Data Quality Evaluation update
(draft updates to 2016 HEER Office TGM Section 4.2.7.3)

Draft updates to Section 4.2.7.3 of the 2016 edition of the Hawaii DU-MIS guidance (HIDOH April 2020)

X.1 Data Quality Evaluation

X.1.1 Review of Sample Collection and Processing Methods

Data verification is a completeness check that all specified activities involved in data collection and processing have been completed and documented and that the necessary records (objective evidence) are available to proceed to data validation. For example, if the sampling design called for Multi Increment (MI) samples to be prepared by combing 50 increments of soil from a targeted DU but only 30 increments were taken, this would be documented during the data verification evaluation.

The quality of the sample data generated must be reviewed to determine if the data are reliable to answer the risk and/or remediation-based questions prepared at the beginning of the project. This requires a review the sampling plan design and the methods used to collect the samples. The precision and reproducibility of the data generated must also be reviewed.

A checklist summary of each topic is provided in Table X-1. The table is not intended to comprehensive for all aspects of the investigation and should be modified as appropriate on a site-specific basis. Refer to the noted sections of this guidance document and related appendices for detailed information on each topic. Deviations from the recommended methods should be discussed in the investigation report and resulting limitations of the data collected described and considered in the report recommendations. Methods to help minimize data error when the sample collection and analysis conditions noted in Table X-1 cannot be met are discussed in the associated appendices.

Table X-1. Sample data quality and usability checklist.

Acceptable?	Site Investigation Stage
	Conceptual Site Model and DU Designation
	• Site history and potential sources and type of contamination well understood?
	• Site investigation questions used to designate Decision Units for testing clearly stated and based on risk and/or optimization of anticipated remediation requirements?
	• Questions and decision statements developed for individual Decision Units presented?
	• Area and total volume of soil associated with each Decision Unit noted and acceptable for intended purposes?
	• To-scale map depicting location and size of Decision Units provided?
	Field Sample Collection
	• Summary of sample collection methods provided, including approximate final mass of each sample?
	• Multi Increment samples prepared by collecting and combining a minimum of increments appropriate to chemical present and nature of contamination?
	• Increments appropriately spaced and collected (Section?
	• Complete, unobstructed access to all portions of the DU soil available for sample collection?
	• Core-shaped increments collected?
	• Samples to be tested for volatile chemicals preserved in methanol in the field or met requirements for alternative preservation and testing methods?
	• Minimum sample mass of 2-3 kilograms met (minimum 300 grams for samples to be tested for volatile contaminants)?
	• Triplicate Multi Increment Samples collected in at least 10% of Decision Units (minimum 1 set) to test total data precision?
	Laboratory Processing and Testing
	• Samples to be tested for non-volatile chemicals air-dried and sieved to target particle size for each specific Decision Unit?
	• Analytical subsample collected using a sectoral splitter or manually collected from at least 30 points?
	• Minimum 30-gram analytical subsample mass extracted for <2mm particle size soil?
	• Minimum 10-gram analytical subsample mass extracted for <250µm particle size soil?
	• Triplicate analytical subsamples collected from at least 10% of samples submitted (minimum 1 set)?
	• Holding times met?
	• Analytical quality control and quality assessment criteria met (e.g., spikes, blanks, etc.; refer also to USEPA 2002)?
	Replicate Sample Collection and Data Precision Evaluation
	• Replicate field sample and laboratory subsample data meet data precision requirements?

	• Source of error for replicate data that exceed an RSD of 35% determined?
	• Laboratory subsampling error identified and subsamples recollected after grinding of primary sample or larger subsample mass collected?
	• Data adjusted or new samples collected for DUs with replicate data that exceed an RSD of 50%?

X.1.2 Review of Replicate Data Precision

The total precision of MIS sample data is evaluated based on a comparison of data for replicate samples collected from the same Decision Unit. Replicate sample data can only be used to evaluate the total precision of the overall sample collection and testing method. The term "precision" is different from the term "accuracy." Precision describes the reproducibility of the overall sampling method. The accuracy of the data with respect to the true mean concentration of the contaminant in the subject Decision Unit area and volume of soil can only be known by extracting the chemical from the entire volume of soil and measuring the mass.

This is routinely done in mining operations (e.g., extraction of gold from crushed ore) but not as part of most environmental investigation and remediation projects, although error in sample data can sometimes be estimated as part of an in situ remediation project. The true error in the data therefore can never be determined. The potential for significant error in environmental can, however, be assess based on a review of how the samples were collected, processed and tested (perhaps the most important step) and a review of the precision of replicate sample data sets.

Statistical evaluation of replicate sample data involves a two-step procedure. The first step is to calculate the relative standard deviation (RSD) of the contaminant concentration for the data set. The RSD reflects the precision of the total sampling method, including combined field and laboratory error. The lower the RSD, the more precise the sampling method used and the more reproducible and reliable the data for individual DU where replicate samples were not collected.

As summarized in Table X-2, an RSD for replicate sample data $\leq$ 35% suggests that the sampling method has good reproducibility and, assuming the samples were properly collected and processed, the data can be used for reliable decision making. An RSD >35% but $\leq$50% indicates less reliable but in most cases still acceptable for decision making, given the typical safety factor built into risk-based action levels. An RSD >50% but <100% indicates poor data precision and the need to either retest affected DUs using samples with a greater number of increments and total, bulk mass or, if deemed acceptable by a risk assessor, use the RSD to upwardly adjust data for DUs where replicate samples were not collected to reflect a hypothetical, "maximum" concentration of the contaminant for. An RSD >100% indicates very poor data precision and the likely need to resample the affected DUs.

Review replicate subsample data from the laboratory to determine if laboratory error appears to account for most of the total error in the sample data. Note that high RSDs can become unavoidable as contaminant concentrations approach the laboratory method reporting and detection limits. Consultation with a risk assessor trained in Multi Increment sampling methods is required to determine if the collection of additional samples is necessary. Replicate sample RSDs also typically increase as the magnitude of contamination increases. Sample data that significantly exceed target action levels is generally acceptable for decision making even though the RSD of the replicate data indicate very poor precision.

The collection of a minimum of 50 increments per sample and a minimum, bulk sample mass of 1-2kg is normally reliable to achieve a replicate sample RSD of <35%. The collection of 2-3 kg samples is, however, recommended for soil that might contain high-concentration nuggets of contamination. Examples include soil impacted with lead shot, chips of lead-based paint and PCBs in the form of tarry balls or fragments of caulking or sealants.

Table X-2. Recommendations for assessment of data quality based on the relative standard deviation of replicate samples.

Replicate Sample Data Precision	**Use of DU Data for Decision Making**
Good (RSD≤35%)	• Data for DUs where replicate samples were not collected can be assumed to be representative without adjustment; • Compare unadjusted MIS data directly with target action values (use arithmetic mean of replicate sample data). • Collection of followup, confirmation samples for DUs where remedial action is necessary *not required* if data for Boundary DUs meet target action levels.
Moderate (35%<RSD≤50%)	• Data for DUs where replicate samples were not collected have lower confidence but are adequate for comparison to action levels or use in a risk assessment without adjustment; • Review and discuss sampling methods and laboratory processing and analysis methods and summarize potential sources of error in reports for future reference (e.g., inadequate increment collection methods, insufficient number of increments, inadequate laboratory processing, etc); • Compare unadjusted MIS data directly with target action values (use the arithmetic mean of replicate sample data); • Collection of followup, more reliable confirmation samples for DUs where remedial action is necessary *required* even if data for Boundary DUs meet target action levels (e.g., number of increments and total sample mass increased; laboratory processing steps improved, etc.).
Poor (50%<RSD≤100%)	• Data for DUs where replicate samples were not collected are not reliably representative of the DU mean; • Review and discuss field sampling methods and laboratory processing and summarize potential sources of error in reports for future reference; • If the majority of the total error is due to subsampling or (less likely) analysis in the laboratory, require the laboratory to reprocess and retest the samples, including milling of samples if necessary, with additional replicate subsamples collected and tested to reassess precision; • If replicate sample data precision is still poor, consider retesting affected DUs using samples with a greater number of increments and total, bulk mass; OR, If determined acceptable by a risk assessor trained in Multi Increment sampling methods: • For DUs with replicate sample data, compare of the highest reported concentration of the contaminant to the action or cleanup level;

	• For DUs without replicate sample data, adjust the reported contaminant concentration upwards by the RSD calculated for the DU with replicate sample data; • Additional evidence of data acceptance (or rejection) should be provided for decision-making purposes, including site history and potential for contamination above the level of concern, adequacy of methods used in collecting, processing and analyzing samples, closeness of data to action levels and safety margins built into the action levels, and other information as available and pertinent. • Collection of additional confirmation sampling in DUs where remedial action is necessary required, using samples with a greater number of increments and total, bulk mass and the collection of replicate samples.
Very Poor (RSD>100%)	• Data for all DUs are not reliably representative of the DU mean, including data for DUs where replicate samples were collected; • If the majority of the total error is due to subsampling or (less likely) analysis in the laboratory, require the laboratory to reprocess and retest the samples, including milling of samples if necessary, with additional replicate subsamples collected and tested to reassess precision; • Review and discuss field sampling methods and laboratory processing and analysis methods and summarize potential sources of error in reports for future reference; • Retesting is not required for DUs where the need for remediation is already clear from the data and other field evidence. • Consider the collection of new samples in DUs using the following approach: a) If known, designate suspected source areas as separate DUs for individual characterization, b) Collect a minimum of 75 increments per sample; c) Ensure a minimum, 2-3 kg final sample mass; d) Collect replicate samples in all anticipated high-concentration and high-risk DUs; • As an alternative, consult with a risk assessor trained in Multi Increment sampling methods regarding the safety level incorporated into the target action level or cleanup level and the need to resample high exposure risk areas (e.g., all sample data an order of magnitude or more below action levels). • Additional evidence of data acceptance (or rejection) should be provided for decision-making purposes, including site history and potential for contamination above the level of concern, adequacy of methods used in collecting, processing and analyzing samples, closeness of data to action levels and safety margins built into the action levels, and other information as available and pertinent. • Collect replicate confirmation samples in all DUs requiring remediation.

X.2.1.3 Consideration of 95% UCLs

Basis and Comparability of 95% UCLs for Discrete Sample Data

The direct comparison of unadjusted data as described above for a properly collected, MI sample is acceptable for decision making provided that sample quality DQOs are met. While a 95% UCL could in practice be calculated for a set of replicate. MI samples, this would be unrelated to use of a 95% UCL for a single set of discrete samples. A 95% UCL is calculated for a single set of discrete sample data in order to address uncertainty in estimation of the mean due to variability between individual data points. This is appropriate, given the higher potential for error in discrete sample data as described earlier in this document.

Under an MI sampling approach, this uncertainty is addressed through the preparation of a single sample that meets minimum increment number, total mass and processing and testing requirements with respect to Gy's sampling theory for particulate matter. Decades of experience in the mining industry had demonstrated that this approach provides a far superior and reliable estimate of the mean than traditional, discrete sampling methods (Pitard 2019).

The collection and evaluation of replicate, MI sample data to assess the precision of the overall sampling method represents an additional step in the data quality evaluation process not included in traditional, discrete sample investigations. A comparable test would require comparison of 95% UCLs based on the collection and comparison of multiple, replicate sets of discrete sample data from the same DU. This of course is never done. The range of potential (and unavoidable) error associated with the 95% UCL calculated for a single set of discrete sample data is therefore always unknown (refer to Brewer et al. 2017b). The presence of this hidden error in a 95% UCL calculated for a single set of discrete sample data is highlighted by the sometimes-high variability in data for replicate, MI samples. Multi Increment sampling methods identify and as needed address this error; discrete sampling methods do not.

Practitioners of Gy's sampling theory in the mining industry routinely use unadjusted, ISM-type data for decision making if the Relative Standard Deviation of replicate sample data is less than 35%. This supports the objective that data generated using the sampling method likely follow a normal distribution, as intended. Resampling is generally called for when the RSD exceeds 35% due to the potential error in the data. A 95% UCL is never calculated or used, since this defeats the purpose of Gy's sampling theory and can lead to erroneous decision making. This in part reflects the high level of precision and data accuracy required for marketing of commodities such as gold or iron in crushed ore. Differences in the actual mass of the commodity extracted from the ore of just a few percent could bankrupt a company.

Applicability and Calculation of 95% UCLs for ISM Data

The calculation of a 95% UCL for replicate, MI sample data is not an integral part of Gy's Theory of Sampling and was strongly discouraged in conversations with Francis Pitard and a group of international, sampling statisticians during the World Conference on Sampling and Blending in Beijing, China, in 2018 (Pitard 2018, personal communication; see also Pitard 2019). Environmental action levels normally include a

significant margin of safety, often up to an order of magnitude or more. Relatively small error in estimation of mean contaminant concentration based on data for a single MI sample is therefore normally acceptable, provided that the sample was properly collected, processed and tested.

Although not routinely required by the HEER Office, some risk assessors may nonetheless prefer the use of a 95% UCL calculated from replicate MI sample data in order to document overall data precision and as an added measure of confidence that the true mean of the DU does not exceed a targeted action level or risk. Examples include action levels for contaminants that include only a minimal safety margin and the need to more conservatively address risk in anticipated high-exposure areas. This and the specific statistical test(s) to be used to calculate a 95% UCL should be discussed with the HEER Office project manager at the beginning of systematic planning process and incorporated into decision statements for individual DUs. A recommendation by the risk assessor for the collection of replicate samples and use of a 95% UCL for comparison to action levels or direct estimation of risk is likely to be applicable to only a small subset of the DUs associated with a given project.