INTEGRATED DISPOSAL FACILITY
ADDENDUM HA.a
VISUAL SAMPLING PLAN REPORT DOCUMENTATION
CHANGE CONTROL LOG

Change Control Logs ensure that changes to this unit are performed in a methodical, controlled, coordinated, and transparent manner. Each unit addendum will have its own change control log with a modification history table. The "**Modification Number**" represents Ecology's method for tracking the different versions of the permit. This log will serve as an up to date record of modifications and version history of the unit.

Modification History Table

Modification Date	Modification Number

This page intentionally left blank.

INTEGRATED DISPOSAL FACILITY
ATTACHMENT HA.a
VISUAL SAMPLING PLAN REPORT DOCUMENTATION

This page intentionally left blank.

ATTACHMENT HA.a
VISUAL SAMPLING PLAN REPORT DOCUMENTATION

TABLE OF CONTENTS

This page intentionally left blank.

HA.a.1 VISUAL SAMPLING PLAN REPORT DOCUMENTATION FOR THE INTEGRATED DISPOSAL FACILITY TREATMENT PAD[1]

Systematic Sampling Locations for Comparing a Median with a Fixed Threshold (Monparametric – Multi-Agency Radiation Survey and Site Investigation Manual [MARSSIM])

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 (Table HA.a-1) summarizes the sampling design developed. A figure that shows sampling locations in the field (Figure HA.a-1) and a table that lists sampling location coordinates (Table HA.a-2) are also provided below.

Table HA.a-1 Summary of Sampling Design

Primary Objective of Design	Compare a site mean or median to a fixed threshold
Type of Sampling Design	Nonparametric
Sample Placement (Location) in the Field	Systematic with a random start location
Working (Null) Hypothesis	The median (mean) value at the site exceeds the threshold
Formula for calculating number of sampling locations	Sign Test - MARSSIM version
Calculated number of samples	10
Number of samples adjusted for EMC	10
Number of samples with MARSSIM Overage	12
Number of samples on map[a]	12
Number of selected sample areas[b]	1
Specified sampling area[c]	7396.00 ft^2
Size of grid / Area of grid cell[d]	26.6773 feet / 616.333 ft^2
Grid pattern	Triangular

[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]Size of grid / Area of grid gives the linear and square dimensions of the grid used to systematically place samples. If there was more than one sample area, this represents the largest dimensions used.

[1]This report was automatically produced* by Visual Sample Plan (VSP) software version 7.12a.
This design was last modified 6/17/2019 12:41:28 pm.
Software and documentation available at http://vsp.pnnl.gov.

*The report contents may have been modified or reformatted by end-user of software.

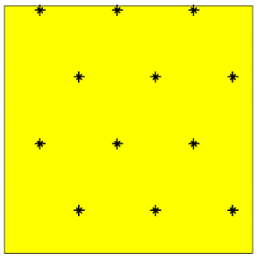

Figure HA.a-1

Table HA.a-2 Area: Treatment Pad

X Coord	Y Coord	Label	Value	Type	Historical	Sample Area
25.7776	14.9933			Systematic		
52.4549	14.9933			Systematic		
79.1323	14.9933			Systematic		
12.4389	38.0965			Systematic		
39.1163	38.0965			Systematic		
65.7936	38.0965			Systematic		
25.7776	61.1998			Systematic		
52.4549	61.1998			Systematic		
79.1323	61.1998			Systematic		
12.4389	84.3030			Systematic		
39.1163	84.3030			Systematic		
65.7936	84.3030			Systematic		

Primary Sampling Objective

The primary purpose of sampling at this site is to compare a site median or mean value with a fixed threshold. The working hypothesis (or "null" hypothesis) is that the median (mean) value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the median (mean) value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative one, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach

A nonparametric systematic sampling approach with a random start was used to determine the number of samples and to specify sampling locations. A nonparametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that typical parametric assumptions may not be true.

Both parametric and nonparametric equations rely on assumptions about the population. Typically, however, nonparametric 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.

VSP offers many options to determine the locations at which measurements are made or samples are collected and subsequently measured. For this design, systematic grid point sampling was chosen. Locating the sample points systematically provides data that are all equidistant apart. This approach does not provide as much information about the spatial structure of the potential contamination as simple random sampling does. Knowledge of the spatial structure is useful for geostatistical analysis. However, it ensures that all portions of the site are equally represented. Statistical analyses of systematically collected data are valid if a random start to the grid is used.

Nuclides

The following table (Table HA.a-3) summarizes the analyzed nuclides.

Table HA.a-3 Nuclides Analyzed by Study

Nuclide	$DCGL_W$	$DCGL_{EMC}$
Analyte 1	1	

Number of Total Samples: Calculation Equation and Inputs

The equation used to calculate the number of samples is based on a Sign test (see PNNL 13450 for discussion). For this site, the null hypothesis is rejected in favor of the alternative one if the median (mean) is sufficiently smaller than the threshold. The number of samples to collect is calculated so that if the inputs to the equation are true, the calculated number of samples will cause the null hypothesis to be rejected.

The formula used to calculate the number of samples is:

$$n = \frac{(Z_{1-a} + Z_{1-\beta})^2}{4(SignP - 0.5)^2}$$

where

$$SignP = \Phi\left(\frac{\Delta}{S_{total}}\right)$$

- $\Phi(z)$ is the cumulative standard normal distribution on $(-\infty,z)$ (see PNNL-13450 for details),
- n is the number of samples,
- S_{total} is the estimated standard deviation of the measured values including analytical error,
- Δ is the width of the gray region,
- α is the acceptable probability of incorrectly concluding the site median (mean) is less than the threshold,
- β is the acceptable probability of incorrectly concluding the site median (mean) exceeds the threshold,
- $Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
- $Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

Note: MARSSIM suggests that the number of samples should be increased by at least 20% to account for missing or unusable data and uncertainty in the calculated value of n. VSP allows a user-supplied percent overage as discussed in MARSSIM (EPA 2000, p. 5-33).

For each nuclide in the Nuclides Analyzed by Study table (Table HA.a-3), the values of these inputs that result in the calculated number of sampling locations are:

Table HA.a-3

Nuclide	n^a	n^b	n^c	Parameter					
				S_{total}	Δ	α	β	$Z_{1-\alpha}$[d]	$Z_{1-\beta}$[e]
Analyte 1	10	10	12	0.45	0.6	0.05	0.2	1.64485	0.841621

[a]The number of samples calculated by the formula.

[b]The number of samples increased by EMC calculations.

[c]The final number of samples increased by the MARSSIM Overage of 20%.

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

[e]This value is automatically calculated by VSP based upon the user defined value of β.

Performance

The following figure (Figure HA.a-2) is a performance goal diagram, described in U.S. Environmental Protection Agency's (EPA) QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true median (mean) values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.

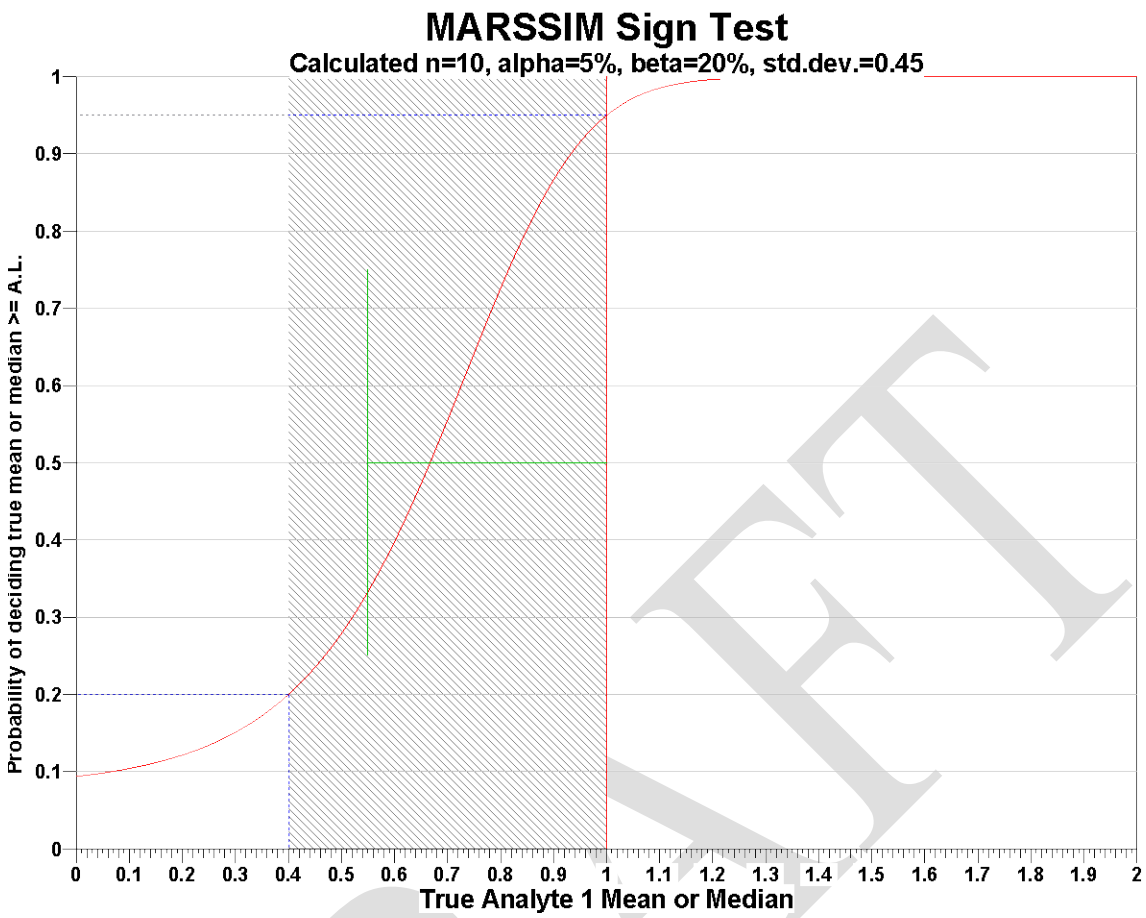

Figure HA.a-2

Statistical Assumptions

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

1. The computed sign test statistic is normally distributed,
2. The variance estimate, S^2, is reasonable and representative of the population being sampled,
3. The population values are not spatially or temporally correlated, and
4. The sampling locations will be selected probabilistically.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the gridded sample locations were selected based on a random start.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that $\mu >$ action level and alpha (%), probability of mistakenly concluding that $\mu <$ action level. The following table (Table HA.a-4) shows the results of this analysis.

Table HA.a-4 Number of Samples

AL=1		α=5		α=10		α=15	
		s=0.9	s=0.45	s=0.9	s=0.45	s=0.9	s=0.45
LBGR=90	β=15	1103	280	825	209	659	167
	β=20	948	240	692	176	542	138
	β=25	826	209	587	149	449	114
LBGR=80	β=15	280	75	209	56	167	45
	β=20	240	64	176	47	138	36
	β=25	209	56	149	40	114	30
LBGR=70	β=15	128	36	95	27	77	22
	β=20	110	32	81	23	63	18
	β=25	95	27	69	20	52	15

s = Standard Deviation

LBGR = Lower Bound of Gray Region (% of Action Level)

β = Beta (%), Probability of mistakenly concluding that $\mu >$ action level

α = Alpha (%), Probability of mistakenly concluding that $\mu <$ action level

AL = Action Level (Threshold)

Note: Values in table are not adjusted for EMC.

1

HA.a.2 VISUAL SAMPLING PLAN REPORT DOCUMENTATION FOR THE INTEGRATED DISPOSAL FACILITY STORAGE PAD[2]

Systematic Sampling Locations for Comparing a Median with a Fixed Threshold (Nonparametric - MARSSIM)

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 (Table HA.a-5) summarizes the sampling design developed. A figure that shows sampling locations in the field (Figure HA.a-3) and a table that lists sampling location coordinates (Table HA.a-6) are also provided below.

Table HA.a-5 Summary of Sampling Design

Primary Objective of Design	Compare a site mean or median to a fixed threshold
Type of Sampling Design	Nonparametric
Sample Placement (Location) in the Field	Systematic with a random start location
Working (Null) Hypothesis	The median (mean) value at the site exceeds the threshold
Formula for calculating number of sampling locations	Sign Test - MARSSIM version
Calculated number of samples	10
Number of samples adjusted for EMC	10
Number of samples with MARSSIM Overage	12
Number of samples on map[a]	18
Number of selected sample areas[b]	1
Specified sampling area[c]	18396.00 ft^2
Size of grid / Area of grid cell[d]	35.061 feet / 1064.58 ft^2
Grid pattern	Triangular

[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]Size of grid / Area of grid gives the linear and square dimensions of the grid used to systematically place samples. If there was more than one sample area, this represents the largest dimensions used.

[2]This report was automatically produced* by Visual Sample Plan (VSP) software version 7.12a.
This design was last modified 6/17/2019 12:59:49pm.
Software and documentation available at http://vsp.pnnl.gov

*The report contents may have been modified or reformatted by end-user of software.

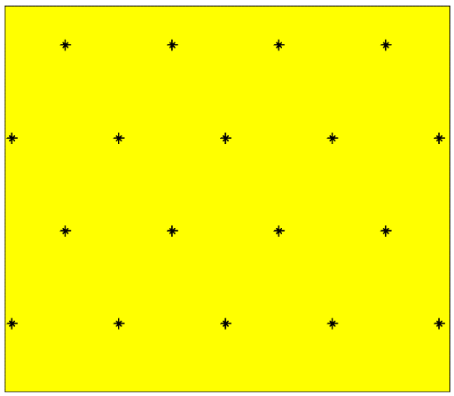

Figure HA.a-3

Table HA.a-6 Area: Storage Pad

X Coord	Y Coord	Label	Value	Type	Historical	Sample Area
2.4850	22.1127			Systematic		
37.5460	22.1127			Systematic		
72.6071	22.1127			Systematic		
107.6681	22.1127			Systematic		
142.7291	22.1127			Systematic		
20.0155	52.4764			Systematic		
55.0765	52.4764			Systematic		
90.1376	52.4764			Systematic		
125.1986	52.4764			Systematic		
2.4850	82.8402			Systematic		
37.5460	82.8402			Systematic		
72.6071	82.8402			Systematic		
107.6681	82.8402			Systematic		
142.7291	82.8402			Systematic		
20.0155	113.2039			Systematic		
55.0765	113.2039			Systematic		
90.1376	113.2039			Systematic		
125.1986	113.2039			Systematic		

Primary Sampling Objective

The primary purpose of sampling at this site is to compare a site median or mean value with a fixed threshold. The working hypothesis (or "null" hypothesis) is that the median (mean) value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the median (mean) value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative one, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach

A nonparametric systematic sampling approach with a random start was used to determine the number of samples and to specify sampling locations. A nonparametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that typical parametric assumptions may not be true.

Both parametric and nonparametric equations rely on assumptions about the population. Typically, however, nonparametric 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 nonparametric equation was used.

VSP offers many options to determine the locations at which measurements are made or samples are collected and subsequently measured. For this design, systematic grid point sampling was chosen. Locating the sample points systematically provides data that are all equidistant apart. This approach does not provide as much information about the spatial structure of the potential contamination as simple random sampling does. Knowledge of the spatial structure is useful for geostatistical analysis. However, it ensures that all portions of the site are equally represented. Statistical analyses of systematically collected data are valid if a random start to the grid is used.

Nuclides

The following table summarizes the analyzed nuclides.

Table HA.a-7 Nuclides Analyzed by Study

Nuclide	$DCGL_W$	$DCGL_{EMC}$
Analyte 1	1	

Number of Total Samples: Calculation Equation and Inputs

The equation used to calculate the number of samples is based on a Sign test (see PNNL 13450 for discussion). For this site, the null hypothesis is rejected in favor of the alternative one if the median (mean) is sufficiently smaller than the threshold. The number of samples to collect is calculated so that if the inputs to the equation are true, the calculated number of samples will cause the null hypothesis to be rejected.

The formula used to calculate the number of samples is:

$$n = \frac{(Z_{1-a} + Z_{1-\beta})^2}{4(SignP - 0.5)^2}$$

where

$$SignP = \Phi\left(\frac{\Delta}{S_{total}}\right)$$

- $\Phi(z)$ is the cumulative standard normal distribution on (-∞,z) (see PNNL-13450 for details),
- n is the number of samples,
- S_{total} is the estimated standard deviation of the measured values including analytical error,
- Δ is the width of the gray region,
- α is the acceptable probability of incorrectly concluding the site median (mean) is less than the threshold,

β is the acceptable probability of incorrectly concluding the site median (mean) exceeds the threshold,

$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is 1-α,

$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is 1-β.

Note: MARSSIM suggests that the number of samples should be increased by at least 20% to account for missing or unusable data and uncertainty in the calculated value of n. VSP allows a user-supplied percent overage as discussed in MARSSIM (EPA 2000, p. 5-33).

For each nuclide in the Nuclides Analyzed by Study table (Table HA.a-7), the values of these inputs that result in the calculated number of sampling locations are:

Table HA.a-8

Nuclide	n[a]	n[b]	n[c]	Parameter					
				S_{total}	Δ	α	β	$Z_{1-\alpha}$[d]	$Z_{1-\beta}$[e]
Analyte 1	10	10	12	0.45	0.6	0.05	0.2	1.64485	0.841621

[a]The number of samples calculated by the formula.

[b]The number of samples increased by EMC calculations.

[c]The final number of samples increased by the MARSSIM Overage of 20%.

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

[e]This value is automatically calculated by VSP based upon the user defined value of β.

Performance

The following figure (Figure HA.a-4) is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true median (mean) values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.

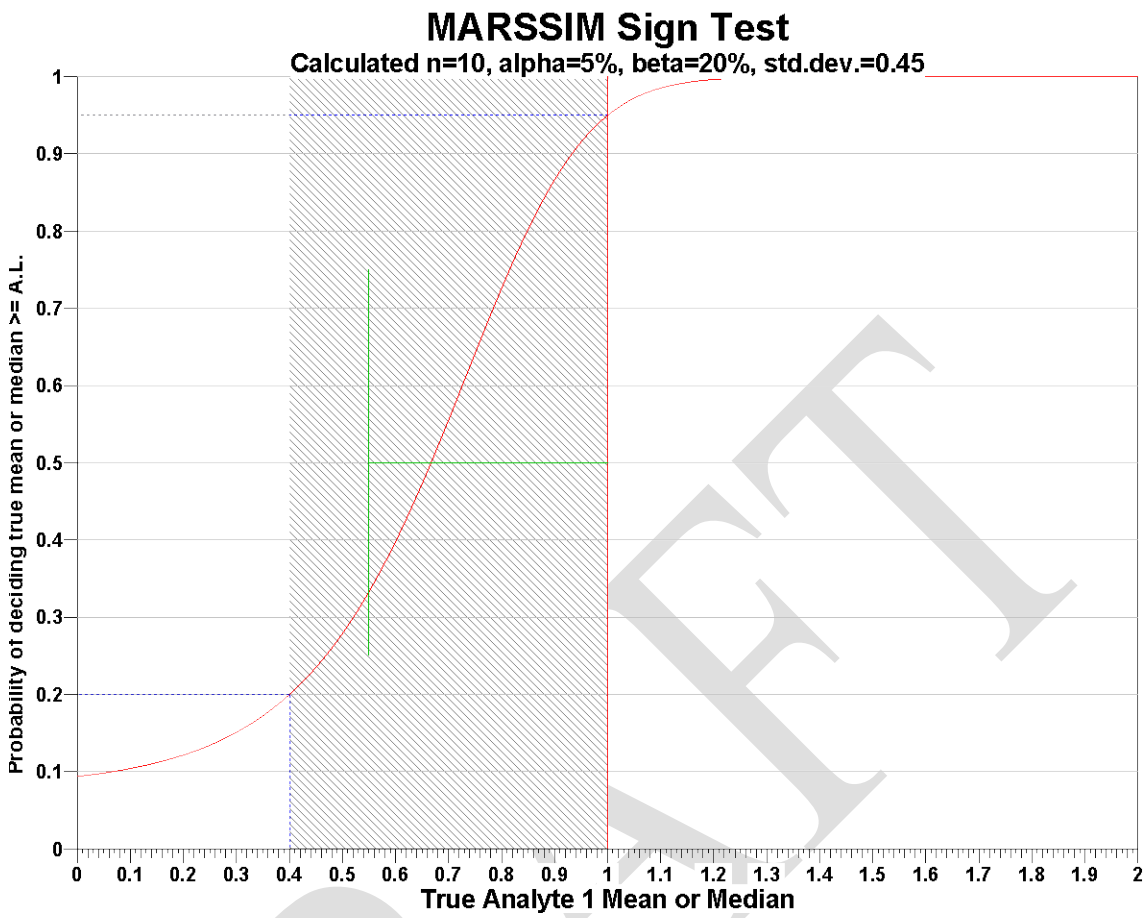

Figure HA.a-4

Statistical Assumptions

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

1. The computed sign test statistic is normally distributed,
2. The variance estimate, S^2, is reasonable and representative of the population being sampled,
3. The population values are not spatially or temporally correlated, and
4. The sampling locations will be selected probabilistically.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the gridded sample locations were selected based on a random start.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that $\mu >$ action level and alpha (%), probability of mistakenly concluding that $\mu <$ action level. The following table (Table HA.a-9) shows the results of this analysis.

Table HA.a-9 Number of Samples

AL=1		α=5		α=10		α=15	
		s=0.9	s=0.45	s=0.9	s=0.45	s=0.9	s=0.45
LBGR=90	β=15	1103	280	825	209	659	167
	β=20	948	240	692	176	542	138
	β=25	826	209	587	149	449	114
LBGR=80	β=15	280	75	209	56	167	45
	β=20	240	64	176	47	138	36
	β=25	209	56	149	40	114	30
LBGR=70	β=15	128	36	95	27	77	22
	β=20	110	32	81	23	63	18
	β=25	95	27	69	20	52	15

s = Standard Deviation

LBGR = Lower Bound of Gray Region (% of Action Level)

β = Beta (%), Probability of mistakenly concluding that μ > action level

α = Alpha (%), Probability of mistakenly concluding that μ < action level

AL = Action Level (Threshold)

Note: Values in table are not adjusted for EMC.

1