United States Environmental Protection Agency | Office of Solid Waste and Emergency Response Washington, D.C. 20460 | Publication 9285.7-08I May 1992

EPA

Supplemental Guidance to RAGS: Calculating the Concentration Term

Office of Emergency and Remedial Response
Hazardous Site Evaluation Division, OS-230

Intermittent Bulletin
Volume 1 Number 1

The overarching mandate of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) is to protect human health and the environment from current and potential threats posed by uncontrolled releases of hazardous substances. To help meet this mandate, the U.S. Environmental Protection Agency's (EPA's) Office of Emergency and Remedial Response has developed a human health risk assessment process as part of its remedial response program. This process is described in *Risk Assessment Guidance for Superfund: Volume I — Human Health Evaluation Manual* (RAGS/HHEM). Part A of RAGS/HHEM addresses the baseline risk assessment, and describes a general approach for estimating exposure to individuals from hazardous substance releases at Superfund sites.

This bulletin explains the concentration term in the exposure/intake equation to remedial project managers (RPMs), risk assessors, statisticians, and other personnel. This bulletin presents the general intake equation as presented in RAGS/HHEM Part A, discusses basic concepts concerning the concentration term, describes generally how to calculate the concentration term, presents examples to illustrate several important points, and, lastly, identifies where to get additional help.

THE CONCENTRATION TERM

How is the concentration term used?

RAGS/HHEM Part A presents the Superfund risk assessment process in four "steps": (1) data collection and evaluation; (2) exposure assessment; (3) toxicity assessment; and (4) risk characterization. The concentration term is calculated for use in the exposure assessment step. Highlight 1 presents the general equation Superfund uses for calculating exposure, and illustrates that the concentration term (C) is one of several parameters needed to estimate contaminant intake for an individual.

For Superfund assessments, the concentration term (C) in the intake equation is an estimate of the arithmetic average concentration for a contaminant based on a set of site sampling results. Because of the uncertainty associated with estimating the true average concentration at a site, the 95 percent upper confidence limit (UCL) of the arithmetic mean should be used for this variable. The 95 percent UCL provides reasonable confidence that the true site average will not be underestimated.

Why use an average value for the concentration term?

An estimate of average concentration is used because:

Supplemental Guidance to RAGS is a bulletin series on risk assessment of Superfund sites. These bulletins serve as supplements to *Risk Assessment Guidance for Superfund: Volume I — Human Health Evaluation Manual*. The information presented is intended as guidance to EPA and other government employees. It does not constitute rulemaking by the Agency, and may not be relied on to create a substantive or procedural right enforceable by any other person. The Government may take action that is at variance with these bulletins.

Highlight 1
GENERAL EQUATION FOR ESTIMATING EXPOSURE TO A SITE CONTAMINANT

$$I = C \times \frac{CR \times EFD}{BW} \times \frac{1}{AT}$$

where:

I = intake (i.e., the quantitative measure of exposure in RAGS/HHEM)
C = contaminant concentration
CR = contact (intake) rate
EFD = exposure frequency and duration
BW = body weight
AT = averaging time

(1) carcinogenic and chronic noncarcinogenic toxicity criteria[1] are based on lifetime average exposures; and

(2) average concentration is most representative of the concentration that would be contacted at a site over time.

For example, if you assume that an exposed individual moves randomly across an exposure area, then the spatially averaged soil concentration can be used to estimate the true average concentration contacted over time. In this example, the average concentration contacted over time would equal the spatially averaged concentration over the exposure area. While an individual may not actually exhibit a truly random pattern of movement across an exposure area, the assumption of equal time spent in different parts of the area is a simple but reasonable approach.

When should an average concentration be used?

The two types of exposure estimates now being required for Superfund risk assessments, a reasonable maximum exposure (RME) and an average, should both use an average concentration. To be protective, the overall estimate of intake (see Highlight 1) used as a basis for action at Superfund sites should be an estimate in the high end of the intake/dose distribution. One high-end option is the RME used in the Superfund program. The RME, which is defined as the highest exposure that could reasonably be expected to occur for a given exposure pathway at a site, is intended to account for both uncertainty in the contaminant concentration and variability in exposure parameters (e.g., exposure frequency, averaging time). For comparative purposes, Agency guidance (U.S. EPA, *Guidance on Risk Characterization for Risk Managers and Risk Assessors*, February 26, 1992) states that an average estimate of exposure also should be presented in risk assessments. For decision-making purposes in the Superfund program, however, RME is used to estimate risk.[2]

Why use an estimate of the arithmetic mean rather than the geometric mean?

The choice of the arithmetic mean concentration as the appropriate measure for estimating exposure derives from the need to estimate an individual's long-term average exposure. Most Agency health criteria are based on the long-term average daily dose, which is simply the sum of all daily doses divided by the total number of days in the averaging period. This is the definition of an arithmetic mean. The

[1] When acute toxicity is of most concern, a long-term average concentration generally should not be used for risk assessment purposes, as the focus should be to estimate short-term peak concentrations.

[2] For additional information on RME, see RAGS/HHEM Part A and the National Oil and Hazardous Substances Pollution Contingency Plan (NCP), *55 Federal Register* 8710, March 8, 1990.

arithmetic mean is appropriate regardless of the pattern of daily exposures over time or the type of statistical distribution that might best describe the sampling data. The geometric mean of a set of sampling results, however, bears no logical connection to the cumulative intake that would result from long-term contact with site contaminants, and it may differ appreciably from — and be much lower than — the arithmetic mean. Although the geometric mean is a convenient parameter for describing central tendencies of lognormal distributions, it is not an appropriate basis for estimating the concentration term used in Superfund exposure assessments. The following simple example may help clarify the difference between the arithmetic and geometric mean when used for an exposure assessment:

> Assume the daily exposure for a trespasser subject to random exposure at a site is 1.0, 0.01, 1.0, 0.01, 1.0, 0.01, 1.0, and 0.01 units/day over an 8-day period. Given these values, the cumulative exposure is simply their summation, or 4.04 units. Dividing this by 8 days of exposure results in an arithmetic mean of 0.505 units/day. This is the value we would want to use in a risk assessment for this individual, not the geometric mean of 0.1 units/day. Viewed another way, multiplication of the geometric mean by the number of days equals 0.8 units, considerably lower than the known cumulative exposure of 4.04 units.

UCL AS AN ESTIMATE OF THE AVERAGE CONCENTRATION

What is a 95 percent UCL?

The 95 percent UCL of a mean is defined as a value that, when calculated repeatedly for randomly drawn subsets of site data, equals or exceeds the true mean 95 percent of the time. Although the 95 percent UCL of the mean provides a conservative estimate of the average (or mean) concentration, it should not be confused with a 95th percentile of site concentration data (as shown in Highlight 2).

Why use the UCL as the average concentration?

Statistical confidence limits are the classical tool for addressing uncertainties of a distribution average. The 95 percent UCL of the arithmetic mean concentration is used as the average concentration because it is not possible to know the true mean. The 95 percent UCL therefore accounts for uncertainties due to limited sampling data at Superfund sites. As sampling data become less limited at a site, uncertainties decrease, the UCL moves closer to the true mean, and exposure evaluations using either the mean or the UCL produce similar results. This concept is illustrated in Highlight 2.

Should a value other than the 95 percent UCL be used for the concentration?

A value other than the 95 percent UCL can be used provided the risk assessor can document that high coverage of the true population mean occurs (i.e., the value equals or exceeds the true population mean with high probability). For exposure areas with limited amounts of data or extreme variability in measured or modeled data, the UCL can be greater than the highest measured or modeled concentration. In these cases, if additional data cannot practicably be obtained, the highest measured or modeled value could be used as the concentration term. Note, however, that the true mean still may be higher than this maximum value (i.e., the 95 percent UCL indicates a higher mean is possible), especially if the most contaminated portion of the site has not been sampled.

CALCULATING THE UCL

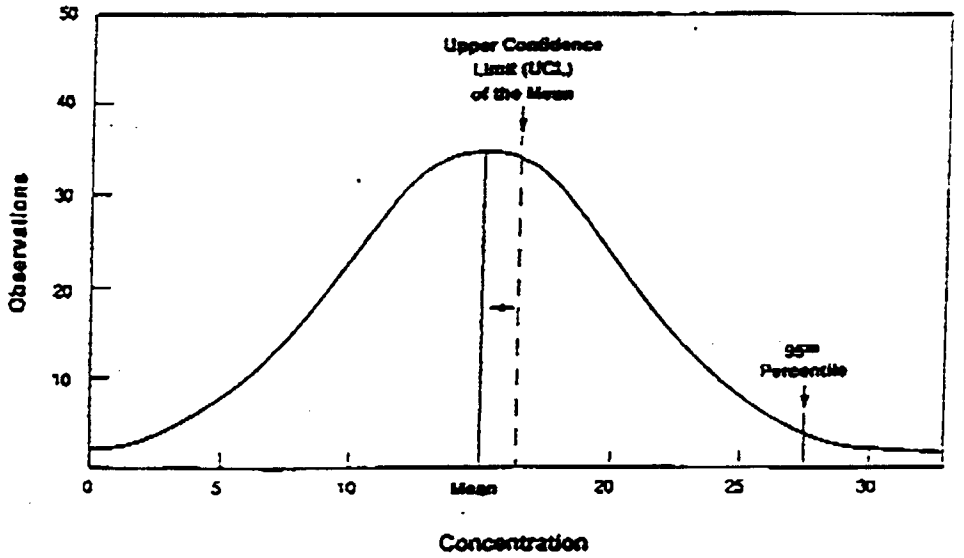
How many samples are necessary to calculate the 95 percent UCL?

Sampling data from Superfund sites have shown that data sets with fewer than 10 samples per exposure area provide poor estimates of the mean concentration (i.e., there is a large difference between the sample mean and the 95 percent UCL), while data sets with 10 to 20 samples per exposure area provide somewhat better estimates of the mean, and data sets with 20 to 30 samples provide fairly consistent estimates of the mean (i.e., the 95 percent UCL is close to the sample mean). Remember that, in general, the UCL approaches the true mean as more samples are included in the calculation.

Should the data be transformed?

EPA's experience shows that most large or "complete" environmental contaminant data sets

Highlight 2
COMPARISON OF UCL AND 95th PERCENTILE



As sample size increases, the UCL of the mean moves closer to the true mean, while the 95th percentile of the distribution remains at the upper end of the distribution.

from soil sampling are lognormally distributed rather than normally distributed (see Highlights 3 and 4 for illustrations of lognormal and normal distributions). In most cases, it is reasonable to assume that Superfund soil sampling data are lognormally distributed. Because transformation is a necessary step in calculating the UCL of the arithmetic mean for a lognormal distribution, the data should be transformed by using the natural logarithm function (i.e., calculate ln(x), where x is the value from the data set). However, in cases where there is a question about the distribution of the data set, a statistical test should be used to identify the best distributional assumption for the data set. The W-test (Gilbert 1987) is one statistical method that can be used to determine if a data set is consistent with a normal or lognormal distribution. In all cases, it is valuable to plot the data to better understand the contaminant distribution at the site.

How do you calculate the UCL for a lognormal distribution?

To calculate the 95 percent UCL of the arithmetic mean for a lognormally distributed data set, first transform the data using the natural logarithm function as discussed previously (i.e., calculate ln(x)). After transforming the data, determine the 95 percent UCL for the data set by completing the following four steps:

(1) Calculate the arithmetic mean of the transformed data (which is also the log of the geometric mean);

(2) Calculate the standard deviation of the transformed data;

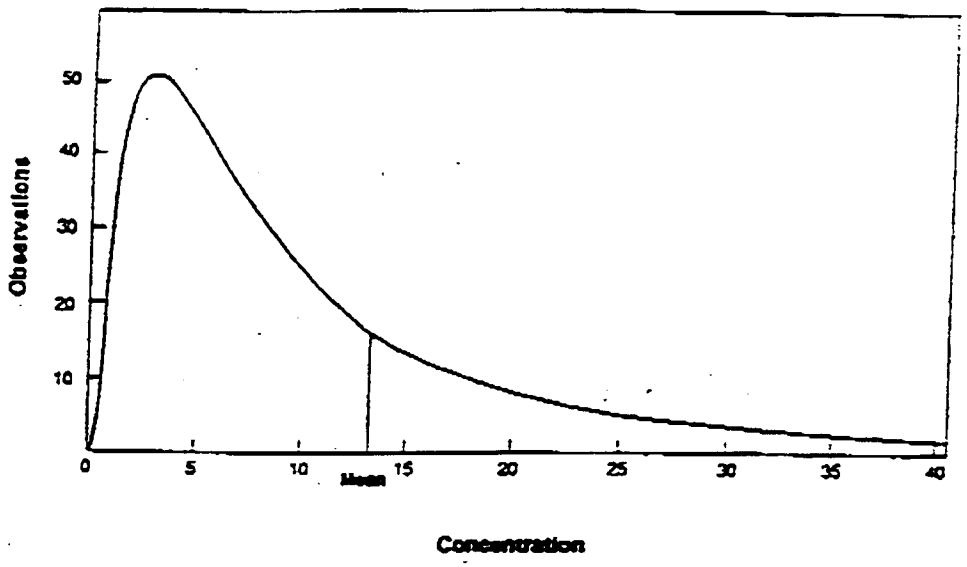
(3) Determine the H-statistic (e.g., see Gilbert 1987); and

(4) Calculate the UCL using the equation shown in Highlight 5.

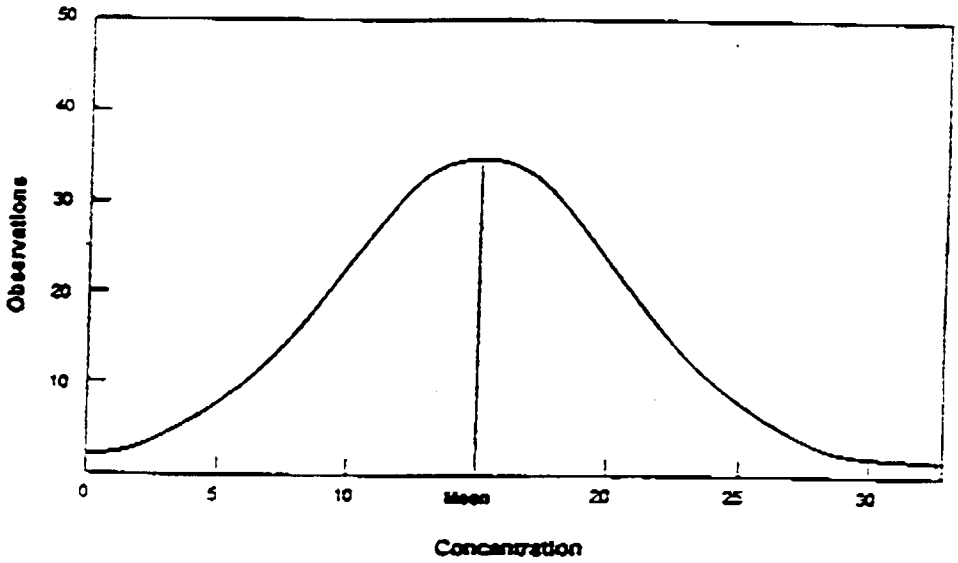
How do you calculate the UCL for a normal distribution?

If a statistical test supports the assumption that the data set is normally distributed, calculate the 95 percent UCL by completing the following four steps:

Highlight 3
EXAMPLE OF A LOGNORMAL DISTRIBUTION



Highlight 4
EXAMPLE OF A NORMAL DISTRIBUTION



Highlight 5
CALCULATING THE UCL OF THE ARITHMETIC MEAN FOR A LOGNORMAL DISTRIBUTION

$$UCL = e^{(\bar{x} + 0.5s^2 + sH/\sqrt{n-1})}$$

where:

UCL	=	upper confidence limit
e	=	constant (base of the natural log, equal to 2.718)
$\bar{x}$	=	mean of the transformed data
s	=	standard deviation of the transformed data
H	=	H-statistic (e.g., from table published in Gilbert 1987)
n	=	number of samples

Highlight 6
CALCULATING THE UCL OF THE ARITHMETIC MEAN FOR A NORMAL DISTRIBUTION

$$UCL = \bar{x} + t(s/\sqrt{n})$$

where:

UCL	=	upper confidence limit
$\bar{x}$	=	mean of the untransformed data
s	=	standard deviation of the untransformed data
t	=	Student-t statistic (e.g., from table published in Gilbert 1987)
n	=	number of samples

(1) Calculate the arithmetic mean of the untransformed data;

(2) Calculate the standard deviation of the untransformed data;

(3) Determine the one-tailed t-statistic (e.g., see Gilbert 1987); and

(4) Calculate the UCL using the equation presented in Highlight 6.

Use caution when applying normal distribution calculations if there is a possibility that heavily contaminated portions of the site have not been adequately sampled. In such cases, a UCL from normal distribution calculations could fall below the true mean, even if a limited data set at a site appears normally distributed.

EXAMPLES

The examples shown in Highlights 7 and 8 address the exposure scenario where an individual at a Superfund site has equal opportunity to contact soil in any sector of the contaminated area over time. Even though the examples address only soil exposures, the UCL approach is applicable to all exposure pathways. Guidance and examples for other exposure pathways will be presented in forthcoming bulletins.

Highlight 7 presents a simple data set and provides a stepwise demonstration of transforming the data — assuming a lognormal distribution — and calculating the UCL. Highlight 8 uses the same data set to show the difference between the UCLs that would result from assuming normal and lognormal distribution of the data. These

Highlight 7

EXAMPLE OF DATA TRANSFORMATION AND CALCULATION OF UCL

This example shows the calculation of a 95 percent UCL of the arithmetic mean concentration for chromium in soil at a Superfund site. This example is applicable only to a scenario in which a spatially random exposure pattern is assumed. The concentrations of chromium obtained from random sampling in soil at this site (in mg/kg) are 10, 13, 20, 36, 41, 59, 67, 110, 110, 136, 140, 160, 200, 230, and 1300. Using these data, the following steps are taken to calculate a concentration term for the intake equation:

(1) Plot the data and inspect the graph. (You may need the help of a statistician for this part [as well as other parts] of the calculation of the UCL.) The plot (not shown, but similar to Highlight 3) shows a skew to the right, consistent with a lognormal distribution.

(2) Transform the data by taking the natural log of the values (i.e., determine ln(x)). For this data set, the transformed values are: 2.30, 2.56, 3.00, 3.58, 3.71, 4.08, 4.20, 4.70, 4.70, 4.91, 4.94, 5.08, 5.30, 5.44, and 7.17.

(3) Apply the UCL equation in Highlight 5, where:

$\bar{x} = 4.38$
$s = 1.25$
$H = 3.163$ (based on 95 percent)
$n = 15$

The resulting 95 percent UCL of the arithmetic mean is thus found to equal $e^{(6.218)}$, or 502 mg/kg.

Highlight 8

COMPARING UCLS OF THE ARITHMETIC MEAN ASSUMING DIFFERENT DISTRIBUTIONS

In this example, the data presented in Highlight 7 are used to demonstrate the difference in the UCL that is seen if the normal distribution approach were inappropriately applied to this data set (i.e., if, in this example, a normal distribution is assumed).

ASSUMED DISTRIBUTION:	Normal	Lognormal
TEST STATISTIC:	Student-t	H-statistic
95 PERCENT UCL (mg/kg):	325	502

examples demonstrate the importance of using the correct assumptions.

WHERE CAN I GET MORE HELP?

Additional information on Superfund's policy and approach to calculating the concentration term and estimating exposures at waste sites can be obtained in:

- U.S. EPA, *Risk Assessment Guidance for Superfund: Volume I — Human Health Evaluation Manual (Part A)*, EPA/540/1-89/002, December 1989.
- U.S. EPA, *Guidance for Data Useability in Risk Assessment*, EPA/540/G-90/008 (OSWER Directive 9285.7-05), October 1990.
- U.S. EPA, *Risk Assessment Guidance for Superfund (Part A — Baseline Risk Assessment) Supplemental Guidance/ Standard Exposure Factors*, OSWER Directive 9285.6-03, May 1991.

Useful statistical guidance can be found in many standard textbooks, including:

- Gilbert, R.O., *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York, New York, 1987.

Questions or comments concerning the concentration term can be directed to:

- Toxics Integration Branch
  Office of Emergency and Remedial Response
  401 M Street SW
  Washington, DC 20460
  Phone: 202-260-9486

EPA staff can obtain additional copies of this bulletin by calling EPA's Center for Environmental Research Information at FTS 684-7562 (513-569-7652). Others can obtain copies by contacting NTIS at 800-336-4700 (703-487-4650 in the Washington, DC area).

EPA

United States
Environmental Protection
Agency (OS-230)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

First-Class Mail
Postage and Fees Paid
EPA
Permit No. G-35