# Part 4: MEASURING CONTAMINANT CONCENTRATIONS IN SOIL

The Soil Screening Guidance includes a sampling strategy for implementing the soil screening process. Section 4.1 presents the sampling approach for surface soils. This approach provides a simple decision rule based on comparing the maximum contaminant concentrations of composite samples with surface soil screening levels (the Max test) to determine whether further investigation is needed for a particular exposure area (EA). In addition, this section presents a more complex strategy (the Chen test) that allows the user to design a site-specific quantitative sampling strategy by varying decision error limits and soil contaminant variability to optimize the number of samples and composites. Section 4.2 provides a subsurface soil sampling strategy for developing SSLs and applying the screening procedure for the volatilization and migration to ground water exposure pathways.

Section 4.3 describes the technical details behind the development of the SSL sampling strategy, including analyses and response to public and peer-review comments received on the December 1994 draft guidance.

The sampling strategy for the soil screening process is designed to achieve the following objectives:

- Estimate mean concentrations of contaminants of concern for comparison with SSLs
- Fill in the data gaps in the conceptual site model necessary to develop SSLs.

The soils of interest for the first objective differ according to the exposure pathway being addressed. For the direct ingestion, dermal, and fugitive dust pathways, EPA is concerned about surface soils. The sampling goal is to determine average contaminant concentrations of surface soils in exposure areas of concern. For inhalation of volatiles, migration to ground water and, in some cases, plant uptake, subsurface soils are the primary concern. For these pathways, the average contaminant concentration through each source is the parameter of interest.

The second objective (filling in the data gaps) applies primarily to the inhalation and migration to ground water pathways. For these pathways, the source area and depth as well as average soil properties within the source are needed to calculate the pathway-specific SSLs. Therefore, the sampling strategy needs to address collection of these site-specific data.

Because of the difference in objectives, the sampling strategies for the ingestion pathway and for the inhalation and migration to ground water pathways are addressed separately. If both surface and subsurface soils are a concern, then surface soils should be sampled first because the results of surface soil analyses may help delineate source areas to target for subsurface sampling.

At some sites, a third sampling objective may be appropriate. As discussed in the Soil Screening Guidance, SSLs may not be useful at sites where background contaminant levels are above the SSLs. Where sampling information suggests that background contaminant concentrations may be a concern, background sampling may be necessary. *Methods for Evaluating the Attainment of Cleanup Standards - Volume 3: Reference-Based Standards for Soil and Solid Media* (U.S. EPA, 1994e) provides further information on sampling soils to determine background conditions at a site.

In order to accurately represent contaminant distributions at a site, EPA used the Data Quality Objectives (DQO) process (Figure 4) to develop a sampling strategy that will satisfy Superfund program objectives. The DQO process is a systematic data collection planning process developed by EPA to ensure that the right type, quality, and quantity of data are collected to support EPA decision making. As shown in Sections 4.1.1 through 4.1.6, most of the key outputs of the DQO process already have been developed as part of the Soil Screening Guidance. The DQO activities addressed in this section are described in detail in the *Data Quality Objectives for Superfund: Interim Final Guidance* (U.S. EPA, 1993b) and the *Guidance for the Data Quality Objectives Process* (U.S. EPA, 1994c). Refer to these documents for more information on how to complete each DQO activity or how to develop other, site-specific sampling strategies.

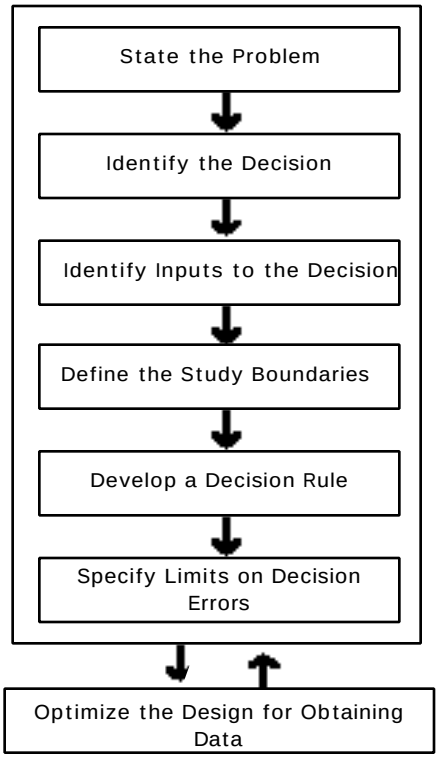



**Figure 4. The Data Quality Objectives process.**

## 4.1 Sampling Surface Soils

A sampling strategy for surface soils is presented in this section, organized by the steps of the DQO process. The first five steps of this process, from defining the problem through developing the basic decision rule, are summarized in Table 21, and are described in detail in the first five subsections. The details of the two remaining steps of the DQO process, specifying limits on decision errors and optimizing the design, have been developed separately for two alternative hypothesis testing procedures (the Max test and the Chen method) and are presented in four (4.1.6, 4.1.7, 4.1.9, and 4.1.10) subsections. In addition, a data quality assessment (DQA) follows the DQO process step for optimizing the design. The DQA ensures that site-specific error limits are achieved. Sections 4.1.8 and 4.1.11 describe the DQA for the Max and Chen tests, respectively. The technical details behind the development of the surface soil sampling design strategy are explained in Section 4.3.

**4.1.1 State the Problem.** In screening, the problem is to identify the contaminants and exposure areas (EAs) that do not pose significant risk to human health so that future investigations can be focused on the areas and contaminants of concern at a site.

The main site-specific activities involved in this first step of the DQO process include identifying the data collection planning team (including technical experts and key stakeholders) and specifying the available resources. The list of technical experts and stakeholders should contain all key personnel who are involved with applying the Soil Screening Guidance at the site. Other activities in this step include developing the conceptual site model (CSM), identifying exposure scenarios, and preparing a summary description of the surface soil contamination problem. The User's Guide (U.S. EPA, 1996) describes these activities in with more detail.

**4.1.2 Identify the Decision.** The decision is to determine whether the mean surface soil concentrations exceed surface soil screening levels for specific contaminants within EAs. If so, the EA must be investigated further. If not, no further action is necessary under CERCLA for the specific contaminants in the surface soils of those EAs.

## Table 21. Sampling Soil Screening DQOs for Surface Soils

| DQO Process Steps | Soil Screening Inputs/Outputs |
|---|---|
| **State the Problem** | |
| Identify scoping team | Site manager and technical experts (e.g., toxicologists, risk assessors, statisticians, soil scientists) |
| Develop conceptual site model (CSM) | CSM development (described in Step 1 of the User's Guide, U.S. EPA, 1996) |
| Define exposure scenarios | Direct ingestion and inhalation of fugitive particulates in a residential setting; dermal contact and plant uptake for certain contaminants |
| Specify available resources | Sampling and analysis budget, scheduling constraints, and available personnel |
| Write brief summary of contamination problem | Summary of the surface soil contamination problem to be investigated at the site |
| **Identify the Decision** | |
| Identify decision | Do mean soil concentrations for particular contaminants (e.g., contaminants of potential concern) exceed appropriate screening levels? |
| Identify alternative actions | Eliminate area from further study under CERCLA<br>or<br>Plan and conduct further investigation |
| **Identify Inputs to the Decision** | |
| Identify inputs | Ingestion and particulate inhalation SSLs for specified contaminants<br>Measurements of surface soil contaminant concentration |
| Define basis for screening | Soil Screening Guidance |
| Identify analytical methods | Feasible analytical methods (both field and laboratory) consistent with program-level requirements |
| **Define the Study Boundaries** | |
| Define geographic areas of field investigation | The entire NPL site (which may include areas beyond facility boundaries), except for any areas with clear evidence that no contamination has occurred |
| Define population of interest | Surface soils (usually the top 2 centimeters, but may be deeper where activities could redistribute subsurface soils to the surface) |
| Divide site into strata | Strata may be defined so that contaminant concentrations are likely to be relatively homogeneous within each stratum based on the CSM and field measurements |
| Define scale of decision making | Exposure areas (EAs) no larger than 0.5 acre each (based on residential land use) |
| Define temporal boundaries of study | Temporal constraints on scheduling field visits |
| Identify practical constraints | Potential impediments to sample collection, such as access, health, and safety issues |
| **Develop a Decision Rule** | |
| Specify parameter of interest | "True mean" (μ) individual contaminant concentration in each EA. (since the determination of the "true mean" would require the collection and analysis of many samples, the "Max Test" uses another sample statistic, the maximum composite concentration). |
| Specify screening level | Screening levels calculated using available parameters and site data (or generic SSLs if site data are unavailable). |
| Specify "if..., then..." decision rule | If the "true mean" EA concentration exceeds the screening level, then investigate the EA further. If the "true mean" is less than the screening level, then no further investigation of the EA is required under CERCLA. |

**4.1.3 Identify Inputs to the Decision.** This step of the DQO process requires identifying the inputs to the decision process, including the basis for further investigation and the applicable analytical methods. The inputs for deciding whether to investigate further are the ingestion, dermal, and fugitive dust inhalation SSLs calculated for the site contaminants as described in Part 2 of this document, and the surface soil concentration measurements for those same contaminants. Therefore, the remaining task is to identify Contract Laboratory Program (CLP) methods and/or field methods for which the quantitation limits (QLs) are less than the SSLs. EPA recommends the use of field methods, such as soil gas surveys, immunoassays, or X-ray fluorescence, where applicable and appropriate as long as quantitation limits are below the SSLs. At least 10 percent of field samples should be split and sent to a CLP laboratory for confirmatory analysis (U.S. EPA, 1993d).

**4.1.4 Define the Study Boundaries.** This step of the DQO process defines the sample population of interest, subdivides the site into appropriate exposure areas, and specifies temporal or practical constraints on the data collection. The description of the population of interest must include the surface soil depth.

**Sampling Depth.** When measuring soil contamination levels at the surface for the ingestion and inhalation pathways, the top 2 centimeters is usually considered surface soil, as defined by *Urban Soil Lead Abatement Project* (U.S. EPA 1993f). However, additional sampling beyond this depth may be appropriate for surface soils under a future residential use scenario in areas where major soil disturbances can reasonably be expected as a result of landscaping, gardening, or construction activities. In this situation, contaminants that were at depth can be moved to the surface. Thus, it is important to be cognizant of local residential construction practices when determining the depth of surface soil sampling and to weigh the likelihood of that area being developed.

**Subdividing the Site.** This step involves dividing the site into areas or strata depending on the likelihood of contamination and identifying areas with similar contaminant patterns. These divisions can be based on process knowledge, operational units, historical records, and/or prior sampling. Partitioning the site into such areas and strata can lead to a more efficient sampling design for the entire site.

For example, the site manager may have documentation that large areas of the site are unlikely to have been used for waste disposal activities. These areas would be expected to exhibit relatively low variability and the sampling design could involve a relatively small number of samples. The greatest intensity of sampling effort would be expected to focus on areas of the site where there is greater uncertainty or greater variability associated with contamination patterns. When relatively large variability in contaminant concentrations is expected, more samples are required to determine with confidence whether the EA should be screened out or investigated further.

Initially, the site may be partitioned into three types of areas:

1. Areas that are not likely to be contaminated
2. Areas that are known to be highly contaminated
3. Areas that are suspected to be contaminated and cannot be ruled out.

Areas that are not likely to be contaminated generally will not require further investigation if this assumption is based on historical site use information or other site data that are reasonably complete and accurate. (However, the site manager may also want take a few samples to confirm this assumption). These may be parts of the site that are within the legal boundaries of the property but

were completely undisturbed by hazardous-waste-generating activities. All other areas need investigation.

Areas that are known to be highly contaminated (i.e., sources) are targeted for subsurface sampling. The information collected on source area and depth is used to calculate site-specific SSLs for the inhalation and migration to ground water pathways (see Section 4.2 for more information).

Areas that are suspected to be contaminated (and cannot be ruled out for screening) are the primary subjects of the surface soil investigation. If a geostatistician is available, a geostatistical model may be used to characterize these areas (e.g., kriging model). However, guidance for this type of design is beyond the scope of the current guidance (see Chapter 10 of U.S. EPA, 1989a).

**Defining Exposure Areas.** After the site has been partitioned into relatively homogeneous areas, each region that is targeted for surface soil sampling is then subdivided into EAs. An EA is defined as that geographical area in which an individual may be exposed to contamination over time. Because the SSLs were developed for a residential scenario, EPA assumes the EA is a suburban residential lot corresponding to 0.5 acre. For soil screening purposes, each EA should be 0.5 acre or less. To the extent possible, EAs should be constructed as square or rectangular areas that can be subdivided into squares to facilitate compositing and grid sampling. If the site is currently residential, then the EA should be the actual residential lot size. The exposure areas should not be laid out in such a way that they unnecessarily combine areas of high and low levels of contamination. The orientation and exact location of the EA, relative to the distribution of the contaminant in the soil, can lead to instances where sampling of the EA may lead to results above the mean, and other instances, to results below the mean. Try to avoid straddling contaminant "distribution units" within the 0.5 acre EA.

The sampling strategy for surface soils allows investigators to determine mean soil contaminant concentration across an EA of interest. An arithmetic mean concentration for an EA best represents the exposure to site contaminants over a long period of time. For risk assessment purposes, an individual is assumed to move randomly across an EA over time, spending equivalent amounts of time in each location. Since reliable information about specific patterns of nonrandom activity for future use scenarios is not available, random exposure appears to be the most reasonable assumption for a residential exposure scenario. Therefore, spatially averaged surface soil concentrations are used to estimate mean exposure concentrations.

Because all the EAs within a given stratum should exhibit similar contaminant concentrations, one site-specific sampling design can be developed for all EAs within that stratum. As discussed above, some strata may have relatively low variability and other strata may have relatively high variability. Consequently, a different sampling design may be necessary for each stratum, based upon the stratum-specific estimate of the contaminant variability.

### 4.1.5 Develop a Decision Rule.

Ideally, the decision rule for surface soils is:

> If the mean contaminant concentration within an EA exceeds the screening level, then investigate that EA further.

This "screening level" is the actual numerical value used to compare against the site contamination data. It may be identical to the SSL, or it may be a multiple of the SSL (e.g., 2 SSL) for a hypothesis test designed to achieve specified decision error rates in a specified region above and below the SSL. In addition, another sample statistic (e.g., the maximum concentration) may be used as an estimate of the mean for comparison with the "screening level."

**4.1.6 Specify Limits on Decision Errors for the Max Test.** Sampling data will be used to support a decision about whether an EA requires further investigation. Because of variability in contaminant concentrations within an EA, practical constraints on sample sizes, and sampling or measurement error, the data collected may be inaccurate or nonrepresentative and may mislead the decision maker into making an incorrect decision. A decision error occurs when sampling data mislead the decision maker into choosing a course of action that is different from or less desirable than the course of action that would have been chosen with perfect information (i.e., with no constraints on sample size and no measurement error).

EPA recognizes that data obtained from sampling and analysis are never perfectly representative and accurate, and that the costs of trying to achieve near-perfect results can outweigh the benefits. Consequently, EPA acknowledges that uncertainty in data must be tolerated to some degree. The DQO process controls the degree to which uncertainty in data affects the outcomes of decisions that are based on those data. This step of the DQO process allows the decision maker to set limits on the probabilities of making an incorrect decision.

The DQO process utilizes hypothesis tests to control decision errors. When performing a hypothesis test, a presumed or baseline condition, referred to as the "null hypothesis" ($H_o$), is established. This baseline condition is presumed to be true unless the data conclusively demonstrate otherwise, which is called "rejecting the null hypothesis" in favor of an alternative hypothesis. For the Soil Screening Guidance, the baseline condition, or $H_o$, is that **the site needs further investigation.**

When the hypothesis test is performed, two possible decision errors may occur:

1. Decide not to investigate an EA further (i.e., "walk away") when the correct decision (with complete and perfect information) would be to "investigate further"

2. Decide to investigate further when the correct decision would be to "walk away."

Since the site is on the NPL, site areas are presumed to need further investigation. Therefore, the data must provide clear evidence that it would be acceptable to "walk away." This presumption provides the basis for classifying the two types of decision errors. The "incorrectly walk away" decision error is designated as the Type I decision error because one has incorrectly rejected the baseline condition (null hypothesis). Correspondingly, the "unnecessarily investigate further" decision error is designated as the Type II decision error.

To complete the specification of limits on decision errors, Type I and Type II decision error probability limits must be defined in relation to the SSL. First a "gray region" is specified with respect to the mean contaminant concentration within an EA. The gray region represents the range of contaminant levels near the SSL, where uncertainty in the data (i.e., the variability) can make the decision "too close to call." In other words, when the average of the data values is very close to the SSL, it would be too expensive to generate a data set of sufficient size and precision to resolve what the correct determination should be. (i.e., Does the average concentration fall "above" or "below" the SSL?)

The Soil Screening Guidance establishes a default range for the width and location of the "gray region": from one-half the SSL (0.5 SSL) to two times the SSL (2 SSL). By specifying the upper edge of the gray region as twice the SSL, it is possible that exposure areas with mean values slightly higher than the SSL may be screened from further study. However, EPA believes that the exposure scenario

and assumptions used to derive SSLs are sufficiently conservative to be protective in such cases.

On the lower side of the gray region, the consequences of decision errors at one-half the SSL are primarily financial. If the lower edge of the gray region were to be moved closer to the SSL, then more exposure areas that were truly below the SSL would be screened out, but more money would be spent on sampling to make this determination. If the lower edge of the gray region were to be moved closer to zero, then less money could be spent on sampling, but fewer EAs that were truly below the SSLs would be screened out, leading to unnecessary investigation of EAs. The Superfund program chose the gray region to be one-half to two times the SSL after investigating several different ranges. This range for the gray region represents a balance between the costs of collecting and analyzing soil samples and making incorrect decisions. While it is desirable to estimate exactly the exposure area mean, the number of samples required are much more than project managers are generally willing to collect in a "screening" effort. Although some exposure areas will have contaminant concentrations that are between the SSL and twice the SSL and will be screened out, human health will still be protected given the conservative assumptions used to derive the SSLs.

The Soil Screening Guidance establishes the following goals for Type I and Type II decision error rates:

- Prob ("walk away" when the true EA mean is 2 SSL) = 0.05
- Prob ("investigate further" when the true EA mean is 0.5 SSL) = 0.20.

This means that there should be no more than a 5 percent chance that the site manager will "walk away" from an EA where the true mean concentration is 2 SSL or more. In addition, there should be no more than a 20 percent chance that the site manager will unnecessarily investigate an EA when the mean is 0.5 SSL or less.

These decision error limits are general goals for the soil screening process. Consistent with the DQO process, these goals may be adjusted on a site-specific basis by considering the available resources (i.e., time and budget), the importance of screening surface soil relative to other potential exposure pathways, consequences of potential decision errors, and consistency with other relevant EPA guidance and programs.

Table 22 summarizes this step of the DQO process for the Max test, specifying limits on the decision error rates, and the final step of the DQO process for the Max test, optimizing the design. Figure 5 illustrates the gray region for the decision error goals: a Type I decision error rate of 0.05 (5 percent) at 2 SSL and a Type II decision error rate of 0.20 (20 percent) at 0.5 SSL.

**4.1.7 Optimize the Design for the Max Test.** This section provides instructions for developing an optimum sampling strategy for screening surface soils. It discusses compositing, the selection of sampling points for composited and uncomposited surface soil sampling, and the recommended procedures for determining the sample sizes necessary to achieve specified limits on decision errors using the Max test.

## Table 22. Sampling Soil Screening DQOs for Surface Soils under the Max Test

| DQO Process Steps | Soil Screening Inputs/Outputs |
|---|---|
| **Specify Limits on Decision Errors*** | |
| Define baseline condition (null hypothesis) | The EA needs further investigation |
| Define the gray region** | From 0.5 SSL to 2 SSL |
| Define Type I and Type II decision errors | Type I error: Do not investigate further ("walk away from") an EA whose true mean exceeds the screening level of 2 SSL<br>Type II error: Investigate further when an EA's true mean falls below the screening level of 0.5 SSL |
| Identify consequences | Type I error: potential public health consequences<br>Type II error: unnecessary expenditure of resources to investigate further |
| Assign acceptable probabilities of Type I and Type II decision errors | Goals:<br>Type I: 0.05 (5%) probability of not investigating further when “true mean” of the EA is 2 SSL<br>Type II: 0.20 (20%) probability of investigating further when “true mean” of the EA is 0.5 SSL |
| Define QA/QC goals | CLP precision and bias requirements<br>10% CLP analyses for field methods |
| **Optimize the Design** | |
| Determine how to best estimate “true mean” | Samples composited across the EA estimate the EA mean ($\bar{x}$). Use maximum composite concentration as a conservative estimate of the true EA mean. |
| Determine expected variability of EA surface soil contaminant concentrations | A conservatively large expected coefficient of variation (CV) from prior data for the site, field measurements, or data from other comparable sites and expert judgment. A minimum default CV of 2.5 should be used when information is insufficient to estimate the CV. |
| Design sampling strategy by evaluating costs and performance of alternatives | Lowest cost sampling design option (i.e., compositing scheme and number of composites) that will achieve acceptable decision error rates |
| Develop planning documents for the field investigation | Sampling and Analysis Plan (SAP)<br>Quality Assurance Project Plan (QAPjP) |

* Since the DQO process controls the degree to which uncertainty in data affects the outcome of decisions that are based on that data, specifying limits on decision errors will allow the decision maker to control the probability of making an incorrect decision when using the DQOs.

** The gray region represents the area where the consequences of decision errors are minor (and uncertainty in sampling data makes decisions too close to call).

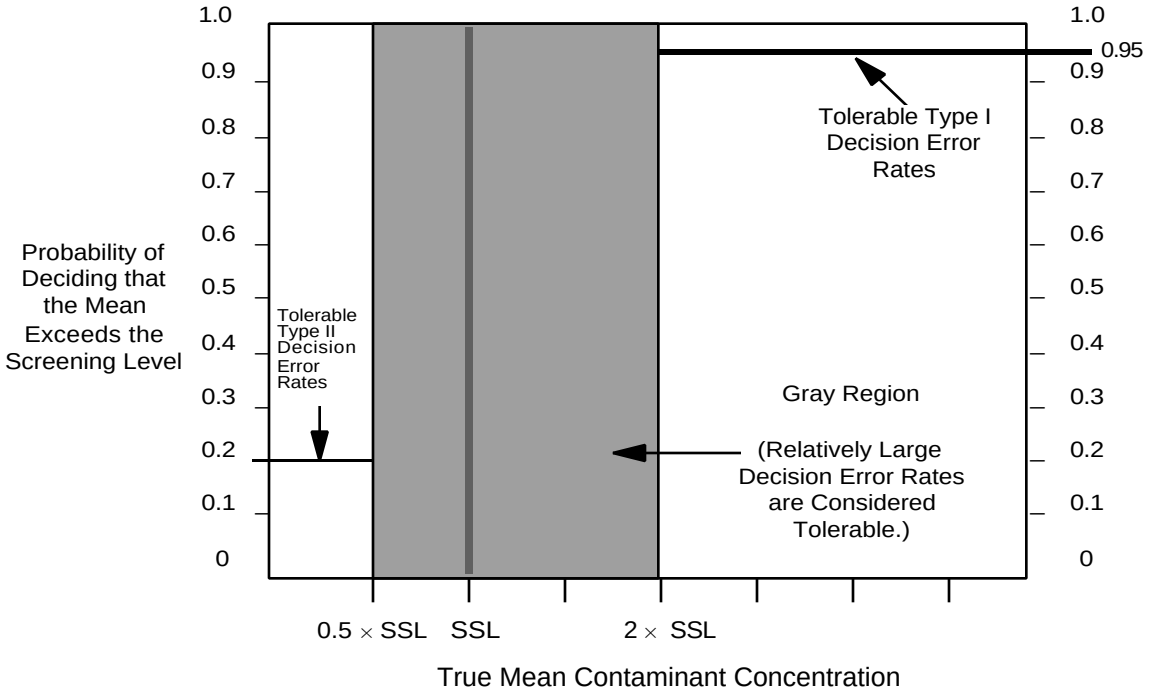



**Figure 5. Design performance goal diagram.**

Note that the size, shape, and orientation of sampling volume (i.e., "support") for heterogenous media have a significant effect on reported measurement values. For instance, particle size has a varying affect on the transport and fate of contaminants in the environment and on the potential receptors. Because comparison of data from methods that are based on different supports can be difficult, defining the sampling support is important in the early stages of site characterization. This may be accomplished through the DQO process with existing knowledge of the site, contamination, and identification of the exposure pathways that need to be characterized. Refer to *Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies* (U.S. EPA, 1992f) for more information about soil sampling support.

The SAP developed for surface soils should specify sampling and analytical procedures as well as the development of QA/QC procedures. To identify the appropriate analytical procedures, the screening levels must be known. If data are not available to calculate site-specific SSLs, then the generic SSLs in Appendix A should be used.

**Compositing.** Because the objective of surface soil screening is to ensure that the mean contaminant concentration does not exceed the screening level, the physical "averaging" that occurs during compositing is consistent with the intended use of the data. Compositing allows a larger number of locations to be sampled while controlling analytical costs because several discrete samples are physically mixed (homogenized) and one or more subsamples are drawn from the mixture and submitted for analysis. If the individual samples in each composite are taken across the EA, each composite represents an estimate of the EA mean.

A practical constraint to compositing in some situations is the heterogeneity of the soil matrix. The

efficiency and effectiveness of the mixing process may be hindered when soil particle sizes vary widely or when the soil matrix contains foreign objects, organic matter, viscous fluids, or sticky material. Soil samples should not be composited if matrix interference among contaminants is likely (e.g., when the presence of one contaminant biases analytical results for another).

Before individual specimens are composited for chemical analysis, the site manager should consider homogenizing and splitting each specimen. By compositing one portion of each specimen with the other specimens and storing one portion for potential future analysis, the spatial integrity of each specimen is maintained. If the concentration of a contaminant in a composite sample is high, the splits of the individual specimens from which it was composed can be analyzed discretely to determine which individual specimen(s) have high concentrations of the contaminant. This will permit the site manager to determine which portion within an EA is contaminated without making a repeat visit to the site.

**Sample Pattern.** The Max test should only be applied using composite samples that are representative of the entire EA. However, the Chen test (see Section 4.1.9) can be applied with individual, uncomposited samples. There are several options for developing a sampling pattern for compositing that produce samples that should be representative. If individual, uncomposited samples will be analyzed for contaminant concentrations, the N sample points can be selected using either (1) simple random sampling (SRS), (2) stratified SRS, or (3) systematic grid sampling (square or rectangular grid) with a random starting point (SyGS/rs). Step-by-step procedures for selecting SRS and SyGS/rs samples are provided in Chapter 5 of the U.S. EPA (1989a) and Chapter 5 of U.S. EPA (1994e). If stratified random sampling is used, the sampling rate must be the same in every sector, or stratum of the EA. Hence, the number of sampling points assigned to a stratum must be directly proportional to the surface area of the stratum.

Systematic grid sampling with a random starting point is generally preferred because it ensures that the sample points will be dispersed across the entire EA. However, if the boundaries of the EA are irregular (e.g., around the perimeter of the site or the boundaries of a stratum within which the EAs were defined), the number of grid sample points that fall within the EA depends on the random starting point selected. Therefore, for these irregularly shaped EAs, SRS or stratified SRS is recommended. Moreover, if a systematic trend of contamination is suspected across the EA (e.g., a strip of higher contamination), then SRS or stratified SRS is recommended again. In this case, grid sampling would be likely to result in either over- or under representation of the strip of higher contaminant levels, depending on the random starting point.

For composite sampling, the sampling pattern used to locate the discrete sample specimens that form each composite sample (N) is important. The composite samples should be formed in a manner that is consistent with the assumptions underlying the sample size calculations. In particular, each composite sample should provide an unbiased estimate of the mean contaminant concentration over the entire EA. One way to construct a valid composite of C specimens is to divide the EA into C sectors, or strata, of equal area and select one point at random from each sector. If sectors (strata) are of unequal sizes, the simple average is no longer representative of the EA as a whole.

Five valid sampling patterns and compositing schemes for selecting N composite samples that each consist of C specimens are listed below:

1. Select an SRS consisting of C points and composite all specimens associated with these points into a sample. Repeat this process N times, discarding any points that were used in a previous sample.

2. Select an SyGS/rs of C points and composite all specimens associated with the points in this sample. Repeat this process N times, using a new randomly selected starting point each time.

3. Select a single SyGS/rs of C×N points and use the systematic compositing scheme that is described in Highlight 3 to form N composites, as illustrated in Figure 6.

4. Select a single SyGS/rs of C×N points and use the random compositing scheme that is described in Highlight 4 to form N composites, as illustrated in Figure 7.

5. Select a stratified random sample of C×N points and use a random compositing scheme, as described in Highlight 5, to form N composites, as illustrated in Figure 8.

Methods 1, 2, and 5 are the most statistically defensible, with method 5 used as the default method in the Soil Screening Guidance. However, given the practical limits of implementing these methods, either method 3 or 4 is generally recommended for EAs with regular boundaries (e.g., square or rectangular). As noted above, if the boundaries of the EA are irregular, SyGS/rs sampling may not result in exactly C×N sample points. Therefore, for EAs with irregular boundaries, method 5 is recommended. Alternatively, a combination of methods 4 and 5 can be used for EAs that can be partitioned into C sectors of equal area of which K have regular boundaries and the remaining C - K have irregular boundaries.

Additionally, compositing within sectors to indicate whether one sector of the EA exceeds SSLs is an option that may also be considered. See Section 4.3.6 for a full discussion.

**Sample Size.** This section presents procedures to determine sample size requirements for the Max test that achieve the site-specific decision error limits discussed in Section 4.1.6. The Max test is based on the maximum concentration observed in N composite samples that each consist of C individual specimens. The individual specimens are selected so that each of the N composite samples is representative of the site as a whole, as discussed above. Hence, this section addresses determining the sample size pair, C and N, that achieves the site-specific decision error limits. Directions for performing the Max test in a manner that is consistent with DQOs established for a site are presented later in this section.

Table 23 presents the probabilities of Type I errors at 2 SSL and Type II errors at 0.5 SSL (the boundary points of the gray region discussed in Section 4.1.6) for several sample size options when the variability for concentrations of individual measurements across the EA ranges from 100 percent to 400 percent (CV = 1.0 to 4.0). Two choices for the number, C, of specimens per composite are shown in this table: 4 and 6. Fewer than four specimens per composite is not considered sufficient for the Max test. Fewer than four specimens per composite does not achieve the decision error limit goals for the level of variability generally encountered at CERCLA sites. More than six specimens may be more than can be effectively homogenized into a composite sample.

The number, N, of composite samples shown in Table 23 ranges from 4 to 9. Fewer than four samples is not considered sufficient because, considering decision error rates from simulation results (Section 4.3), the Max text should be based on at least four independent estimates of the EA mean. More than nine composite samples per EA is generally unlikely for screening surface soils at Superfund sites. However, additional sample size options can be determined from the simulation results reported in Appendix I.

***Highlight 3: Procedure for Compositing of Specimens from a Grid Sample Using a Systematic Scheme (Figure 6)***

1. Lay out a square or triangular grid sample over the EA, using a random start. Step-by-step procedures can be found in Chapter 5 of U.S. EPA (1989a). The number of points in the grid should be equal to C×N, where C is the desired number of specimens per composite and N is the desired number of composites.

2. Divide the EA into C sectors (strata) of equal area and shape such that each sector contains the same number of sample points. The number of sectors (C) should be equal to the number of specimens in each composite (since one specimen per area will be used in each composite) and the number of points within each sector, N, should equal the desired number of composite samples.

3. Label the points within one sector in any arbitrary fashion from 1 to N. Use the same scheme for each of the other sectors.

4. Form composite number 1 by compositing specimens with the '1' label, form composite number 2 by compositing specimens with the '2' label, etc. This leads to N composite samples that are subjected to chemical analysis.

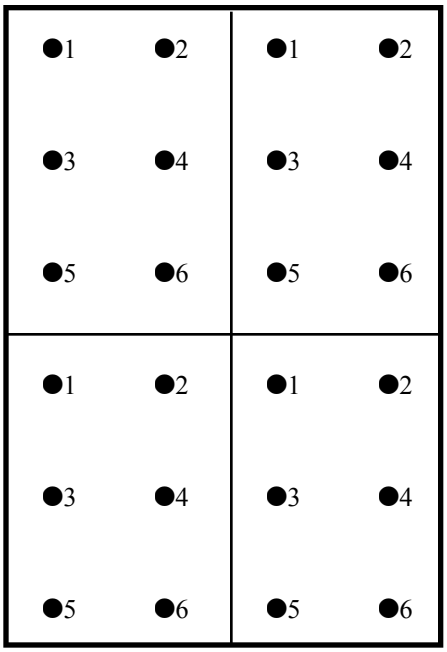



**Figure 6. Systematic (square grid points) sample with systematic compositing scheme (6 composite samples consisting of 4 specimens).**

***Highlight 4: Procedure for Compositing of Specimens from a Grid Sample Using a Random Scheme (Figure 7)***

1. Lay out a square or triangular grid sample over the EA, using a random start. Step-by-step procedures can be found in Chapter 5 of U.S. EPA (1989a). The number of points in the grid should be equal to C×N, where C is the desired number of specimens per composite and N is the desired number of composites.

2. Divide the EA into C sectors (strata) of equal area and shape such that each sector contains the same number of sample points. The number of sectors (C) should be equal to the number of specimens in each composite (since one specimen per area will be used in each composite) and the number of points within each sector, N, should equal the desired number of composite samples.

3. Use a random number table or random number generator to establish a set of labels for the N points within each sector. This is done by first labeling the points in a sector in an arbitrary fashion (say, points A, B, C,...) and associating the first random number with point A, the second with point B, etc. Then rank the points in the sector according to the set of random numbers and relabel each point with its rank. Repeat this process for each sector.

4. Form composite number 1 by compositing specimens with the '1' label, form composite number 2 by compositing specimens with the '2' label, etc. This leads to N composite samples that are subjected to chemical analysis.

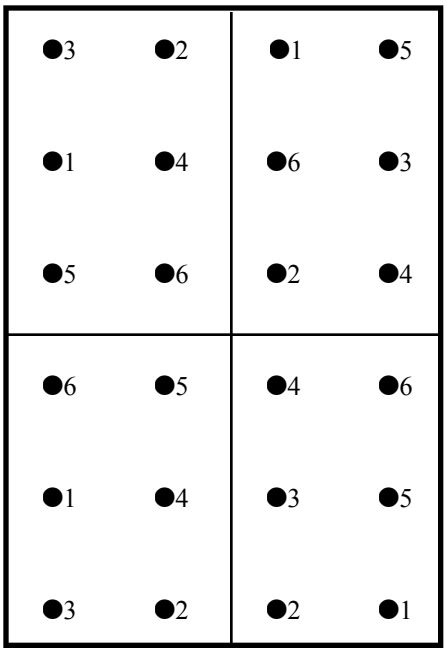



**Figure 7. Systematic (square grid points) sample with random compositing scheme (6 composite samples consisting of 4 specimens).**

***Highlight 5: Procedure for Compositing of Specimens from a Stratified Random Sample Using a Random Scheme (Figure 8)***

1. Divide the EA into C sectors (strata) of equal area, where C is equal to the number of specimens to be in each composite (since one specimen per stratum will be used in each composite).

2. Within each stratum, choose N random locations, where N is the desired number of composites. Step-by-step procedures for choosing random locations can be found in Chapter 5 of U.S. EPA (1989a).

3. Use a random number table or random number generator to establish a set of labels for the N points within each sector. This is done by first labeling the points in a sector in an arbitrary fashion (say, points A, B, C,...) and associating the first random number with point A, the second with point B, etc. Then rank the points in the sector according to the set of random numbers and relabel each point with its rank. Repeat this process for each sector.

4. Form composite number 1 by compositing specimens with the '1' label, form composite number 2 by compositing specimens with the '2' label, etc. This leads to N composite samples that are subjected to chemical analysis.

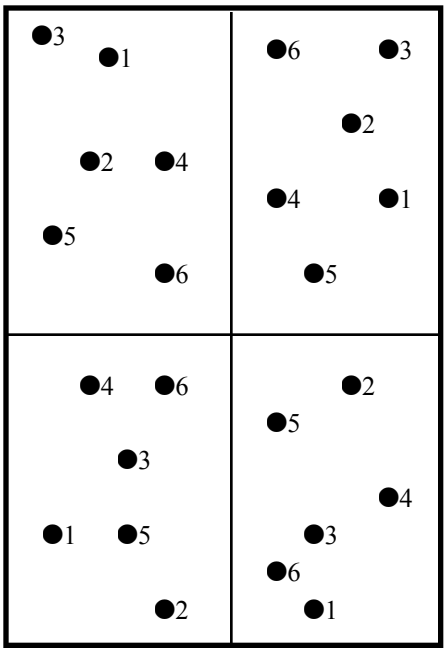



**Figure 8. Stratified random sample with random compositing scheme (6 composite samples consisting of 4 specimens).**

## Table 23. Probability of Decision Error at 0.5 SSL and 2 SSL Using Max Test

| | CV=1.0[a] | | CV=1.5 | | CV=2.0 | | CV=2.5 | | CV=3.0 | | CV=3.5 | | CV=4.0 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample Size[b] | $E_{0.5}$[c] | $E_{2.0}$[d] | $E_{0.5}$ | $E_{2.0}$ | $E_{0.5}$ | $E_{2.0}$ | $E_{0.5}$ | $E_{2.0}$ | $E_{0.5}$ | $E_{2.0}$ | $E_{0.5}$ | $E_{2.0}$ | $E_{0.5}$ | $E_{2.0}$ |
| | C = 4 specimens per composite[e] | | | | | | | | | | | | | |
| 4 | <.01 | 0.08 | 0.02 | 0.11 | 0.09 | 0.13 | 0.14 | 0.19 | 0.19 | 0.20 | 0.24 | 0.26 | 0.25 | 0.30 |
| 5 | <.01 | 0.05 | 0.02 | 0.06 | 0.11 | 0.10 | 0.15 | 0.10 | 0.26 | 0.17 | 0.26 | 0.18 | 0.31 | 0.25 |
| 6 | <.01 | 0.03 | 0.02 | 0.04 | 0.11 | 0.06 | 0.21 | 0.08 | 0.28 | 0.11 | 0.31 | 0.11 | 0.35 | 0.16 |
| 7 | <.01 | 0.01 | 0.03 | 0.02 | 0.12 | 0.04 | 0.25 | 0.05 | 0.31 | 0.08 | 0.36 | 0.09 | 0.41 | 0.15 |
| 8 | <.01 | 0.01 | 0.03 | 0.01 | 0.16 | 0.02 | 0.25 | 0.04 | 0.36 | 0.05 | 0.42 | 0.07 | 0.41 | 0.09 |
| 9 | <.01 | 0.01 | 0.05 | 0.01 | 0.16 | 0.01 | 0.28 | 0.03 | 0.36 | 0.04 | 0.44 | 0.07 | 0.48 | 0.08 |
| | C = 6 specimens per composite | | | | | | | | | | | | | |
| 4 | <.01 | 0.08 | <.01 | 0.11 | 0.03 | 0.12 | 0.08 | 0.16 | 0.15 | 0.17 | 0.26 | 0.20 | 0.23 | 0.27 |
| 5 | <.01 | 0.05 | <.01 | 0.06 | 0.04 | 0.09 | 0.11 | 0.09 | 0.17 | 0.13 | 0.22 | 0.15 | 0.25 | 0.20 |
| 6 | <.01 | 0.03 | 0.01 | 0.04 | 0.06 | 0.04 | 0.14 | 0.06 | 0.19 | 0.09 | 0.25 | 0.09 | 0.29 | 0.12 |
| 7 | <.01 | 0.01 | 0.01 | 0.02 | 0.06 | 0.02 | 0.14 | 0.04 | 0.23 | 0.06 | 0.29 | 0.08 | 0.37 | 0.08 |
| 8 | <.01 | 0.01 | 0.01 | 0.01 | 0.06 | 0.02 | 0.15 | 0.02 | 0.25 | 0.03 | 0.30 | 0.04 | 0.40 | 0.06 |
| 9 | <.01 | 0.01 | 0.01 | 0.01 | 0.06 | 0.01 | 0.18 | 0.02 | 0.28 | 0.03 | 0.34 | 0.03 | 0.39 | 0.04 |

[a] The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA, including measurement error.

[b] Sample size (N) = number of composite samples.

[c] $E_{0.5}$ = Probability of requiring further investigation when the EA mean is 0.5 SSL.

[d] $E_{2.0}$ = Probability of not requiring further investigation when the EA mean is 2.0 SSL.

[e] C = number of specimens per composite sample, where each composite consists of points from a stratified random or systematic grid sample from across the entire EA.

NOTE: All decision error rates are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the quantitation limit (i.e., SSL/100), and half the EA has concentrations that follow a gamma distribution (a conservative distributional assumption).

The error rates shown in Table 23 are based on the simulations presented in Appendix I. These simulations are based on the following assumptions:

1. Each of the N composite samples is based on C specimens selected to be representative of the EA as a whole, as specified above (C = number of sectors or strata).

2. One-half the EA has concentrations below the quantitation limit (which is assumed to be SSL/100).

3. One-half the EA has concentrations that follow a gamma distribution (see Section 4.3 for additional discussion).

4. Each chemical analysis is subject to a 20 percent measurement error.

The error rates presented in Table 23 are based on the above assumptions which make them robust for most potential distributions of soil contaminant concentrations. Distribution assumptions 2 and 3 were used because they were found in the simulations to produce high error rates relative to other potential contaminant distributions (see Section 4.3). If the proportion of the site below the quantitation limit (QL) is less than half or if the distribution of the concentration measurements is some other distribution skewed to the right (e.g., lognormal), rather than gamma, then the error rates achieved are likely to be no worse than those cited in Table 23. Although the actual contaminant distribution may be different from those cited above as the basis for Table 23, only extensive investigations will usually generate sufficient data to determine the actual distribution for each EA.

Using Table 23 to determine the sample size pair (C and N) needed to achieve satisfactory error rates with the Max test requires an *a priori* estimate of the coefficient of variation for measurements of the contaminant of interest across the EA. The coefficient of variation (CV) is the ratio of the standard deviation of contaminant concentrations for individual, uncomposited specimens divided by the EA mean concentration. As discussed in Section 4.1.4, the EAs should be constructed within strata expected to have relatively homogeneous concentrations so that an estimate of the CV for a stratum may be applicable for all EAs in that stratum. The site manager should use a conservatively large estimate of the CV for determining sample size requirements because additional sampling will be needed if the data suggest that the true CV is greater than that used to determine the sample sizes.

Potential sources of information for estimating the EA or stratum means, variances, and CVs include the following (in descending order of desirability):

- Data from a pilot study conducted at the site
- Prior sampling data from the site
- Data from similar sites
- Professional judgment.

For more information on estimating variability, see Section 6.3.1 of U.S. EPA (1989a).

**4.1.8 Using the DQA Process: Analyzing Max Test Data.** This section provides guidance for analyzing the data for the Max test.

The hypothesis test for the Max test is very simple to implement, which is one reason that the Max test is attractive as a surface soil screening test. If $x_1$, $x_2$, ..., $x_N$ represent concentration measurements for N composite samples that each consist of C specimens selected so that each

composite is representative of the EA as a whole (as described in Section 4.1.7), the Max test is implemented as follows:

> If Max ($x_1$, $x_2$, ..., $x_N$) ≥ 2 SSL, then investigate the EA further;
> If Max ($x_1$, $x_2$, ..., $x_N$) < 2 SSL, and the data quality assessment (DQA) indicates that the sample size was adequate, then no further investigation is necessary.

In addition, the step-by-step procedures presented in Highlight 6 must be implemented to ensure that the site-specific error limits, as discussed in Section 4.1.6, are achieved.

If the EA mean is below 2 SSL, the DQA process may be used to determine if the sample size was sufficiently large to justify the decision to not investigate further. To use Table 23 to check whether the sample size is adequate, an estimate of the CV is needed for each EA. The first four steps of Highlight 6, the DQA process for the Max test, present a process for the computation of a sample CV for an EA based on the N composite samples that each consist of C specimens.

However, the sample CV can be quite large when all the measurements are very small (e.g., well below the SSL) because CV approaches infinity as the EA sample mean ($\bar{x}$) approaches zero. Thus, when the composite concentration values for an EA are all near zero, the sample CV may be questionable and therefore unreliable for determining if the original sample size was sufficient (i.e., it could lead to further sampling when the EA mean is well below 2 SSL). To protect against unnecessary additional sampling in such cases, compare all composites against the equation given in Step 5 of Highlight 6. If the maximum composite sample concentration is below the value given by the equation, then the sample size may be assumed to be adequate and no further DQA is necessary.

To develop Step 5, EPA decided that if there were no compositing (C=1) and all the observations (based on a sample size appropriate for a CV of 2.5) were less than the SSL, then one can reasonably assume that the EA mean was not greater than 2 SSL. Likewise, because the standard error for the mean of C specimens, as represented by the composite sample, is proportional to $1/\sqrt{C}$, the comparable condition for composite observations is that one can reasonably assume that the EA mean was not greater than 2 SSL when all composite observations were less than SSL/$\sqrt{C}$. If this is the case for an EA sample set, the sample size can be assumed to be adequate and no further DQA is needed. Otherwise (when at lease one composite observation is not this small), use Table 23 with the sample CV for the EA to determine whether a sufficient number of samples were taken to achieve DQOs.

In addition to being simple to implement, the Max test is recommended because it provides good control over the Type I error rates at 2 SSL with small sample sizes. It also does not need any assumptions regarding observations below the QL. Moreover, the Max test error rates at 2 SSL are fairly robust against alternative assumptions regarding the distribution of surface soil concentrations in the EA. The simulations in Appendix I show that these error rates are rather stable for lognormal or Weibull contaminant concentration distributions and for different assumptions about portions of the site with contaminant concentrations below the QL.

***Highlight 6: Directions for Data Quality Assessment for the Max Test***

Let $x_1, x_2, ..., x_N$ represent contaminant concentration measurements for N composite samples that each consist of C specimens selected so that each composite is representative of the EA as a whole. The following describes the steps required to ensure that the Max test achieves the DQOs established for the site.

STEP 1: The site manager determines the Type I error rate to be achieved at 2 SSL and the Type II error rate to be achieved at 0.5 SSL, as described in Section 4.1.6.

STEP 2: Calculate the sample mean $\bar{x} = \left[ \sum_{i=1}^{N} x_i \right] \frac{1}{N}$

STEP 3: Calculate the sample standard deviation

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left( x_i - \bar{x} \right)^2}$$

STEP 4: Calculate the sample estimate of the coefficient of variation, CV, for individual concentration measurements from across the EA.

$$CV = \frac{\sqrt{C}\ s}{\bar{x}}$$

NOTE: This is a conservation approximation of the CV for individual measurements.

STEP 5: If Max $(x_1, x_2, ..., x_N) < \frac{SSL}{\sqrt{C}}$, then no further data quality assessment is needed and the EA needs no further investigation.

Otherwise proceed to Step 6.

STEP 6: Use the value of the sample CV calculated in Step 4 as the true CV of concentrations to determine which column of Table 23 is applicable for determining sample size requirements. Using the error limits established in Step 1, determine the sample size requirements from this table. If the required sample size is greater than that implemented, further investigation of the EA is necessary. The further investigation may consist of selecting a supplemental sample and repeating the Max test with the larger, combined sample.

A limitation of the Max test is that it does not provide as good control over the Type II error rates at 0.5 SSL as it does for Type I error rates at 2 SSL. In fact, for a fixed number, C, of specimens per composite, the Type II error rate increases as the number of composite samples, N, increases. As the sample size increases, the likelihood of observing an unusual sample with the maximum exceeding 2 SSL increases. However, the Type II error rate can be decreased by increasing the number of

specimens per composite. This unusual performance of the Max test as a hypothesis testing procedure occurs because the rejection region is fixed below 2 SSL and thus does not depend on the sample size (as it does for typical hypothesis testing procedures).

**4.1.9 Specify Limits on Decision Errors for Chen Test.** Although the Max test is adequate and appropriate for selecting a sample size for site screening, there are other alternate methods of screening surface soils. One such alternate method is the Chen test. In general, the Chen test differs from the Max test in its basic assumption about site contamination and the purpose of soil sampling. Because of this variation, these two methods have different null hypotheses and different decision error types.

There are two formulations of the statistical hypothesis test concerning the true (but unknown) mean contaminant concentration, $\mu$, that achieve the Soil Screening Guidance decision error rate goals specified in Section 4.1.6. They are:

1. Test the null hypothesis, $H_0$: $\mu \geq 2$ SSL, versus the alternative hypothesis, $H_1$: $\mu < 2$ SSL, at the 5 percent significance level using a sample size chosen to achieve a Type II error rate of 20 percent at 0.5 SSL.

2. Test the null hypothesis, $H_0$: $\mu \leq 0.5$ SSL, versus the alternative hypothesis, $H_1$: $\mu > 0.5$ SSL, at the 20 percent significance level using a sample size chosen to achieve a Type II error rate of 5 percent at 2 SSL.

The first formulation of the problem (which is commonly used in the Superfund program) has the advantage that the error rate that has potential public health consequences is controlled directly via the significance level of the test. The error rate that has primarily cost consequences can be reduced by increasing the sample size above the minimum requirement. However, EPA has identified a new test procedure, the Chen test (Chen, 1995), which requires the second formulation but is less sensitive to assumptions regarding the distribution of the contaminant measurements than the Land procedure used in the December 1994 draft Technical Background Document (see Section 4.3). This section provides guidance regarding application of the Chen test and is, therefore, based on the second formulation of the hypothesis test.

A disadvantage of the second formulation is its performance when the true EA mean is between 0.5 SSL and the SSL. In this case, as the sample size increases, the test indicates the decision to investigate further, even though the mean is less than the SSL. In fact, no test procedure with feasible sample sizes performs well when the true EA mean is in the "gray region" between 0.5 SSL and 2 SSL (see Section 4.3). Whenever large sample sizes are feasible, one should modify the problem statement and test the null hypothesis, $H_0$: $\mu \leq$ SSL, instead of $H_0$: $\mu \leq 0.5$ SSL. One would then develop appropriate DQOs for this modified hypothesis test (e.g., significance level of 20 percent at the SSL and 5 percent probability of decision error at 2 SSL).

When the true mean of an EA is compared with the screening level, there are two possible decision errors that may occur: (1) decide not to investigate an EA further (i.e., "walk away") when the correct decision would be to "investigate further"; and (2) decide to investigate further when the correct decision would be to "walk away." For the Chen test, the "incorrectly walk away" decision error is designated as the Type II decision error because it occurs when we incorrectly accept the null hypothesis. Correspondingly, the "unnecessarily investigate further" decision error is designated as the Type I decision error because it occurs when we incorrectly reject the null hypothesis.

As discussed in Section 4.1.6, the Soil Screening Guidance specifies a default gray region for decision errors from 0.5 SSL to 2 SSL and sets the following goals for Type I and Type II error rates:

- Prob ("investigate further" when the true EA mean is 0.5 SSL) = 0.20
- Prob ("walk away" when the true EA mean is 2 SSL) = 0.05.

Table 24 summarizes this step of the DQO process for the Chen test, specifying limits on the decision error rates, and the final step of the DQO process, optimizing the design.

**4.1.10 Optimize the Design Using the Chen Test.** This section includes guidance on developing an optimum sampling strategy for screening surface soils. It discusses compositing, the selection of sampling points for composited and uncomposited surface soil sampling, and the recommended procedures for determining the sample sizes necessary to achieve specified limits on decision errors using the Chen test.

Note that the size, shape, and orientation of sampling volume (i.e., "support") for heterogenous media have a significant effect on reported measurement values. For instance, particle size has a varying affect on the transport and fate of contaminants in the environment and on the potential receptors. Because comparison of data from methods that are based on different supports can be difficult, defining the sampling support is important in the early stages of site characterization. This may be accomplished through the DQO process with existing knowledge of the site, contamination, and identification of the exposure pathways that need to be characterized. Refer to *Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies* (U.S. EPA, 1992f) for more information about soil sampling support.

The SAP developed for surface soils should specify sampling and analytical procedures as well as the development of QA/QC procedures. To identify the appropriate analytical procedures, the screening levels must be known. If data are not available to calculate site-specific SSLs, then the generic SSLs in Appendix A should be used.

**Compositing.** Because the objective of surface soil screening is to ensure that the mean contaminant concentration does not exceed the screening level, the physical "averaging" that occurs during compositing is consistent with the intended use of the data. Compositing allows a larger number of locations to be sampled while controlling analytical costs because several discrete samples are physically mixed (homogenized) and one or more subsamples are drawn from the mixture and submitted for analysis. If the individual samples in each composite are taken across the EA, each composite represents an estimate of the EA mean.

A practical constraint to compositing in some situations is the heterogeneity of the soil matrix. The efficiency and effectiveness of the mixing process may be hindered when soil particle sizes vary widely or when the soil matrix contains foreign objects, organic matter, viscous fluids, or sticky material. Soil samples should not be composited if matrix interference among contaminants is likely (e.g., when the presence of one contaminant biases analytical results for another).

## Table 24. Sampling Soil Screening DQOs for Surface Soils under Chen Test

| DQO Process Steps | Soil Screening Inputs/Outputs |
|---|---|
| **Specify Limits on Decision Errors** | |
| Define baseline condition (null hypothesis) | EA needs no further investigation |
| Define gray region | From 0.5 SSL to 2 SSL |
| Define Type I and Type II decision errors | Type I error: Investigate further when an EA's true mean concentration is below 0.5 SSL<br>Type II error: Do not investigate further ("walk away from") when an EA true mean concentration is above 2 SSL |
| Identify consequences | Type I error: unnecessary expenditure of resources to investigate further<br>Type II error: potential public health consequences |
| Assign acceptable probabilities of Type I and Type II decision errors | Goals:<br>Type I: 0.20 (20%) probability of investigating further when EA mean is 0.5 SSL<br>Type II: 0.05 (5%) probability of not investigating further when EA mean is 2 SSL |
| **Optimize the Design** | |
| Determine expected variability of EA surface soil contaminant concentrations | A conservatively large expected coefficient of variation (CV) from prior data for the site, field measurements, or data from other comparable sites and expert judgment |
| Design sampling strategy by evaluating costs and performance of alternatives | Lowest cost sampling design option (i.e., compositing scheme and number of composites) that will achieve acceptable decision error rates |
| Develop planning documents for the field investigation | Sampling and Analysis Plan (SAP)<br>Quality Assurance Project Plan (QAPjP) |

Before individual specimens are composited for chemical analysis, the site manager should consider homogenizing and splitting each specimen. By compositing one portion of each specimen with the other specimens and storing one portion for potential future analysis, the spatial integrity of each specimen is maintained. If the concentration in a composite is high, the splits of the individual specimens of which it was composed can be analyzed subsequently to determine which individual specimen(s) have high concentrations. This will permit the site manager to determine which portion within an EA is contaminated without making a repeat visit to the site.

**Sample Pattern.** The Chen test can be applied using composite samples that are representative of the entire EA or with individual uncomposited samples.

Systematic grid sampling (SyGS) generally is preferred because it ensures that the sample points will be dispersed across the entire EA. However, if the boundaries of the EA are irregular (e.g., around the perimeter of the site or the boundaries of a stratum within which the EAs were defined), the number of grid sample points that fall within the EA depends on the random starting point selected. Therefore, for these irregularly shaped EAs, SRS or stratified SRS is recommended. Moreover, if a systematic trend of contamination is suspected across the EA (e.g., a strip of higher contamination),

then SRS or stratified SRS is recommended again. In this case, grid sampling would be likely to result in either over- or under representation of the strip of higher contaminant levels, depending on the random starting point.

For composite sampling, the sampling pattern used to locate the C discrete sample specimens that form each composite sample is important. The composite samples must be formed in a manner that is consistent with the assumptions underlying the sample size calculations. In particular, each composite sample must provide an unbiased estimate of the mean contaminant concentration over the entire EA. One way to construct a valid composite of C specimens is to divide the EA into C sectors, or strata, of equal area and select one point at random from each sector. If sectors (strata) are of unequal sizes, the simple average is no longer representative of the EA as a whole.

Valid sampling patterns and compositing schemes for selecting N composite samples that each consist of C specimens include the following:

1. Select an SRS consisting of C points and composite all specimens associated with these points into a sample. Repeat this process N times, discarding any points that were used in a previous sample.

2. Select an SyGS/rs of C points and composite all specimens associated with the points in this sample. Repeat this process N times, using a new randomly selected starting point each time.

3. Select a single SyGS/rs of CN points and use the systematic compositing scheme that is described in Highlight 3 to form N composites, as illustrated in Figure 6.

4. Select a single SyGS/rs of C×N points and use the random compositing scheme that is described in Highlight 4 to form N composites, as illustrated in Figure 7.

5. Select a stratified random sample of C×N points and use a random compositing scheme, as described in Highlight 5, to form N composites, as illustrated in Figure 8.

Methods 1, 2, and 5 are the most statistically defensible, with method 5 used as the default method in the Soil Screening Guidance. However, given the practical limits of implementing these methods, either method 3 or 4 is generally recommended for EAs with regular boundaries (e.g., square or rectangular). As noted above, if the boundaries of the EA are irregular, SyGS/rs sampling may not result in exactly C×N sample points. Therefore, for EAs with irregular boundaries, method 5 is recommended. Alternatively, a combination of methods 4 and 5 can be used for EAs that can be partitioned into C sectors of equal area of which K have regular boundaries and the remaining C - K have irregular boundaries.

**Sample Size.** This section provides procedures to determine sample size requirements for the Chen test that achieve the site-specific decision error limits discussed in Section 4.1.6. The Chen test is an upper-tail test for the mean of positively skewed distributions, like the lognormal (Chen, 1995). It is based on the mean concentration observed in a simple random sample, or equivalent design, selected from a distribution with a long right-hand tail.

The Chen procedure is a hypothesis testing procedure that is robust among the family of right-skewed distributions (see Section 4.3). That is, decision error rates for a given sample size are relatively insensitive to the particular right-skewed distribution that generated the data. This

robustness is important in the context of surface soil screening because the number of surface soil samples will usually not be sufficient to determine the distribution of the concentration measurements.

The procedures presented above for selecting composited or uncomposited simple random or systematic grid samples can all be used to generate samples for application of the Chen test. The Chen procedure is based on a simple random sample, or one that can be analyzed as if it were an SRS. Directions for performing the Chen test in a manner that is consistent with the DQOs that have been established for a site are presented later.

Tables 25 through 30 provide the sample sizes required for the Chen test performed at the 10, 20, or 40 percent levels of significance (probability of Type I error at 0.5 SSL) and achieve, at most, a 5 or 10 percent probability of (Type II) error at 2 SSL. The Type II error rates at 2 SSL are based on the simulations presented in Appendix I. These simulations are based on the following assumptions:

1. Each of the N composite samples is based on C specimens selected to be representative of the EA as a whole, as specified above.

2. One-half the EA has concentrations below the quantitation limit (which is assumed to be SSL/100).

3. One-half the EA has concentrations that follow a gamma distribution.

4. Measurements below the QL are replaced by 0.5 QL for computation of the Chen test statistic.

5. Each chemical analysis is subject to a 20 percent measurement error.

Distributional assumptions 2 and 3 were used as the basis for the Type II error rates at 2 SSL (shown in Tables 25 through 30) because they were found in the simulations to produce high error rates relative to other potential contaminant distributions. If the proportion of the site below the QL is less than half or if the distribution of the concentration measurements is some other right-skewed distribution (e.g., lognormal), rather than gamma, then the Type II error rates achieved are likely to be no worse than those cited in Tables 25 through 30. No sample sizes, N, less than four are shown in these tables (irrespective of the number of specimens per composite) because consideration of the simulation results presented in Section 4.3 has led to a program-level decision that at least four separate analyses are required to adequately characterize the mean of an EA. No sample sizes in excess of nine are presented because of a program-level decision that more than nine samples per exposure area is generally unlikely for screening surface soils at Superfund sites. However, additional sample size options can be determined from the simulations reported in Appendix I.

When using Tables 25 through 30 to determine the sample size pair (C and N) needed to achieve satisfactory error rates with the Chen test, investigators must have an *a priori* estimate of the CV for measurements of the contaminant of interest across the EA. As previously discussed for the Max test, the site manager should use a conservatively large estimate of the CV for determining sample size requirements because additional sampling will be required if the data suggest that the true CV is greater than that used to determine the sample sizes.

## Table 25. Minimum Sample Size for Chen Test at 10 Percent Level of Significance to Achieve a 5 Percent Chance of "Walking Away" When EA Mean is 2.0 SSL, Given Expected CV for Concentrations Across the EA

| Number of specimens per composite[b] | Coefficient of variation (CV)[a] | | | | |
|---|---|---|---|---|---|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
| 2 | 7 | 9 | >9 | >9 | >9 |
| 3 | 5 | 7 | 9 | >9 | >9 |
| 4 | 4 | 6 | 8 | >9 | >9 |
| 5 | 4 | 5 | 6 | 8 | >9 |
| 6 | 4 | 4 | 5 | 7 | 9 |

[a]The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA and includes measurement error.
[b]Each composite consists of points from a stratified random or systematic grid sample across the entire EA.
NOTE: Sample sizes are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the limit of detection, and that half the EA has concentrations following a gamma distribution (a conservative distributional assumption).

## Table 26. Minimum Sample Size for Chen Test at 20 Percent Level of Significance to Achieve a 5 Percent Chance of "Walking Away" When EA Mean is 2.0 SSL, Given Expected CV for Concentrations Across the EA

| Number of specimens per composite[b] | Coefficient of variation (CV)[a] | | | | | |
|---|---|---|---|---|---|---|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 |
| 1 | 9 | >9 | >9 | >9 | >9 | >9 |
| 2 | 5 | 7 | >9 | >9 | >9 | >9 |
| 3 | 4 | 5 | 7 | 9 | >9 | >9 |
| 4 | 4 | 4 | 6 | 7 | >9 | >9 |
| 5 | 4 | 4 | 4 | 6 | 8 | >9 |
| 6 | 4 | 4 | 4 | 5 | 8 | 9 |

[a]The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA and includes measurement error.
[b]Each composite consists of points from a stratified random or systematic grid sample across the entire EA.
NOTE: Sample sizes are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the limit of detection, and that half the EA has concentrations following a gamma distribution (a conservative distributional assumption).

## Table 27. Minimum Sample Size for Chen Test at 40 Percent Level of Significance to Achieve a 5 Percent Chance of "Walking Away" When EA Mean is 2.0 SSL, Given Expected CV for Concentrations Across the EA

| Number of specimens per composite[b] | Coefficient of variation (CV)[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
| 1 | 5 | 9 | >9 | >9 | >9 | >9 | >9 |
| 2 | 4 | 4 | 8 | 9 | >9 | >9 | >9 |
| 3 | 4 | 4 | 5 | 7 | >9 | >9 | >9 |
| 4 | 4 | 4 | 4 | 5 | 8 | >9 | >9 |
| 5 | 4 | 4 | 4 | 5 | 6 | 9 | >9 |
| 6 | 4 | 4 | 4 | 4 | 5 | 8 | 9 |

[a]The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA and includes measurement error.
[b]Each composite consists of points from a stratified random or systematic grid sample across the entire EA.
NOTE: Sample sizes are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the limit of detection, and that half the EA has concentrations following a gamma distribution (a conservative distributional assumption).

## Table 28. Minimum Sample Size for Chen Test at 10 Percent Level of Significance to Achieve a 10 Percent Chance of "Walking Away" When EA Mean is 2.0 SSL, Given the Expected CV for Concentrations Across the EA

| Number of specimens per composite[b] | Coefficient of variation (CV)[a] | | | | | |
|---|---|---|---|---|---|---|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 |
| 2 | 6 | 7 | >9 | >9 | >9 | >9 |
| 3 | 4 | 5 | 7 | >9 | >9 | >9 |
| 4 | 4 | 4 | 6 | 7 | >9 | >9 |
| 5 | 4 | 4 | 5 | 6 | 8 | >9 |
| 6 | 4 | 4 | 4 | 5 | 7 | 9 |

[a]The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA and includes measurement error.
[b]Each composite consists of points from a stratified random or systematic grid sample across the entire EA.
NOTE: Sample sizes are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the limit of detection, and that half the EA has concentrations following a gamma distribution (a conservative distributional assumption).

## Table 29. Minimum Sample Size for Chen Test at 20 Percent Level of Significance to Achieve a 10 Percent Chance of "Walking Away" When EA Mean is 2.0 SSL, Given Expected CV for Concentrations Across the EA

| Number of specimens per composite[b] | Coefficient of variation (CV)[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
| 1 | 7 | 9 | >9 | >9 | >9 | >9 | >9 |
| 2 | 4 | 5 | 8 | >9 | >9 | >9 | >9 |
| 3 | 4 | 4 | 5 | 8 | >9 | >9 | >9 |
| 4 | 4 | 4 | 4 | 5 | 8 | >9 | >9 |
| 5 | 4 | 4 | 4 | 5 | 6 | 8 | >9 |
| 6 | 4 | 4 | 4 | 4 | 5 | 7 | 9 |

[a]The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA and includes measurement error.
[b]Each composite consists of points from a stratified random or systematic grid sample across the entire EA.
NOTE: Sample sizes are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the limit of detection, and that half the EA has concentrations following a gamma distribution (a conservative distributional assumption).

## Table 30. Minimum Sample Size for Chen Test at 40 Percent Level of Significance to Achieve a 10 Percent Chance of "Walking Away" When EA Mean is 2.0 SSL, Given Expected CV for Concentrations Across the EA

| Number of specimens per composite[b] | Coefficient of variation (CV)[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
| 1 | 4 | 7 | 9 | >9 | >9 | >9 | >9 |
| 2 | 4 | 4 | 5 | 8 | 9 | >9 | >9 |
| 3 | 4 | 4 | 4 | 5 | 7 | 9 | >9 |
| 4 | 4 | 4 | 4 | 4 | 5 | 7 | >9 |
| 5 | 4 | 4 | 4 | 4 | 5 | 6 | 8 |
| 6 | 4 | 4 | 4 | 4 | 4 | 5 | 6 |

[a]The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA and includes measurement error.
[b]Each composite consists of points from a stratified random or systematic grid sample across the entire EA.
NOTE: Sample sizes are based on 1,000 simulations that assume that each composite is representative of the entire EA, that half the EA has concentrations below the limit of detection, and that half the EA has concentrations following a gamma distribution (a conservative distributional assumption).

Given an *a priori* estimate of the CV of concentration measurements in the EA, the site manager can use Table 26 to determine a sample size option that achieves the decision error goals for surface soil screening presented in Section 4.1.6 (i.e., not more than 20 percent chance of error at 0.5 SSL and not more than 5 percent at 2 SSL). For example, suppose that the site manager expects that the maximum true CV for concentration measurements in an EA is 2. Then Table 26 shows that six composite samples, each consisting of four specimens, will be sufficient to achieve the decision error limit goals.

**4.1.11 Using the DQA Process: Analyzing Chen Test Data.** Step-by-step instructions for using the Chen test to analyze data from both discrete random samples and pseudo-random samples (e.g., composite samples constructed as described previously) are provided in Highlight 7. This method for analyzing the data is a robust procedure for an upper-tailed test for the mean of a positively skewed distribution. As explained by Chen (1995), this procedure is a robust generalization of the familiar Student's t-test; it further generalizes a method developed by Johnson (1978) for asymmetric distributions.

The only assumption necessary for valid application of the Chen procedure is that the sample be a random sample from a right-skewed distribution. This robustness within the broad family of right-skewed distributions is appropriate for screening surface soil because the distribution of concentrations within an EA may depart from the common assumption of lognormality.

Computation of the Chen test statistic, as shown in Highlight 7, requires that concentration values be available for all N individual or composite samples analyzed for the contaminant of interest. If an analytical test result is reported below the quantitation limit, it should be used in the computations. For results below detection, substitute one-half the QL.

A disadvantage of the Chen procedure is that the hypothesis, "the EA needs no further investigation," must be treated as the alternative hypothesis, rather than as the null hypothesis. As a result, the Type I error rate at 0.5 SSL is controlled via the significance level of the test, rather than the error rate at 2 SSL, which may have public health consequences. Hence, if the sample sizes (C and N) are based on an assumed CV that is too small, the desired error rate at 2 SSL is likely not to be achieved. Therefore, it is important to perform the data quality assurance check specified in Steps 6 through 8 of Highlight 7 to ensure that the desired error rate at 2 SSL is achieved. Moreover, it is important that the site manager base the initial EA sample sizes on a conservatively large estimate of the CV so that this process will not result in the need for additional sampling.

**4.1.12 Special Considerations for Multiple Contaminants.** If the surface soil samples collected for an EA will be tested for multiple contaminants, be aware that the expected CVs for the different contaminants may not all be identical. A conservative approach is to base the sample sizes for all contaminants on the largest expected CV.

**4.1.13 Quality Assurance/Quality Control Requirements.** Regardless of the sampling approach used, the Superfund quality assurance program guidance must be followed to ensure that measurement error rates are documented and within acceptable limits (U.S. EPA, 1993d).

***Highlight 7: Directions for the Chen Test Using Simple Random Sample Scheme***

Let $x_1, x_2, \ldots, x_N$, represent concentration measurements for N random sampling points or N pseudo-random sampling points (i.e., from a design that can be analyzed as if it were a simple random sample). The following describes the steps for a one-sample test for $H_o$: $\mu \leq 0.5$ SSL at the $100\alpha\%$ significance level that is designed to achieve a $100ß\%$ chance of incorrectly accepting $H_o$ when $\mu = 2$ SSL.

STEP 1: Calculate the sample mean $$\bar{x} = \left[\sum_{i=1}^{N} x_i\right] \frac{1}{N}$$

STEP 2: Calculate the sample standard deviation

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}$$

STEP 3: Calculate the sample skewness

$$b = N\,\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{(N-1)\,(N-2)\,s^3}$$

STEP 4: Calculate the Chen test statistic, $t_2$, as follows:

$$a = \frac{b}{6\sqrt{N}}$$

$$t = \frac{\bar{x} - 0.5\ \text{SSL}}{s\ /\ \sqrt{N}}$$

$$t_2 = t + a\left(1 + 2t^2\right) + 4a^2\left(t + 2t^3\right)$$

STEP 5: Compare $t_2$ to $z_\alpha$, the $100(1-\alpha)$ percentile of the standard normal probability distribution.

If $t_2 > z_\alpha$, the null hypothesis is rejected, and the EA needs further investigation.

If $t_2 \leq z_\alpha$, there is insufficient evidence to reject the null hypothesis. Proceed to Step 6 to determine if the sample size is sufficient to achieve a $100ß\%$ or less chance of incorrectly accepting the $H_o$ when $\mu = 2$ SSL.

***Highlight 7: Directions for the Chen Test Using Simple Random Sample Scheme (continued)***

STEP 6: Let C represent the number of specimens composited to form each of the N samples, where each of $x_1, x_2, ..., x_N$ is a composite sample consisting of C specimens selected so that each composite is representative of the EA as a whole. (If each of $x_1, x_2, ..., x_N$ is an individual random or pseudo-random sampling point, then C = 1.)

If $\text{Max}(x_1, x_2, ..., x_N) < \frac{SSL}{\sqrt{C}}$, then no further data quality assessment is needed and the EA needs no further investigation.

Otherwise proceed to Step 7.

STEP 7: Calculate the sample estimate of the coefficient of variation, CV, for individual concentration measurements from across the EA.

$$CV = \frac{\sqrt{C}\ s}{\bar{x}}$$

NOTE: This calculation ignores measurement error, which results in conservatively large sample size requirements.

STEP 8: Use the value of the sample CV calculated in Step 7 as the true CV of concentrations in Tables 25 through 30 to determine the minimum sample size, $N^*$, necessary to achieve a 100ß% or less chance of incorrectly accepting $H_o$ when $\mu = 2$ SSL.

If $N \geq N^*$, the EA needs no further investigation.

If $N < N^*$, further investigation of the EA is necessary. The further investigation may consist of selecting a supplemental sample and repeating this hypothesis testing procedure with the larger, combined sample.

**4.1.14 Final Analysis.** After either the Max test or the Chen test has been performed for each EA of interest (0.5 acre or less) at an NPL site, the pattern of decisions for individual EAs (to "walk away" or to "investigate further") should be examined. If some EAs for which the decision was to "walk away" are surrounded by EAs for which the decision was to "investigate further," it may be more efficient to identify an area including all these EAs for further study and develop a global investigation strategy.

**4.1.15 Reporting.** The decision process for surface soil screening should be thoroughly documented as part of the RI/FS process. This documentation should include a map of the site

(showing the boundaries of the EAs and the sectors, or strata, within EAs that were used to select sampling points within the EAs); documentation of how composite samples were formed and the number of composite samples that were analyzed for each EA; the raw analytical data; the results of all hypothesis tests; and the results of all QA/QC analyses.

## 4.2 Sampling Subsurface Soils

Subsurface soil sampling is conducted to estimate the mean concentrations of contaminants in each source at a site for comparison to inhalation and migration to ground water SSLs. Measurements of soil properties and estimates of the area and depth of contamination in each source are also needed to calculate SSLs for these pathways. Table 31 shows the steps in the DQO process necessary to develop a sampling strategy to meet these objectives. Each of these steps is described below.

**4.2.1 State the Problem.** Contaminants present in subsurface soils at the site may pose significant risk to human health and the environment through the inhalation of volatiles or by the migration of contaminants through soils to an underlying potable aquifer. The problem is to identify the contaminants and source areas that do not pose significant risk to human health through either of these exposure pathways so that future investigations may be focused on areas and contaminants of true concern.

Site-specific activities in this step include identifying the data collection planning team (including technical experts and key stakeholders) and specifying the available resources (i.e., the cost and time available for sampling). The list of technical experts and stakeholders should contain all key personnel who are involved with applying SSLs to the site. Other activities include developing the conceptual site model and identifying exposure scenarios, which are fully addressed in the *Soil Screening Guidance: User's Guide* (U.S. EPA, 1996).

**4.2.2 Identify the Decision.** The decision is to determine whether mean soil concentrations in each source area exceed inhalation or migration to ground water SSLs for specific contaminants. If so, the source area will be investigated further. If not, no further action will be taken under CERCLA.

**4.2.3 Identify Inputs to the Decision.** Site-specific inputs to the decision include the average contaminant concentrations within each source area and the inhalation and migration ground water SSLs. Calculation of the SSLs for the two pathways of concern also requires site-specific measurements of soil properties (i.e., bulk density, fraction organic carbon content, pH, and soil texture class) and estimates of the areal extent and depth of contamination.

A list of feasible sampling and analytical methods should be assembled during this step. EPA recommends the use of field methods where applicable and appropriate. Verify that Contract Laboratory Program (CLP) methods and field methods for analyzing the samples exist and that the analytical method detection limits or field method detection limits are appropriate for the site-specific or generic SSL. The *Sampler's Guide to the Contract Laboratory Program* (U.S. EPA, 1990) and the *User's Guide to the Contract Laboratory Program* (U.S. EPA, 1991d) contain further information on CLP methods.

## Table 31. Soil Screening DQOs for Subsurface Soils

| DQO Process Steps | Soil Screening Inputs/Outputs |
|---|---|
| **State the Problem** | |
| Identify scoping team | Site manager and technical experts (e.g., toxicologists, risk assessors, hydrogeologists, statisticians). |
| Develop conceptual site model (CSM) | CSM development (described in Step 1 of the User's Guide, U.S. EPA, 1996). |
| Define exposure scenarios | Inhalation of volatiles and migration of contaminants from soil to potable ground water (and plant uptake for certain contaminants). |
| Specify available resources | Sampling and analysis budget, scheduling constraints, and available personnel. |
| Write brief summary of contamination problem | Summary of the subsurface soil contamination problem to be investigated at the site. |
| **Identify the Decision** | |
| Identify decision | Do mean soil concentrations for particular contaminants (e.g., contaminants of potential concern) exceed appropriate SSLs? |
| Identify alternative actions | Eliminate area from further action or study under CERCLA<br>or<br>Plan and conduct further investigation. |
| **Identify Inputs to the Decision** | |
| Identify decision | Volatile inhalation and migration to ground water SSLs for specified contaminants<br>Measurements of subsurface soil contaminant concentration |
| Define basis for screening | Soil Screening Guidance |
| Identify analytical methods | Feasible analytical methods (both field and laboratory) consistent with program-level requirements. |
| **Specify the Study Boundaries** | |
| Define geographic areas of field investigation | The entire NPL site (which may include areas beyond facility boundaries), except for any areas with clear evidence that no contamination has occurred. |
| Define population of interest | Subsurface soils |
| Define scale of decision making | Sources (areas of contiguous soil contamination, defined by the area and depth of contamination or to the water table, whichever is more shallow). |
| Subdivide site into decision units | Individual sources delineated (area and depth) using existing information or field measurements (several nearby sources may be combined into a single source). |
| Define temporal boundaries of study | Temporal constraints on scheduling field visits. |
| Identify (list) practical constraints | Potential impediments to sample collection, such as access, health, and safety issues. |
| **Develop a Decision Rule** | |
| Specify parameter of interest | Mean soil contaminant concentration in a source (as represented by discrete contaminant concentrations averaged within soil borings). |
| Specify screening level | SSLs calculated using available parameters and site data (or generic SSLs if site data are unavailable). |
| Specify "if..., then..." decision rule | If the mean soil concentration exceeds the SSL, then investigate the source further. If the mean soil boring concentration is less than the SSL, then no further investigation is required under CERCLA. |

## Table 31. (continued)

| | |
|---|---|
| **Specify Limits on Decision Errors** | |
| Define QA/QC goals | CLP precision and bias requirements<br>10% CLP analyses for field methods |
| **Optimize the Design** | |
| Determine how to estimate mean concentration in a source | For each source, the highest mean soil core concentration (i.e., depth-weighted average of discrete contaminant concentrations within a boring). |
| Define subsurface sampling strategy by evaluating costs and site-specific conditions | Number of soil borings per source area; number of sampling intervals with depth. |
| Develop planning documents for the field investigation | Sampling and Analysis Plan (SAP)<br>Quality Assurance Project Plan (QAPjP) |

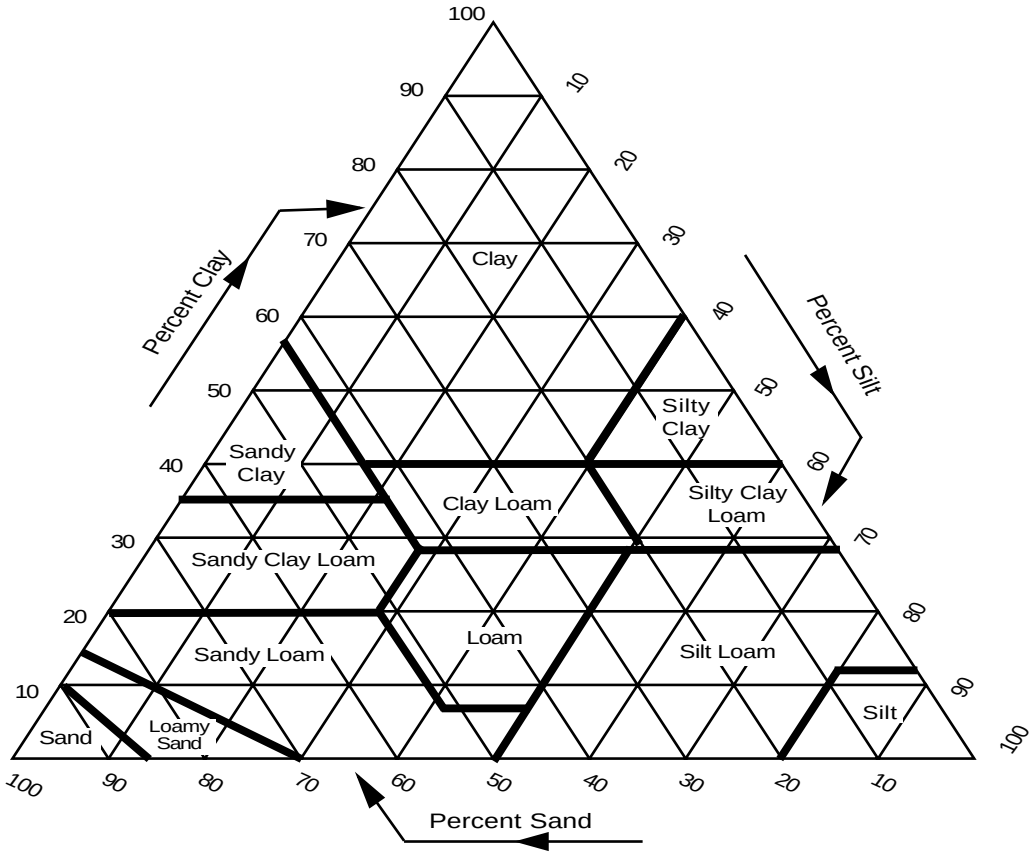
Field methods will be useful in defining the study boundaries (i.e., area and depth of contamination) during site reconnaissance and during the sampling effort. For example, soil gas survey is an ideal method for determining the extent of volatile contamination in the subsurface. EPA expects field methods will become more prevalent and useful because the design and capabilities of field portable instrumentation are rapidly evolving. Documents on standard operating procedures (SOPs) for field methods are available through NTIS and should be referenced in soil screening documentation if these methods are used.

Soil parameters necessary for SSL calculation are soil texture, bulk density, and soil organic carbon. Some of these parameters can be measured in the field, others require laboratory measurement. Although laboratory measurements of these parameters cannot be obtained under the Superfund Contract Laboratory Program, they are readily available from soil testing laboratories across the country.

Note that the size, shape, and orientation of sampling volume (i.e., "support") for heterogenous media have a significant effect on reported measurement values. For instance, particle size has a varying affect on the transport and fate of contaminants in the environment and on the potential receptors. Comparison of data from methods that are based on different supports can be difficult. Defining the sampling support is important in the early stages of site characterization. This may be accomplished through the DQO process with existing knowledge of the site, contamination, and identification of the exposure pathways that need to be characterized. Refer to *Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies* (U.S. EPA, 1992f) for more information about soil sampling support.

**Soil Texture.** The soil texture class (e.g., loam, sand, silt loam) is necessary to estimate average soil moisture conditions and to estimate infiltration rates. A soil's texture classification is determined from a particle size analysis and the U.S. Department of Agriculture (USDA) soil textural triangle shown at the top of Figure 9. This classification system is based on the USDA soil particle size classification at the bottom of Figure 9. The particle size analysis method in Gee and Bauder (1986) can provide this particle size distribution also. Other particle size analysis methods may be used as long as they provide the same particle size breakpoints for sand/silt (0.05 mm) and silt/clay (0.002 mm). Field methods are an alternative for determining soil textural class; an example from Brady (1990) is also presented in Figure 9.

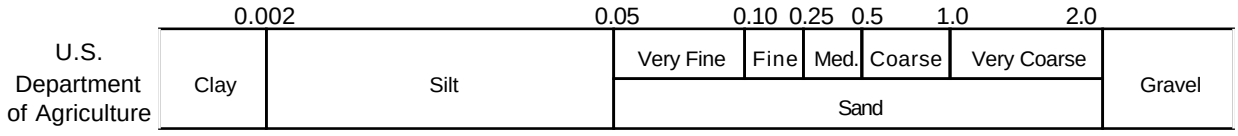
**Figure 9: U.S. Department of Agriculture soil texture classification.**




**Criteria Used with the Field Method for Determining Soil Texture Classes** (Source: Brady, 1990)

| Criterion | Sand | Sandy loam | Loam | Silt loam | Clay loam | Clay |
|---|---|---|---|---|---|---|
| 1. Individual grains visible to eye | Yes | Yes | Some | Few | No | No |
| 2. Stability of dry clods | Do not form | Do not form | Easily broken | Moderately easily broken | Hard and stable | Very hard and stable |
| 3. Stability of wet clods | Unstable | Slightly stable | Moderately stable | Stable | Very stable | Very stable |
| 4. Stability of "ribbon" when wet soil rubbed between thumb and fingers | Does not form | Does not form | Does not form | Broken appearance | Thin, will break | Very long, flexible |

**Particle Size, mm**




Source: USDA.

**Dry Bulk Density.** Dry soil bulk density ($\rho_b$) is used to calculate total soil porosity and can be determined for any soil horizon by weighing a thin-walled tube soil sample (e.g., Shelby tube) of known volume and subtracting the tube weight to estimate field bulk density (ASTM D 2937). A moisture content determination (ASTM 2216) is then made on a subsample of the tube sample to adjust field bulk density to dry bulk density. The other methods (e.g., ASTM D 1556, D 2167, D 2922) are not generally applicable to subsurface soils. ASTM soil testing methods are readily available in the *Annual Book of ASTM Standards, Volume 4.08, Soil and Rock; Building Stones*, which is available from ASTM, 100 Barr Harbor Drive, West Conshohocken, PA, 19428.

**Organic Carbon and pH.** Soil organic carbon is measured by burning off soil carbon in a controlled-temperature oven (Nelson and Sommers, 1982). This parameter is used to determine soil-water partition coefficients from the organic carbon soil-water partition coefficient, $K_{oc}$. Soil pH is used to select site-specific partition coefficients for metals and ionizing organic compounds (see Part 5). This simple measurement is made with a pH meter in a soil/water slurry (McLean, 1982) and may be measured in the field using a portable pH meter.

### 4.2.4 Define the Study Boundaries.

As discussed in Section 4.1.4, areas that are known to be highly contaminated (i.e., sources) are targeted for subsurface sampling. The information collected on source area and depth is used to calculate site-specific SSLs for the inhalation and migration to ground water pathways. Contamination is defined by the lower of the CLP practical quantitation limit for each contaminant or the SSL. For the purposes of this guidance, source areas are defined by area and depth as contiguous zones of contamination. However, discrete sources that are near each other may be combined and investigated as a single source if site conditions warrant.

### 4.2.5 Develop a Decision Rule.

The decision rule for subsurface soils is:

> If the mean concentration of a contaminant within a source area exceeds the screening level, then investigate that area further.

In this case "screening level" means the SSL. As explained in Section 4.1.5, statistics other than the mean (e.g., the maximum concentration) may be used as estimates of the mean in this comparison as long as they represent valid or conservative estimates of the mean.

### 4.2.6 Specify Limits on Decision Errors.

EPA recognizes that data obtained from sampling and analysis can never be perfectly representative or accurate and that the costs of trying to achieve near-perfect results can outweigh the benefits. Consequently, EPA acknowledges that uncertainty in data must be tolerated to some degree. The DQO process attempts to control the degree to which uncertainty in data affects the outcomes of decisions that are based on data.

The sampling intensity necessary to accurately determine the mean concentration of subsurface soil contamination within a source with a specified level of confidence (e.g., 95 percent) is impracticable for screening due to excessive costs and difficulties with implementation. Therefore, EPA has developed an alternative decision rule based on average concentrations within individual soil cores taken in a source:

> If the mean concentration within **any** soil core taken in a source exceeds the screening level, then investigate that source further.

For each core, the mean core concentration is defined as the depth-weighted average concentration within the zone of contamination (see Section 4.2.7). Since the soil cores are taken in the area(s) of highest contamination within each source, the highest average core concentration among a set of core samples serves as a conservative estimate of the mean source concentration. Because this rule is not a statistical decision, it is not possible to statistically define limits on decision errors.

Standard limits on the precision and bias of sampling and analytical operations conducted during the sampling program do apply. These are specified by the Superfund quality assurance program requirements (U.S. EPA, 1993d), which must be followed during the subsurface sampling effort.

If field methods are used, at least 10 percent of field samples should be split and sent to a CLP laboratory for confirmatory analysis (U.S. EPA, 1993d).

Although the EPA does not require full CLP sample tracking and quality assurance/quality control (QA/QC) procedures for measurement of soil properties, routine EPA QA/QC procedures are recommended, including a Quality Assurance Project Plan (QAPjP), chain-of-custody forms, and duplicate analyses.

**4.2.7 Optimize the Design.** Within each source, the Soil Screening Guidance suggests taking two to three soil cores using split spoon or Shelby tube samplers. For each soil core, samples should begin at the ground surface and continue at approximately 2-foot intervals until no contamination is encountered or to the water table, whichever is shallower. **Subsurface sampling depths and intervals can be adjusted at a site to accommodate site-specific information on surface and subsurface contaminant distributions and geological conditions** (e.g., large vadose zones in the West).

The number and location of subsurface soil sampling (i.e., soil core) locations should be based on knowledge of likely surface soil contamination patterns and subsurface conditions. This usually means that core samples should be taken directly beneath areas of high surface soil contamination. Surface soils sampling efforts and field measurements (e.g., soil gas surveys) taken during site reconnaissance will provide information on source areas and high contaminant concentrations to help target subsurface sampling efforts. Information in the CSM also will provide information on areas likely to have the highest levels of contamination. Note that there may be sources buried in subsurface soils that are not discernible at the surface. Information on past practices at the site included in the CSM can help identify such areas. Surface geophysical methods also can aid in identifying such areas (e.g., magnetometry to detect buried drums).

The intensity of the subsurface soil sampling needed to implement the soil screening process typically will not be sufficient to fully characterize the extent of subsurface contamination. In these cases, conservative assumptions should be used to develop hypotheses on likely contaminant distributions (e.g., the assumption that soil contamination extends to the water table). Along with knowledge of subsurface hydrogeology and stratigraphy, geostatistics can be a useful tool in developing subsurface contaminant distributions from limited data and can provide information to help guide additional sampling efforts. However, instructions on the use of geostatistics is beyond the scope of this guidance.

Samples for measuring soil parameters should be collected when taking samples for measuring contaminant concentrations. If possible, consider splitting single samples for contaminant and soil parameter measurements. Many soil testing laboratories have provisions in place for handling and testing contaminated samples. However, if testing contaminated samples is a problem, samples may be taken from clean areas of the site as long as they represent the same soil texture and series and are

taken from the same depth as the contaminant concentration samples.

The SAP developed for subsurface soils should specify sampling and analytical procedures as well as the development of QA/QC procedures. To identify the appropriate analytical procedures, the screening levels must be known. If data are not available to calculate site-specific SSLs, then the generic SSLs in Appendix A should be used.

Finally, soil investigation for the migration to ground water pathway should not be conducted independently of ground water investigations. Contaminated ground water may indicate the presence of a nearby source area, with contaminants leaching from soil into the aquifer.

**4.2.8 Analyzing the Data.** The mean soil contaminant concentration for each soil core should be compared to the SSL for the contaminant. The soil core average should be obtained by averaging analyses results for the discrete samples taken along the entire soil core within the zone of contamination (compositing will prevent the evaluation of contaminant concentration trends with depth).

If each subsurface soil core segment represents the same subsurface soil interval (e.g., 2 feet), then the average concentration from the surface to the depth of contamination is the simple arithmetic average of the concentrations measured for core samples representative of each of the 2-foot segments from the surface to the depth of contamination or to the water table. However, if the intervals are not all of the same length (e.g., some are 2 feet while others are 1 foot or 6 inches), then the calculation of the average concentration in the total core must account for the different lengths of the intervals.

If $c_i$ is the concentration measured in a core sample representative of a core interval of length $l_i$, and the n-th interval is considered to be the last interval in the source area (i.e., the n-th sample represents the depth of contamination), then the average concentration in the core from the surface to the depth of contamination should be calculated as the following depth-weighted average ($\overline{c}$),

$$\overline{c} = \frac{\sum_{i=1}^{n} l_i c_i}{\sum_{i=1}^{n} l_i} \qquad \textbf{(61)}$$

If the leach test option is used, a sample representing the average contaminant concentration within the zone of contamination should be formed for each soil core by combining discrete samples into a composite sample for the test. The composites should include only samples taken within the zone of contamination (i.e., clean soil below the lower limit of contamination should not be mixed with contaminated soil).

As with any Superfund sampling effort, all analytical data should be reviewed to ensure that Superfund quality assurance program requirements are met (U.S. EPA, 1993d).

**4.2.9 Reporting.** The decision process for subsurface soil screening should be thoroughly documented. This documentation should contain as a minimum: a map of the site showing the contaminated soil sources and any areas assumed not to be contaminated, the soil core sampling points within each source, and the soil core sampling points that were compared with the SSLs; the depth and area assumed for each source and their basis; the average soil properties used to calculate

SSLs for each source; a description of how samples were taken and (if applicable) how composite samples were formed; the raw analytical data; the average soil core contaminant concentrations compared with the SSLs for each source; and theresults of all QA/QC analyses.

## 4.3 Basis for the Surface Soil Sampling Strategies: Technical Analyses Performed

This section describes a series of technical analyses conducted to support the sampling strategy for surface soils outlined in the Soil Screening Guidance. Section 4.3.1 describes the sample design procedure presented in the December 1994 draft guidance (U.S. EAP, 1994h). The remaining sections describe the technical analyses conducted to develop the final SSL sampling strategy. Section 4.3.2 describes an alternative, nonparametric procedure that EPA considered but **rejected** for the soil screening strategy.

Section 4.3.3 describes the simulations conducted to support the selection of the Max test and the Chen test in the final Soil Screening Guidance. These simulation results also can be used to determine sample sizes for site conditions not adequately addressed by the tables in Section 4.1. Quantitation limit and multiple comparison issues are discussed in Sections 4.3.4 and 4.3.5, respectively. Section 4.3.6 describes a limited investigation of compositing samples within individual EA sectors or strata.

**4.3.1 1994 Draft Guidance Sampling Strategy.** The DQO-based sampling strategy in the 1994 draft Soil Screening Guidance assumed a lognormal distribution for contaminant levels over an EA and derived sample size determinations from lognormal confidence interval procedures by C. E. Land (1971). This section summarizes the rationale for this approach and technical issues raised by peer review.

For the 1994 draft Soil Screening Guidance, EPA based the surface soil SSL methodology on the comparison of the arithmetic mean concentration over an EA with the SSL. As explained in Section 4.1, this approach reflects the type of exposure to soil under a future residential land use scenario. A person moving randomly across a residential lot would be expected to experience an average concentration of contaminants in soil.

Generally speaking, there are few nonparametric approaches to statistical inference about a mean unless a symmetric distribution (e.g., normal) is assumed, in which case the mean and median are identical and inference about the median is the same as inference about the mean. However, environmental contaminant concentration distributions over a surface area tend to be skewed with a long right tail, so symmetry is not plausible. In this case the main options for inference about means are inherently parametric, i.e., they are based on an assumed family of probability distributions.

In addition to being skewed with a long right tail, environmental contaminant concentration data must be positive because concentration measurements cannot be negative. Several standard two-parameter probability models are nonnegative and skewed to the right, including the gamma, lognormal, and Weibull distributions. The properties of these distributions are summarized in Chapter 12 of Gilbert (1987).

The lognormal distribution is the distribution most commonly used for environmental contaminant data (see, e.g., Gilbert, 1987, page 164). The lognormal family can be easy to work with in some respects, due to the work of Land (1971, 1975) on estimating confidence intervals for lognormal parameters, which are also described in Gilbert (1987).

The equation for estimating the Land upper confidence limit (UL) for a lognormal mean has the form

$$UL = \exp\left( \bar{y} + \frac{s_y^2}{2} + \frac{s_y H}{\sqrt{n - 1}} \right) \quad \textbf{(62)}$$

where $\bar{y}$ and $s_y$ are the average and standard deviation of the sample log concentrations. The lower confidence limit (LL) has a similar form. The factor H depends on $s_y$ and n and is tabulated in Gilbert (1987) and Land (1975). If the data truly follow a lognormal distribution, then the Land confidence limits are exact (i.e., the coverage probability of a 95 percent confidence interval is 0.95).

The problem formulation used to develop SSL DQOs in the 1994 draft Soil Screening Guidance tested the null hypothesis $H_0$: $\mu \geq 2$ SSL versus the alternative hypothesis $H_1$: $\mu < 2$ SSL, with a Type I error rate of 0.05 (at 2 SSL), and a Type II error rate of 0.20 at 0.5 SSL ($\mu$ represents the true EA mean). That is, the probability of incorrectly deciding not to investigate further when the true mean is 2 SSL was set not to exceed 0.05, and the probability of incorrectly deciding to investigate further when the true mean is 0.5 SSL was not to exceed 0.20.

This null hypothesis can be tested at the 5 percent level of significance by calculating Land's upper 95 percent confidence limit for a lognormal mean, if one assumes that the true EA concentrations are lognormally distributed. The null hypothesis is rejected if the upper confidence limit falls below 2 SSL.

Simulation studies of the Land procedure were used to obtain sample size estimates that achieve these DQOs for different possible values of the standard deviation of log concentrations. Additional simulation studies were conducted to calculate sample sizes and to investigate the properties of the Land procedure in situations where specimens are composited.

All of these simulation studies assumed a lognormal distribution of site concentrations. If the underlying site distribution is lognormal, then the composites, viewed as physical averages, are not lognormal (although they may be approximately lognormal). Hence, correction factors are necessary to apply the Land procedure with compositing, if the individual specimen concentrations are assumed lognormal. The correction factors were also developed through simulations. The correction factors are multiplied by the sample standard deviation, $s_y$, before calculating the confidence limit and conducting the test.

Procedures for estimating sample sizes and testing hypotheses about the site mean using the Land procedure, with and without compositing, are described in the 1994 draft Technical Background Document (U.S. EPA, 1994i).

A peer review of the draft Technical Background Document identified several issues of concern:

- The use of a procedure relying strongly on the assumption of a lognormal distribution
- Quantitation limit issues
- Issues associated with multiple hypothesis tests where multiple contaminants are present in site soils.

The first issue is of concern because the small sample sizes appropriate for surface soil screening will not provide sufficient data to validate this assumption. To address this issue, EPA considered several alternative approaches and performed extensive analyses. These analyses are described in Sections 4.3.2 and 4.3.3. Section 4.3.3 describes extensive simulation studies involving a variety of distributions that were done to compare the Land, Chen, and Max tests and to develop the latter two as options for soil screening.

**4.3.2 Test of Proportion Exceeding a Threshold.** One of the difficulties noted for the Land test, described in Section 4.3.1, is its strong reliance on an assumption of lognormality (see Section 4.3.3). Even in cases where the assumption may hold, there will rarely be sufficient information to test it.

A second criticism of applying the Land test (or another test based on estimating the mean) is that values must be substituted for values reported as less than a quantitation limit (<QL). (As noted in Section 4.3.4, how one does this substitution is of little relevance if the SSL is much larger than the QL. However, even if a moderate proportion of the data values fall below the QL and are censored, then the lognormal distribution may not be a good model for the observed concentrations.)

A third criticism of using the Land test for screening is its requirement for large sample sizes when the contaminant variability across the EA is expected to be large (e.g., a large coefficient of variation). Because of these drawbacks to applying the Land procedure, EPA considered alternative, nonparametric procedures. One such alternative that was considered is the test described below.

For a given contaminant, let P represent the proportion of all possible sampling units across the EA for which the concentration exceeds 2 SSL. In essence, P represents the proportion of the EA with true contaminant levels above 2 SSL. A nonparametric test involving P was developed as follows.

Let $P_0$ be a fixed proportion of interest chosen in such a way that if that proportion (or more) of the EA has contamination levels above 2 SSL, then that EA should be investigated further. One way to obtain a rough equivalence between the test for a mean greater than 2 SSL and a test involving P is to choose $1\text{-}P_0$ to correspond to the percentile of the lognormal distribution at which the mean occurs. One can show that this is equivalent to choosing

$$P_0 = 1 - \Phi\left[0.5\sigma\right] = 1 - \Phi\left[0.5\sqrt{1n\left(1+CV^2\right)}\right] \qquad \textbf{(63)}$$

where

| | | |
|---|---|---|
| $\sigma$ | = | assumed standard deviation of the logarithms of the concentrations |
| CV | = | assumed coefficient of variation of the contaminant concentrations |
| $\Phi$ | = | distribution function of the standard normal distribution. |

Here, the fixed proportion $P_0$ will be less than one-half. The hypotheses are framed as

$H_0$: $P \geq P_0$ (EA needs further investigation)

versus

$H_1$: $P < P_0$ (EA does not need further investigation).

The test is based on concentration data from a grid sample of N points in the EA (without compositing). Let p represent the proportion of these n points with observed concentrations greater than or equal to 2 SSL. The test is carried out by choosing a critical value, $p_c$, to meet the desired Type I error rate, that is,

$$\alpha = \text{Prob}\ (p < p_c \mid P = P_0) = 0.05. \quad \textbf{(64)}$$

The sample size should be chosen to satisfy the Type II error rate at some specified alternative value $P_1$, where $P_1 < P_0$. For example, to have an 80 percent power at $P_1$:

$$1-\beta = \text{Prob}\ (p < p_c \mid P = P_1) = 0.80. \quad \textbf{(65)}$$

If the same type of rationale for choosing $P_0$ (corresponding to 2 SSL) is used to make $P_1$ correspond to 0.5 SSL, then one would choose

$$P_1 = 1 - \Phi\ [0.5\ \sigma + 1.386/\sigma]. \quad \textbf{(66)}$$

Sample sizes for this test were developed based on the preceding formulation and were found to be approximately the same as those required by the Land procedure, though they tended to be slightly higher than the Land sample sizes for small $\sigma$, and slightly smaller for large $\sigma$.

The major advantage of this test, in contrast to the Land procedure, for example, is its generality; the only assumption required is that random sampling be used to select the sample points. Its principal disadvantages are:

- Compositing of samples cannot be included (since the calculation of p requires the count of the number of units with observed levels at or above 2 SSL).
- The test does not deal directly with the mean contaminant level at the EA, which is the fundamental parameter for risk calculations.
- Because the test does not depend directly on the magnitude of the concentrations, it is possible that the test will give misleading results relative to a test based on a mean. This can occur, for example, when only a small portion of the EA has very high levels (i.e., a hot spot). In that case, the observed p will converge for increasing n to that proportion of the EA that is contaminated; it would do the same if the concentration levels in that same portion were just slightly above 2 SSL. A test based on a mean for large samples, however, is able to distinguish between these two situations; by its very nature, a test based on a proportion of measurements exceeding a single threshold level cannot.

For these reasons, the test described here based on the proportion of observations exceeding 2 SSL was not selected for inclusion in the current guidance.

**4.3.3 Relative Performance of Land, Max, and Chen Tests.** A simulation study was conducted to compare the Land, Chen, and Max tests and to determine sample sizes necessary to achieve DQOs. This section describes the design of the simulation study and summarizes its results. Detailed output from the simulations is presented in Appendix I.

**Treatment of Data Below the Quantitation Limit.** Review of quantitation limits for 110 chemicals showed that for more than 90 percent of the chemicals, the quantitation limit was less than 1 percent of the ingestion SSL. In such cases, the treatment of values below the QL is not expected to have much effect, as long as all data are used in the analysis, with concentrations assigned to results below the QL in some reasonable way. In the simulations, the QL was assumed to be SSL/100 and any simulated value below the QL was set equal to 0.5 QL. This is a conservative assumption based on the comparison of ingestion SSLs with QLs.

**Decision Rules.** For the **Land procedure**, as discussed in Section 4.3.1, the null hypothesis $H_0$: $\mu \geq 2$ SSL (where $\mu$ represents the true mean concentration for the EA) can be tested at the 5 percent level by calculating Land's upper 95 percent confidence limit for a lognormal mean. The null hypothesis is rejected (i.e., surface soil contaminant concentrations are less than 2 SSL), if this upper confidence limit falls below 2 SSL. This application of the Land (1971) procedure, as described in the draft 1994 Guidance, will be referred to as the "SSL DQOs" and the "original Land procedure."

For the **Max test**, one decides to walk away if the maximum concentration observed in composite samples taken from the EA does not exceed 2 SSL. As indicated in Section 4.1.6, it is viewed as providing a test of the original null hypothesis, $H_0$: $\mu \geq 2$ SSL. The Max test does not inherently control either type of error rate (i.e., its critical region is always the region below 2 SSL, not where concentrations below a threshold that achieve a specified Type I error rate). However, control of error rates for the Max test can be achieved through the DQO process by choice of design (i.e., by choice of the number N of composite samples and choice of the number C of specimens per composite).

The **Chen test** requires that the null hypothesis have the form $H_0$: $\mu \leq \mu_0$, with the alternative hypothesis as $H_1$: $\mu > \mu_0$ (Chen, 1995). Hypotheses or DQOs of this form are referred to as "flipped hypotheses" or "flipped DQOs" because they represent the inverse of the actual hypothesis for SSL decisions. In the simulations, the Chen method was applied with $\mu_0 = 0.5$ SSL at significance levels (Type I error rates) of 0.4, 0.3, 0.2, 0.1, 0.05, 0.025, and 0.01. In this formulation, a Type I error occurs if one decides incorrectly to investigate further when the true site mean, $\mu$, is at or below 0.5 SSL.

The two formulations of the hypotheses are equivalent in the sense that both allow achievement of soil screening DQOs. That is, working with either formulation, it is possible to control the probability of incorrectly deciding to walk away when the true site mean is 2 SSL and to also control the probability of incorrectly deciding to investigate further when the true site mean is 0.5 SSL.

In addition to the original Land procedure, the Chen test, and the Max test, the simulations also include the Land test of the flipped null hypothesis $H_0$: $\mu \leq 0.5$ SSL at the 10 percent significance level. This Land test of the flipped hypothesis was included to investigate how interchanging the null and alternative hypotheses affected sample sizes for the Land and Chen procedures.

**Simulation Distributions.** In the following description of the simulations, parameter acronyms used as labels in the tables of results are indicated by capital letters enclosed in parentheses.

Each distribution used for simulation is a mixture of a lower concentration distribution and a higher concentration distribution. The lower distribution represents the EA in its natural (unpolluted) state, and the higher distribution represents contaminated areas. Typically, all measurements of pollutants in uncontaminated areas are below the QL. Accordingly, the lower distribution is assumed to be completely below the QL. For the purposes of this analysis, it is unnecessary to specify any other aspect of the lower distribution, because any measurement below the QL is set equal to 0.5 QL.

A parameter between 0 and 1, called the mixing proportion (MIX), specifies the probability allocated to the lower distribution. The remaining probability (1-MIX) is spread over higher values according to either a lognormal, gamma, or Weibull distribution. The parameters of the higher distribution are chosen so that the overall mixture has a given true EA mean (MU) and a given coefficient of variation (CV). Where s is the sample standard deviation, $\bar{x}$ is the sample mean, and C is the number of specimens per composite sample, CV is defined as:

$$CV = \frac{s}{\bar{x}} \quad \text{or} \quad CV = \frac{\left(\sqrt{C}\right)s}{\bar{x}}.$$

The following parameter values were used in the simulations:

EA mean (MU) = 0.5 SSL or 2 SSL

EA coefficient of variation (CV) = 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, or 6 (i.e., 100 to 600 percent)

Number of specimens per composite (C) = 1, 2, 3, 4, 5, 6, 8, 9, 12, or 16

Number of composites chemically analyzed (N) = 4, 5, 6, 7, 8, 9, 12, or 16.

The true EA mean was set equal to 0.5 SSL or 2 SSL in order to estimate the two error rates of primary concern. Most CVs encountered in practice probably will lie between 1 and 2.5 (i.e., variability between 100 and 250 percent). This expectation is based on data from the Hanford site (see Hardin and Gilbert, 1993) and the Piazza Road site (discussed in Section 4.3.6). EPA believes that the most practical choices for the number of specimens per composite will be four and six. In some cases, compositing may not be appropriate (the case C = 1 corresponds to no compositing). EPA also believes that for soil screening, a practical number of samples chemically analyzed per EA lies below nine, and that screening decisions about soils in each EA should not be based on fewer than four chemical analyses.

For a given CV, there is a theoretical limit to how large the mixing proportion can be. The values of the mixing proportion used in the simulations are shown below as a function of CV. The case MIX = 0 corresponds to an EA characterized by a gamma, lognormal, or Weibull distribution. A value of MIX near 1 indicates an EA where all concentrations are below the QL except those in a small portion of the EA. Neither of these extremes implies an extreme overall mean. If MIX = 0, the contaminating (higher) distribution can have a low mean, resulting in a low overall mean. If MIX is near 1 (i.e., a relatively small contamination area), a high overall mean can be obtained if the mean of the distribution of contaminant concentrations is high enough.

| CV | Values of MIX used in the simulations |
|---|---|
| 1.0 | 0, 0.49 |
| 1.5 | 0, 0.50 |
| 2.0 | 0, 0.50, 0.75 |
| 2.5 | 0, 0.50, 0.85 |
| 3.0 | 0, 0.50, 0.85 |
| 3.5 | 0, 0.50, 0.90 |
| 4.0 | 0, 0.50, 0.90 |
| 5.0 | 0, 0.50, 0.95 |
| 6.0 | 0, 0.50, 0.95 |

**Treatment of Measurement Error.** Measurement errors were assumed to be normally distributed with mean 0 (i.e., unbiased measurements) and standard deviation equal to 20 percent of the true value for each chemically analyzed sample. (Earlier simulations included measurement error standard deviations of 10 percent and 25 percent. The difference in results between these two cases was negligible.)

**Number of Simulated Samples.** Unique combinations of the simulation parameters considered (i.e., 2 values of the EA mean, 10 values for the number of specimens per composite, 8 values for the number of composite samples, 25 combinations of CV and MIX, and 3 contamination models—lognormal, gamma, Weibull), result in a total of 12,000 simulation conditions. One thousand simulated random samples were generated for each of the 12,000 cases obtained by varying the simulation parameters as described above. The average number of physical samples simulated from an EA for a hypothesis test (i.e., the product CN) was 56.

The following 10 hypothesis tests were applied to each of the 12 million random samples:

- Chen test at significance levels of 0.4, 0.3, 0.2, 0.1, 0.05, 0.025, and 0.01
- Original Land test of the null hypothesis $H_0$: $\mu \geq 2$ SSL at the 5 percent significance level
- Land test of the flipped null hypothesis $H_0$: $\mu \leq 0.5$ SSL at the 10 percent significance level
- Maximum test.

These simulations involved generation of approximately 650 million random numbers.

**Simulation Results.** A complete listing of the simulation results, with 150 columns and 59 lines per page, requires 180 pages and is available from EPA on a 3.5-inch diskette.

Representative results for gamma contamination data, with eight composite samples that each consist of six specimens, are shown in Table 32. The gamma contamination model is recommended for determining sample size requirements because it was consistently seen to be least favorable, in the sense that it required higher sample sizes to achieve DQOs than either of the lognormal or Weibull

models. Hence, sample sizes sufficient to protect against a gamma distribution of contaminant concentrations are also protective against a lognormal or Weibull distribution.

## Table 32. Comparison of Error Rates for Max Test, Chen Test (at .20 and .10 Significance Levels), and Original Land Test, Using 8 Composites of 6 Samples Each, for Gamma Contamination Data

| MU/SSL | MIX | Max test | 0.20 Chen test | 0.10 Chen test | Land test |
|---|---|---|---|---|---|
| **C=6 N=8 CV=4** | | | | | |
| 0.5 | .00 | .35 | .18 | .09 | .99 |
| 0.5 | .50 | .40 | .22 | .11 | .99 |
| 0.5 | .90 | .40 | .19 | .09 | .98 |
| 2.0 | .00 | .06 | .10 | .18 | .00 |
| 2.0 | .50 | .06 | .11 | .18 | .00 |
| 2.0 | .90 | .04 | .16 | .29 | .01 |
| **C=6 N=8 CV=3** | | | | | |
| 0.5 | .00 | .24 | .18 | .10 | .93 |
| 0.5 | .50 | .25 | .19 | .10 | .94 |
| 0.5 | .85 | .23 | .22 | .11 | .99 |
| 2.0 | .00 | .04 | .03 | .06 | .00 |
| 2.0 | .50 | .03 | .03 | .05 | .00 |
| 2.0 | .85 | .03 | .06 | .12 | .00 |
| **C=6 N=8 CV=2** | | | | | |
| 0.5 | .00 | .07 | .22 | .11 | .57 |
| 0.5 | .50 | .06 | .19 | .09 | .68 |
| 0.5 | .75 | .04 | .19 | .10 | .85 |
| 2.0 | .00 | .02 | .00 | .00 | .01 |
| 2.0 | .50 | .02 | .00 | .01 | .00 |
| 2.0 | .75 | .01 | .00 | .01 | .00 |
| **C=6 N=8 CV=1** | | | | | |
| 0.5 | .00 | .00 | .20 | .10 | .01 |
| 0.5 | .49 | .00 | .20 | .12 | .12 |
| 2.0 | .00 | .01 | .00 | .00 | .02 |
| 2.0 | .49 | .01 | .00 | .00 | .00 |

MU = True EA Mean - see subsection entitled “Simulation Distributions” in Section 4.3.3.
MIX = Mixing Proportion - see subsection entitled “Simulation Distributions” in Section 4.3.3
C = Number of specimens in a composite.
N = Number of composites analyzed.
CV = EA coefficient of variation $\frac{(\sqrt{C})s}{\bar{x}}$

where s = sample standard deviation and $\bar{x}$ = mean sample concentration

Table 32 shows that the original Land method is unable to control the error rates at 0.5 SSL for gamma distributions. This limitation of the Land method was seen consistently throughout the results for all nonlognormal distributions tested. This limitation led to removal of the Land procedure from the Soil Screening Guidance.

Earlier simulation results for gamma and Weibull distributions did not censor results below the QL and used pure unmixed distributions. In these cases, as the sample size N increased, with all other factors fixed, the Land error rates at 0.5 SSL increased toward 1. Normally, the expectation is that as the sample size increases, information increases, and error rates decrease.

When using data from a Weibull or gamma distribution, the Land confidence interval endpoints converge to a value **that does not equal** the true site mean, $\mu_x$ , and results in an increase in error rates. This phenomenon is easily demonstrated, as follows. Let X denote the concentration random variable, let Y = ln(X) denote its logarithm. Let $\mu_y$ and $\sigma_y$ denote the mean and standard deviation of logarithms of the soil concentrations. Then, as the sample size increases, the Land confidence interval endpoints (UL and LL) converge to

$$UL = LL = \exp\left(\mu_y + \frac{\sigma_y^2}{2}\right) . \qquad \textbf{(67)}$$

If X is lognormally distributed, this expression is the mean of X. If X has a Weibull or gamma distribution, this expression is **not** the mean of X. This inconsistency accounts for the increase in error rates with sample size.

Table 32 also shows the fundamental difference between the Max test and the Chen test. For the Max test, the probability of error in deciding to walk away when the EA mean is 2.0 SSL is fairly stable, ranging from 0.01 to 0.06 across the different values of the CV. On the other hand, these error rates vary more across the CV values for the Chen test (e.g., from 0.00 to 0.29 for Chen test at the 0.10 significance level). This occurs because the Chen test is designed to control the other type of error rate (at 0.5 SSL). The Max test is presented in the 1995 Soil Screening Guidance (U.S. EPA, 1995c) because of its simplicity and the stability of its control over the error rate at 2 SSL.

Table 33 shows error rate estimates for four to nine composite samples that each consist of four, six, or eight specimens for EAs with CVs of 2, 2.5, 3, or 3.5, and assuming a gamma distribution. Table 33 should be adequate for most SSL planning purposes. However, more complete simulation results are reported in Appendix I.

Planning for CVs at least as large as 2 is recommended because it is known that CVs greater than 2 occur in practice (e.g., for two of seven EAs in the Piazza Road simulations reported in Section 4.3.6). One conclusion that can be drawn from Table 33 is that composite sample sizes of four are often inadequate. Further support for this conclusion is reported in the Piazza Road simulations discussed in Section 4.3.6.

**Conclusions.** The primary conclusions from the simulations are:

- For distributions other than lognormal, the Land procedure is prone to decide to investigate further at 0.5 SSL, when the correct decision is to walk away. It is therefore unsuitable for surface soil screening.

- Both the Max test and the Chen test perform acceptably under a variety of distributional assumptions and are potentially suitable for surface soil screening.

## Table 33. Error Rates of Max Test and Chen Test at .2 (C20) and .1 (C10) Significance Level for CV = 2, 2.5, 3, 3.5

| | | CV = 2.0 | | | CV = 2.5 | | | CV = 3.0 | | | CV = 3.5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **N** | **MU/SSL** | **Max** | **C20** | **C10** | **Max** | **C20** | **C10** | **Max** | **C20** | **C10** | **Max** | **C20** | **C10** |
| **C = 4** | | | | | | | | | | | | | |
| 4 | 0.5 | .09 | .20 | .11 | .14 | .18 | .09 | .19 | .18 | .08 | .24 | .20 | .10 |
| 4 | 2.0 | .13 | .08 | .16 | .19 | .17 | .28 | .20 | .21 | .33 | .26 | .29 | .42 |
| 5 | 0.5 | .11 | .21 | .10 | .15 | .18 | .09 | .26 | .20 | .08 | .26 | .20 | .09 |
| 5 | 2.0 | .10 | .05 | .11 | .10 | .09 | .18 | .17 | .19 | .30 | .18 | .23 | .36 |
| 6 | 0.5 | .11 | .21 | .12 | .21 | .20 | .10 | .28 | .21 | .11 | .31 | .19 | .09 |
| 6 | 2.0 | .06 | .03 | .08 | .08 | .08 | .14 | .11 | .13 | .23 | .11 | .18 | .28 |
| 7 | 0.5 | .12 | .20 | .10 | .25 | .22 | .11 | .31 | .20 | .09 | .36 | .18 | .10 |
| 7 | 2.0 | .04 | .03 | .05 | .05 | .04 | .09 | .08 | .11 | .18 | .08 | .14 | .23 |
| 8 | 0.5 | .16 | .19 | .09 | .25 | .20 | .09 | .36 | .20 | .10 | .42 | .20 | .09 |
| 8 | 2.0 | .02 | .02 | .03 | .04 | .03 | .07 | .05 | .08 | .14 | .07 | .13 | .21 |
| 9 | 0.5 | .16 | .21 | .11 | .28 | .20 | .09 | .36 | .18 | .09 | .44 | .22 | .12 |
| 9 | 2.0 | .01 | .01 | .02 | .03 | .03 | .06 | .04 | .07 | .13 | .07 | .12 | .20 |
| **C = 6** | | | | | | | | | | | | | |
| 4 | 0.5 | .03 | .20 | .12 | .08 | .21 | .12 | .15 | .20 | .10 | .16 | .17 | .08 |
| 4 | 2.0 | .14 | .03 | .08 | .16 | .08 | .17 | .17 | .14 | .24 | .20 | .19 | .33 |
| 5 | 0.5 | .04 | .20 | .10 | .11 | .17 | .09 | .17 | .20 | .10 | .22 | .20 | .10 |
| 5 | 2.0 | .09 | .02 | .05 | .09 | .04 | .10 | .13 | .10 | .18 | .15 | .13 | .24 |
| 6 | 0.5 | .06 | .20 | .11 | .14 | .21 | .10 | .19 | .20 | .10 | .25 | .20 | .10 |
| 6 | 2.0 | .04 | .01 | .02 | .06 | .03 | .07 | .09 | .07 | .14 | .09 | .10 | .19 |
| 7 | 0.5 | .06 | .20 | .09 | .12 | .19 | .10 | .23 | .22 | .10 | .29 | .21 | .10 |
| 7 | 2.0 | .02 | .00 | .01 | .05 | .02 | .04 | .06 | .06 | .10 | .08 | .09 | .14 |
| 8 | 0.5 | .06 | .19 | .09 | .15 | .20 | .10 | .25 | .19 | .10 | .30 | .19 | .10 |
| 8 | 2.0 | .02 | .00 | .01 | .02 | .01 | .03 | .03 | .03 | .05 | .04 | .06 | .11 |
| 9 | 0.5 | .06 | .20 | .10 | .18 | .22 | .11 | .28 | .20 | .11 | .34 | .19 | .09 |
| 9 | 2.0 | .01 | .00 | .01 | .02 | .01 | .02 | .03 | .02 | .04 | .03 | .05 | .09 |
| **C = 8** | | | | | | | | | | | | | |
| 4 | 0.5 | .02 | .21 | .13 | .06 | .19 | .10 | .10 | .21 | .10 | .14 | .18 | .08 |
| 4 | 2.0 | .12 | .02 | .05 | .15 | .04 | .09 | .17 | .09 | .17 | .19 | .14 | .25 |
| 5 | 0.5 | .03 | .22 | .11 | .05 | .20 | .11 | .11 | .20 | .10 | .17 | .19 | .09 |
| 5 | 2.0 | .07 | .01 | .02 | .09 | .02 | .06 | .09 | .04 | .10 | .12 | .08 | .17 |
| 6 | 0.5 | .02 | .18 | .09 | .08 | .21 | .11 | .13 | .19 | .10 | .20 | .20 | .10 |
| 6 | 2.0 | .04 | .00 | .01 | .06 | .01 | .02 | .07 | .04 | .07 | .08 | .07 | .13 |
| 7 | 0.5 | .03 | .20 | .11 | .09 | .20 | .11 | .18 | .21 | .11 | .22 | .20 | .11 |
| 7 | 2.0 | .03 | .00 | .00 | .04 | .01 | .01 | .04 | .02 | .04 | .05 | .05 | .09 |
| 8 | 0.5 | .04 | .20 | .10 | .11 | .21 | .11 | .17 | .21 | .10 | .26 | .19 | .10 |
| 8 | 2.0 | .02 | .00 | .00 | .02 | .01 | .01 | .04 | .01 | .03 | .03 | .03 | .06 |
| 9 | 0.5 | .04 | .21 | .11 | .11 | .21 | .10 | .20 | .19 | .10 | .30 | .23 | .12 |
| 9 | 2.0 | .01 | .00 | .00 | .02 | .00 | .01 | .01 | .00 | .01 | .02 | .02 | .04 |

MU = True EA Mean - see subsection entitled “Simulation Distributions” in Section 4.3.3.
MIX= Mixing Proportion - see subsection entitled “Simulation Distributions” in Section 4.3.3
C = Number of specimens in a composite.
N = Number of composites analyzed.
CV = EA coefficient of variation $\frac{(\sqrt{C})s}{\bar{x}}$

where s = sample standard deviation and $\bar{x}$ = mean sample concentration

**4.3.4 Treatment of Observations Below the Limit of Quantitation.** Test procedures that are based on estimating a mean contaminant level for an EA, such as the Land and Chen procedures, make use of each measured concentration value. For this reason, the use of all reported concentration measurements in such calculations should be considered regardless of their magnitude—that is, even if the measured levels fall below a quantitation level. One argument for this approach is that the QL is itself an estimate. Another is that some value will have to be substituted for any censored data point (i.e., a point reported as <QL), and the actual measured value is at least as accurate as a substituted value.

The peer review of the Draft Soil Screening Guidance raised the following issue:

> If such censored values do occur in a data set, what values should be used?

There is a substantial amount of literature on this subject and a variety of sophisticated approaches. In the context of SSLs, however, a simple approach is recommended. Consistent with general Superfund guidance, each observation reported as "<QL" shall be replaced with 0.5 QL for computation of the sample mean.

The evidence suggests that the ingestion SSL generally will be 2 orders of magnitude or more greater than the QL for most contaminants. In these cases, the results of soil screening will be insensitive to alternative procedures that could be used to substitute values for observations reported as "<QL." When the SSL is not much greater than the QL (e.g., SSL < 50 QL), the outcome of the soil screening could be affected by the procedure used to substitute for "<QL" values.

The most conservative approach would be to substitute the concentration represented by the QL itself for all observations reported as "<QL." In the context of the SSLs, however, the simple approach of using 0.5 QL is suggested. This will be sufficiently conservative given the conservative factors underlying the SSLs.

**4.3.5 Multiple Hypothesis Testing Considerations.** The Soil Screening Guidance addresses the following hypothesis testing problem for each EA:

> $H_0$: mean concentration of a given chemical ≥ 2 SSL

versus

> $H_1$: mean concentration of a given chemical < 2 SSL.

The default value for the probability of a Type I error is $\alpha = 0.05$, while the default value for the power of the test at 0.5 SSL is $1-\beta = 0.80$. The test is applied separately for each chemical, so that these probabilities apply for each individual chemical. Thus, there is an 80 percent probability of walking away from an EA (i.e., rejecting $H_0$) when only one chemical is being tested and its true mean level is 0.5 SSL and a 5 percent probability of walking away if its true mean level is 2 SSL.

However, the Soil Screening Guidance does not explicitly address the following issues:

> What is the composite probability of walking away from an EA if there are multiple contaminants?

and

> If such probabilities are unacceptable, how should one compensate when testing for multiple contaminants within a single EA?

The answer to the first question cannot be determined, in general, since the concentrations of the various contaminants will often be dependent on one another (e.g., this would be expected if they originated from the same source of contamination). The joint probability of walking away can be determined, however, if one makes the simplifying assumption that the contaminant concentrations for the different chemicals are independent (uncorrelated). In that case, the probability of walking away is simply the product of the individual rejection probabilities.

For two chemicals (Chemical A and Chemical B, say), this is:

Pr{walking away from EA} = Pr{reject $H_0$ for Chemical A} × Pr{reject $H_0$ for Chemical B}.

While these joint probabilities must be regarded as approximate, they nevertheless serve to illustrate the effect on the error rates when dealing with multiple contaminants.

Assume (for illustrative purposes only) that the probabilities for rejecting the null hypothesis (walking away from the EA) for each single chemical appear as follows:

| True concentration | Probability of rejecting $H_0$ |
|---|---|
| 0.2 SSL | 0.95 |
| 0.5 SSL | 0.80 (default 1-ß) |
| 0.7 SSL | 0.60 |
| 1.0 SSL | 0.50 |
| 1.5 SSL | 0.20 |
| 2.0 SSL | 0.05 (default α) |

Let C(A) denote the concentration of Chemical A divided by the SSL, and let P(A) denote the corresponding probability of rejecting $H_0$. Define C(B) and P(B) similarly for Chemical B. Assuming independence, the joint probabilities of rejecting the null hypothesis (walking away) are as shown in Table 34.

## Table 34. Probability of "Walking Away" from an EA When Comparing Two Chemicals to SSLs

| Chemical A | | Chemical B | | | | | |
|---|---|---|---|---|---|---|---|
| C(A) | P(A) | C(B) = 0.2 P(B) =.95 | C(B) = 0.5 P(B) = .80 | C(B) = 0.7 P(B) = .60 | C(B) = 1.0 P(B) = .50 | C(B) = 1.5 P(B) = .20 | C(B) = 2.0 P(B) = .05 |
| 0.2 | 0.95 | 0.90 | 0.76 | 0.57 | 0.48 | 0.19 | 0.05 |
| 0.5 | 0.80 | 0.76 | 0.64 | 0.48 | 0.40 | 0.16 | 0.04 |
| 0.7 | 0.60 | 0.57 | 0.48 | 0.36 | 0.30 | 0.12 | 0.03 |
| 1.0 | 0.50 | 0.48 | 0.40 | 0.30 | 0.25 | 0.10 | 0.03 |
| 1.5 | 0.20 | 0.19 | 0.16 | 0.12 | 0.10 | 0.04 | 0.01 |
| 2.0 | 0.05 | 0.05 | 0.04 | 0.03 | 0.03 | 0.01 | <0.01 |

These probabilities demonstrate that the test procedure will tend to be very conservative if multiple chemicals are involved—that is, **all** of the chemical concentrations must be quite low relative to their SSL in order to have a high probability of walking away from the EA. On the other hand, there will be a high probability that further investigation will be called for if the mean concentration for even a single chemical is twice the SSL.

A potential problem occurs when there are several chemicals under consideration and when all or most of them have levels slightly below the SSL (e.g., near 0.5 SSL). For instance, if each of six independent chemicals had levels at 0.5 SSL, the probability of rejecting the null hypothesis would be 80 percent for each such chemical, but the probability of walking away from the EA would be only $(0.80)^6 = 0.26$.

If the same samples are being analyzed for multiple chemicals, then the original choice for the number of such samples ideally should have been based on the worst case (i.e., the chemical expected to have the largest variability). In this case, the probability of correctly rejecting the null hypothesis at 0.5 SSL for the chemicals with less variability will be higher. The overall probability of walking away will be greater than shown above if all or some of the chemicals have less variability than assumed as the basis for determining sample sizes. Here, the sample size will be large enough for the probability of rejecting the null hypothesis at 0.5 SSL to be greater than 0.80 for these chemicals.

The probability values assumed above for deciding that no further investigation is necessary for individual chemicals, which are the basis for these conclusions, are equally applicable for the Land, Chen, and Max tests. They simply represent six hypothetical points of the power curves for these tests (from 0.2 SSL to 2.0 SSL). Therefore, the conclusions are equally applicable for each of the hypothesis testing procedures that have been considered in the current guidance for screening surface soils.

If the surface soil concentrations are positively correlated, as expected when dealing with multiple chemicals, then it is likely that either all the chemicals of concern have relatively high concentrations or they all have relatively low concentrations. In this case, the probability of making the correct decision for an EA would be greater than that suggested by the above calculations that assume independence of the various chemicals.

However, the potential problem of several chemicals having concentrations near 0.5 SSL is not precluded by assuming positive correlations. In fact, it suggests that if the EA average for one chemical is near 0.5 SSL, then the average for others is also likely to be near 0.5 SSL, which is exactly the situation where the probability of **not** walking away from the EA can become large because there is a high probability that $H_0$ will be rejected for at least one of these chemicals.

An alternative would be to use multiple hypothesis testing procedures to control the overall error rate for the set of chemicals (i.e., the set of hypothesis tests) rather than the separate error rates for the individual chemicals. Guidance for performing multiple hypothesis tests is beyond the scope of the current document. Obtain the advice of a statistician familiar with multiple hypothesis testing procedures if the overall error rates for multiple chemicals is of concern for a particular site. The classical statistical guidance regarding this subject is *Simultaneous Statistical Inference* (Miller, 1991).

**4.3.6 Investigation of Compositing Within EA Sectors.** If one decides that an EA needs further investigation, then it is natural to inquire which portion(s) of the EA exceed the screening level. This is a different question than simply asking whether or not the EA average soil concentration exceeds the SSL. Conceivably, this question may require additional sampling, chemical

analysis, and statistical analysis. A natural question is whether this additional effort can be avoided by forming composites within sectors (subareas) of the EA. The sector with the highest estimated concentration would then be a natural place to begin a detailed investigation.

The simulations to investigate the performance of rules to decide whether further investigation is required, reported in Section 4.3.3, make specific assumptions about the sampling design. It is assumed that N composite samples are chemically analyzed, each consisting of C specimens selected to be statistically representative of the entire EA. The key point, in addition to random sampling, is that composites must be formed **across** sectors rather than **within** sectors. This assumption is necessary to achieve composite samples that are representative of the EA mean (i.e., have the EA mean as their expected value).

If compositing is limited to sectors, such as quadrants, then each composite represents its sector, rather than the entire EA. The simulations reported in Section 4.3.3, and sample sizes based on them, do not apply to this type of compositing. This does not necessarily preclude compositing within sectors for both purposes, i.e., to test the hypothesis about the EA mean and also to indicate the most contaminated sector. However, little is known about the statistical properties of this approach when applying the Max test, which would depend on specifics of the actual spatial distribution of contaminants for a given EA. Because of the lack of extensive spatial data sets for contaminated soil, there is limited basis for determining what sample sizes would be adequate for achieving desired DQOs for various sites. However, one spatial data set was available and used to investigate the performance of compositing within sectors at one site.

**Piazza Road Simulations.** Data from the Piazza Road NPL site were used to investigate the properties of tests of the EA mean based on compositing within sectors, as compared to compositing between sectors. The investigation of a single site cannot be used to validate a given procedure, but it may indicate whether further investigation of the procedure is worthwhile.

Seven nonoverlapping 0.4-acre EAs were defined within the Piazza Road site. Each EA is an 8-by-12 grid composed of 14'x14' squares. The data consist of a single dioxin measurement of a composite sample from each small square. These measurements are regarded as true values for the simulations reported in this section. Measurement error was incorporated in the same fashion as for the simulations reported in Section 4.3.3.

Each of the seven EAs was subdivided into four 4-by-6 sectors, six 4-by-4 sectors, eight 4-by-3 sectors, twelve 2-by-4 sectors, and sixteen 2-by-3 sectors. Results are presented here for the cases of four, six, and eight sectors because composites of more than eight specimens are expected to be used rarely, if at all.

Table 35 presents the "true" mean and CV for each EA, computed from all 96 measurements within the 0.4-acre EA. The CVs range from 1.0 to 2.2. Note that two of the seven CVs equal or exceed 2 at this site. This supports EPA's belief that at many sites it is prudent, when planning sample size requirements for screening, to assume a CV of at least 2.5 and to consider the possibility of CVs as large as 3 or 3.5.

As data on variability within EAs for different sites and contaminant conditions accrue over time, it will be possible to base the choice of procedures on a larger, more comprehensive database, rather than just a single site.

Appendix J contains results of simulations from the seven Piazza Road EAs. Sampling with

replacement from each sector was used, because this was felt to be more consistent with the planned compositing. To estimate the error rates at 0.5 SSL and 2 SSL for each EA, the SSL was defined so that the site mean first was regarded as 0.5 SSL and then was regarded as 2 SSL.

**Notation for Results from Piazza Road Simulations.** The following notation is used in Appendix J. The design variable (DES) indicates whether compositing was within sector (DES=W) or across sectors (DES=X). As in Section 4.3.3, C denotes the number of specimens per composite, and N denotes the number of composite samples chemically analyzed. Results in Appendix J are for the Chen test at the 10 percent significance level and for the Max test. The true mean and CV are shown in the header for each EA.

**Table 35. Means and CVs for Dioxin Concentrations for 7 Piazza Road Exposure Areas**

| EA | Mean of EA | CV of EA | N |
|---|---|---|---|
| 1 | 2.1 | 1.0 | 96 |
| 2 | 2.4 | 1.6 | 96 |
| 3 | 5.1 | 1.1 | 96 |
| 4 | 4.0 | 1.2 | 96 |
| 5 | 9.3 | 2.0 | 96 |
| 6 | 15.8 | 2.2 | 96 |
| 7 | 2.8 | 1.4 | 96 |

**Results and Conclusions from Piazza Road Simulations.** Although the results from a single site cannot be assumed to apply to all sites, the following observations can be made based on the Piazza Road simulations reported in Appendix J.

- The error rate at 0.5 SSL for the Chen test, using compositing across sectors (DES=X), is generally close to the nominal rate of 0.10. For compositing within sectors (DES=W), the error rate for Chen at 0.5 SSL is generally much lower than the nominal rate.

- Except for plans involving only four analyses (N = 4), the error rate at 2 SSL is always below 0.05 for the Chen test. For the Max test, the error rate at 2 SSL fluctuated between 0 and 16 percent. The error rate at 2 SSL is smaller for the Chen test at the 10 percent significance level than for the Max test in virtually all cases. The only two exceptions to this are for compositing within sector (DES=W) in EA No. 6.

- This observation provides further support for the conclusion drawn from the simulations reported in Section 4.3.3: plans involving only four analyses can result in high error rates in determining the mean contaminant concentration of an EA with the Max test. In most cases the error rates of concern to EPA (at 2 SSL) are 0.10 or larger.

- In general, error rates estimated from Piazza Road simulations for compositing across sectors are at least as small as would be predicted on the basis of the simulation results reported in Section 4.3.3.

- The simulation results show that compositing within sectors using the Max test may be an option for site managers who want to know whether one sector of an EA is more contaminated than the other. However, use of the Max test when compositing within sectors may lead the site manager to draw conclusions about the mean contaminant concentration in that sector only, not across the entire EA.