DAVID Y. IGE
GOVERNOR OF HAWAII

LIBBY CHAR, MD
ACTING DIRECTOR OF HEALTH

STATE OF HAWAI'I
DEPARTMENT OF HEALTH
P. O. BOX 3378
HONOLULU, HI 96801-3378

In reply, please refer to:
File: RB-249-2020

Date: November 23, 2020

To: Interested Parties

Through: Fenix Grange, Supervisor
Hazard Evaluation and Emergency Response Office

From: Roger Brewer, PhD, Senior Environmental Scientist
Hazard Evaluation and Emergency Response

Subject: Critique and Use of ITRC *Incremental Sampling Methodology* Guidance Document in Hawai´i

This technical memorandum presents a critique of the 2020 document entitled *Incremental Sampling Methodology* (ISM) published by the Interstate Technical and Regulatory Council (ITRC 2020). The document, which serves as an update to a 2012 edition with the same title, describes the causes behind long-known limitations of "discrete" sampling methods traditionally used in the environmental industry for characterization of soil, sediment and related particulate media. ISM-type sampling methods introduced in the document were specifically developed to address shortcomings of discrete sampling approaches and provide a more reliable and risk-based approach to site characterization.

The "ISM" investigation and sampling methodology described in the ITRC document is similar in concept to "Decision Unit (DU)" and Multi Increment Sample (MIS)" site investigation methods (DU-MIS) published in the Hawai´i Department of Health (HIDOH), Hazard Evaluation and Emergency Response (HEER) Office *Technical Guidance Manual* (TGM; HIDOH 2016 and updates). Research into DU-MIS investigation methods was initiated in Hawai´i in 2004. Initial guidance was published in 2009 (referenced in the 2012 ITRC ISM document) and most recently updated in 2016. Additional updates are anticipated in 2021. The guidance reflects the field experience of both consultants and HEER Office staff over 15 years at hundreds of diverse sites and the collection of tens of thousands of "MI" samples.

A summary of comments is provided below. Overall, the discussions of discrete sample data unreliability and the implementation of ISM-type sampling methods in field and laboratory are well done and serve as a useful introduction to Gy's Theory of Sampling. Unfortunately, discussions of "statistical concepts," "data evaluation" and "computer simulations" in subsequent sections of the document, prepared by a separate work group, contain numerous omissions, misstatements and outright errors and for the most part should be avoided.

Detailed comments are provided in the attachment to this memorandum. **It is important that environmental investigations carried out in Hawai´i be based on DU-MIS guidance**

presented in the HEER Office *Technical Guidance Manual*. The term "Multi Increment®" sample, in that format, should be used to describe the samples collected (Multi Increment is a trademark of Envirostat, Inc.). The term "ISM" or reference to the ITRC document should not be used in reports, since this could call into question the familiarity of the preparer with guidance presented in the HEER Office TGM. As described below, while the introduction, field implementation and laboratory processing sections of the ITRC ISM document are in general well-presented and comparable to the HEER Office DU-MIS guidance, significant errors in discussions of the statistical basis, evaluation and reliability of ISM-type sampling methods and data negate use of the document as a stand-alone reference for work in this state.

Summary of Comments

Field and Laboratory Implementation

Section 1 (Introduction) of the ITRC document provides a useful overview of the nature of contaminant distribution in soil and the basis of often hidden error associated with reliance on discrete soil data. Although less detailed than the Hawai´i guidance, Section 2 (Nature of Soil Sampling), Section 4 (Field Implementation) and Section 5 (Laboratory Processing) of the ITRC document provide a good, general overview of the implementation of ISM-type sampling methods in the field and at the laboratory. A reasonably strong, basic understanding of Gy's Theory of Sampling by the authors as well as experience in the field and laboratory is evident in these sections.

Readers are likely, however, to find the interchangeable use of the term "Sampling Unit" for more specific "Source Area DU" and "Boundary Area DU" throughout these and other sections of the document to be somewhat confusing. Use of the additional term "Sampling Unit" is unnecessary in Hawaii's experience and could lead to the collection of unreliable sample data.

Furthermore, a premature recommendation for the default collection of only 30-increment samples is made in some parts of the document. This is likely to lead to failed replicate sample data under many site scenarios in Hawaii's experience. The default number of increments recommended was increased from 30 to 50 in the 2016 update of the HEER Office TGM, based on discussions with consultants and experience in the field by HEER Office staff. The collection of 75-increment samples is recommended for testing of soils with high-concentration "nuggets" of contaminants. As few as 30 increments per sample might be adequate for water-applied contaminants, but should be verified with replicate sample data.

Statistical Concepts, Data Evaluation and Use in Risk Assessments

Variability in the range of training and experience within the ITRC ISM update group is more stark and problematic in Section 3.2 (Statistical Concepts), Section 3.3 (Planning for the Use of ISM Data), Section 6 (Data Quality Evaluation) and Section 8 (ISM for Risk Assessment) of the ITRC document. In spite of the titles, concise and critical discussions of the statistical underpinnings and decades of field research of ISM-type sampling methods developed by the mining industry are completely lacking in these sections. The authors instead erroneously attempt to discuss the nature and hypothetical precision of ISM-type sampling methods in terms of classic, "inferential" statistics utilized in the past by statisticians and risk assessors to interpret discrete sample data – the same methods that Gy's Theory of Sampling demonstrated to be unreliable and not applicable to particulate matter such as soil and sediment. Offers to put the authors in touch with internationally known sampling statisticians and leading researchers in

Gy's Theory of Sampling after reviewing early drafts of the ISM update were unfortunately ignored.

Of particular concern, and what could unfortunately serve as a "poison pill" for reference to the document in general, is the erroneous, implied requirement to collect replicate ISM samples in every DU (at significant added cost and effort) and utilize a 95% UCL for comparison to screening levels and assessment of risk, as is traditionally done with discrete sample data. As discussed in the attachment, the authors fail to point out that data for a single, properly collected and tested, ISM-type sample are specifically intended to replace the use of a 95% UCL calculated for a single set of discrete sample data. Both methods are utilized to capture and represent heterogeneity of contaminant distribution within a targeted area.

The first method relies on calculation of a mean and 95% UCL from a single set of discrete samples, with little to no consideration of the sample collection and testing method. Such methods have never been demonstrated in detailed field studies to generate consistently reliable and accurate data (see Brewer et al. 2020a,b). ISM-type sampling methods, with a statistical basis in Gy's Theory of Sampling and decades of field research (see Pitard 2019), focus on the careful collection of a single sample that meets minimum increment number and bulk mass requirements and strict processing requirements at the laboratory. In addition, replicate ISM samples are collected from a subset of DUs in order to periodically test the precision of the overall sampling method – a key test of data reliability that is markedly absent in discrete sampling guidance and the reason that the unreliability of discrete sample data has been largely overlooked in the environmental industry for the past 30+ years.

Calculation and use of a 95% UCL based on collection of independent, replicate ISM samples from the same Decision Unit, as implied by the authors of the noted sections of the ITRC document, is unnecessary and fails to comprehend the fact that the sampling method incorporated already accounts for and minimizes error associated with discrete samples. The purpose of collecting independent, replicate samples from a portion (e.g., 10%) of Decision Units as part of an ISM investigation is intended only to test the precision of the overall sampling method. If the Relative Standard Deviation of the replicate sets of data meet DQO requirements, then no further adjustment of the data is required. If DQO requirements are not met, then the recollection of samples is often warranted, although in some cases use of the maximum concentration of the replicate data set might be acceptable.

The equivalent in terms of a traditional, discrete sampling investigation would be to calculate and utilize a second, 95% UCL based on the collection of multiple sets of independent, discrete sample data. This has never been recommended or required by the USEPA or any other environmental agency to our knowledge. The authors' recommendation to apply a 95% UCL for investigations based on ISM-type sample data is likely based on investigations which utilize discrete sampling methods. In the case of the latter, this might indeed be necessary, given the potentially wide variability of averages generated for independent sets of discrete samples demonstrated by field research carried out but the HEER Office.

To conclude, the implied, default requirement in the ITRC ISM update for the utilization of a 95% UCL for comparison to risk-based screening levels or quantitative assessment of risk is unfounded and reflects the lack of training and practical experience in Gy's Theory of Sampling by the authors. Calls to routinely do so and examples provided in the noted sections should be overlooked by readers for what is otherwise a useful, introductory document on the need for the

environmental industry to transition away from outdated, unreliable, discrete sampling methods to more reliable, science-based and field tested, ISM-type sampling methods.

Computer Simulations

A similar lack of an adequate understanding of Gy's Theory of Sampling, field experience and misapplication of classic, inferential statistics to complex particulate media such as soil is especially evident in the computer simulations presented in Appendix B to the ITRC document. While perhaps useful in a qualitative sense to emphasize the need to include a large number of increment points in sample data, the misstatements, errors and lapses in this section of the document are too numerous to highlight in this review.

Especially problematic errors in the appendix include:

- Erroneous suggestion that risk-based or remediation-based DUs should be subdivided for testing in order to better understand contaminant concentration variability within the DU (DUs are specifically designated at a scale appropriate for decision making and within which the variability of contaminant distribution is no longer important);
- Erroneous suggestion that an individual increment represent "…an independent random sample" of the DU area and volume of material (an increment is a "specimen" of the population of interest and in no means represents an independent "sample" of that population);
- Erroneous suggestion that simple random increment collection methods provide more representative data than systematic random collection methods (field experience consistently demonstrates a higher precision, reproducibility and as demonstrated in research by the mining industry overall accuracy of systematic, random sampling methods for particulate matter);
- Failure to take into consideration inherent error associated with the random variability of data for discrete sample data in computer-generated, isoconcentration maps (among problems leads to the generation of seemingly isolated but entirely false "hot spots" and "cold spots" and failed remedial efforts; see Brewer et al. 2017a,b));
- Pre-mature conclusion that 30-increment ISM samples are routinely adequate to capture heterogeneity within a DU and provide reliable data (50-increment samples necessary under many site scenarios; see above and below discussions);
- Erroneous conclusion that a high RSD (and thus low precision) for replicate ISM samples is preferable to ensure representative sample data (completely against Gy's Theory of Sampling and primary intent of replicate samples to demonstrate acceptable precision of the overall sampling method).

Readers are advised to overlook this section of the ITRC document. Refer to Sections 3, 4 and 5 of the HIDOH DU-MIS guidance (HIDOH 2016) and information and references provided in Brewer et al. (2017a,b) for additional information and guidance on these topics.

Conclusion

The sections of the document that review the nature of contaminants in soil, the causes behind the unreliability of discrete sample data and the implementation of ISM-type sampling methods in the field in the ITRC document are for the most part well down and worthy of review. Although the authors of the sections that discuss "statistical concepts" and "use of ISM data in risk assessment" are no doubt well-versed in classic, inferential statistics and experienced in the

use of these methods to attempt to make sense of discrete soil sample data, the misstatements, lapses and outright errors in these sections effectively negate use of the document as a stand-alone reference. The point of Gy's Theory of Sampling and the development of ISM-type sampling methods is, however, that these methods are too simplistic for testing of particulate materials as complex as soil. An open, documented discussion of these sections with presentations from outside experts experience in the use of Gy's Theory of Sampling to characterize particulate matter should have been included a critical part of preparation of the ITRC document. That this did not happen is highly unfortunate and diminishes the utility of an otherwise very useful and much needed, additional source of guidance for the environmental industry.

Attachment: Critique of ITRC ISM Document

Attachment - Critique of ITRC ISM Document

Part 1: Terminology (Sections 1 and 2)

General

The introductory sections of the ITRC document do a good job of distinguishing between traditional "discrete" sampling methodologies and more reliable, ISM-type methodologies similar in structure to Hawaii's DU-MIS guidance. Under a traditional, discrete sample investigation approach, multiple individual and relatively small (e.g., 100-200g) samples of soil or sediment are collected from single points across a project site and individually tested. Decisions are then made on the need for additional sample collection and/or assessment of risk. The extent of contamination above levels of potential concern is determined based on a comparison of individual data points to screening levels. A 95% UCL or in some cases the maximum concentration of the contaminant reported for individual samples is normally used to assess risk for a specified, exposure area.

Discrete sample data for single points as well as mean values for multiple points have been demonstrated to be unreliable due to the inherent random nature of contaminant distribution in soil (e.g., Brewer et al. 2017a,b). This can lead to the need for repeated mobilizations to collect additional samples with no clear end point as well as an under or over estimation of mean contaminant concentrations and associated risk.

ISM-type sampling methods described in the ITRC document, based on Pierre Gy's Theory of Sampling, are specifically intended to replace and improve the reliability and efficiency of past, discrete sampling methods. Decisions Units are normally designated to address investigation questions regarding risk and/or optimization of anticipated remediation. Multiple DUs, often five to ten or more, are normally used to initially characterize a site. Under these methods, a single sample of soil, sediment or similar particulate media is prepared for each DU by collecting and combining multiple "increments" of material from a pre-designated, well-thought-out, DU area and volume of area of material. The sample must meet minimum increment number requirements (e.g., 50) as well as minimum bulk sample mass requirements (e.g., 1-3 kg). Replicate (triplicate) samples are collected in a subset of the DUs (e.g., 10%) and used to test the overall precision of the sampling method.

Decision statements regarding followup actions to be carried out based on the results of DU-specific MIS or ISM data are prepared *prior* to sample collection. As described in the ITRC document, DU-MIS and ISM data provide far superior reliability than traditional, discrete sample data. Decisions regarding risk, remediation or the acceptability of no further action are can normally be made quickly made without minimal need to collect additional sample data. This approach reduces both the time and cost to complete investigation and cleanup actions and also helps to minimize future liability due to missed contamination.

Note that ISM-type samples are referred in some places of the ITRC document to as "structured composites." While useful to assist newcomers in the basic idea of ISM-type samples, this is technically incorrect. Under Gy's Theory of Sampling, the term "composite" refers to the combination of material from two otherwise different populations or "DUs" (see Section 4.4.11 of the HEER Office TGM). This is in stark contrast to Multi Increment and ISM-type sampling methods, where combining of material from separate DUs is specifically not allowed. Multi Increment and ISM-type samples are not true "composites," they are simply "samples."

Past USEPA guidance confusingly uses the term "composite" in both manners – samples prepared from a single area and samples prepared by mixing soil or sediment from distinct, separate areas. As discussed in the HEER TGM and recorded webinars, the latter has very significant implications for investigations carried out under regulations such as the Toxic Substances Control Act (TSCA), where the term "composite" does in fact refer to the intentional mixing of soil from otherwise distinct "DU" areas. Data for true, composite samples are required to be divided by the number of "samples" (points) included in the composite for comparison to TSCA-mandated cleanup levels. Erroneous reference to a Multi Increment or ISM sample collected from a single DU as a "composite" has caused significant and unnecessary financial and legal problems for responsible parties and should be avoided.

Part 2: Introduction and Implementation (Sections 1, 2, 3.1, 3.4, 4, 5 and Appendix A)

Discussions of the nature of contaminants in soil, the hidden fallibility of traditional, discrete sample data and the basic implementation of DU and ISM investigation methods in the field and in the laboratory in the first two sections of the ITRC document are in general very well written and will provide a useful introduction to those just getting started in the field. The case studies provided in Appendix A are likewise useful, but as discussed below in several instances suffer from confusion over use of the term "Sampling Unit" and an erroneous, implied requirement in latter parts of the document for the use of a 95% UCL for all final decision making.. The Cost-Benefit analysis in Section 3.4 is likewise well done. Users of ISM-type investigation methods are encouraged to look beyond the potential moderate increase in time and cost to design and carry out the initial investigation and consider end-of project costs as well as the unquantifiable cost of lawsuits should overlooked contamination be discovered in the future.

Sampling Units

One drawback beginning in Chapter 2 and carried through much of the rest of the document is use of the vague and largely unnecessary term "Sampling Unit." In most cases, this term is used interchangeably with the more descriptive and appropriate terms "Source Area DU" and "Boundary Area DU" introduced earlier in the same chapter. This will likely cause some confusion among field workers.

On comparison and as depicted in Figure 1 and Figure 2 at the end of this technical memorandum, the basic division of a project site into individual areas for characterization is not significantly different between the HIDOH guidance and the ITRC guidance. Figure 1 and Figure 2 present diagrams of risk-based versus remediation-based DU categories utilized in the two documents. The terminology is somewhat different and the ITRC document collectively refers to Source Area and Boundary Area DUs as "Nature and Extent" DUs, but the overall intent is essentially identical.

Figure 3, drawn from examples in the ITRC document where the term "Sampling Unit" was used, depicts the division of a large, purely risk-based, "Exposure Area" DU into smaller, Source Area and Boundary DUs for more detailed characterization and optimization of remedial actions, based on an initial conclusion that the mean contaminant concentration for the Exposure Area DU as a whole would likely exceed screening levels. Several of the examples in the ITRC document present similar scenarios, but unnecessarily refer to what should be more descriptively classified as "Source Area" and "Boundary Area" DUs as less clear, "Sampling Units." In the figure, smaller Source Area DUs are designated in areas of anticipated, especially heavy contamination that might require more expensive remediation. Boundary DUs are also

designated along the perimeters of Source Area DUs in areas of anticipated clean soil to serve as confirmation samples. Boundary DUs are designated beneath Source Area DUs and tested either before (e.g., using a direct-push rig) or after remedial actions (e.g., collection of samples from excavation walls and floors). Under ideal circumstances, no additional confirmation sampling would be required after the completion of any necessary remedial actions.

One concern is the implication in some sections of the document that it is acceptable to collect less-than-adequate samples from Sampling Units (e.g., inadequate number of increments or bulk sample mass) for "screening purposes." It is important that such data not be used for final decision making regarding risk or the need for remediation. Both the ITRC document (Section 2.5.1.2) and the HIDOH guidance (Section 4.3) discuss the use of existing, discrete sample data to assist in DU designation. This can be very useful in the field, particularly if target contaminants include heavy metals and a portable XRF is available (refer to Section 8 of HIDOH TGM). In most cases, however, it is a waste of time and money to collect inadequate ISM-type samples from what would otherwise be designated as a formal DU when the added time and effort in the field to collect a proper, reliably representative sample from the area would be minimal. Reliance on poor sample data can also lead to significant, future liability for responsible parties.

There is, in Hawaii's experience, no need to refer to an area designated for collection of an MI (ISM) sample as a "Sampling Unit." More specific terms noted in Figures 1 and 2 should instead be utilized. If the rationale for the collection of a sample from a targeted area and decisions to be made based on the resulting data are unclear, then adequate thought has not been put into the ultimate objectives of the site investigation and a sample should not be collected. Sample collection purely for the purposes of ill-defined "screening" should be avoided. Doing so can lead to premature, erroneous conclusions regarding risk and lead to unforeseen and unnecessary legal and financial problems for the responsible party and other affected entities.

Gy's Theory of Sampling

Section 2.6, "Managing Heterogeneity to Ensure Sample Representativeness," discusses key aspects of Gy's Theory of Sampling and sources of error in sampling of particulate matter such as soil and sediment. Refer to Minnitt et al. (2007) as well as other references included in the HEER Office's field study of discrete sample data reliability (Brewer et al. 2017a,b) for more in-depth overviews of this topic.

The discussion of individual sources of error and ways to control these errors can seem complicated. In the field, however, the process is relatively straightforward (refer to noted section of the ITRC ISM document; see also Section 4 of the HIDOH *Technical Guidance Manual* and related training webinars):

1. Designate a well-thought-out, risk- or remediation-based "Decision Unit" area and volume of soil for testing;
2. Define the specific particle size fraction(s) to be tested (e.g., <2mm and/or <150µm);
3. Prepare a single, 1-2kg sample by collecting and combining 50+, core-shaped "increments" of material in a systematic random fashion within the targeted DU (2-3 kg samples recommended for soil with high-concentration nuggets of contaminants, e.g., lead shot or PCB-based material);
4. Process the sample in a manner that isolates the targeted particle-size fraction for subsampling;

5. Collect an ISM-type subsample with a minimum mass of 10 to 30 grams in a systematic random fashion from a large number of points within the processed sample manner and in a similar manner that the sample was collected in the field.

This approach is normally sufficient to address primary errors associated with the collection of representative field samples, including:

- Fundament Error (addresses minimum sample mass requirement);
- Grouping and Segregation Error (addresses distributional heterogeneity within the targeted DU as a whole and compositional heterogeneity between individual particles); and
- Increment Delimitation and Extraction Error (addresses how individual increments are collected in the field for primary samples and how subsamples are collected from processed samples for testing at the laboratory).

Note that the equations used to address Fundament Error dictate that a sample with a mass as low as 10 grams could in theory be representative of the <2 mm particle size fraction for very large volumes of particulate matter (e.g., Pitard 2019). In practice, however, a minimum sample mass of 1-3 kg is typically necessary to overcome error associated with the physical collection of a sample in the field.

Default Number of Sample Increments
Section 2.5.2 of the document states that "…a default of 30 increments per field sample is sufficient to control random contaminant heterogeneity in most case." This is based partially on overly simplistic computer simulations presented in Appendix B of the ISM document. A default of 30-increment, Multi Increment samples was recommended in the 2009 edition of Hawaii's *Technical Guidance Manual*. Since that time, however, the experience of consultants as well as HIDOH field staff at hundreds of different types of sites clearly indicates that replicate sample data for 30-increment samples will commonly fail RSD requirements, even in cases where the heterogeneity of contaminants in soil was assumed to be relatively low.

The default, minimum-recommended number of increments per sample was increased to 50 in the 2016 update of Hawaii's DU-MIS guidance, with caveats on the need for 75+ increment samples in some circumstances (e.g., lead-based paint or PCBs). The additional field effort in terms of the overall cost of the project usually makes this cost-beneficial. HIDOH often concurs with the collection of 30-increment subsurface samples due to cost and access limitations, but replicate samples fail RSD limits much more often than 50-increment samples. This often results in a requirement to resample if the soil is disturbed in the future and the imposition of restrictions on future site use.

Part 3: Statistical Concepts, Data Quality Evaluation and Risk Assessment (Sections 3.2, 3.3, 6, 8 and Appendix B)

Section 3.2 of the ISM document, "Statistical Concepts," disappointingly reverts to a review of basic, inferential statistical methods used to evaluate a single set of discrete sample data – the same methods that Gy's Theory of Sampling was specifically developed to replace. The resulting misstatements, errors and overall lack of understanding of the statistical foundation of ISM-type sampling methods as well as the inherent error in discrete sample data are too numerous for in-depth discussion in this critique.

Most critical is the erroneous, implied requirement to collect replicate samples from all DUs and calculate a 95% UCL for comparison to screening levels and assessment of risk - a misguided holdover from past attempts to make sense of a single set of discrete sample data. This mistake is unfortunately perpetuated throughout the other sections of the document, particularly Section 3.3 - Planning for the Use of ISM Data, Section 6 - Data Quality Evaluation, Section 8 - ISM for Risk Assessment and the computer simulations presented in Appendix B. This mistaken requirement will likely be seen as a "poison pill" by many readers new to ISM-type sampling methods and discourage the necessary and much belated transition within the environmental industry from unreliable, discrete sampling methods to ISM-type methods. A brief review of the origin and intended use of 95% UCLs in environmental investigations is therefore warranted.

Origin of the Use of a 95% UCL

Use of a 95% UCL in environmental investigations evolved from challenges that risk assessors faced at the time that sampling guidance was being prepared in the 1990s and early 2000s. The complex nature of error associated with testing of particulate matter like soil and sediment succinctly described under Gy's Theory of Sampling (originally available only in French) were largely unknown in the environmental industry at this time. Risk assessors were instead deluged with tables of seemingly conflicting and uninterpretable, discrete sample data and failed laboratory quality control data with no opportunity to collect additional samples in the field.

In order to draw some conclusion from the data, risk assessors turned to traditional, inferential statistical tests used to manage variability and extract means (e.g., age, weight, etc.) from seemingly similar populations, such as samples of people (Figure 4). This led to the development of USEPA's ProUCL program as more sophisticated tests for estimations of means from highly variable data sets were compiled. Tests that called for the collection of additional, discrete samples in order to improve the precision of the estimated mean were routinely ignored due to a lack of time and resources. Risk assessors instead did the best that they could with the data provided. In order to address variability between individual sample points and the lack independent, replicate sets of discrete sample data to test the overall precision of the sampling method, calculation and use of a 95% UCL for single sets of discrete sample data became routine.

Consideration of a 95% UCL for estimation of a mean for a single set of discrete sample data was appropriate and indoctrinated in USEPA and other agency risk assessment guidance. Statistical tests, however, only address the precision of the test employed to estimate a mean for the data set provided. The tests are unable to directly assess the representativeness of the data set itself.

Data for individual sample points also cannot be assumed to reflect true "exposure point concentrations" for child or adult receptors. Concentration varies with the mass of soil tested (see Brewer et al. 2017a,b). Environmental workers not familiar with Gy's Theory of Sampling often get hung up in a fallacious search for the "maximum" concentration of a contaminant in soil. This is actually an easy question to answer. If the contaminant is present in the soil, then at some scale, for example at the scale of individual particles or coatings on particles, the maximum will necessarily be 100%. If the contaminant is not present, then the maximum concentration is 0%.

Any concentration between 1% and 100% reported by the laboratory reflects the "dilution" of the contaminant in the subsamples tested with clean particles of soil. The range of concentrations reported for testing of 0.5-gram subsamples of soil would be anticipated to be much broader than

the range of concentrations reported for 30-gram subsamples. Neither of the data sets directly reflect the range of concentrations representative of exposure to 100 to 200 milligram masses of soil – the incidental soil ingestion rates typically utilized in risk assessments for adults and children. The 95% UCL can only be assumed to represent the mean concentration of the contaminant for the total area and volume of soil included in the exposure area DU.

These facts have traditionally been overlooked in the environmental industry, where use of 95% UCLs routinely leads to premature claims of total data precision. Reliance on 95% UCLs estimated from a single set of discrete sample data indeed masks hidden error in data representativeness related to how samples were collected in the field and processed for analysis at the laboratory. As highlighted in Section 1.3 of the ITRC document, decades of research by the mining industry and more recently the environmental industry clearly document that this – not laboratory analysis, is the greatest source of error in data for particulate media such as soil and sediment (e.g., Pitard 2019; see also references in Brewer et al. 2017a,b). This error would have been quickly identified in the form of a high variability of contaminant means had independent, replicate sets of discrete sample data been collected from the same area and compared to assess overall sampling precision.

Such a test is a required part of DU-MIS and ISM investigation methods but is virtually unheard of as part of investigations that rely on discrete sample data. While no doubt shocking to environmental professionals who have been working in the industry for decades, including those of us who initiated research on DU-MIS methods in Hawai′i over 15 years ago, it is a fact that the routine reliability and reproducibility of discrete sample data for both site characterization and risk assessment purposes has never been demonstrated in the field (see Brewer et al. 2017a,b).

Replacement of Discrete Sample 95% UCLs with Gy's Theory of Sampling
Limitations in reliance on statistical evaluation of a single set of discrete sample data were well known in the mining industry by the late 1800s. Errors were routinely huge, as mining assayers compared the mass of a commodity (e.g., gold) physically extracted from a body of crushed ore material to the mass initially estimated based on the mean of individual discrete samples. Multiple sources of sample collection and processing error were identified over time and combined into what became known as Pierre Gy's Theory of Sampling (Pitard 2019; see also references provided in Brewer et al. 2017b). Sampling methods in the environmental industry related to this research eventually evolved into what came to be referred to as "DU-MIS" in the HIDOH guidance and "ISM" in the ITRC guidance (Figure 4).

It is important for readers of the ITRC document to understand that unadjusted data for a single, ISM-type sample collected in accordance with Gy's Theory of Sampling are specifically intended *to replace* less reliable use of a 95% UCL extracted from single set of discrete sample data (see Figure 4). This critical distinction between the two sampling methods is strikingly absent from the noted sections of the document. Implicit requirements in Section 8 for the collection of replicate samples from all DUs in order to meet past USEPA recommendations for the use of a 95% UCL in decision making fails to understand this background and distinction between the two sampling methods. This could lead to the very false impression that a 95% UCL calculated for ten discrete soil samples, directly representative of only a spoonful or small handful of soil, is superior to a single, 1-3kg, 50-increment ISM sample collected and processed in accordance with Gy's Theory of Sampling and as described in early section of the document (see Brewer et al. 2017a,b).

For example, Section 6.2.1.1 (Single DU Sample Result), correctly states that "Single ISM results do not allow for the calculation of a DU-specific CI or quantification of the precision of the estimate" but fails to note that the same has always been true for single sets of discrete samples (see Figure 4). Section 3.2.4.4 misleadingly claims that "Both ISM and discrete sampling designs can be used to obtain defensible estimates of DU means." The latter would only be true if individual, discrete samples were collected, processed and tested in accordance with Gy's Theory of Sampling *and* if replicate sets of discrete sample data were collected to assess the overall precision of the sampling method.

In practice, neither is done. Methods to control field error and ensure adequate data precision based on Gy's Theory of Sampling are directly incorporated into DU-MIS and ISM sampling methods (e.g., DU specification, number of sample increments and total sample mass, increment collection, sample processing and minimum subsample mass, etc.). Mistakes in these steps of the site investigation process are where most errors in environmental data occur. Similar, rigorous methods to control field and laboratory sampling error are entirely missing from traditional, discrete sampling methods. There is no comparison in the reliability and quality of the resulting data.

Application of 95% UCLs to ISM-Type Sample Data
Proper collection of ISM samples in accordance with Gy's Theory of Sampling negates the need to test the total precision of every sample collected as part of a site investigation, provided that replicate data are collected in a subset of DUs (e.g., 10%). The primary purpose of the replicate sample data is to recheck the overall precision of the sampling method employed, including the number of increments per sample, increment collection method, bulk sample mass and laboratory processing and testing. If the Relative Standard Deviation of the replicate sample data meet DQO requirements then the data can be used as-is. If not, then the potential sources of error must be identified and samples either recollected of the adequacy of the sample data otherwise reviewed in terms of the safety margin built into screening levels and target risks.

Routine use of 95% UCLs calculated from replicate ISM sample data has no basis in Gy's Theory of Sampling, just as consideration of a 95% UCL based on the collection of *replicate sets* of discrete sample data has no basis in USEPA and related risk-assessment guidance. As discussed in Section 4.2.7.1 of the HEER Office TGM, the periodic collection and testing of replicate, ISM sample data from the same DU is similar to "batch" collection and testing of replicate subsamples from a subset of samples submitted to a laboratory for analysis (e.g., 1 in 10 or 1 in 20). This is done only to verify the precision of the test method employed. No adjustment of data is required. If DQOs for data quality are not met, then sample and subsample collection methods should be improved. There is no need to collect replicate field samples from every DU in the field or replicate subsamples from every sample tested by a laboratory provided that precision of replicate data for a subset of field DUs or laboratory samples is periodically tested. More important, and as discussed in Section 6.1 of the ITRC ISM document and Appendix A to this critique, is a thorough review of sample collection in the field and sample processing and testing in the laboratory – the overwhelming source of error in data for particulate matter such as soil and sediment.

To be fair, the HEER Office made a similar mistake in a recommendation to calculate and refer to the 95% UCL of replicate MI sample data in the 2016 edition of the DU-MIS guidance when the RSD for replicate sample data exceeded 50% (HIDOH 2016; refer to Section 4). While highly favorable of the Hawaii's guidance in meetings in meetings with HEER Office staff in

2018, Francis Pitard, considered to be the world's leading expert in sampling of particulate matter (see Pitard 2019), and his colleagues adamantly opposed the use of a 95% UCL for replicate sample data as part of Gy's Theory of Sampling and requested that this be omitted in future updates. Emphasis should instead be placed on ensuring that samples are collected, processed and analyzed in an appropriate manner, with the collection of new samples considered when the RSD for an associated set of replicate data exceeds targeted, data quality limits.

Planned updates to the Data Quality Evaluation section of the HEER Office TGM (Section 4.2.7.3) are presented in Appendix A. Included is a checklist of data quality review in terms of how a sample was collected, processed and tested. A similar checklist is included in Section 6.1.3 of the ISM document. The updated recommendations in Appendix A can be used for projects carried out in Hawai′i but should be discussed with project managers ahead of time.

Additional Information

Refer to publications by Francis Pitard and his colleagues referenced in the HEER Office TGM (HIDOH 2016, and updates) and in the HEER Office field study of discrete sample data reliability (Brewer et al. 2017a,b) for additional information on the history and basis of Gy's Theory of Sampling and ISM-type site investigation strategies. Links to recorded DU-MIS related presentations and training webinars are posted to the HEER Office webinar webpage (https://health.hawaii.gov/heer/guidance/heer-webinars/).

Consultants involved in environmental investigations in Hawai´i are also strongly encouraged to take classes on sampling theory taught by Chuck Ramsey of EnviroStat Inc., one of the leading experts on "sampling of anything" who has trained a number of HIDOH staff and local consultants and regularly assists in updates to the HEER Office's DU-MIS guidance.

Contact Roger Brewer (roger.brewer@doh.hawaii.gov) with the HEER Office for additional information on the background and use of Hawaii's DU-MIS investigation guidance. Suggestions for edits and updates to the guidance are welcome. The next update of the guidance is anticipated in early to mid 2021.

Appendix A: Draft Data Quality Evaluation update (HEER Office TGM Section 4)

References

Brewer, R., Peard, J., and Heskett, M. 2017a. A critical review of discrete soil sample reliability: Part 1 – Field study results. Soil and Sediment Contamination. Vol 26, No 1. Available from: http://dx.doi.org/10.1080/15320383.2017.1244171

Brewer, R., Peard, J., and Heskett, M. 2017b. A critical review of discrete soil sample reliability: Part 2 – Implications. Soil and Sediment Contamination. Vol 26, No 1. Available from: http://dx.doi.org/10.1080/15320383.2017.1244172

HIDOH, 2016, *Technical Guidance Manual*: Hawai´i Department of Health, Office of Hazard Evaluation and Emergency Response, http://www.Hawai′idoh.org/

HIDOH, 2017, *Evaluation of Environmental Hazards at Sites with Contaminated Soil and Groundwater* – Hawaii Edition (Fall 2017): Hawai'i Department of Health, Office of

Hazard Evaluation and Emergency Response. https://health.hawaii.gov/heer/guidance/ehe-and-eals/

ITRC, 2020, *Incremental Sampling Methodology*: Interstate Technology and Regulatory Council, October 2020 (update of 2012 edition), https://ism-2.itrcweb.org/.

Minnitt, R.C.A., Rice, P.M. and Spangenberg, C. (2007) Part 1: Understanding the components of the fundamental sampling error: a key to good sampling practice. J. South. Afr. Inst. Min. Metall. 107. p. 505-511.

Pitard, F.F., 2019. *Theory of Sampling and Sampling Practice* (3rd Edition): CRC Press, Inc., Boca Raton, FL.

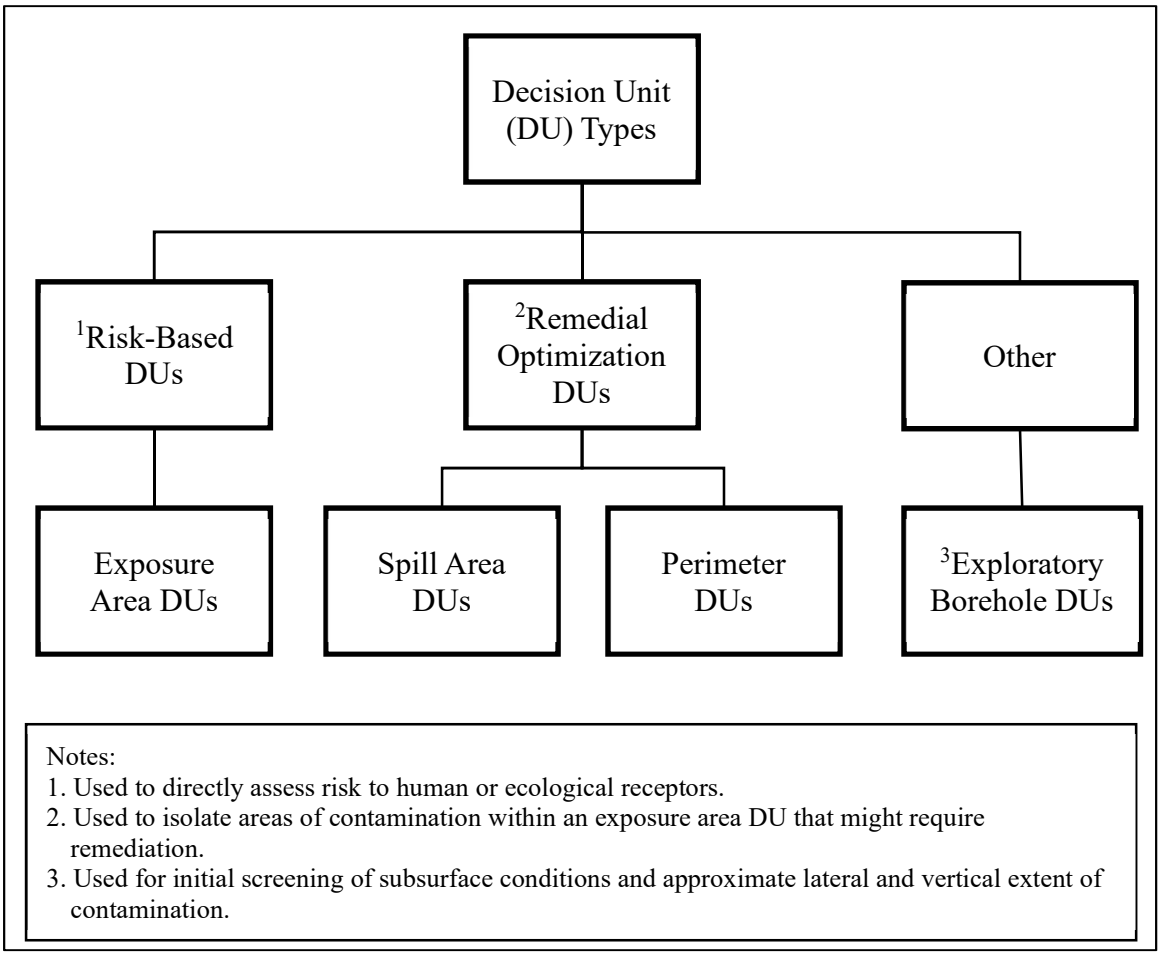
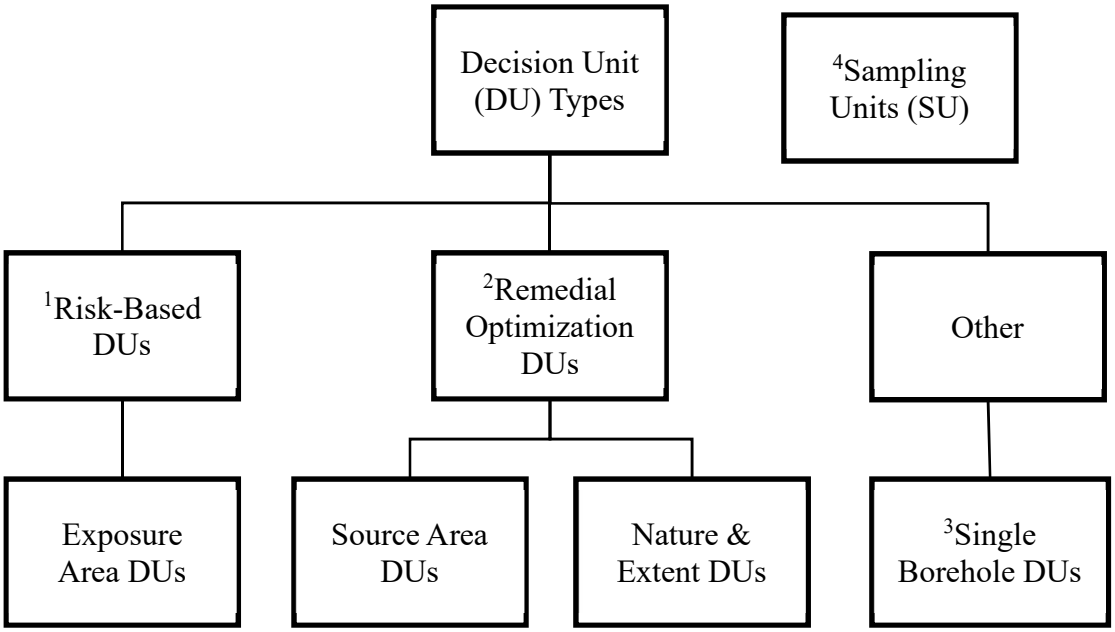


Figure 1. Hawai´i DU-MIS Decision Unit designation scheme.



Notes:
1. Used to directly assess risk to human or ecological receptors.
2. Used to isolate areas of contamination within an exposure area DU that might require remediation.
3. Used for initial screening of subsurface conditions and approximate lateral and vertical extent of contamination (referred to in case studies but not described in text of ISM document; former term used in earlier editions of the HIDOH DU-MIS guidance).
4. Refer to comments on Chapter 2 of the ISM document. Additional category of normally small areas and volumes of soil to be tested using field screening methods and for which individual decision will not be made. Primarily used to assist in the designation of Decision Units. Term not used in HIDOH DU-MIS guidance but comparable to use of a field screening method (e.g., portable XRC) to test a grid of "large mass" discrete samples and identify apparent source areas within a site for more detailed characterization (HEER Office TGM Section 8.4). Care should be taken to avoid overuse of the term and confusion with areas more appropriately referred to as "Source Area DUs" and "Nature and Extent" DUs.

Figure 2. ITRC ISM Decision Unit designation scheme.

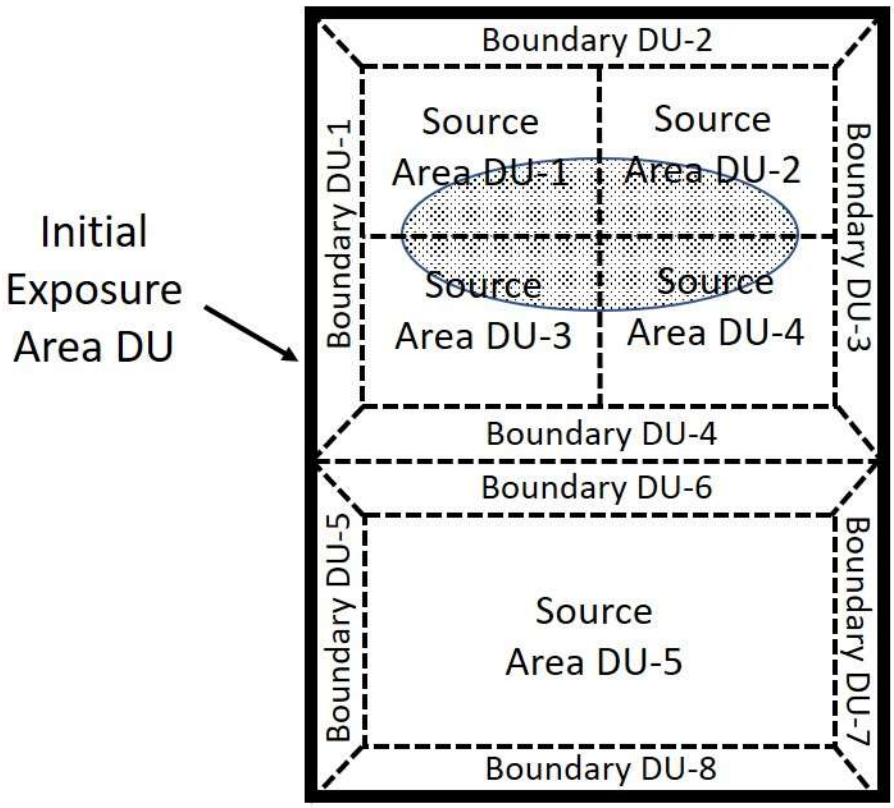


Figure 3. Hypothetical division of a purely risk-based, Exposure Area DU into smaller, Source Area DUs (assumed to be contaminated) and Boundary Area DUs (assumed to be clean) for optimization of anticipated remedial actions.

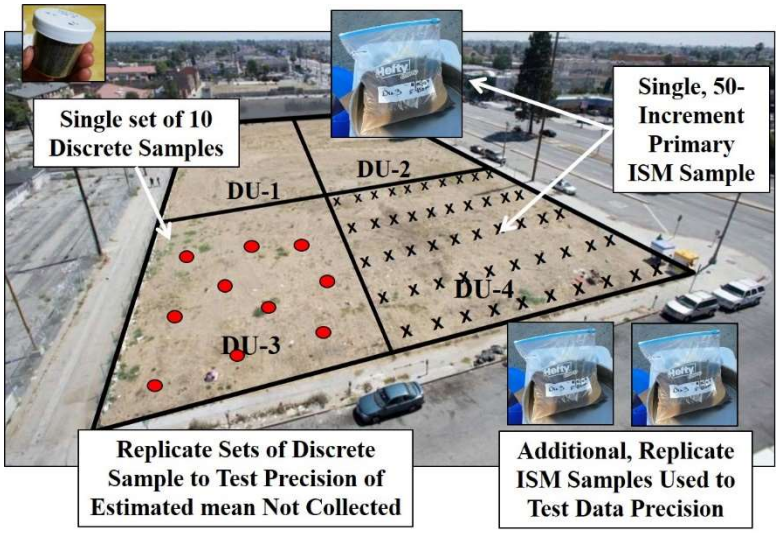


Figure 4. Comparison of use of a 95% UCL for a single set of discrete samples to estimate a contaminant mean to use of a single ISM sample to estimate a mean. The collection of replicate ISM samples from a portion of the DUs in order to test the precision of primary sample data is a required part of DU-MIS and ISM sampling methods. No such requirement exists for the use of discrete sample data in risk assessments. The precision of the estimated mean and 95% UCL is therefore always untested and unknown.

Appendix A

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 Decision Unit Designation
	• Site history and potential sources and type of contamination well understood?
	• Site investigation questions used to designate DU for testing clearly stated and based on risk and/or optimization of anticipated remediation?
	• Questions and decision statements developed for individual DU presented?
	• Area and total volume of soil associated with each DU noted and acceptable for intended purposes?
	• To-scale map depicting location and size of DU 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 DUs (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 DU?
	• Analytical subsample collected using a sectoral splitter or manually collected from at least 30 points?
	• Minimum 30-gram analytical subsample mass extracted for <2mm particles?
	• Minimum 10-gram analytical subsample mass extracted for <250µm particles?
	• 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. Much more stringent RSD requirements for data precision, as low as 5% or less, are required in some cases Resampling is generally called for when RSD limits are exceeded. 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.