APPENDIX H – VISUAL SAMPLING PLAN DESIGN REPORT

VSP Sample Design Report for Calculating a One-Sided Confidence Interval for the Population Mean Using Simple Random Sampling

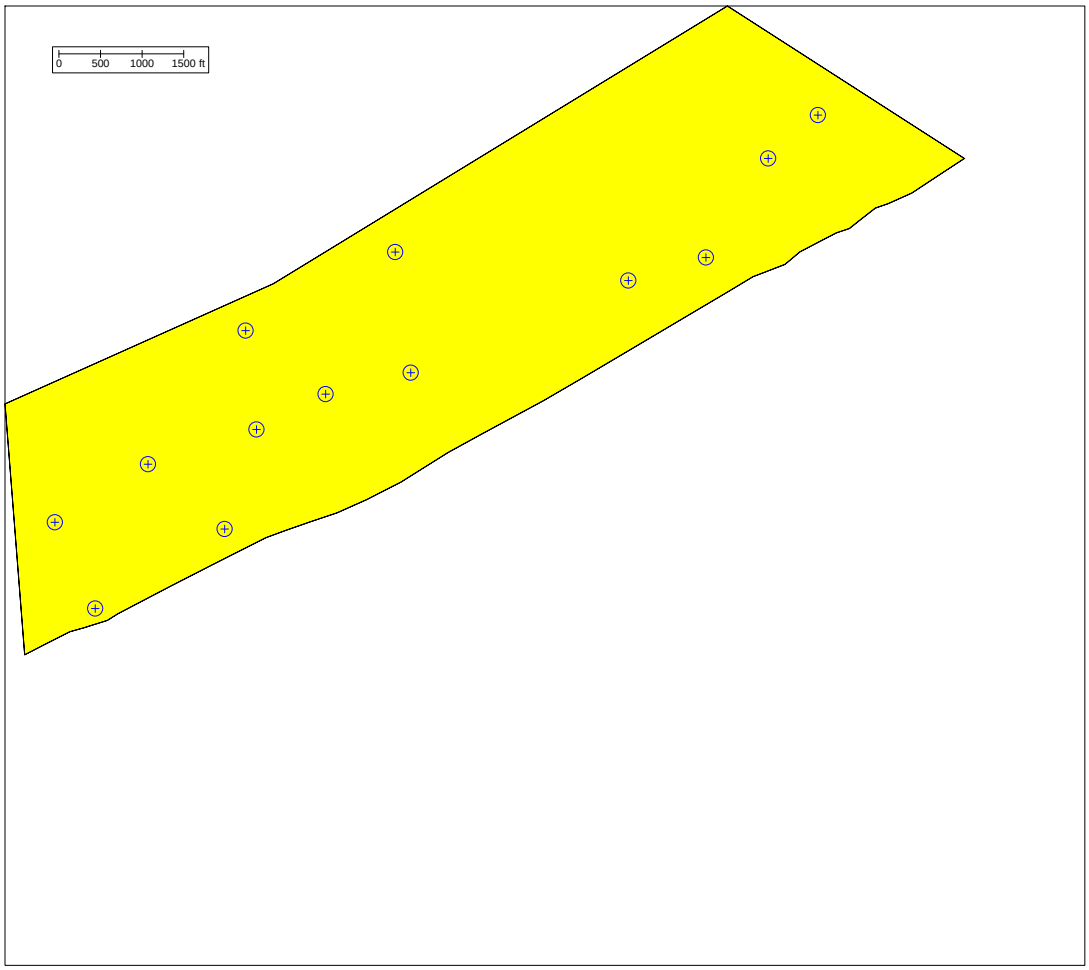
Summary

This report summarizes the sampling design used, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design developed. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Construct a Confidence Interval on the True Mean
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Formula for calculating number of sampling locations	Confidence Limits using Student's t-distribution
Calculated total number of samples	13
Number of samples on map [a]	13
Number of selected sample areas [b]	1
Specified sampling area [c]	32737508.52 ft^2
Total cost of sampling [d]	$2,570.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	**Y Coord**	**Label**	**Value**	**Type**	**Historical**
355086.6696	4478506.6842			Random	
356753.4280	4480639.7893			Random	
359556.1732	4480296.7862			Random	
352663.2171	4477391.3725			Random	
353148.8946	4476355.0051			Random	
361833.5604	4482286.0953			Random	
354956.4125	4479694.6848			Random	
354704.3989	4477309.8984			Random	
353783.0164	4478088.9754			Random	
356939.2373	4479188.8947			Random	
361235.9092	4481764.3539			Random	
360488.6264	4480573.7179			Random	
355917.5911	4478932.5259			Random	

Primary Sampling Objective

The primary purpose of sampling at this site is to construct a confidence interval on the true population mean value. After the samples are collected and analyzed, the resulting sample values can be used to construct a one-sided confidence

interval. Once the confidence interval is computed (which will be an upper threshold), you can have the specified confidence that the true population mean is less than the upper threshold.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are true. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric equations rely on assumptions about the population. Typically, however, non-parametric equations require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than if a non-parametric equation was used.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a confidence interval calculation using the Student's t-distribution. The formula used to calculate the number of samples is:

$$n = \left[\frac{t_{1-\alpha,df}\, S_{total}}{d} \right]^2$$

where
n is the recommended minimum sample size for the study area,
S_{total} is the estimated standard deviation due to both sampling and analytical variability,
α is the maximum acceptable probability that the true mean will not lie in the confidence interval (the confidence level is 1-α),
d is the width of the confidence interval,
$t_{1-\alpha,df}$ is the value of the Student's t-distribution with $df=n-1$ degrees of freedom such that the proportion of the distribution less than $t_{1-\alpha}$ is 1-α.

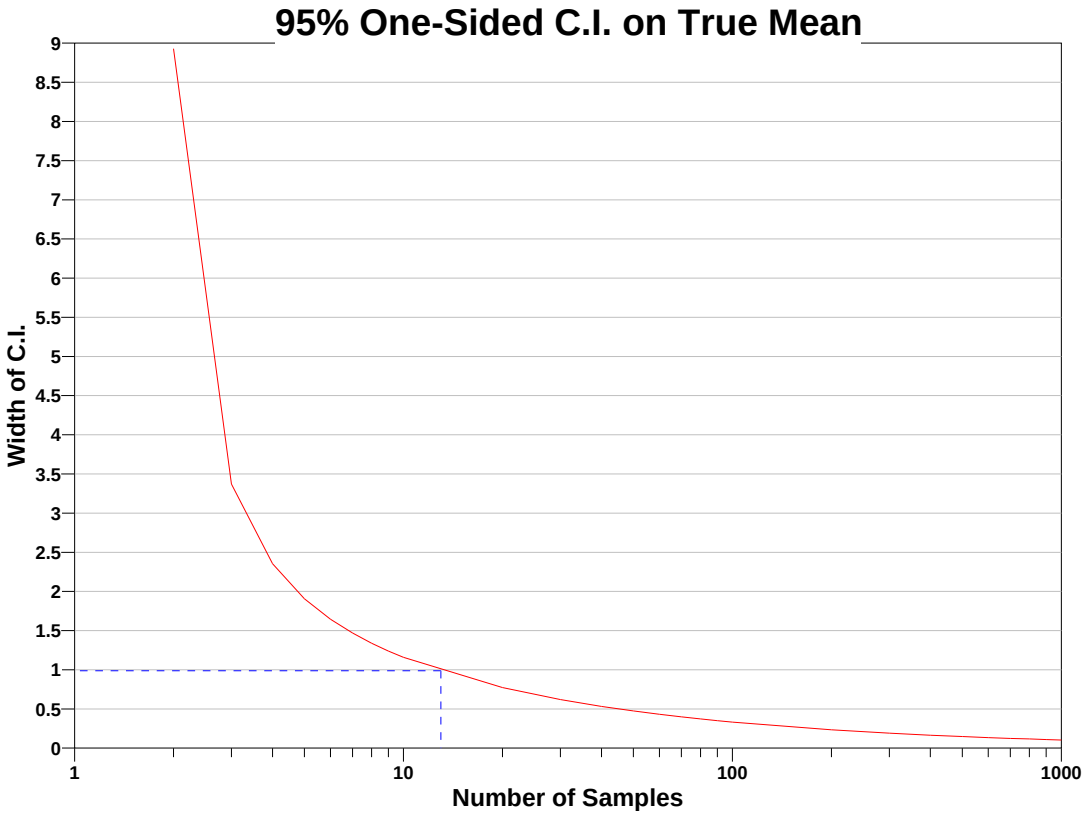
Because n appears on both sides of the equation (on the right side it appears in the degrees of freedom of the t-statistic), the equation must be solved iteratively. VSP does this automatically using the iteration scheme in Gilbert (1987, pg. 32).

The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter			
		S	**d**	**α**	**$t_{1-\alpha,df}$**
Pb	13	2 ppm	1 ppm	5%	1.78229 [a]

[a] This value is automatically calculated by VSP based upon the user defined value of α

The following figure is a graph representing the relationship between the width of the confidence interval and the number of samples. The blue dashed line illustrates the specified maximum desirable confidence interval width. Where this dashed line intersects the red curve is the number of samples calculated by VSP.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed,
2. the population values are not spatially or temporally correlated, and
3. the sampling locations will be selected randomly.

The first two assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, confidence level (1-α) (%) and width of confidence interval. The following table shows the results of this analysis.

Number of Samples						
	d=0.5		d=1		d=1.5	
	s=4	s=2	s=4	s=2	s=4	s=2
CL=99	350	90	90	25	42	14
CL=97	229	59	59	17	28	9
CL=95	176	46	46	13	22	8
CL=93	141	37	37	11	18	7
CL=91	117	31	31	9	15	6

s = Standard Deviation
CL = Confidence Level (1-α) (%)
d = Width of Confidence Interval

Cost of Sampling

The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the

estimated total cost of sampling and analysis at this site is $2,570.00, which averages out to a per sample cost of $197.69. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**13 Samples**
Field collection costs		$40.00	$520.00
Analytical costs	$150.00	$150.00	$1,950.00
Sum of Field & Analytical costs		**$190.00**	**$2,470.00**
Fixed planning and validation costs			$100.00
Total cost			**$2,570.00**

Recommended Data Analysis Activities

Post data collection activities generally follow those outlined in EPA's Guidance for Data Quality Assessment (EPA, 2000). The data analysts will become familiar with the context of the problem and goals for data collection and assessment. The data will be verified and validated before being subjected to statistical or other analyses. Graphical and analytical tools will be used to verify to the extent possible the assumptions of any statistical analyses that are performed as well as to achieve a general understanding of the data. The data will be assessed to determine whether they are adequate in both quality and quantity to support the primary objective of sampling.

Because the primary objective for sampling for this site is to compute a confidence interval, the data should be assessed in this context. Assuming the data are adequate, at least one statistical test should be done to evaluate whether the data are normally distributed. Appropriate confidence intervals for the mean value should then be calculated. Results of the exploratory and quantitative assessments of the data should be reported, along with conclusions that may be supported by them.