qualifiers (or flags) has been defined by the CLP for use by the laboratories to denote

problems with the analytical data. These qualifiers and their potential use in risk assessment are discussed in RAGS (EPA 1989a).

5.6.2 Combining the Assessment of Sampling and Analysis

Once the quality of the sampling and analysis effort has been assessed using the five data quality indicators, combine the results to determine the overall assessment of a particular indicator across sampling and analysis. Combining the assessment for completeness, comparability, and representativeness is discussed in this section as a qualitative procedure. Statistical models are available for combining data sets with different variability and bias. The risk assessor should consult a chemist or statistician if the magnitude of the sampling and analysis effort warrants the use of a formal statistical treatment of comparability.

The basic model for estimating total variability across sampling and analysis components is presented in Exhibit 69. An example of a non-statistical approach to combining the assessment results is given in Exhibit 70. Using this approach, each data quality indicator is assessed to determine whether a problem exists in either sampling or analysis. This assessment leads to different combinations of problem determination. For example, completeness may have been a problem in sampling [YES] but not a problem in analysis [NO]; the combination is [YES/NO].

Basic guidance is given on the combinations of sampling and analysis once assessment patterns based on the determination of a problem have been established. This guidance is qualitative in nature and is presented to assist in organizing the data assessment problem for the application of professional judgment. If the assessment pattern is [NO/NO], the issue of combining results is not a problem. Conversely, if the pattern is [YES/YES], the issue of combining results is an issue of the effects of the combined magnitudes. Instances of combined sampling and analysis problems for a single indicator will have significant effects on the risk assessment uncertainty. The most complicated assessment pattern to interpret is encountered when a problem occurs in one area but not in another (e.g., in sampling but not in analysis). This situation is briefly discussed for each indicator in the following sections.

EXHIBIT 69. BASIC MODEL FOR ESTIMATING TOTAL VARIABILITY ACROSS SAMPLING AND ANALYSIS COMPONENTS

$$\sigma_t^2 = \sigma_m^2 + \sigma_p^2$$

where σ_t = total variability
σ_m = measurement variability
σ_p = population variability

$$\sigma_m^2 = \sigma_s^2 + \sigma_h^2 + \sigma_{ss}^2 + \sigma_a^2 + \sigma_b^2$$

where σ_s = sampling variability (standard deviation)
σ_h = handling, transportation and storage variability
σ_{ss} = preparation variability (subsampling variability)
σ_a = laboratory analytical variability
σ_b = between batch variability

NOTE: It is assumed that the data are normally distributed or that a normalizing data transformation has been performed.

Source: EPA 1990c.

21-002-069

EXHIBIT 70. COMBINING DATA QUALITY INDICATORS FROM SAMPLING AND ANALYSIS INTO A SINGLE ASSESSMENT OF UNCERTAINTY

Data Quality Indicators	Assessment of Problems		Combined Sampling and Analytical Determination *
	Sampling	Analytical	
Completeness	YES NO	YES NO	YES/YES YES/NO NO/YES
Comparability	YES NO	YES NO	YES/YES YES/NO NO/YES
Representativeness	YES NO	YES NO	YES/YES YES/NO NO/YES
Precision	YES NO	YES NO	YES/YES YES/NO NO/YES
Accuracy	YES NO	YES NO	YES/YES YES/NO NO/YES

* The combination [NO/NO] indicates that the data quality indicator will not affect the level of uncertainty in data useability.

21-002-070

Completeness. A sample is considered incomplete for all analytes. Analytical incompleteness is usually related to particular analytes. In this instance [YES/YES], the effect on the risk assessment will vary according to chemical. For some chemicals, the data points will be lost in both sampling and analysis.

The effects of a loss in the number of sample points for a particular chemical can be substantial. For example, if collection of 10 samples was planned and one sample could not be collected because of site access problems, one was broken in transport, and the laboratory experienced analysis problems with three samples for the chemical of potential concern causing the data to be rejected, then only five data points remain.

If the assessment pattern is [YES/NO], the effects are distributed across all chemicals involved in the risk assessment. If the pattern is [NO/YES], the effects are localized to the particular chemical affected.

Comparability. Comparability problems in sampling are primarily due to different sampling designs and time periods. Seasonal variations are treated like spatial variations because the risk assessment is calculated as risk over time. Data can be averaged and considered as a single data set. For analytical data, comparability problems are related primarily to the use of different methods and laboratories. A pattern of [YES/YES] will indicate that the risk assessor will have considerable difficulty in combining the various data sets into a single assessment of risk. In situations of [YES/NO] or [NO/YES], the problem of sampling comparability is more difficult to resolve. Models exist for determining comparability between methods and integrating results across laboratories. These models involve the general statistical approach to confirming data sets with different but known variability and bias (Taylor 1987).

Representativeness. Representativeness in sampling is critical to the risk assessment. Non-representativeness affects both false negatives (chemicals not identified) and estimates of concentration and, therefore, affects estimates of RME. Analytical representativeness involves the question of whether the analytical results represent the sample collected. For example, holding times and sample preservation can cause the analytical results not to be representative of the sample collected. These questions should be treated separately in the discussion of effects.

Precision. The contribution to imprecision from sampling variability often exceeds that from analytical variability in the measurement process. If precision is a problem in both sampling and analysis, the risk assessor should focus on the impact of sampling variability on the estimate of RME. Analytical variability will be minimal in comparison to the effects of sampling variability unless the sampling variability is untypically low and the analytical variability is untypically high.

Accuracy. The assessment of accuracy in sampling is focused primarily on recoveries from spiked or performance evaluation samples. Analytical performance and potential blank contamination are reflected in analytical spike recoveries. If the pattern is [YES/YES] for accuracy, this may require assessment of calibration, or of potential blank contaminants, and integration of their possible effects by comparison of results from laboratory and field QC samples.

If the accuracy pattern is [NO/YES], then the issue is analytical performance. Low variability in sampling as measured by low coefficients of variation for chemicals of potential concern should increase the risk assessor's concern over an analytical accuracy problem.

High sampling variability (CV>25%) will greatly reduce the effects of analytical bias on the level of certainty of the risk assessment.