

A SIMULATION STUDY OF EXTRAPOLATION UNCERTAINTY IN EXPOSURE ASSESSMENT– USE OF PILOT STUDY RESULTS FOR SITE INVESTIGATION

Philip Goodrum and Emma Mendelsohn

INTRODUCTION

This document summarizes the methods and findings from a simulation study intended to guide environmental sampling for site investigations for which there is an interest in applying extrapolation concepts. For large sites, it is not uncommon for there to be many decision units (DUs), such as areas where we would like to assess exposure and risk (e.g., exposure units), or compliance with remediation action objectives (e.g., remediation units). A key objective for large site investigations is to balance the potential for a decision error together with the practical constraints of collecting and measuring environmental samples in hundreds or even thousands of subareas.

A common strategy for optimizing sampling designs is to implement adaptive sampling techniques, whereby the investigation is divided into a series of sequential steps or phases. While many different adaptive sampling methods are available (Bujewski and Johnson 1996; USDOE 2001; USEPA 2002; USEPA 2003; Crumbling 2004), the fundamental idea is that results from each phase can be used to guide subsequent steps, following a logical decision framework that satisfies multiple criteria. Key elements of successful frameworks include:

- Proper application of random sampling methods and statistics concepts so that results from the sample can be used to inform estimates of population parameters of interest;
- Clear and straightforward methods for implementation, reproducibility, and communication; and
- Decision rules for determining if no further action is warranted, grounded in well-defined tolerances for decision error rates plus any established regulatory compliance requirements.

The motivation for this simulation study is to provide straightforward guidance, based on thousands of simulated sampling programs, regarding the conditions when results from a Pilot Study can be extrapolated to the entire Site with acceptable error rates.

This simulation study is specifically intended to guide investigations that include the following sampling design options:

1. Use of either discrete sampling or incremental sampling methods (ISM) (ITRC, 2012).
2. Reliance on the 95 percent upper confidence limit for the arithmetic mean (95UCL), as the primary basis for assessing compliance with an established action level (AL).
3. Division of a large site into multiple DUs.
4. Use of a Pilot Study to guide a conclusion regarding the overall Site condition (including unsampled areas). The Pilot Study consists of a random selection of a subset of all the DUs across the Site.
5. An option for ISM to use $r=3$ replicates to calculate a 95UCL in some of the Pilot Study DUs, and $r=1$ in others, borrowing information on estimated variance from the DUs for which $r=3$ were obtained.
6. A decision error rate that constrains the number of DUs (or equivalently, the proportion of the total Site) for which the mean concentration is likely to be less than or equal to an AL.

SITE INVESTIGATION SCENARIOS

A range of site investigation scenarios were explored to represent plausible environmental conditions and realistic sampling schemes. All simulations applied simple random sampling methods. The software tools, probability distributions, and major assumptions used to define the scenarios are summarized in Table 1.

It should be noted that the environmental medium that is sampled here is not specified. It may be easiest to envision a surface soil sampling program, but the concepts and findings should be more broadly applicable to any medium that can be evaluated using sampling designs that involve defined DUs.

Table 1. Scenarios evaluated in this simulation study.

Sampling Method	Lower Areas	Higher Areas	P(High) ^a	AL ^b	Pilot % of Site ^c	ISM DU % r=3 ^d
Discrete ^e	Log(100, 50) (CV=0.5)	Log(600, 300) (CV=0.5)	U(5%, 25%)	U(10, 6000)	U(10%, 90%)	NA
	Log(100, 150) (CV=1.5)	Log(600, 900) (CV=1.5)	U(5%, 25%)	U(10, 6000)	U(10%, 90%)	NA
	Log(100, 300) (CV=3.0)	Log(600, 1200) (CV=3.0)	U(5%, 25%)	U(10, 6000)	U(10%, 90%)	NA
ISM ^f	Log(100, 50) (CV=0.5)	Log(600, 300) (CV=0.5)	U(5%, 25%)	U(10, 6000)	U(10%, 90%)	U(10%, 66%)
	Log(100, 150) (CV=1.5)	Log(600, 900) (CV=1.5)	U(5%, 25%)	U(10, 6000)	U(10%, 90%)	U(10%, 66%)
	Log(100, 300) (CV=1.5)	Log(600, 1200) (CV=1.5)	U(5%, 25%)	U(10, 6000)	U(10%, 90%)	U(10%, 66%)

Notes:

CV = coefficient of variation, equal to the ratio of the standard deviation divided by the arithmetic mean

Log = lognormal distribution defined by an arithmetic mean and standard deviation.

NA = not applicable

U = uniform distribution defined by minimum and maximum

^a Probability of assigning a DU to a “Higher Area” category.

^b Action level varies from 10x less than the mean of the “Lower Area” (i.e., 100/10 =10) to 10x greater than the mean of the “Higher Area” (i.e., 600 x10 = 6,000).

^c Percentage of Site-wide DUs that are randomly assigned to the Pilot Study.

^d Percentage of Pilot Study DUs that are characterized using r=3 replicates (only applies to ISM simulations).

^e Simulations of discrete sampling are executed using R version 3.5.0 (R Core Team 2018).

^f Simulations of ISM sampling are executed using Crystal Ball v.11.1 with Microsoft Excel® (Oracle, 2015)

Definition of Compliance

For these simulations, compliance is defined relative to an AL, which may be a regulatory compliance standard, guidance value, or even a post-remediation risk-based goal. Each simulation defines a “true” condition for every DU, meaning that the arithmetic mean concentration is known. The purpose of the Pilot Study is to determine if the evaluation of a subset of the DUs, based on a simulated sampling program, leads to the correct inference about the Site condition. Therefore, the Pilot Study provides the estimate of compliance. In practice, site investigators do not know the true Site condition and cannot calculate error rates based on sample results. Using numerical simulation, we can define the true condition, and thereby test a wide range of sampling programs to evaluate the frequency of *error rates* – meaning the number of times (or probability) that a conclusion based on the Pilot Study does not match the known conditions.

For the Pilot Study, Site-wide compliance is estimated using a set of calculated 95UCLs, one for each DU that is part of the Pilot Study. For simulations of discrete sampling programs,

each 95UCL is calculated from a set of random observations using the UCL method decision matrix recommended by USEPA as implemented in ProUCL (USEPA, 2013). For simulations of ISM, both the Students t and Chebyshev UCL methods are applied to a set of random observations (called replicates), consistent with guidance from ITRC (ITRC, 2012). We can refer to the outcome of the Pilot Study as “a”, and make one of two inferences about the Site overall:

- $a = 0$ (Site is *likely* not in compliance because at least one 95UCL exceeds the AL)
- $a = 1$ (Site is *likely* in compliance because all 95UCLs are less than or equal to the AL)

Similarly, we can evaluate the outcome of the Site condition, which we can call “b”, and make one of two observations:

- $b = 0$ (Site is not in compliance; the mean of one or more DUs exceeds the AL)
- $b = 1$ (Site is in compliance; all of the DU means are less than or equal to the AL)

A 95UCL is not needed to evaluate the Site condition because we are able to define the full set of true arithmetic mean concentrations as part of the numerical simulation.

For these simulations, a high bar is set for Site-wide compliance. True compliance occurs only if the mean for every DU is less than the AL. Ideally, the Pilot Study would capture the true compliance conditions each time, but in practice there is always a chance of making the wrong decision. The performance of the Pilot Study is evaluated based on the frequency with which the estimate of compliance matches the true Site conditions (i.e., $a=b$).

In a sense, the simulation study is partly an evaluation of the reliability of the “coverage probability” of a 95UCL, meaning the likelihood that the one-sided 95 percent confidence limit exceeds the true mean. If the Pilot Study included the entire Site (i.e., 100% of the DUs), then we would anticipate an error rate (for $a=1$, $b=0$) of no more than 5 percent, because the 95UCL can be expected to underestimate the “true mean” with no greater than 5 percent frequency (on average). Less obvious, however, is the performance as the proportion of the Site included in the Pilot Study decreases, particularly under varying conditions of contamination relative to an AL. This performance is evaluated based on a set of error rates, which are discussed in detail below. In general, the performance of the Pilot Study can be thought of as a function of three factors: 1) the variability in concentrations across the Site; 2) the AL relative to the average concentration at the scale of a DU (as opposed to the Site-wide mean); and 3) the sampling design (e.g., proportion of DUs sampled, discrete or ISM sampling methods). This document discusses the sensitivity of the error rates when these factors are varied across plausible ranges.

Decision Errors and Error Rates

For each iteration of a sampling design applied to a Site scenario, there are four possible outcomes based on the prediction (outcome “a” using the Pilot Study) and the true Site condition (outcome “b”). Each set of outcomes (a, b) is referred to as a “Match” and illustrated both conceptually and with a numerical example in Exhibit 1.

A		Pilot Study	
		NC	C
		NC	C
Site-Wide Condition	NC	1 a=0, b=0	3 a=1, b=0
	C	2 a=0, b=1	4 a=1, b=1
		C = compliance; NC = Not in Compliance	

B		Pilot Study	
		NC	C
		NC	C
Site-Wide Condition	NC	87.9%	5.2%
	C	2.3%	4.6%
		C = compliance; NC = Not in Compliance	

Exhibit 1. (A) Decision logic to evaluate if the conclusions from the Pilot Study and the true Site condition are a Match. Match = 3 (highlighted) is the least desirable outcome from the perspective of health protection. (B) Example outcome illustrating a case when the Pilot Study has a 5.2% error rate for incorrectly concluding the Site is in compliance.

The four possible “Match” results (Pilot Study outcome, and accuracy based on true Site condition), including the two types of errors ($a \neq b$), are discussed below.

- Match = 1, “Not in Compliance ($a=0$), Correct ($b=0$)”

This result can only happen if at least one 95UCL for the Pilot Study exceeds the AL. When the Pilot Study is correct, then there is also at least one exceedance of a true mean. Because the 95UCL is a summary statistic that is intended to overestimate the true mean most of the time, an exceedance of the 95UCL in the Pilot Study does not guarantee an exceedance of a true mean (see “Match = 2” below).

- Match = 2, “Not in Compliance ($a=0$), Error ($b=1$)”, False Noncompliance

This result can happen if the true means of the Site DUs are all less than the AL, but at least one 95UCL in the Pilot Study exceeds the AL. This error occurs because the 95UCL is a summary statistic that is intended to overestimate the true mean most of the time. This error may be costly in terms of time and money because it may

trigger an action that was not necessary. However, this is not the primary error of concern because it does not adversely affect exposure and health risks (see “Match = 3” below).

- *Match = 3, “Compliance ($a=1$), Error ($b=0$)”, False Compliance*

This result is the primary extrapolation error concern. In this situation, the Pilot Study steered us to incorrectly conclude that all true means are less than or equal to the AL when in fact there is at least one DU that fails the compliance criterion. A given sampling design, when implemented in the simulation just once, may or may not trigger this error. The key is to simulate the same sampling program many times, and determine if the frequency of the error is unacceptable. The acceptable error rate is an important risk management decision. This simulation study presents examples using a tolerance level of 5 percent for a Match=3 error.

- *Match = 4, “Compliance ($a=1$), Correct ($b=1$)”*

This result is the “desired” outcome by the field team. The Pilot Study leads us to conclude that the Site is in compliance, and in fact, this is true. The extrapolation was successful and allowed the Site to be characterized without sampling every DU.

SIMULATION LOGIC

This section describes the sequence of simulation steps in more detail. Many of the steps are the same for both discrete and ISM sampling designs. Deviations are noted in the subsections for each approach.

Exhibit 2 illustrates the logic flow of the simulations. The key steps are numbered to facilitate cross referencing to the descriptions below.

Step 1 – Define the Parent Distributions

For all the scenarios, two parent distributions are defined, one that characterizes variability in contamination in *Lower Concentration Areas* and a second that characterizes *Higher Concentration Areas*. The purpose of defining two parent distributions is to introduce some heterogeneity, and to allow some of the DUs to be more influenced by subareas of elevated concentrations (formerly referred to as “hot spots” by some practitioners). Of course, every site is unique, and the challenge of any simulation study, and risk assessment guidance in general, is to capture a plausible range of conditions that represent cases that investigators will most often encounter. Results reported here may not apply to sites that exhibit more extreme conditions.

The Lower areas are characterized by a two-parameter lognormal distribution with a mean of 100 (units are intentionally excluded as they are not pertinent to the simulation study) and standard deviations (SDs) of 50, 150, and 300. The corresponding coefficient of variation (CV, equal to the ratio of SD divided by mean) values are 0.5, 1.5, and 3.0.

The Higher areas are characterized by a two-parameter lognormal distribution with a mean of 600 and SDs of 300, 900, and 1200 (again reflecting CVs of 0.5, 1.5, and 3.0). The Higher areas can be thought of as being more influenced by subareas of elevated concentrations than the Lower areas.

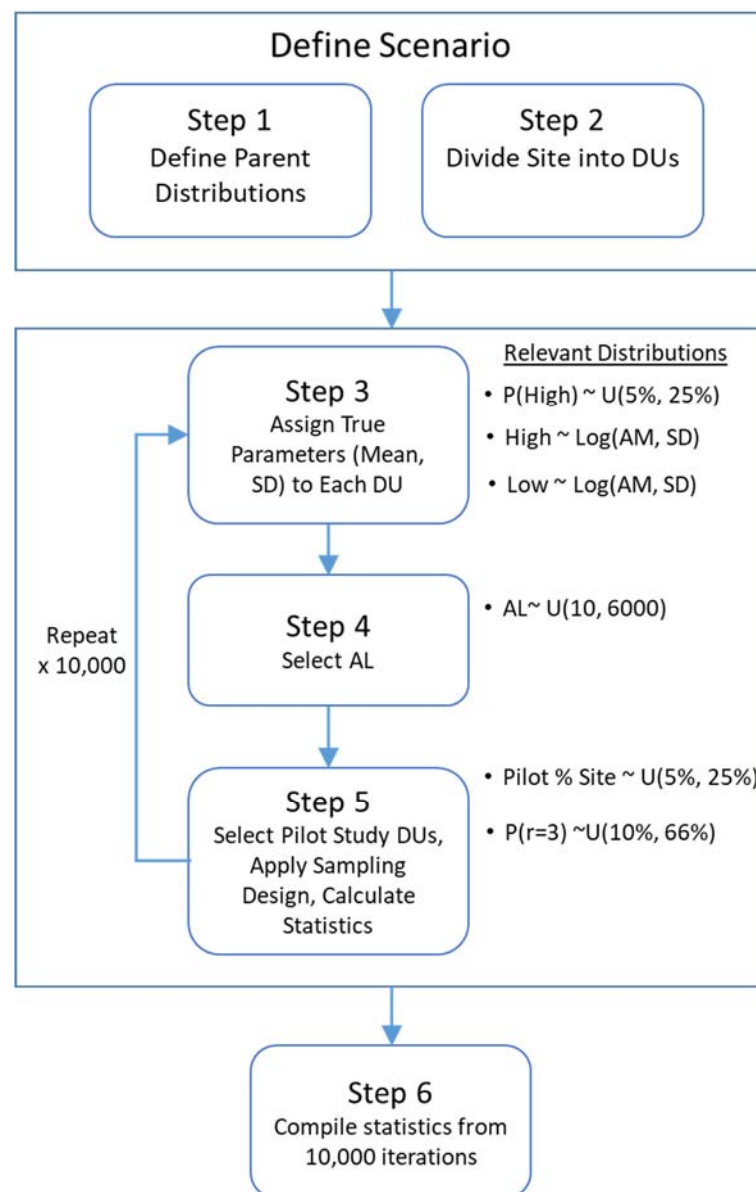


Exhibit 2. Decision logic for simulations.

Step 2 – Divide the Site into DUs

It is assumed that the entire Site can be subdivided into a set of DUs. In practice, the shape and area of the DUs can and will vary. For simplicity, we describe the “average DU” size along with the Site area. The ratio of the Site Area divided by the average DU size determines the number of DUs across the entire study. For these simulations, we have made the following assignments:

- Site Area = 100 acres
- Average DU Size = 1 acre

Therefore, the Site consists of 100 DUs, some portion of which are randomly assigned to the Pilot Study.

Step 3 – Assign Each DU a Set of True Parameters

The contamination in each DU is defined by a two-parameter lognormal distribution, but the set of parameters are allowed to vary. The variability stems from two sources of randomness:

1. Assignment of a DU to “Lower” or “Higher” area. This is accomplished with the variable, $P(\text{High})$, which is the probability of assigning a DU to a Higher Concentration Area. Equivalently, the probability of assigning a DU to a Lower Concentration Area is $(1-P(\text{High}))$.
2. For each DU, randomly draw $n=15$ observations from the appropriate parent distribution (Lower or Higher). Calculate the mean and SD, and consider these to be the “true” parameter values for the lognormal distribution of a DU.

The outcome at this point is the fully characterized Site condition. Next, we need to specify an AL. As discussed below, both the set of “true parameters” and the AL vary with each iteration of the simulation, in an effort to capture a wide range of plausible scenarios.

Step 4 – Select an AL and Calculate the Ratio

Initially, the simulation study was executed with a finite set of fixed ALs. While this does in fact yield some findings, the question remains – can the results be generalized to conditions when the AL is higher or lower? To address this question, the AL was varied with each iteration. The AL is selected at random from a uniform distribution, with parameters that bracket the Low and High parent distributions by an order of magnitude. Specifically, the minimum $AL = 100/10 = 10$, and the maximum $AL = 600 \times 10 = 6,000$.

The AL is used subsequently in graphical summaries of the simulation results to determine if there are patterns with respect to the performance of specific sampling designs.

Step 5 – Apply the Sampling Design for the Pilot Study

With each iteration of the simulation, first a key sampling design parameter is determined for the Pilot Study:

- Pilot Percentage of Site (Pilot Percent) – the fraction of the Site that is characterized by the Pilot Study is assumed to vary between 10% and 90%. This is selected at random using a uniform distribution with these minimum and maximum values. The number of DUs characterized by the Pilot Study is, therefore:

$$\text{DU Count} = (\text{Site Area} / \text{Average DU Size}) \times \text{Pilot Percent}$$

The result for DU Count is rounded to the nearest whole number.

For each of those DUs, the sampling program is implemented once. This means:

- A set of random observations is obtained from the DU-specific lognormal distribution using the true parameters defined in Step 3.
- The 95UCL is calculated for each DU, and the average 95UCL is calculated for the set of DUs.

These simulations are executed using detects only; no nondetects have been introduced, although this level of detail could be included in subsequent studies.

Discrete Sampling

For the discrete sampling program, for each DU, n=15 observations are selected from the specified lognormal distribution. The 95UCL is calculated following the calculation method decision tree recommended by USEPA and implemented in ProUCL (USEPA, 2013). These simulations were performed in R for efficiency.

The ratio of the average 95UCL from the Pilot Study divided by the AL is tracked with each iteration. As with other metrics, the ratio is used subsequently in graphical summaries of the simulation results to determine if there are patterns with respect to the performance of specific sampling designs.

ISM Sampling

ISM sampling involves the collection of increments throughout the DU that are then composited into a replicate (r), which yields an estimate of the mean. By repeating this step multiple times (e.g., r=3), estimates can be obtained for the (group) mean, SD, and 95UCL.

The lognormal distributions defined in Step 3 can still apply for ISM sampling, but with one additional step that is required to generate corresponding lognormal distributions for replicates. The example below is given for the Lower ~ Log(100, 50).

- Collect n=15 observations from Log(100, 50). Calculate the sample mean and SD.

- Define the lognormal distribution of *replicates* as: $\text{Log}(\text{mean}, \text{SD}/\sqrt{30})$. The ratio of SD/\sqrt{n} , also called the standard error, approximates the standard deviation of the means. Each simulation is executed assuming $n=30$ increments are collected for each replicate.

Next, for each DU, determine if the DU is to be “fully characterized” with three replicates ($r=3$) or one replicate ($r=1$). This is accomplished by assigning a probability of $r=3$ from a uniform distribution with a minimum of 10% and maximum of 66%. For example, if the Pilot Study consists of 40 DUs, and the probability of $r=3$ is randomly selected to be 50%, then for this iteration of the simulation, 20 DUs would be defined with $r=3$ and 20 would be defined with $r=1$. Random values for replicates are obtained from the $\text{Log}(\text{mean}, \text{SD}/\sqrt{30})$ presented above.

With ISM, we calculate two 95UCLs, one using the Student’s t UCL and the second using Chebyshev UCL. In both approaches, an estimate of the SD is needed. For each DU with $r=3$, the SD is obtained directly from the set of three random replicate results. For each DU with $r=1$, we use the average CV from the subset of DUs with $r=3$.

The ratio of the average 95UCL from the Pilot Study divided by the AL is also tracked with each iteration. Two ratios are calculated, one specific to the average of the Student’s t UCLs, and one for the Chebyshev UCLs.

Step 6 – Calculate “Match” Result for Each Iteration and Compile Statistics

From the Pilot Study, we determine the outcome as described above. If any 95UCL exceeds the AL, we predict that the Site is not in compliance.

From the true means assigned to all DUs at the Site, determine the number of DUs for which the mean exceeds the AL. If there is at least one exceedance, the Site is not in compliance.

Assign the “Match” results (i.e., 1, 2, 3, or 4), as described above (see Exhibit 1).

Repeat Steps 3-6 many times (e.g., 10,000). Each iteration yields the following information:

- $P(\text{High})$
- Proportion of Site in Pilot Study
- AL and Ratio
- Match result
- Proportion of Pilot Study DUs characterized by $r=3$ (applies to ISM only)

Specific findings from this simulation study are summarized below.

RESULTS

Results from each of the six scenarios listed in Table 1 are presented below, grouped by discrete and ISM sampling designs. The results were analyzed to address the following key questions:

1. Are there specific sample statistics from a Pilot Study that can serve as reliable indicators of decision error rates, particularly for the Match 3 outcome (i.e., false compliance)?
2. Are the consequences of each decision error equal? Or are there conditions when a decision error may be more problematic in terms of the magnitude of the unsampled area that exceeds an AL?
3. Do any of the options for design parameters influence the decision error rates, including: a) Pilot Study percent of Site area; and b) percentage of Pilot Study DUs that are characterized by $r=3$ replicates (ISM only)?
4. Are findings relatively consistent across a range of contamination conditions, including variability ranging from $CV=0.5$ to 3.0 , and the probability that a DU is more highly contaminated?

Match Outcomes without Grouping

Figures 1 and 2 summarize the frequency of each of the four outcomes (i.e., Matches 1 to 4), without any subsetting or grouping of the results. The overall error rates are contingent on the AL relative to the parent distributions (Lower and Higher Concentration Areas). Therefore, these summaries are most informative for examining the sensitivity of error rates (i.e., Match 2 and Match 3) to the magnitude of the underlying variance (as given by $CV=0.5$, 1.5 , and 3.0) and, for ISM sampling, the choice of UCL calculation methods (i.e., Student's t and Chebyshev).

Discrete Sampling

As shown in Figure 1, the Match 2 error (falsely concluding the Site is noncompliant, when in fact none of the DUs exceed the AL) ranges from 3.9% to 36.7% when the CV increases from 0.5 to 3.0 for the Parent lognormal distributions. The Match 3 error (falsely concluding the Site is in compliance, when in fact one or more DUs exceed an AL) ranges from 0.3% to 0.7% for the same Parent lognormals, illustrating that Match 3 errors are insensitive to this range of CV s.

ISM Sampling

As shown in Figure 2, Match 2 error range from 1.3% to 8.0% using Student's t UCLs, and 2.1% to 12.2% using Chebyshev UCLs, across Parent lognormals of CV=0.5 to 3.0. The slight increase in false noncompliance associated with the Chebyshev UCL is to be expected given that this method yields greater values than Student's t UCL.

The Match 3 error (falsely concluding the Site is in compliance, when in fact one or more DUs exceed an AL) ranges from 0.6% to 4.1% for Student's t UCL, and 0.5% to 3.3% for Chebyshev UCL. Similar to the Match 2 results, the chance of falsely concluding compliance is lower with Chebyshev UCLs because they are less likely to underestimate the true mean than the Student's t UCL applied to the same data set.

Across the full set of simulations, the Match 3 errors comprise a relatively less than 5 percent of the outcomes for both discrete sampling and ISM sampling. This is one indication that the decision error of primary concern when extrapolating from a Pilot Study is likely to be low for both sampling designs.

Match Outcomes Grouped by Ratio

It is unclear when examining all of the simulation results together if the error rates exhibit patterns when the contamination varies relative to the AL. To evaluate this question, the results can be binned to examine different increments of concentrations relative to ALs. Figures 3 to 8 summarizes the distribution of Match outcomes grouped by intervals of Ratios ranging from less than 0.1 (meaning that the average 95UCL was more than an order of magnitude less than the AL) to greater than 0.6 or more (meaning that the average 95UCL was a factor of at least 0.6 times the AL). The results are very informative, and provide important insights regarding expected error rates that can be guided by a Pilot Study.

When the average concentrations across the Site approach the AL, the Ratio approaches unity. As this happens, it is very unlikely that all of the 95UCL calculations will be less than the AL. Therefore, the dominant outcome is Match 1 – the Pilot Study will reliably conclude that the Site is not in compliance. Likewise, when the concentrations are much lower than the AL (e.g., an order of magnitude), then the dominant outcome is Match 4 – the Pilot Study will reliably conclude that the Site is in compliance. The most challenging situation is when the average 95UCL is moderately lower than the AL.

Discrete Sampling

Figures 3 to 5 summarize the discrete sampling results for Parent lognormals ranging from CV=0.5 to 3.0. The results consistently demonstrate that the Match 3 error rate is less than 3%. The highest error rates occur when the Ratio is between 0.2 and 0.4. Interestingly the

Match 3 error is lowest when the variability of the Parent lognormal distributions is greatest (CV=3.0, Figure 5).

The Match 2 error rate is relatively high using discrete sampling. For Ratios between 0.2 and 0.3, error rates increase from 37% to 79% when the Parent lognormal CV increases from 0.5 to 3.0.

ISM Sampling

Figures 6 to 8 summarize the ISM sampling results for both Student's t and Chebyshev UCL methods. Similar to discrete sampling, for these simulation scenarios, the most critical window where the error rates are highest (both Match 2 and 3) is when the Ratio is between 0.1 and 0.3. This is the region where approximately 90% of the total (combined) errors occur. The following is noted:

- The probability of a Match 3 error increases with increasing CV of the Parent lognormal for both Student's t and Chebyshev methods. For example, for the interval of Ratios between 0.2 and 0.3, errors using Student's t increase from 7.1% for CV=0.5 to 9.3% for CV=1.5 and 9.2% for CV=3.0. When CV=3.0, the interval with the highest error (10.4%) shifts down to Ratio of 0.1 to 0.2.
- Match 3 errors for Chebyshev UCLs follow the same pattern across Ratios as Student's t, but with Match 3 error rates that are consistently 15% to 30% lower. For example, for scenarios with the greatest Parent lognormal variability (CV=3.0), when the average UCL from the Pilot Study is between a factor of 0.1 to 0.2 times the AL, the Match 3 error is 10.4% using Student's t UCL and 7.5% using Chebyshev UCL.
- The probability of a Match 2 error is consistently greater than that of a Match 3 error for each Ratio interval, but lower for ISM than for discrete sampling. For these simulations, the highest Match 2 error rates range from approximately 16% to 32% for all of the scenarios evaluated. Therefore, the Match 2 error using ISM sampling occurs with about half the frequency compared to that of discrete sampling.

Overall, the average 95UCL compared with the AL is a simple but reliable indicator of the likelihood that the findings from a Pilot Study may lead to a decision error regarding the overall Site.

Distribution of Match 2 and Match 3 Errors

Figures 9 to 12 isolate the Match 2 and Match 3 errors and illustrate how they are distributed across the range of Ratios discussed above. For discrete sampling (Figure 9),

the total errors is dominated by Match 2, with 92% to 97% of the error across all Ratios and Parent lognormal distribution CVs.

For ISM sampling, Match 2 errors are also dominant, comprising approximately 70% of the error for Student's *t* UCLs and 80% of the error for Chebyshev UCLs across all Ratios and Parent lognormal distribution CVs.

Magnitude of the Site that Exceeds an AL

A successful random sampling program will generate results from a Pilot Study that can be extrapolated to unsampled areas. In addition to understanding the frequency with which an error may occur, it is also helpful to know if the total area of the Site that exceeds an AL may vary in a predictable manner when an error occurs. To explore this question, we plotted the Total Area of Exceedance (which is easily obtained from the set of true mean concentrations assigned to each DU) by variables that describe the Site conditions and sampling plan.

Pilot Study Percent of Site

Figure 13 shows the relationship between percent of Site that exceeds the AL and the percent of Site included in the Pilot Study for cases when there is a Match 3 error. As expected, there is a slight inverse relationship for both discrete and ISM sampling, such that increasing the Site area sampled yields lower extrapolation errors. Overall, based on the 95% confidence intervals, the average Match 3 error is much less than 5 percent when at least 20% of the Site is characterized. The 95% prediction intervals answer the question about expected error rates for any single sampling plan (rather than the performance of the sampling design on average). In both cases, the upper 95% prediction intervals (95% PI) are also relatively low. For discrete sampling, upper 95% PI for Match 3 errors decrease from 10% to 5% when the Pilot Study comprises 10% to 30% of the Site. For ISM sampling, the upper 95% PI Match 3 errors are slightly greater, decreasing from 25% to 10% when the Pilot Study comprises 10% to 30% of the Site.

These results are for Parent lognormals with CV=1.5 and comparable findings were observed for CV=0.5 and 3.0.

Probability of Higher Concentration Area

Figure 14 shows that the probability of sampling from a Parent lognormal distribution with six times higher concentrations (on average) does not affect the likelihood of Match 3 errors. This example illustrates results for ISM sampling using Student's *t* UCLs and Parent lognormals with CV=1.5. Match 3 error rates were observed to be insensitive to the probability of assigning a DU to a Higher Concentration Area across all simulations in this study.

Proportion of DUs with R=3 ISM Replicates

Figure 15 illustrates that there is no correlation between the total Site area of exceedance and the option sometimes implemented in ISM sampling designs of collecting $r=3$ replicates in a subset of the DUs. It turns out that using the average CVs from those DUs as a surrogate of the CV for other DUs with just $r=1$ is a reasonable step. This finding was consistent across a range of Parent lognormal distribution CVs and percentages of the site characterized by the Pilot Study.

SUMMARY AND RECOMMENDATIONS

All simulations were executed using random sampling, meaning that every DU has an equal chance of representing the same distribution of concentrations. This condition is what statisticians refer to as the “independent and identically distributed” (i.i.d.) assumption. The range of scenarios were intended to reflect the plausible conditions and sampling design options that would occur at Sites typically encountered by site investigation teams.

Extrapolation uncertainty was examined under conditions when a Pilot study yields a false noncompliance conclusion (Match 2) and a more concerning false compliance conclusion (Match 3). Overall, the average 95UCL compared with the AL is a simple but reliable indicator of the likelihood that the findings from a Pilot Study may lead to a decision error regarding the overall Site.

The simulations were successful in the sense that the outcomes support the following set of guidelines based on sample statistics that can be readily calculated from a Pilot Study:

Guidelines on Expected Match 3 Error Based on Calculated Ratio (Average 95UCL / AL)

1. If the average 95UCL is greater than 0.4 times the AL or less than 0.1 times the AL, there is a negligible probability of making a Match 3 error with either discrete sampling or ISM sampling (with both Student's t and Chebyshev UCL methods).
2. When the average 95UCL is between 0.1 and 0.4 times the AL and the underlying variation in concentrations is low (e.g., $CV=0.5$), Match 3 errors are generally less than 5 percent for both discrete and ISM sampling.
3. When the average 95UCL is between 0.1 and 0.4 times the AL and the underlying variation increases (e.g., $CV=3.0$), the Match 3 error rate remains less than 5 percent for discrete sampling and increases with ISM sampling – approaching 8 percent using Chebyshev UCL and 10 percent using Student's t UCL.

Guidelines on Expected Magnitude of Site Error when Match 3 Occurs

4. When a Match 3 error occurs, some portion of the Site exceeds the AL. The fraction that exceeds the AL decreases as the percent of the Site characterized by the Pilot Study increases. On average, with both discrete and ISM sampling, the fraction of the Site that exceeds the AL is less than 5 percent when at least 10 percent of the Site is sampled in the Pilot Study using both discrete and ISM sampling methods.
5. Of the Match 3 errors that occur, when the Pilot Study characterizes 30 percent or more of the Site, the total Site exceedance is likely to be less than 5 percent. This was found to be true for 95 percent of the discrete sample programs and 90 percent of the ISM sampling programs.

Guidelines on Sampling Design Options

6. The likelihood of a Match 2 error is always greater than a Match 3 error. ISM sampling reduces the Match 2 error rate by a factor of two compared with discrete sampling applied to the same Site conditions.
7. Match 3 error rates using ISM sampling are insensitive to the fraction of the Pilot Study DUs that are characterized with $r=3$ replicates, so long as 95UCLs are calculated for the DUs with $r=1$ replicate using the average CV from the sample statistics for the $r=3$ replicate DUs.

This information is consolidated into Table 2, which is intended to serve as a quick reference for site investigation teams.

Table 2. Summary of steps to manage extrapolation uncertainty using Pilot Study results.

Sampling Design			Site	Average UCL / AL	Error Rate		% of Site > AL when Match 3 Error Occurs	
Sample Type	95UCL Method	Pilot % of Site	CV	Ratio	Match 2	Match 3	Average Program	95% of Programs
Discrete	ProUCL	10%	0.5	≤ 0.10	0%	< 3%	< 3%	< 10%
				0.10 – 0.40	5 - 37%			
				≥ 0.40	≤ 1%			
			1.5	≤ 0.10	≤ 7%			
				0.10 – 0.40	40 - 73%			
				≥ 0.40	≤ 15%			
			3.0	≤ 0.10	≤ 12%			
				0.10 – 0.40	50 – 80%			
				≥ 0.40	≤ 30%			
Discrete	ProUCL	30%	0.5	≤ 0.10	0%	< 3%	< 2%	< 5%
				0.10 – 0.40	5 - 37%			
				≥ 0.40	≤ 1%			
			1.5	≤ 0.10	≤ 7%			
				0.10 – 0.40	40 - 73%			
				≥ 0.40	≤ 15%			
			3.0	≤ 0.10	≤ 12%			
				0.10 – 0.40	50 – 80%			
				≥ 0.40	≤ 30%			
ISM	Student's t	10%	0.5	≤ 0.10	0%	< 1%	< 7%	< 25%
				0.10 – 0.40	1 – 16%	1 – 7%		
				≥ 0.40	< 2%	< 2%		
			1.5	≤ 0.10	< 1%	< 1%		
				0.10 – 0.40	5 – 26%	3 – 10%		
				≥ 0.40	< 1%	< 2%		
			3.0	≤ 0.10	< 3%	< 3%		
				0.10 – 0.40	3 – 25%	3 – 10%		
				≥ 0.40	0%	< 3%		
ISM	Student's t	30%	0.5	≤ 0.10	0%	< 1%	< 3%	< 10%
				0.10 – 0.40	1 – 16%	1 – 7%		
				≥ 0.40	< 2%	< 2%		
			1.5	≤ 0.10	< 1%	< 1%		
				0.10 – 0.40	5 – 26%	3 – 10%		
				≥ 0.40	< 1%	< 2%		
			3.0	≤ 0.10	< 3%	< 3%		
				0.10 – 0.40	3 – 25%	3 – 10%		
				≥ 0.40	0%	< 3%		

Table 2 (continued). Summary of steps to manage extrapolation uncertainty using Pilot Study results.

Sampling Design			Site	Average UCL / AL	Error Rate		% of Site > AL when Match 3 Error Occurs	
Sample Type	95UCL Method	Pilot % of Site	CV	Ratio	Match 2	Match 3	Average Program	95% of Programs
ISM	Chebyshev	10%	0.5	≤ 0.10	0%	0%	< 7%	< 25%
				0.10 – 0.40	1 – 25%	1 – 5%		
				≥ 0.40	< 3%	< 3%		
			1.5	≤ 0.10	< 2%	< 1%		
				0.10 – 0.40	5 – 32%	1 – 6%		
				≥ 0.40	< 1%	0%		
			3.0	≤ 0.10	< 3%	< 3%		
				0.10 – 0.40	3 – 25%	3 – 10%		
				≥ 0.40	0%	< 3%		
ISM	Chebyshev	30%	0.5	≤ 0.10	0%	0%	< 3%	< 10%
				0.10 – 0.40	1 – 25%	1 – 5%		
				≥ 0.40	< 3%	< 3%		
			1.5	≤ 0.10	< 2%	< 1%		
				0.10 – 0.40	5 – 32%	1 – 6%		
				≥ 0.40	< 1%	0%		
			3.0	≤ 0.10	< 3%	< 3%		
				0.10 – 0.40	3 – 25%	3 – 10%		
				≥ 0.40	0%	< 3%		

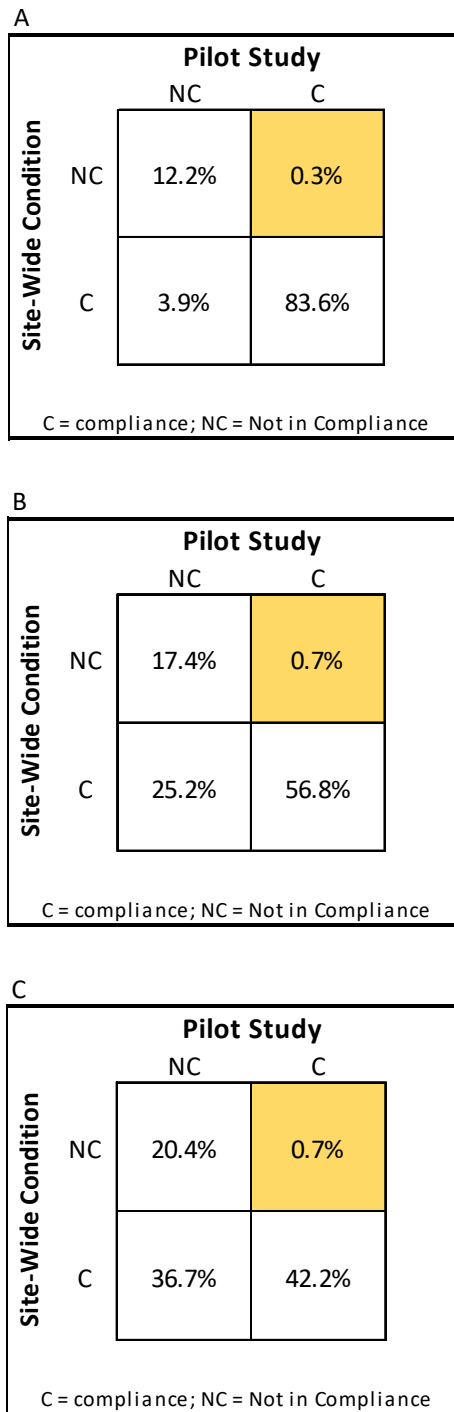


Figure 1. Match results for all 10,000 iterations of Discrete sampling using Parent lognormals with (A) CV=0.5, (B) CV=1.5, and (C) CV=3.0.

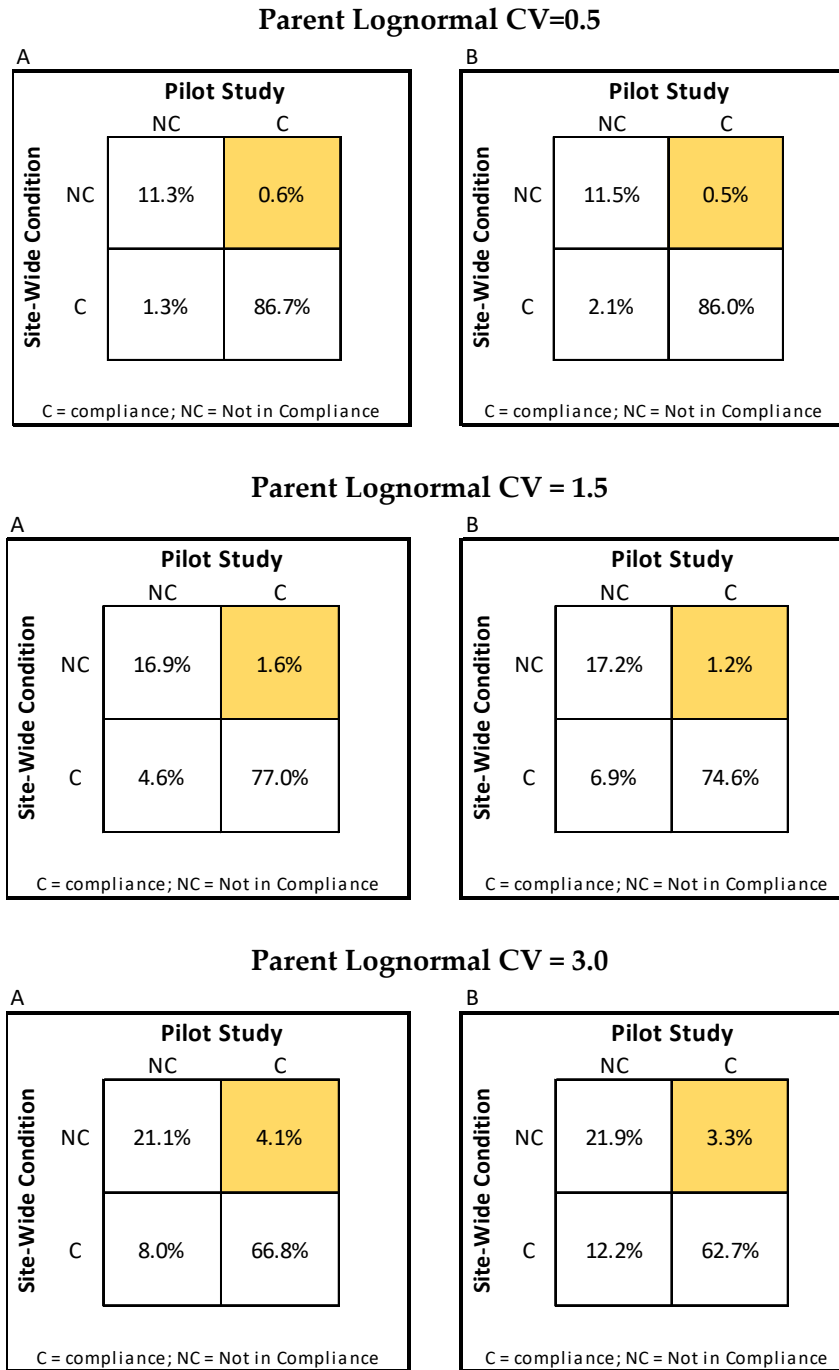


Figure 2. Match results for all 10,000 iterations of ISM sampling using Parent lognormals with CV=0.5, 1.5, and 3.0. For each, (A) Student's t UCL method; (B) Chebyshev UCL method.

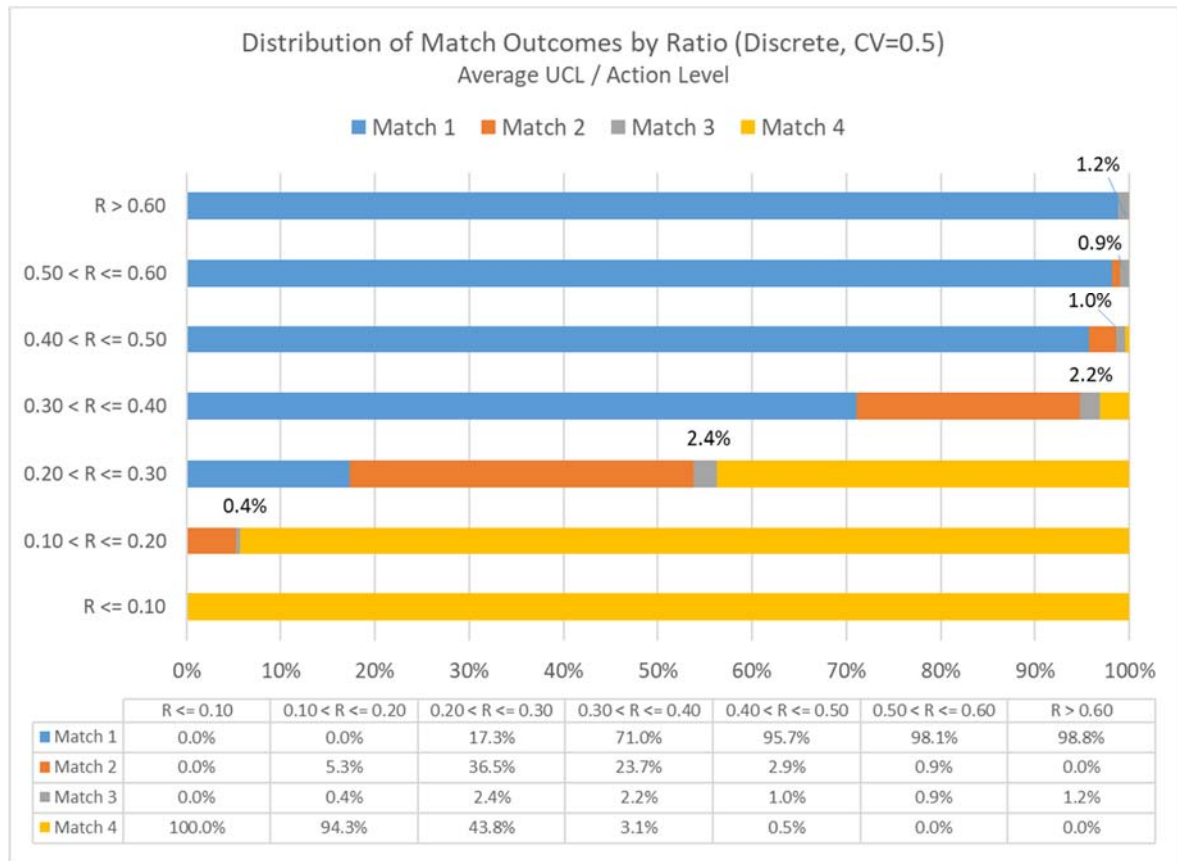


Figure 3. Distribution of simulation outcomes for Discrete sampling, grouping results by the ratio of the average UCL divided by the Action Level. Scenario reflects Parent lognormals using CV=0.5.

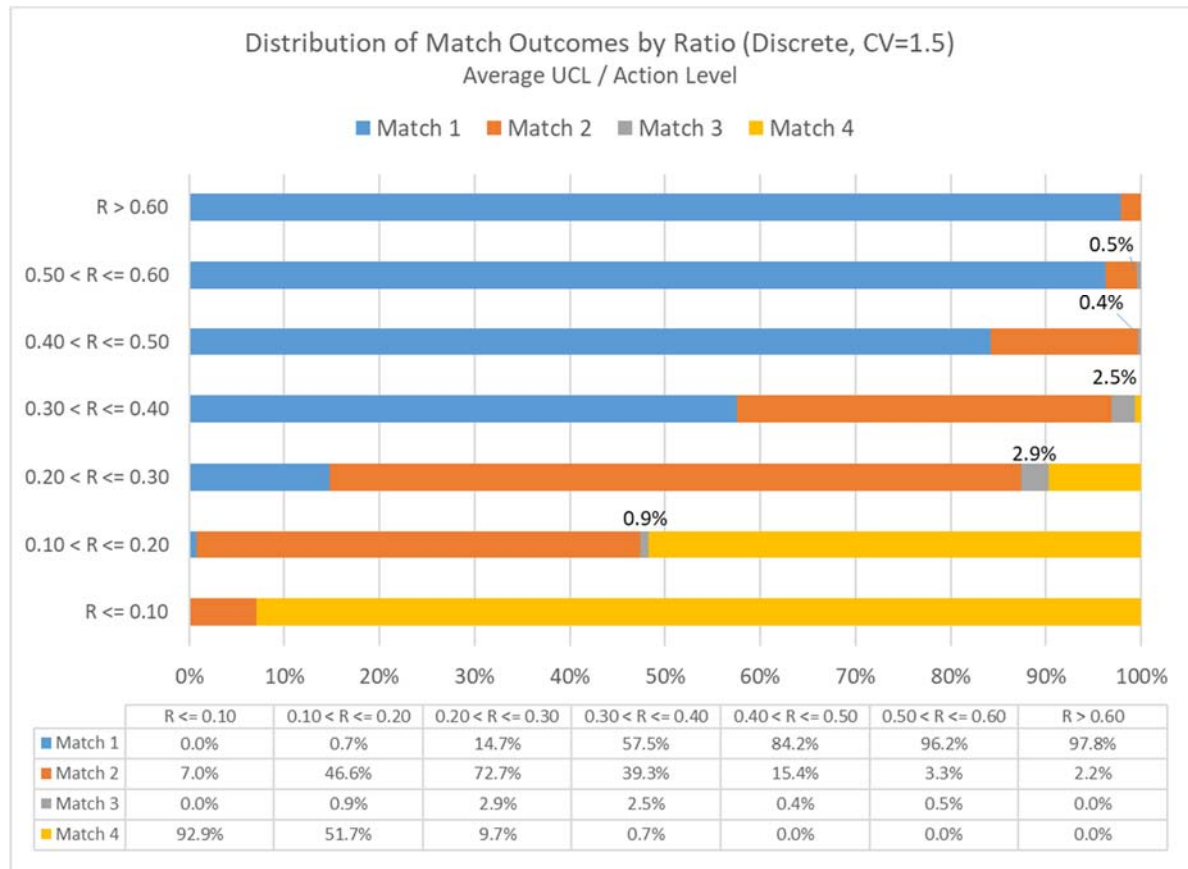


Figure 4. Distribution of simulation outcomes for Discrete sampling, grouping results by the ratio of the average UCL divided by the Action Level. Scenario reflects Parent lognormals using CV=1.5.

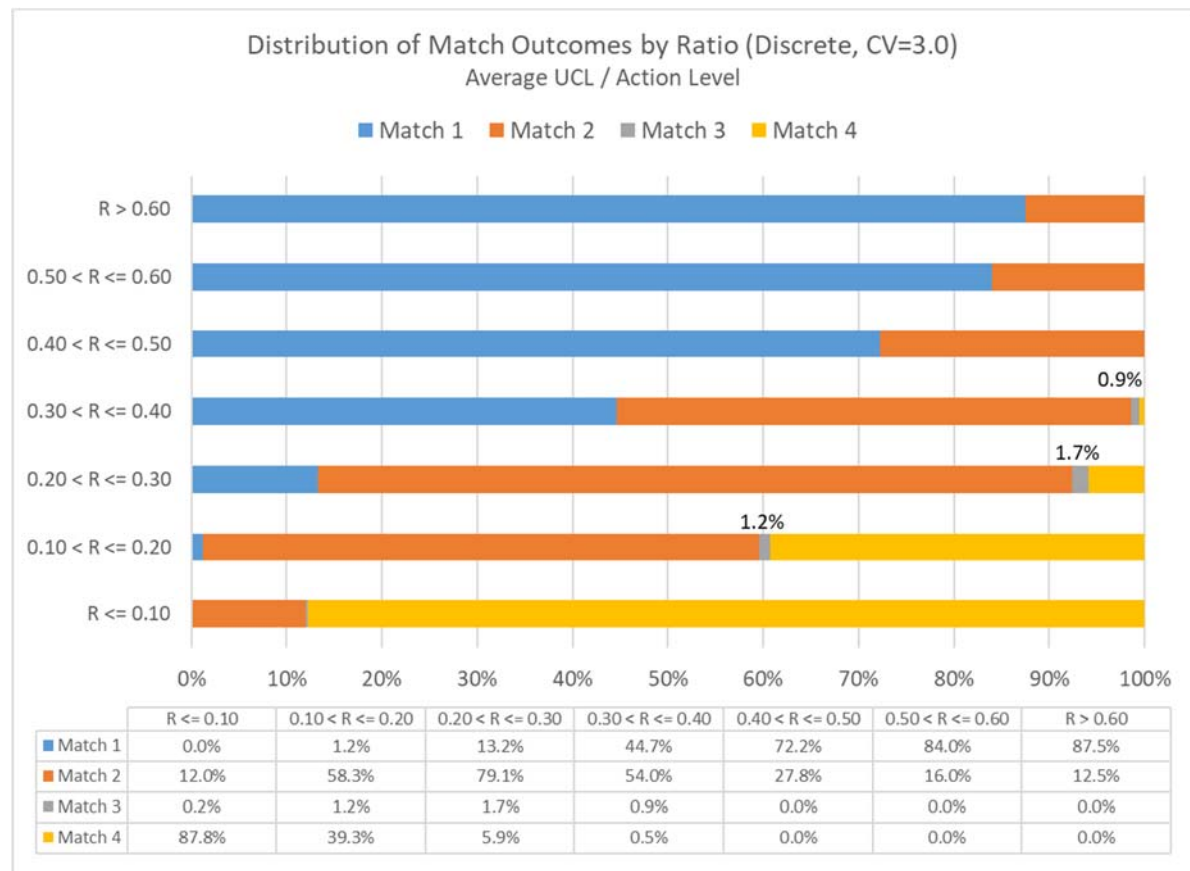


Figure 5. Distribution of simulation outcomes for Discrete sampling, grouping results by the ratio of the average UCL divided by the Action Level. Scenario reflects Parent lognormals using CV=3.0.

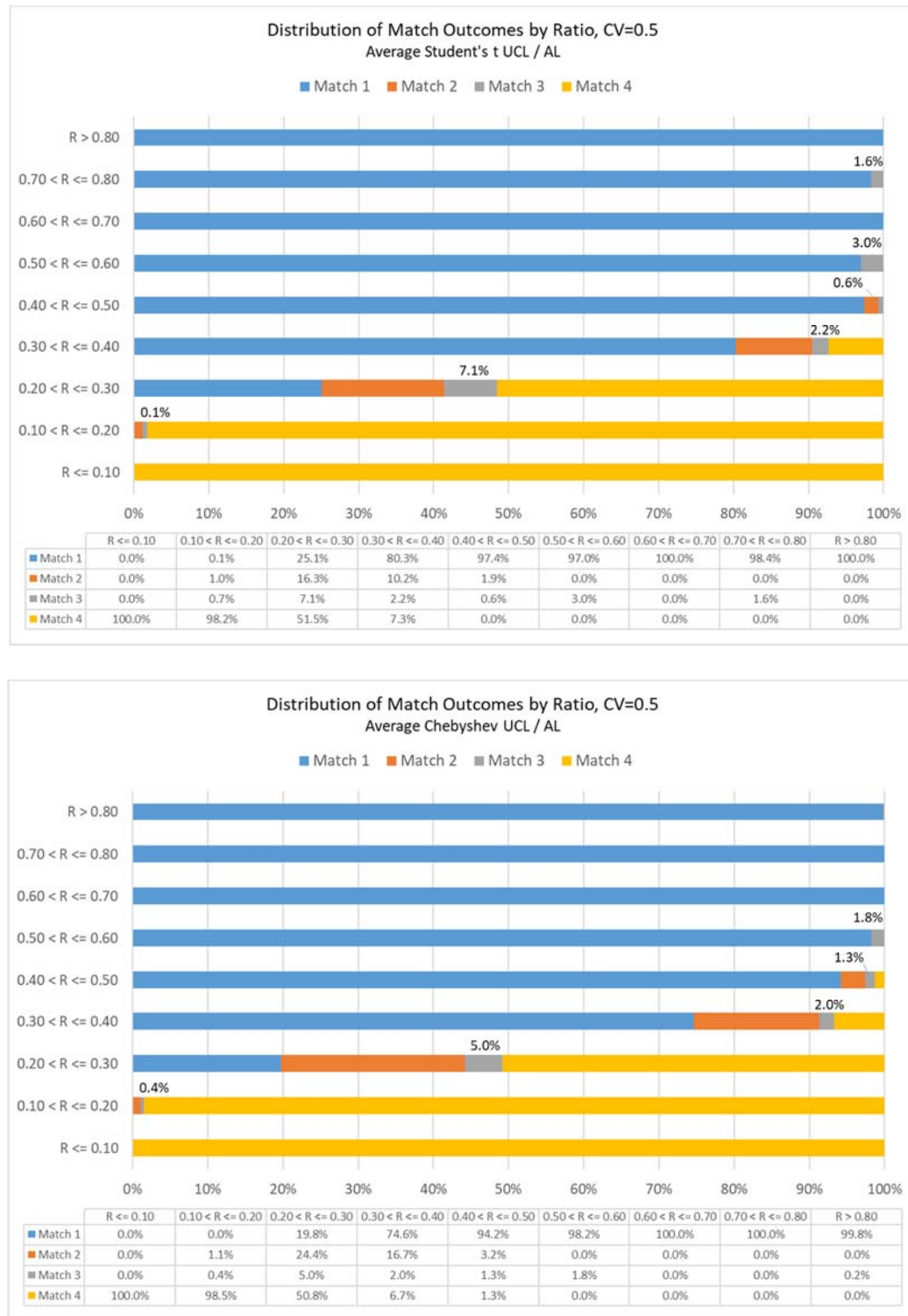


Figure 6. Distribution of simulation outcomes for ISM sampling, grouping results by the ratio of the average UCL divided by the Action Level. Student's t UCL (top figure) and Chebyshev UCL (bottom figure); both scenarios reflect Parent lognormals using CV=0.5.

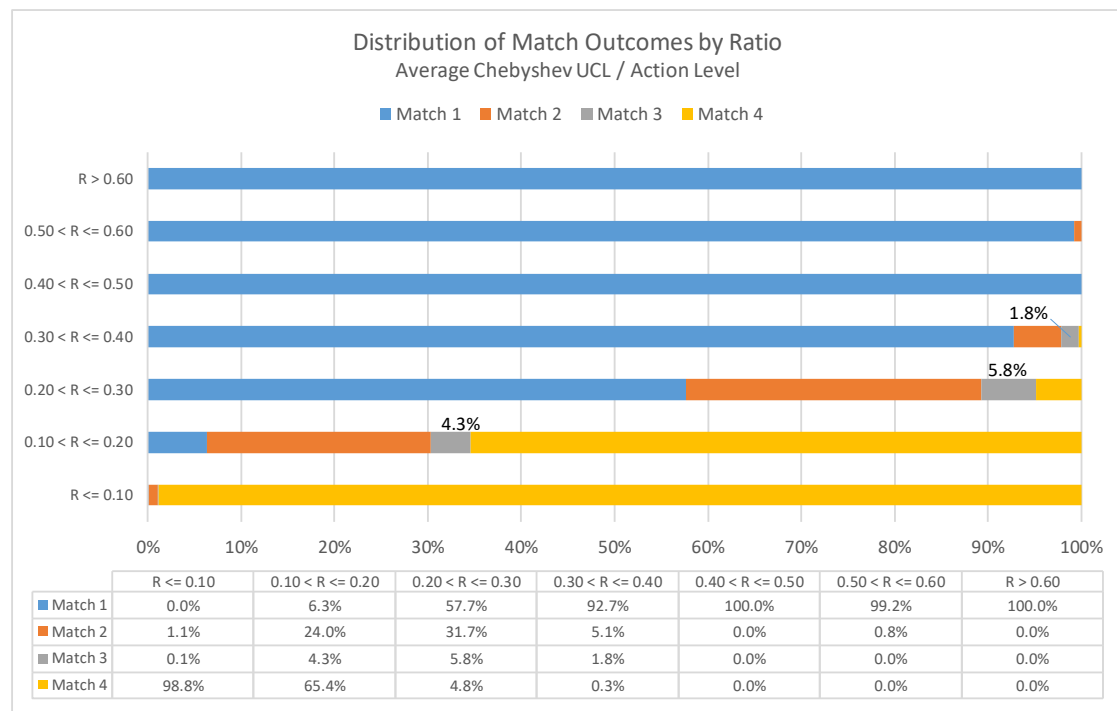
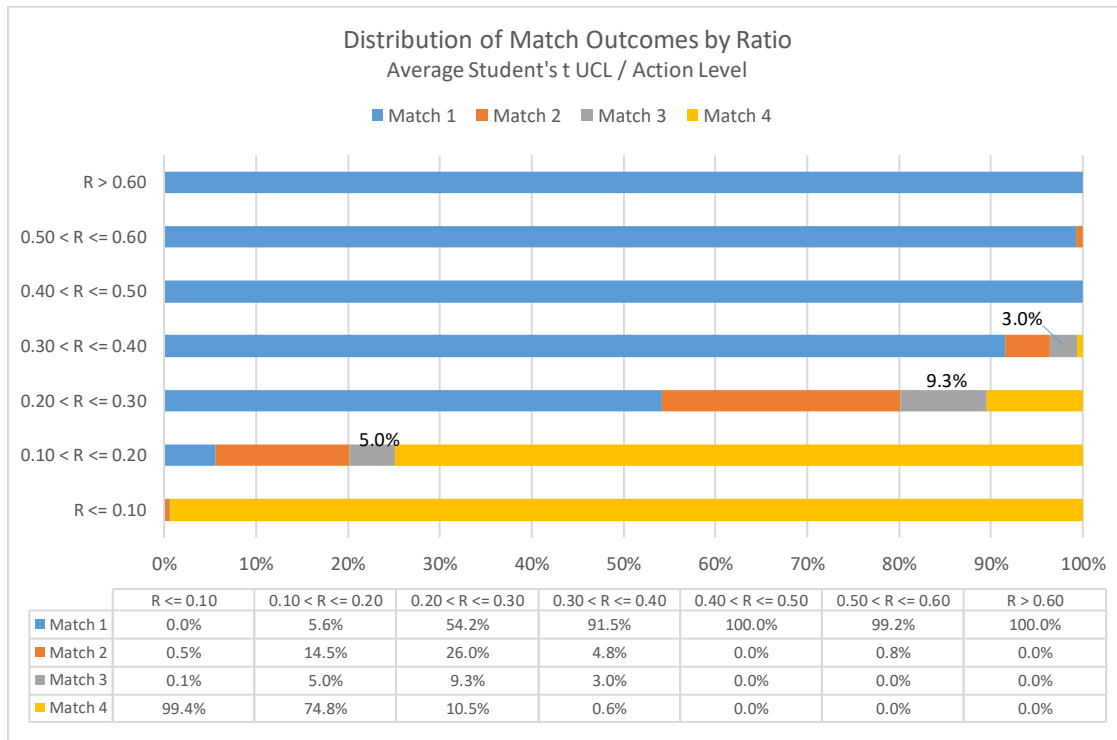


Figure 7. Distribution of simulation outcomes for ISM sampling, grouping results by the ratio of the average UCL divided by the Action Level. Student's t UCL (top figure) and Chebyshev UCL (bottom figure); both scenarios reflect Parent lognormals using CV=1.5.

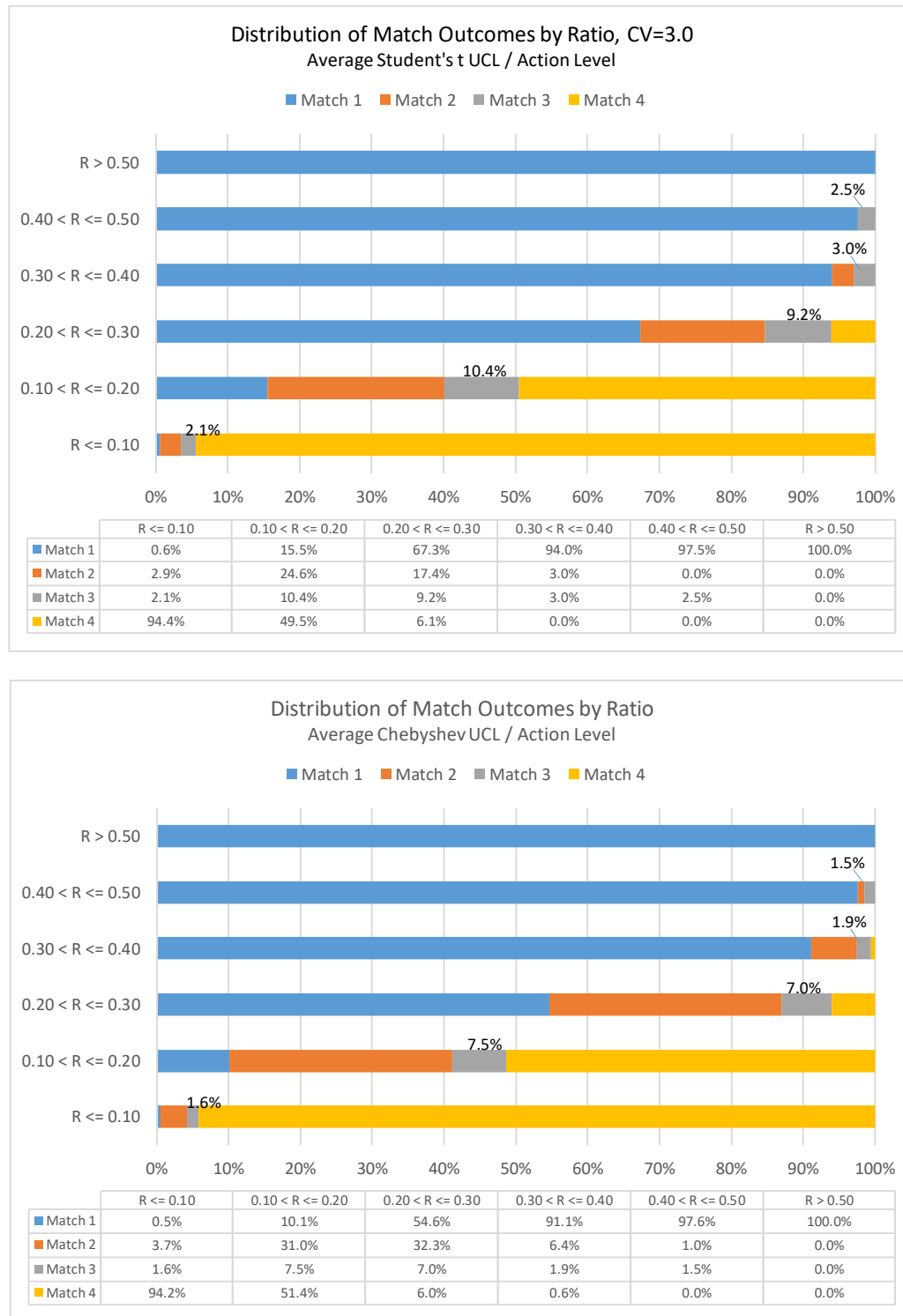


Figure 8. Distribution of simulation outcomes for ISM sampling, grouping results by the ratio of the average UCL divided by the Action Level. Student's t UCL (top figure) and Chebyshev UCL (bottom figure); both scenarios reflect Parent lognormals using CV=3.0.

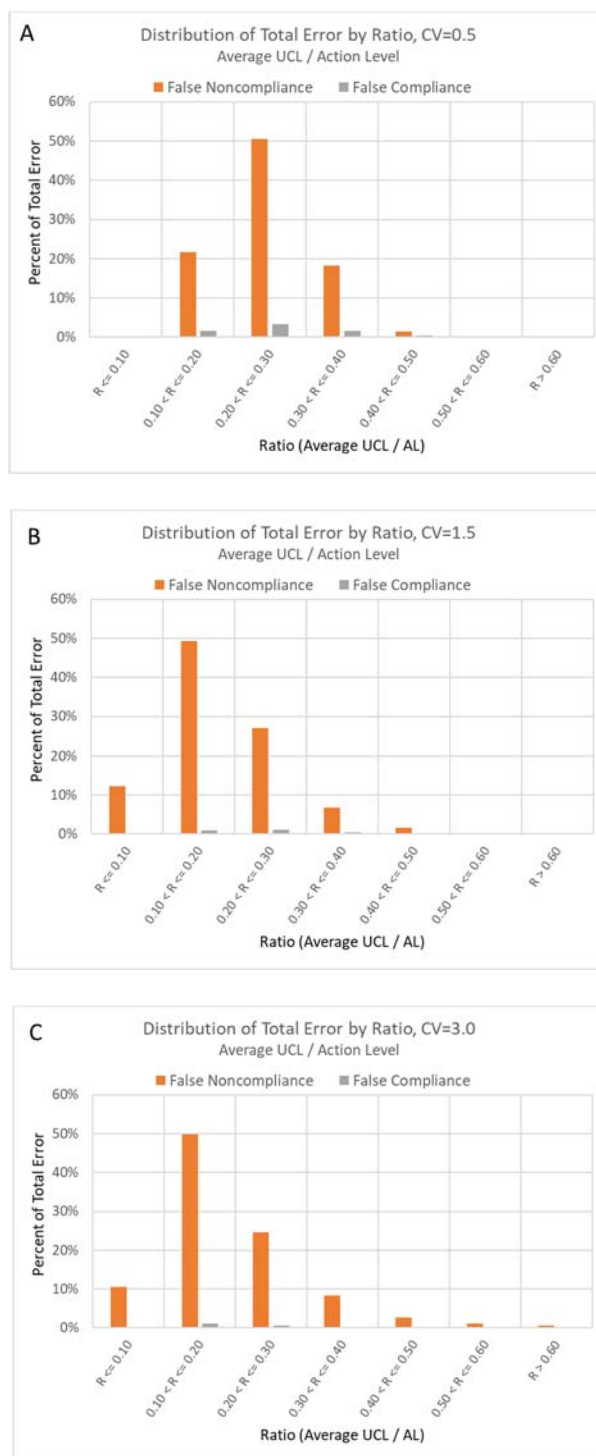
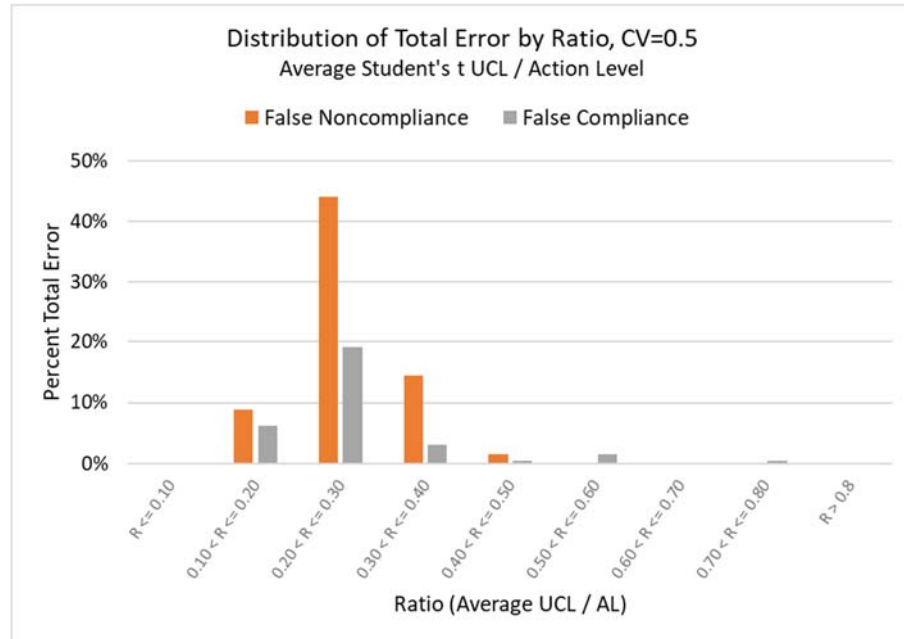


Figure 9. Distribution of total error from Pilot Study simulations of Discrete sampling, including both false conclusions of noncompliance (Match 2) and compliance (Match 3). Results given here are for parent lognormals with (A) CV=0.5; (B) CV=1.5; (C) CV=3.0.

A



B

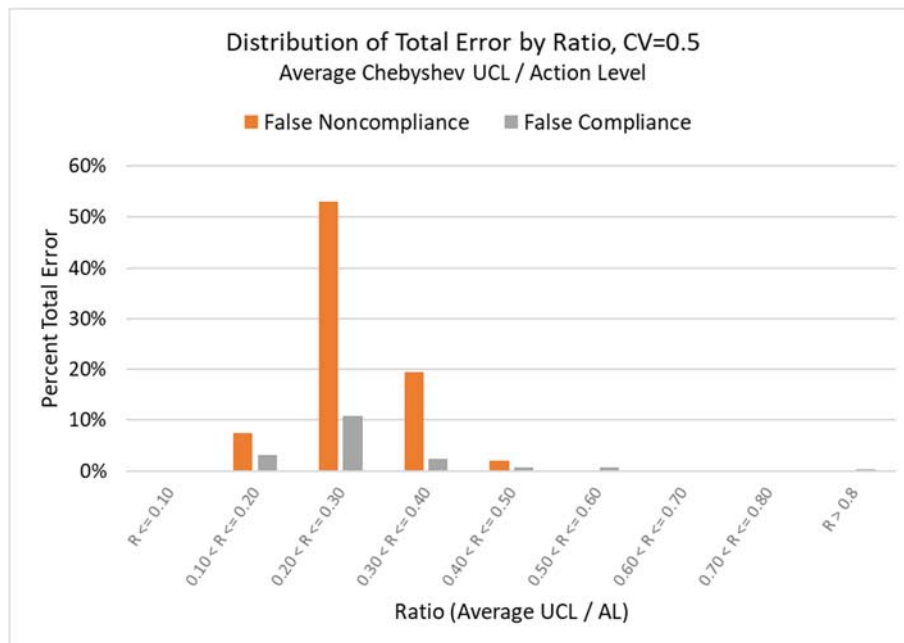
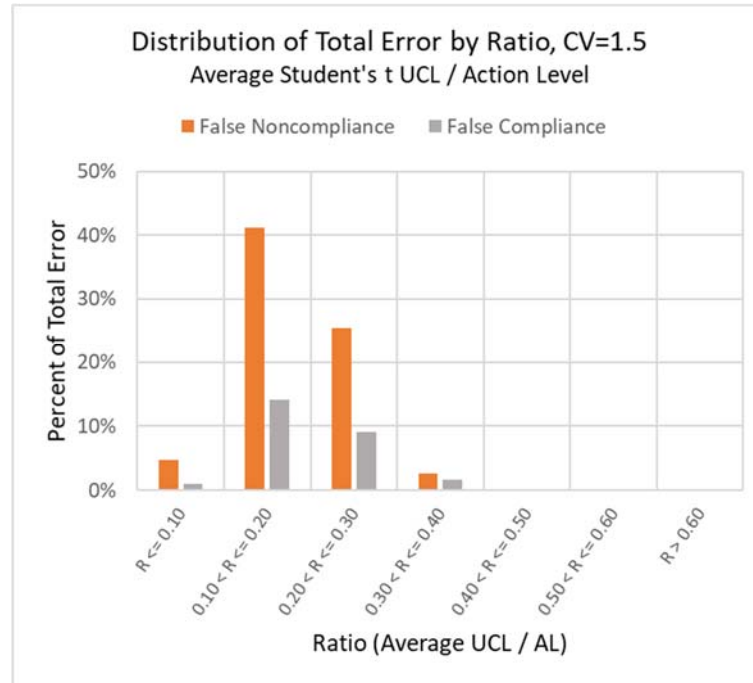


Figure 10. Distribution of total error from Pilot Study simulations of ISM sampling, including both false conclusions of noncompliance (Match 2) and compliance (Match 3). Results given here are for Parent lognormals with CV=0.5. (A) Student's t UCL; (B) Chebyshev UCL.

A



B

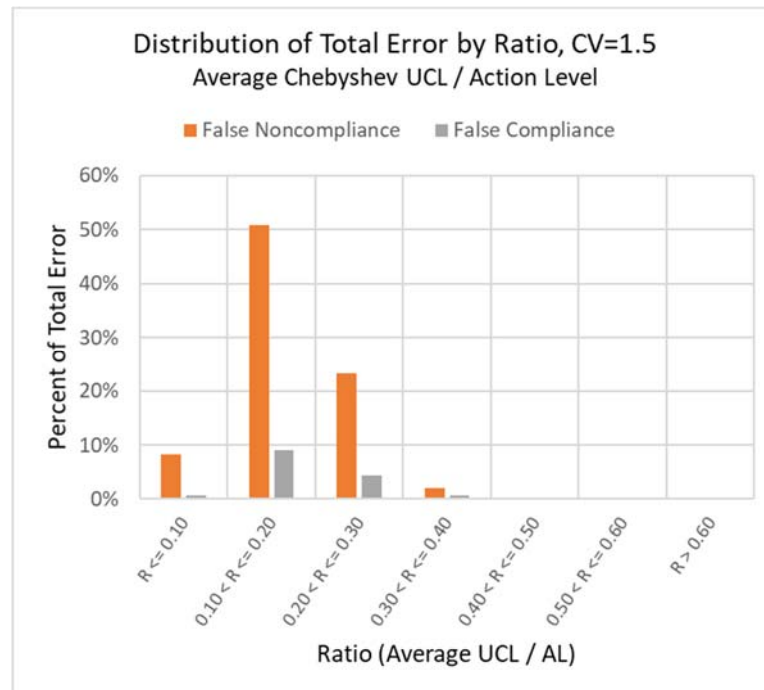
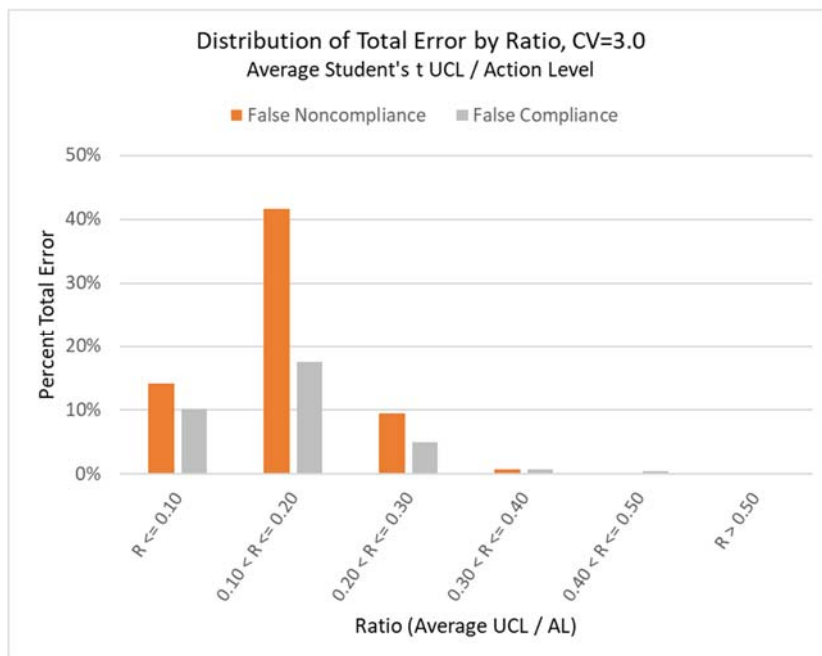


Figure 11. Distribution of total error from Pilot Study simulations of ISM sampling, including both false conclusions of noncompliance (Match 2) and compliance (Match 3). Results given here are for Parent lognormals with CV=1.5. (A) Student's t UCL; (B) Chebyshev UCL.

A



B

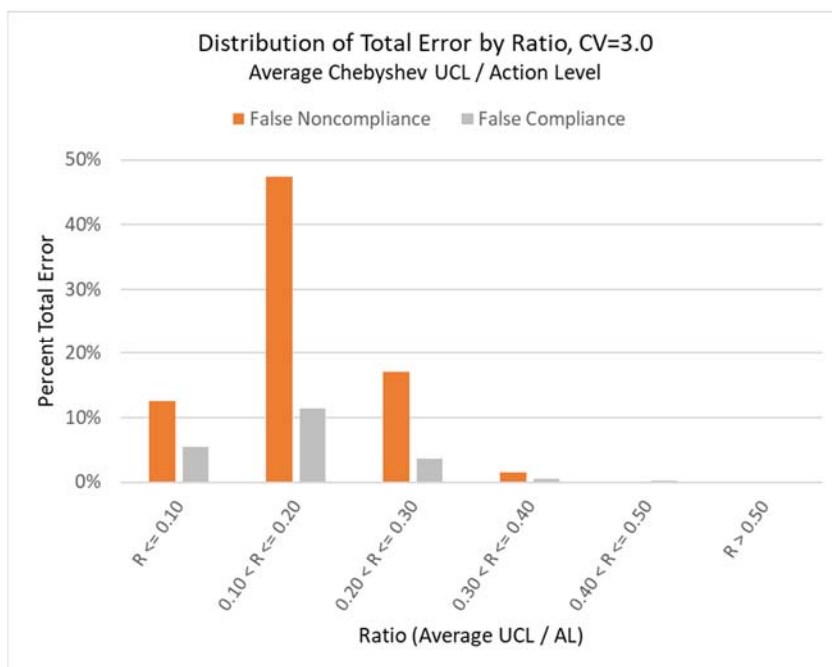
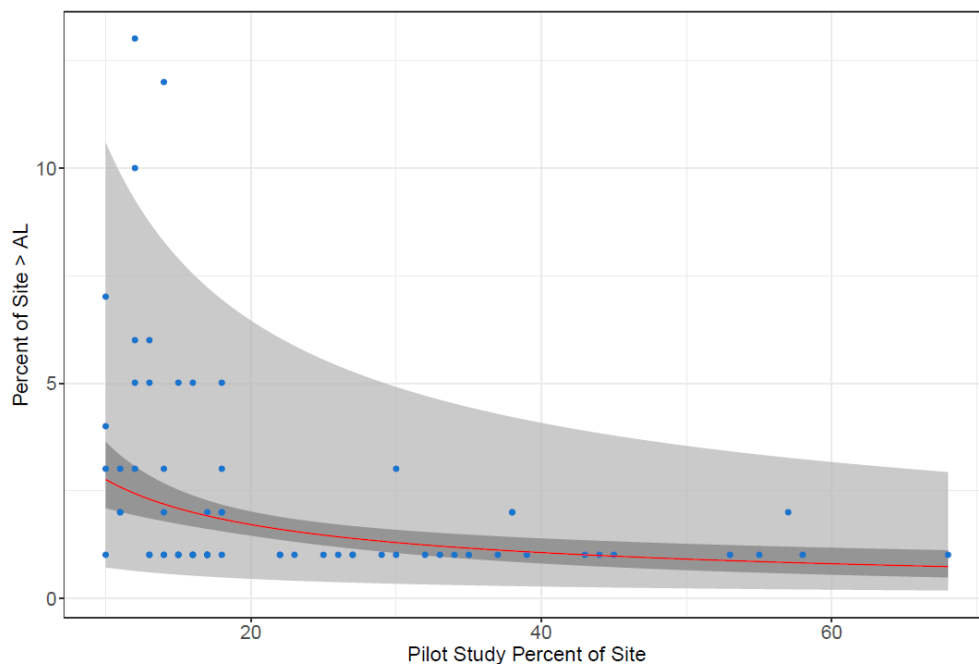


Figure 12. Distribution of total error from Pilot Study simulations of ISM sampling, including both false conclusions of noncompliance (Match 2) and compliance (Match 3). Results given here are for Parent lognormals with CV=3.0. (A) Student's t UCL; (B) Chebyshev UCL.

A



B

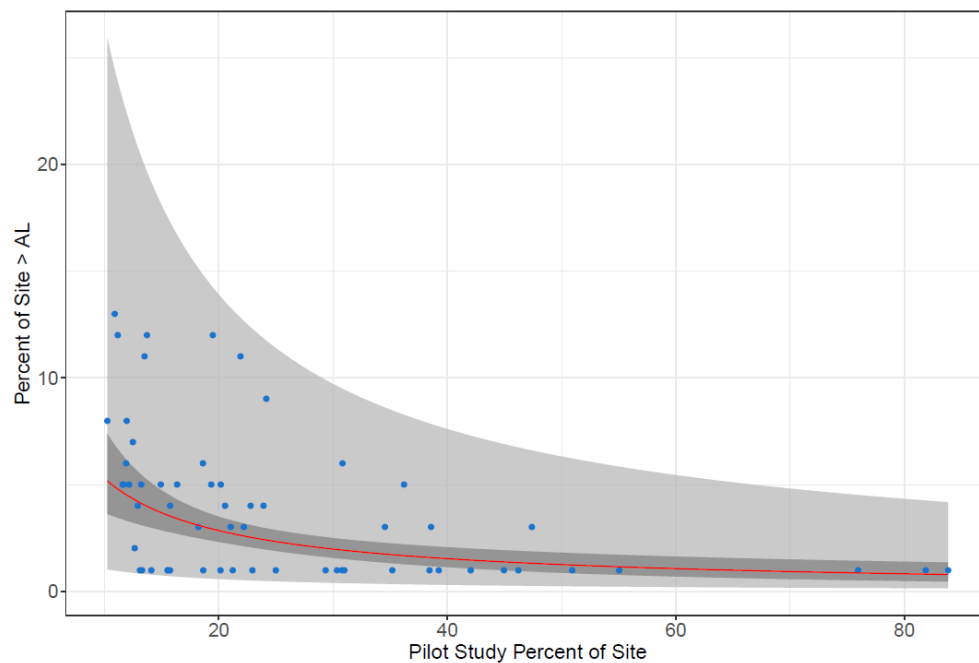


Figure 13. Relationship between percent of Site that exceeds the action level and the percent of Site included in the Pilot Study. Results are for false compliance errors (Match 3) only and Parent lognormal with CV=1.5. (A) Discrete sampling; (B) ISM sampling using Student's t UCL. Dark gray shade is the 95% confidence interval and light gray is the 95% prediction interval.

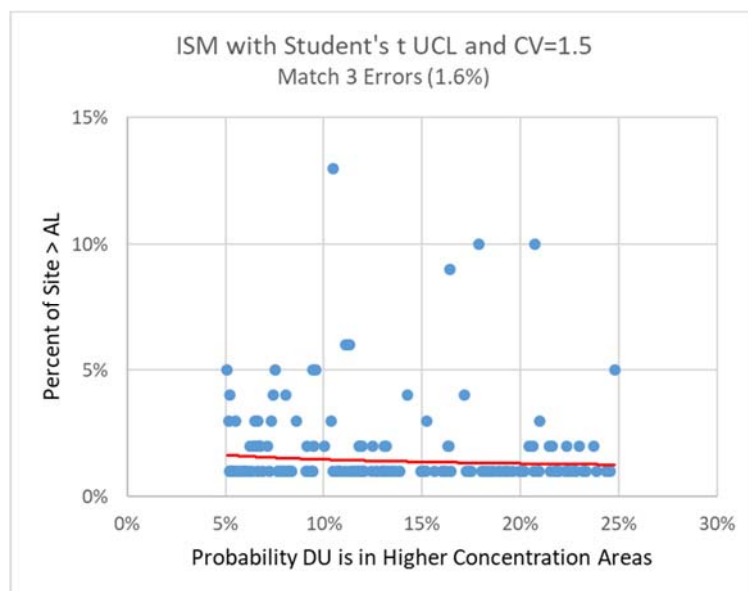


Figure 14. Relationship between percent of Site that exceeds the action level and the probability that any DU is assigned to a Higher Concentration Area. Results are for false compliance errors (Match 3) only, using ISM sampling, Parent lognormals with CV=1.5, and Student's t UCL.

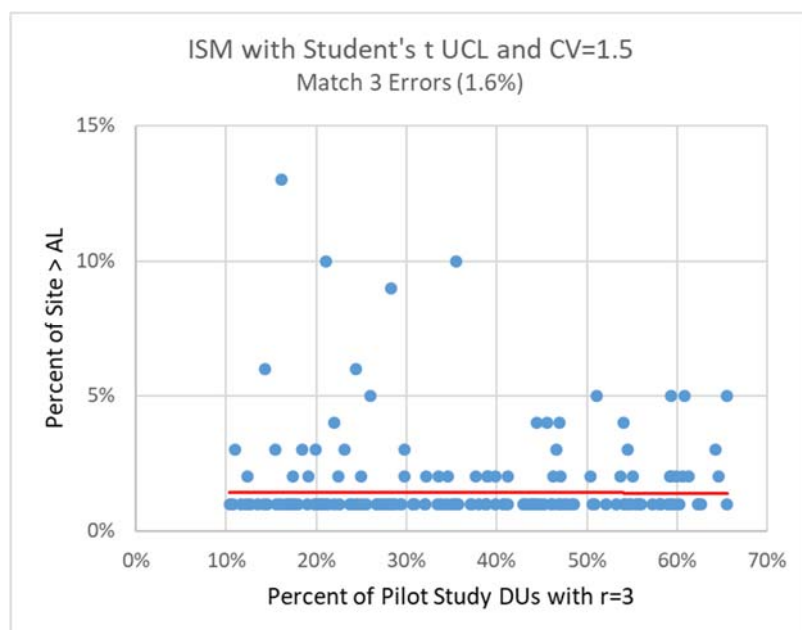


Figure 15. Relationship between percent of Site that exceeds the action level and the percent of Pilot Study DUs that are characterized by r=3 replicates. Results are for false compliance errors (Match 3) only, using ISM sampling, Parent lognormals with CV=1.5, and Student's t UCL.

REFERENCES

Bujewski, G.E. and R.L. Johnson. 1996. Adaptive Sampling Approach to Environmental Site Characterization at Joliet Army Ammunition Plant: Phase II Demonstration. Sandia National Laboratories Report SAND96-0840. Prepared for the Department of Energy. April.

Crumbling, D. 2004. The Triad Approach to Managing Decision Uncertainty for Better Cleanup Projects. United States Environmental Protection Agency, Office of Superfund Remediation and Technology Innovation. Presentation at Region 3 RCRA CA Workshop, September 14.

ITRC (Interstate Technology & Regulatory Council). 2012. Incremental Sampling Methodology. ISM-1. Washington, DC: Interstate Technology & Regulatory Council, Incremental Sampling Methodology Team. Available at: www.itrcweb.org.

Oracle. 2015. Crystal Ball software release 11.1.2.4.400 (32 bit). Available at: <https://www.oracle.com/applications/crystalball/crystal-ball-product.html>.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>.

U.S. DOE (United States Department of Energy). 2001. Adaptive Sampling and Analysis Programs (ASAPs). Characterization, Monitoring, and Sensor Technology, Crosscutting Program and Subsurface Contaminants Focus Area. Office of Environmental Management, Office of Science and Technology. August.

U.S. EPA (United States Environmental Protection Agency). 2002. Guidance on Choosing a Sampling Design for Environmental Data Collection – for Use in Developing a Quality Assurance Project Plan. EPA QA/G-5S. Office of Environmental Information. Washington, DC. EPA/240/R-02/005.

U.S. EPA. 2003. Using the Triad Approach to Streamline Brownfields Site Assessment and Cleanup. Brownfields Technology Primer Series. Office of Solid Waste and Emergency Response, Brownfields Technology Support Center, Washington, DC. June.

U.S. EPA. 2013. ProUCL Version 5.0.00 Technical Guide. Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. Office of Research and Development, Washington, DC. EPA/600/R-07/041.

ACRONYMS AND ABBREVIATIONS

95UCL = 95 percent upper confidence limit for the arithmetic mean

AL = action level

AM = arithmetic mean

CV = coefficient of variation

DU = decision unit

ISM = incremental sampling method

log = lognormal probability distribution

Match = outcome after comparing the decision based on the Pilot Study to the “true” Site condition

r = replicates for an ISM program; each replicate yields a single estimate of the arithmetic mean

SD = standard deviation

U = uniform probability distribution