APPENDIX A

SENSITIVITY ANALYSIS: HOW DO WE KNOW WHAT'S IMPORTANT?

A.0   INTRODUCTION

Sensitivity analysis, as it is applied to risk assessment, is any systematic, common sense technique used to understand how risk estimates and, in particular, risk-based decisions, are dependent on variability and uncertainty in the factors contributing to risk. In short, sensitivity analysis identifies what is "driving" the risk estimates. It is used in both point estimate and probabilistic approaches to identify and rank important sources of variability as well as important sources of uncertainty. The quantitative information provided by sensitivity analysis is important for guiding the complexity of the analysis and communicating important results (see Chapter 6). As such, sensitivity analysis plays a central role in the tiered process for PRA (see Chapter 2). This Appendix focuses on a set of graphical and statistical techniques that can be used to determine which variables in the risk model contribute most to the variation in estimates of risk. This variation in risk could represent variability, uncertainty, or both, depending on the type of risk model and characterization of input variables.

There is a wide array of analytical methods that may be referred to as sensitivity analysis, some of which are very simple and intuitive. For example, a risk assessor may have two comparable studies from which to estimate a reasonable maximum exposure (RME) for childhood soil ingestion. One approach to evaluating this uncertainty would be to calculate the corresponding RME risk twice, each time using a different plausible point estimate for soil ingestion rate. Similarly, in a probabilistic model, there may be uncertainty regarding the choice of a probability distribution. For example, lognormal and gamma distributions may be equally plausible for characterizing variability in an input variable. A simple exploratory approach would be to run separate Monte Carlo simulations with each distribution in order to determine the effect that this particular source of uncertainty may have on risk estimates within the RME range (90th to 99.9^{th} percentile, see Chapter 1).

Sensitivity analysis can also involve more complex mathematical and statistical techniques such as correlation and regression analysis to determine which factors in a risk model contribute most to the variance in the risk estimate. The complexity generally stems from the fact that multiple sources of variability and uncertainty are influencing a risk estimate at the same time, and sources may not act independently. An input variable contributes significantly to the output risk distribution if it is both highly variable *and* the variability propagates through the algebraic risk equation to the model output (i.e., risk). Changes to the distribution of a variable with a high sensitivity could have a profound impact on the risk estimate, whereas even large changes to the distribution of a low sensitivity variable may have a minimal impact on the final result. Information from sensitivity analysis can be important when trying to determine where to focus additional resources. The choice of technique(s) should be determined by the information needs for risk management decision making.

This appendix presents guidance on both practical decision making and theoretical concepts associated with the sensitivity analysis that are commonly applied in risk assessment. An overview of the type of information provided by sensitivity analysis is presented first, followed by guidance on how to decide what method to use in each of the tiers. A straightforward example of applications of Tier 1 and Tier 2 sensitivity analysis methods is shown, followed by a more detailed discussion of the theory and equations associated with the different methods.

EXHIBIT A-1

DEFINITIONS FOR APPENDIX A

Continuous Variables - A random variable that can assume any value within an interval of real numbers (e.g., body weight).

Correlation - A quantitative expression of the statistical association between two variables; usually represented by the Pearson correlation coefficient for linear models, and the Spearman rank correlation coefficient (see below) for nonlinear models.

Discrete Variables - A random variable that can assume any value within a finite set of values (e.g., number of visits to a site in one year) or at most a countably infinite set of values, meaning that you can count observations, but there is no defined upper limit. An example of countably infinite would be the number of dust particles in a volume of air (a Poisson distribution), whereas *uncountably* infinite would be the number of points in a line segment.

Local Sensitivity Analysis - Evaluation of the model sensitivity at some nominal points within the range of values of input variable(s).

Monte Carlo Analysis (MCA) or Monte Carlo Simulation - The process of repeatedly sampling from probability distributions to derive a distribution of outcomes. MCA is one of several techniques that may be used in PRA.

Multiple Regression Analysis - A statistical method that describes the extent, direction, and strength of the relationship between several (usually continuous) independent variables (e.g., exposure duration, ingestion rate) and a single continuous dependent variable (e.g., risk).

Nonparametric Tests - Statistical tests that do not require assumptions about the form of the population probability distribution.

Range Sensitivity Analysis - Evaluation of the model sensitivity across the entire range of values of the input variable(s).

Rank - If a set of values is sorted in ascending order (smallest to largest), the rank corresponds to the relative position of a number in the sequence. For example, the set {7, 5, 9, 12} when sorted gives the following sequence {5, 7, 9, 12} with ranks ranging from 1 to 4 (i.e., rank of 5 is 1, rank of 7 is 2, rank of 9 is 3, and rank of 12 is 4).

Sensitivity Analysis - Sensitivity generally refers to the variation in output of a model with respect to changes in the values of the model's input(s). Sensitivity analysis attempts to provide a ranking of the model inputs based on their relative contributions to model output variability and uncertainty. Common metrics of sensitivity include:

‣ Pearson Correlation Coefficient - A statistic r that measures the strength and direction of linear association between the values of two quantitative variables. The square of the coefficient (r^2) is the fraction of the variance of one variable that is explained by least-squares regression on the other variable, also called the coefficient of determination..

‣ Sensitivity Ratio - Ratio of the change in model output per unit change in an input variable; also called *elasticity*.

‣ Sensitivity Score - A sensitivity ratio that is weighted by some characteristic of the input variable (e.g., variance, coefficient of variation, range).

‣ Spearman Rank Order Correlation Coefficient - A "distribution free" or nonparametric statistic r that measures the strength and direction of association between the ranks of the values (not the values themselves) of two quantitative variables. See Pearson (above) for r^2.

A.1.0   UTILITY OF SENSITIVITY ANALYSIS

As highlighted in Exhibit A-2, sensitivity analysis can provide valuable information for both risk assessors and risk management decision makers throughout the tiered process for PRA. By highlighting important sources of variability and uncertainty in the risk assessment, sensitivity analysis is generally an important component of the overall uncertainty analysis. For example, methods that quantify parameter uncertainty and model uncertainty may yield different estimates of the RME risk. This information can be used to guide the tiered process by supporting decisions to conduct additional analyses or prioritize resource allocations for additional data collection efforts. Results of sensitivity analysis can also facilitate the risk communication process by focusing discussions on the important features of the risk assessment (e.g., constraints of available data, state of knowledge, significant scientific issues, and significant policy choices that were made when alternative interpretations of data existed).

EXHIBIT A-2

UTILITY OF SENSITIVITY ANALYSIS

- **Decision making with the tiered approach-** *e.g., After quantifying parameter uncertainty, we are 95 percent confident that the RME risk is below the risk level of concern— no further analysis is needed. Also—selection of a beta distribution over a lognormal distribution for ingestion rate changes the 95th percentile of the risk distribution by a factor of 10—further evaluation may be needed.*

- **Resource allocation** - *e.g., Two of the 10 exposure variables contribute 90 percent of the variability in the risk estimate.*

- **Risk communication** - *e.g., For input variable X, if we were to use a distribution based on site-specific data instead of a national survey, we would expect a minimal change in the RME risk estimate.*

Decision Making with the Tiered Approach

In general, the type of information provided by a sensitivity analysis will vary with each tier of a PRA. Table A-1 provides an overview of the methods that may be applied in each tier based on the type of information needed. In Tier 1, sensitivity analysis typically involves changing one or more input variables or assumptions and evaluating the corresponding changes in the risk estimates. Ideally, the results for Tier 1 would be useful in deciding which exposure pathways, variables, and assumptions are carried forward for further consideration in subsequent tiers of analysis. By identifying the variables that are most important in determining risk, one can also decide whether point estimates, rather than probability distribution functions (PDFs), can be used with little consequence to the model output. This information is important not only for designing 1-D MCA models of variability, but also for designing more complex analyses of uncertainty discussed in Appendix D (e.g., 2-D MCA models, geostatistical analysis, Bayesian analysis). Section A.2.2 provides an overview of the Tier 1 methods and some insights regarding their limitations. Methods associated with Monte Carlo simulations used in Tiers 2 and 3 can take advantage of the ability to vary multiple inputs simultaneously and account for correlations. Sections A.2.3 and A.3 provide an overview of the sensitivity analysis methods that can be applied in a probabilistic analysis.

Table A-1. Overview of Sensitivity Analysis Methods Applicable in Tiers 1, 2, and 3 of a PRA.

Tier	Goal	SA Method(s)	What to Look For	Rationale
1	Quantify contributions of each exposure pathway to risk, identify major and minor pathways	Calculate % of total risk from each exposure pathway	Exposure pathways that contribute a very small percentage (e.g., < 5%) to total risk	Good preliminary step in Tier 1 for reducing the number of exposure variables to focus on in subsequent tiers.
			Exposure variables that appear in multiple exposure pathways	Risk estimates are likely to be more sensitive to variables that appear in multiple exposure pathways.
1	Identify the form of the dose equation for key pathways	Inspection	Equation is multiplicative or additive	SR values can be determined with minimal effort (see Table A-3). For multiplicative equations, SR=1.0 for all variables in the numerator, and SR is a function of the percent change for all variables in the denominator.
			Equation contains variables with exponents (e.g., powers, square roots)	Output is likely to be more sensitive to variables with exponents greater than 1.0.
1	Quantify contributions of each exposure variable to total risk, identify major and minor variables	Sensitivity Ratio (SR), unweighted	SR = 1.0, or SR is the same for multiple variables	It's likely that this is a multiplicative equation (see above), and the SR approach will not be effective at discriminating among relative contributions. Explore sensitivity further with other methods.
			SR ≠ 1.0	SR may vary as a function of the % change in the input variable. In this situation, it can be informative to explore small deviations (± 5%) and large deviations (min, max) in the input variables.
			SR < 1.0	Implies an inverse relationship between the input and output variables (e.g., inputs in the denominator of a risk equation).

Table A-1. Overview of Sensitivity Analysis Methods Applicable in Tiers 1, 2, and 3 of a PRA.

Tier	Goal	SA Method(s)	What to Look For	Rationale
			SR=0	Variable probably appears in both the numerator and denominator and, therefore, cancels out of the risk equation. Examples include exposure duration (ED) in noncancer risk equations, and body weight (BW) if ingestion rate is expressed as a function of body weight.
1	(cont'd) Quantify contributions of each exposure variable to total risk	Sensitivity Ratio (SR), weighted—also called Sensitivity Score	Differences in SR based on the weighting factor	A more informative approach than unweighted SR value for those variables that have sufficient information to define a weighting factor (e.g., coefficient of variation or range).
2	Quantify relative contributions of exposure pathways to risk	1-D MCA for variability or uncertainty, with outputs specifying % contribution of exposure pathways	Compare mean with high- and low-end percentiles of % contribution to risk	The % contribution of each exposure pathway will vary as a function of the variability (or uncertainty) in the inputs; exposure pathways that appear to be relatively minor contributors on average, or from Tier 1 assessment, may in fact be a major contributor to risk under certain exposure scenarios. The likelihood that a pathway is nonnegligible (e.g., > 5%) can be useful information for risk managers.
2	Quantify relative contributions of exposure variables to risk	1-D MCA for variability or uncertainty, Graphical analysis— scatterplots of inputs and output	Nonlinear relationship	Easy and intuitive approach that may identify relationships that other methods could miss. May suggest transformations of input or output variables (e.g., logarithms, power transformations) that would improve correlation and regression analyses.
		1-D MCA, Correlation Analysis using Pearson and /or Spearman Rank	Very high or low correlation coefficients

Differences between relative rankings based on Pearson and Spearman | Easy to implement with commercial software; rank orders the variables based on the *average* contribution to variance. Differences in magnitude of coefficients are expected between Pearson and Spearman rank approaches, but relative order of importance is likely to be the same. |

Table A-1. Overview of Sensitivity Analysis Methods Applicable in Tiers 1, 2, and 3 of a PRA.

Tier	Goal	SA Method(s)	What to Look For	Rationale
		1-D MCA, Multiple Linear Regression Analysis (e.g., stepwise)	Very high or low regression coefficients R^2 and adjusted R^2 for total model	Easy to implement with commercial software; gives contribution to reduction in residual sum of squares (RSS) For risk equations with large sets of input variables, a small subset of inputs may be able to explain the majority of the variance.
2	Quantify relative contributions of exposure variables to RME risk range	1-D MCA; same as previous step, but for subset of risk distribution (e.g., > 90th percentile)	Difference in relative contributions for entire risk distribution and the RME range of the risk distribution	Variables may contribute differently to the high-end of the risk distribution, especially if the input variables are highly skewed. This situation would warrant a closer look at the assumptions regarding the estimate of the variance, differences in the upper tail (high-end percentiles) for alternative choices of probability distributions, and assumptions associated with truncation limits.
		1-D MCA, Goodness-of-fit, K-S or Chi-square; Sort output as above; perform GoF on input distribution only, comparing subset of input values corresponding with high-end risk to subset corresponding with remainder of risk distribution	GoF result—rejection of null (distributions are the same) suggests the variable may be an important contributing factor to the RME risk estimate	A second method for identifying variables that contribute differently at the high-end of the risk distribution. GoF test results should be interpreted with caution because a Monte Carlo simulation will generally yield large sample sizes (e.g., n=5,000 iterations), which is more likely to result in a positive GoF test (i.e., rejection of the null).
3	Quantify relative contributions of exposure pathways and variables to variability and uncertainty in risk	2-D MCA, same sensitivity analysis methods as Tier 2	For variability, evaluate inner loop values; for parameter uncertainty, evaluate outer loop values	The results of a sensitivity analysis depend on the question that is being asked about the risk estimate—are we interested in variability or uncertainty? The major sources of variability in risk may point to a different set of input variables than the major sources of uncertainty in risk.

Resource Allocation

        Decisions regarding allocation of future resources and data collection efforts to reduce lack of knowledge generally should take into consideration the most influential input factors in the model, and the cost of gaining new information about the factors. Sensitivity analysis is a key feature of determining the expected value of information (EVOI) (see Appendix D). Once a sensitivity analysis is used to identify an input variable as being important, the source of its variability generally should be determined. If an input factor has a significant uncertainty component, further research and/or data collection can be conducted to reduce this uncertainty. Reducing major sources of uncertainty, such as the most relevant probability model for variability or the parameter estimates for the model, will generally improve confidence in the model output, such as the estimated 95th percentile of the risk distribution. An input factor may contribute little to the variability in risk, but greatly to the uncertainty in risk (e.g., the concentration term). Likewise, a variable may contribute greatly to the variability in risk, but, because the data are from a well characterized population, the uncertainty is relatively low (e.g., adult tap water ingestion rate).

        An example of the output from a 2-D MCA of uncertainty and variability (see Appendix D) is shown in Figure A-1. Assume for this example that the decision makers choose the 95th percentile risk as the RME risk, and that a sensitivity analysis is run to identify and quantitatively rank the important source(s) of parameter uncertainty. The bar chart (top panel) in Figure A-1 indicates that the mean soil concentration contributes most to the uncertainty in the 95th percentile risk estimate. In addition, the mean exposure frequency is a greater source of uncertainty than the standard deviation exposure frequency. Since both the sample size and variance impact the magnitude of the confidence limits for an arithmetic mean soil concentration, one way to reduce the confidence limits (i.e., the uncertainty) would be to collect additional soil samples. As shown by the box-and-whisker plots (bottom panel) in Figure A-1, increasing the sample size (from *n*=25 to *n*=50) reduced the 90% confidence limits for the 95th percentile risk to below 1E-05, assuming the additional observations support the same estimate of the mean and standard deviation as the original sample.

        Although the uncertainty in a risk estimate can be reduced by further data collection if the sensitive input distribution represents uncertainty, this is not necessarily true for input distributions that represent variability. For example, variability in the distribution of body weights can be better characterized with additional data, but the coefficient of variation (i.e., standard deviation divided by the mean) will not in general be reduced.

Risk Communication

        Even if additional data are not collected to reduce uncertainty, identifying the exposure factors that contribute most to risk or hazard may be useful for risk communication. For example, assume that the input for exposure frequency has the strongest effect on the risk estimate for a future recreational open space. Further examination of this exposure variable reveals that the wide spread (i.e., variance) of the PDF is a result of multiple users (e.g., mountain bikers, hikers, individuals who bring picnics, etc.) of the open space who may spend very different amounts of time recreating. As a result of this analysis, the decision makers and community may decide to focus remediation efforts on protecting the high-risk subpopulation that is expected to spend the most time in the open space.

        After determining which contaminants, media, and exposure pathways to carry into a PRA, numerical experiments generally should be performed to determine the sensitivity of the output to various distributions and parameter estimates that may be supported by the available information. Variables that

do not strongly affect the risk estimates may be characterized with point estimates without significantly altering the risk estimates. This guidance document does not recommend a quantitative metric or rule of thumb for determining when a variable strongly affects the output; this would generally be determined on a case-by-case basis. A qualitative or quantitative analysis may be used depending on the complexity of the risk assessment at this point. For example, incidental ingestion of soil by children is often an influential factor in determining risk from soil, a factor recognized by risk assessors. This recognition is a *de facto* informal sensitivity analysis. An array of quantitative techniques is also available, ranging from something as simple as comparing the range of possible values (i.e., maximum-minimum) for each variable, to more complex statistical methods such as multiple regression analysis. Several of these methods are discussed in more detail in this appendix.

Often, sufficient information is available to characterize a PDF for a minor variable without significant effort. This situation raises a question of whether the variable should be characterized with a point estimate or a PDF. The results of sensitivity analysis should be viewed as supplemental information, rather than an absolute rule for determining when to use a PDF. There are at least two issues to consider related to risk communication. First, the risk communication process may be facilitated by narrowing the focus of the evaluation to the key factors. More attention can be given to the discussion of key variables quantified by PDFs by describing the minor variables with point estimates. However, the decision to use a point estimate should be balanced by considering a second issue regarding perception and trust. There may be a concern that by reducing sources of variability to point estimates, there would be a reduction (however small) in the variability in risk, especially if multiple small sources of variability add up to a nonnegligible contribution. To address these concerns, it may be prudent to leave the PDFs in the calculations despite the results of a sensitivity analysis.
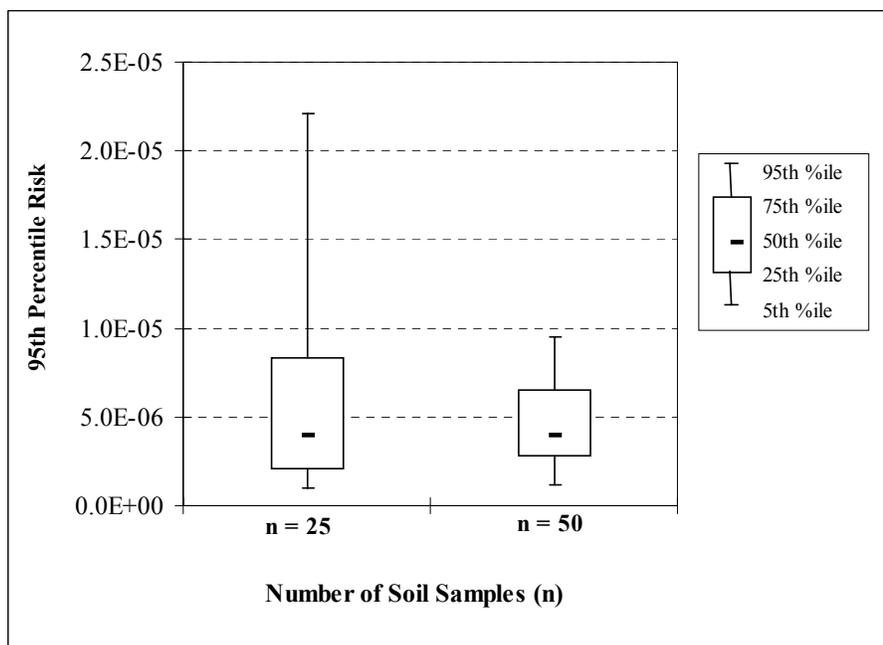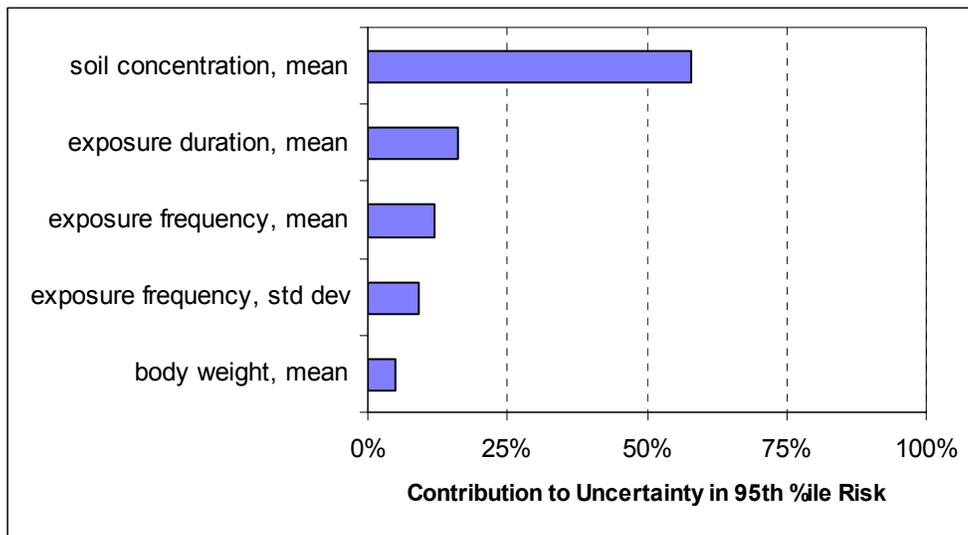
Figure A-1. Results of 2-D MCA in which parameters of input distributions describing variability are assumed to be random values. Results of a sensitivity analysis (top graph) suggest that more than 50% of the uncertainty in the 95th percentile of the risk distribution is due to uncertainty in the arithmetic mean concentration in soil. The bottom graph gives box-and-whisker plots for the 95th percentile of the risk distribution associated with Monte Carlo simulations using different sample sizes (*n*=25 and *n*=50). For this example, the whiskers represent the 5th and 95th percentiles of the distribution for uncertainty, otherwise described as the 90% confidence interval (CI). For *n*=25, the 90% CI is [1.0E-06, 2.2E-05]; for *n*=50, the 90% CI is reduced to [1.2E-06, 9.5E-06]. While increasing *n* did not change the 50th percentile of the uncertainty distribution, it did provide greater confidence that the 95th percentile risk is below $1x10^{-5}$.

A.2.0   COMMON METHODS OF SENSITIVITY ANALYSIS

Of the numerous approaches to sensitivity analysis that are available (see Exhibit A-3), no single approach will serve as the best analysis for all modeling efforts. Often, it will make sense to apply multiple approaches. The best choice(s) for a particular situation will depend on a number of factors, including the nature and complexity of the model and the resources available. A brief description of the more common approaches is provided in this appendix. Sensitivity analysis need not be limited to the methods discussed in this guidance, which focuses on the more common approaches. A large body of scientific literature on various other methods is available (e.g., Iman et al., 1988, 1991; Morgan and Henrion, 1990; Saltelli and Marivort, 1990; Rose et al., 1991; Merz, Small, and Fischbeck, 1992; Shevenell and Hoffman 1993; Hamby, 1994; U.S. EPA, 1997). Any method used, however, generally should be documented clearly and concisely. This includes providing all information needed by a third party to repeat the procedure and corroborate the

> **EXHIBIT A-3**
>
> **SOME KEY INDICES OF SENSITIVITY ANALYSIS**
>
> - Relative contribution of exposure pathways
> - Inspection of risk equation
> - Sensitivity ratios (i.e., elasticity)
> - Sensitivity scores (i.e., weighted sensitivity ratios)
> - Graphical techniques with results of Monte Carlo simulations (e.g., scatter plots)
> - Correlation coefficient (or coefficient of determination, r^2) (e.g., Pearson product moment, Spearman rank)
> - Normalized multiple regression coefficient
> - Goodness-of-fit test for subsets of the risk distribution

results. Relevant information might include the following: exposure pathways and equations; a table with the input variables with point estimates, probability distributions and parameters; and tables or graphs giving the results of the sensitivity analysis and description of the method used. A hypothetical example is presented in this appendix to illustrate how to apply and present the results of selected approaches to sensitivity analysis.

Hypothetical Example of a Noncancer Risk Equation

To illustrate the application of sensitivity analysis concepts to Tier 1 and Tier 2, a hypothetical risk assessment is presented based on the general equation for Hazard Index (HI) given by Equation A-1. Note that HI is equal to the sum of the chemical-specific Hazard Quotient (HQ) values, so technically, this example reflects exposures from a single chemical.

$$HI = \frac{C_i \times I_i \times AF_i \times EF \times ED}{BW \times AT} \times \frac{1}{RfD} \qquad \text{Equation A-1}$$

The terms in Equation A-1 can be defined as follows: concentration in the ith exposure medium (C_i), ingestion or inhalation rate of the ith exposure medium (I_i), absorption fraction of chemical in the ith exposure medium (AF_i), exposure duration (ED), exposure frequency (EF), body weight (BW), averaging time (AT=ED x 365 days/year), and reference dose (RfD).

For this example, HI is calculated as the sum of the exposures to adults from two exposure pathways: tap water ingestion and soil ingestion. Equation A-2 gives the equation for HI while Table A-2 gives the inputs for a point estimate assessment and a probabilistic assessment of variability.

$$HI = \frac{((C_w \times I_w \times AF_w) + (C_s \times I_s \times AF_s)) \times EF \times ED}{BW \times AT} \times \frac{1}{RfD} \quad \text{Equation A-2}$$

Table A-2. Point estimates and probability distributions for input variables used in the hypothetical example of HI associated with occupational exposure via water and soil ingestion.

Input Variable in Equation A-2	Point Estimate		Probability Distribution		Units
	CTE	RME	Type	Parameters	
Concentration in Water (C_w)	40	40	point estimate	40	mg/L
Tap Water Ingestion Rate (I_w)	1.3	2.0	lognormal[1]	[1.3, 0.75]	L/day
Absorption Fraction Water (AF_w)	0.30	0.50	beta[2]	[2.0, 3.0]	unitless
Concentration in Soil (C_s)	90	90	point estimate	90	mg/kg
Soil Ingestion Rate (I_s)	0.05	0.10	uniform	[0, 0.13]	kg/day
Absorption Fraction Soil (AF_s)	0.10	0.30	beta[2]	[1.22 , 4.89]	unitless
Exposure Frequency (EF)	250	350	triangular	[180, 250, 350]	days/yr
Exposure Duration (ED)	1	7	empirical[3]	see below	years
Body Weight (BW)	75	75	lognormal[1]	[74.6, 12.2]	kg
Averaging Time (AT)	365	2555	empirical[4]	ED x 365	days
RfD$_{oral}$[5]	0.5	0.5	point estimate	0.5	mg/kg-day

[1]Parameters of lognormal distribution are [arithmetic mean, standard deviation].
[2]Parameters of beta distribution are [alpha, beta], with range defined by min=0 and max=1.0. Parameter conversions for arithmetic mean and standard deviation are given in Table A-7.
[3]Parameters of empirical cumulative distribution function (ECDF) for ED ~ [min, max, {x}, {p}] = [0, 30, {0.08, 0.18, 0.30, 0.44, 0.61, 0.84, 1.17, 1.72, 3.1, 6.77, 14.15, 23.94}, {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.975, 0.99}]], where x is the array of values and p is the array of corresponding cumulative probabilities.
[4]AT=ED x 365 for noncarcinogenic risks (Hazard Index).
[5]For simplicity, RfD$_{oral}$ is assumed to be applicable to the ingestion of the chemical in both water and soil.

A.2.1   TIER 1 APPROACHES

Approaches for sensitivity analysis in Tier 1 of a PRA are limited to calculations that are based on changing point estimates. They are generally easy to perform and to communicate. As given by Table A-1, goals for the sensitivity analysis in Tier 1 include quantifying the relative contributions of the exposure pathways, identifying potential nonlinear relationships that may exist between input variables and the risk estimate, and rank ordering the relative contribution of exposure variables to variability or uncertainty in the risk estimate. This last goal may be the most difficult to achieve due to the limitations associated with the point estimate methodology. Methods are applied to the hypothetical example presented above (Section A.2.0) in order to demonstrate the inherent limitations of the Tier 1 approaches in some situations.

A.2.1.1 PERCENTAGE CONTRIBUTION OF EXPOSURE PATHWAYS TO TOTAL RISK

For cancer and noncancer risk assessments central tendency exposure (CTE) and RME risk is typically calculated as the sum of risks from multiple exposure pathways. Risks may be dominated by one or two exposure pathways, which can be determined through a simple calculation as shown below. The relative contributions of exposure pathways are likely to differ between the CTE risk and RME risk.

The point estimates in Table A-2 were applied to Equation A-2 to obtain CTE and RME point estimates of HI. Table A-3 gives the percent contributions of soil ingestion and tap water ingestion using Equations A-3 and A-4. Tap water ingestion contributes at least 90% to HI, and the total HI is greater than 1.0 for both CTE and RME point estimates. If 1.0 is the level of concern for HI, and a decision was made to explore variability and uncertainty in a probabilistic analysis, this result might support prioritizing the evaluation of data and assumptions associated with the tap water ingestion pathway.

Table A-3. Percent contribution of exposure pathways to HI for the example in Section A.2.

Exposure Pathway	CTE Point Estimate		RME Point Estimate	
	HI	% of total[2]	HI	% of total
Soil Ingestion	0.02	6 %	0.15	13 %
Tap Water Ingestion	0.28	94 %	1.02	87 %
Total	0.30	100 %	1.17	100 %

[1]Equation A-3: $HI_{total} = HI_{soil} + HI_{water}$
[2]Example using Equation A-4: % of total RME HI for soil ingestion = (0.15 / 1.17) x 100% = 13%.

$$HI_{total} = \sum_{i=1}^{n} HI_i \qquad\qquad \text{Equation A-3}$$

$$Percent\ Contribution_i = \frac{HI_i}{HI_{total}} \times 100\% \qquad\qquad \text{Equation A-4}$$

In this example, the choice of CTE and RME point estimates reflects an effort to explore variability in HI, rather than uncertainty. Even if the concentration terms represent the upper confidence limit on the mean (e.g., 95% UCL), the point estimates chosen to represent the CTE and RME for other exposure variables reflect assumptions about the variability in exposures. There is uncertainty that the choices actually represent the central tendency and reasonable maximum exposures. To explore this uncertainty, alternative choices for CTE and RME may have been selected. This type of exploration of uncertainty in Tier 1 may also be viewed as a form of sensitivity analysis. The percent contribution of exposure pathways could be recalculated, and the sensitivity ratio approaches discussed below may also be applied.

A.2.1.2 INSPECTION OF RISK EQUATION

For many Superfund risk assessments, risk equations can be characterized as relatively simple algebraic expressions involving addition, multiplication, and division of input variables. The term "product-quotient" model is often applied to describe equations such as Equation A-1. For these risk equations, the input variables that are likely to contribute most to the variability or uncertainty in risk can be identified by inspection. In addition, inspection of the risk equation can help to identify which sensitivity analysis methods are unlikely to reveal the relative importance of the input variables. This concept is illustrated by comparing the results of the sensitivity ratio approach (Section A.2.1.3) with the Tier 2 approaches (Section A.2.2) applied to the hypothetical example in Section A.2.0.

Some risk equations can be more complex, involving conditional probabilities, or expressions with exponents (e.g., $y=x^2$, or $y=\exp(1-x)$). In these cases, the Tier 1 sensitivity analysis methods may be effective and highlighting the variables that contribute most to the risk estimates.

A.2.1.3 SENSITIVITY RATIO (SR)

A method of sensitivity analysis applied in many different models in science, engineering, and economics is the **Sensitivity Ratio** (SR), otherwise know as the *elasticity* equation. The approach is easy to understand and apply. The ratio is equal to the percentage change in output (e.g., risk) divided by the percentage change in input for a specific input variable, as shown in Equation A-5.

$$SR = \frac{\left(\dfrac{Y_2 - Y_1}{Y_1}\right) \times 100\%}{\left(\dfrac{X_2 - X_1}{X_1}\right) \times 100\%} \qquad\qquad \text{Equation A-5}$$

where,
- Y_1 = the baseline value of the output variable using baseline values of input variables
- Y_2 = the value of the output variable after changing the value of one input variable
- X_1 = the baseline point estimate for an input variable
- X_2 = the value of the input variable after changing X_1

Risk estimates are considered most sensitive to input variables that yield the highest absolute value for SR. The basis for this equation can be understood by examining the fundamental concepts associated with partial derivatives (see Section A.3.2). In fact, SR is equivalent to the normalized partial derivative (see Equation A-12).

Sensitivity ratios can generally be grouped into two categories—local SR and range SR. For the local SR method, an input variable is varied by a small amount, usually ±5% of the nominal (default) point estimate, and the corresponding change in the model output is observed. For the range sensitivity ratio method, an input variable is varied across the entire range (plausible minimum and maximum values). Usually, the results of local and range SR calculations are the same. When the results differ, the risk assessor can conclude that different exposure variables are driving risk near the high-end (i.e., extreme tails of the risk distribution) than at the central tendency region.

Demonstration of the Limitations of SR Approach

Although SR is a relatively simple and intuitive approach, it does not provide useful information under certain conditions for the more common risk equations. To demonstrate the limitations, first Equation A-5 is applied to the hypothetical example given in Section A.2.0. The results are then extended to a more general case of any of the more common risk models that involve the products of terms (i.e., multiplicative model) or the sum of terms (i.e., additive model).

Table A-4 presents an example of the local SR and range SR approach applied to the set of RME inputs given in Table A-2. For the local SR, each input was increased by 5% (i.e., $\Delta=+5\%$), while for the range SR, each input was increased by 50%. Inputs for exposure frequency were truncated at the maximum value of 365 days/year, which represents a 4.29% increase over the nominal RME value of 350 days/year.

Table A-4. Results of the Sensitivity Ratio (SR) approach applied to the hypothetical example of RME HI given in Section A.2.0. Includes *both* soil ingestion and tap water ingestion pathways.

Input Variable , X in Equation A-2[1]	Nominal RME value (X_1)	Local SR ($\Delta = + 5.0\%$)			Range SR ($\Delta = + 50\%$ or max)		
		X_2	Δ in HI (%)	SR	X_2	Δ in HI (%)	SR
Tap Water Ingestion Rate, I_w (L/day)	2.0	2.1	4.35	0.87	3.0	43.5	0.87
Absorption Fraction Water, AF_w (unitless)	0.50	0.525	4.35	0.87	0.75	43.5	0.87
Soil Ingestion Rate, I_s (kg/day)	0.100	0.105	0.65	0.13	0.150	6.5	0.13
Absorption Fraction Soil, AF_s (unitless)	0.30	0.315	0.65	0.13	0.45	6.5	0.13
Exposure Frequency, EF (days/yr)	350	365[2]	4.29	1.00	365[2]	4.29	1.00
Exposure Duration, ED (years)	7	7.35	0.00	0.00	10.5	0.00	0.00
Body Weight, BW (kg)	75	78.75	- 4.46	- 0.89	112.5	- 33.33	- 0.67

[1]Only input variables that represent variability are included. Concentrations are point estimates of uncertainty. Averaging time is a function of exposure duration. RfD is a fixed point estimate.
[2]Maximum EF of 365 days/yr represents a 4.29% change in the nominal RME value of 350 days/yr.

The following observations can be made from these results:
▸ In decreasing order of sensitivity:

Local SR ($\Delta = 5\%$) rankings: EF > BW > I_w = AF_w > I_s = AF_s > ED

Range SR ($\Delta = 50\%$) rankings: EF > I_w = AF_w > BW > I_s = AF_s > ED

▸ EF is the most sensitive variable with an SR value of 1.0. Since EF is a variable in the numerator for both exposure pathways, this result is to be expected, as will be explained below.

- ► ED yields an SR=0, suggesting it does not contribute to the HI estimate. Upon closer inspection of the risk equation, it is apparent that ED occurs in the numerator of Equation A-2, as well as in the denominator (AT=ED x 365). Thus, ED effectively cancels out of the product quotient model and does not effect the estimate of HI.

- ► BW, the only variable in the denominator of the risk equation, is also the only variable to yield a different SR value when comparing the local and range SR approaches. Thus, BW is the only variable for which SR depends on the percent change in the input (Δ).

- ► BW is the only negative SR value, indicating that HI and BW are inversely related. This is true in general for any variable in the denominator of a product quotient model.

- ► For variables unique to the water ingestion pathway (I_w, AF_w), SR=0.87. Similarly, for variables unique to the soil ingestion pathway (I_s, AF_s), SR=0.13 These SR values are exactly the same as the percent contributions of the tap water ingestion pathway and soil ingestion pathway to HI (see Table A-3).

Since tap water ingestion is the dominant pathway (i.e., 87% of RME HI), a reasonable strategy for the Tier 1 sensitivity ratio approach might be to limit the subsequent probabilistic analysis in Tier 2 to the tap water ingestion pathway; so that input variables unique to the soil ingestion pathway would be characterized by point estimates. For this relatively simple example, this would mean that soil ingestion rate (I_s) and absorption fraction from soil (AF_s) would be described by point estimates instead of PDFs. The question to address would then become—Of the exposure variables in the tap water ingestion pathway, which ones contribute most to HI? A sensitivity ratio approach was applied to the tap water ingestion pathway to address this question. The results are presented in Table A-5.

Table A-5. Results of the Sensitivity Ratio (SR) approach applied to the hypothetical example of RME HI given in Section A.2.0. Includes *only* tap water ingestion pathway.

Input Variable , X in Equation A-2[1]	Nominal RME value (X_1)	Local SR ($\Delta = +5.0\%$)			Range SR ($\Delta = +50\%$ or max)		
		X_2	Δ in HI (%)	SR	X_2	Δ in HI (%)	SR
Tap Water Ingestion Rate, I_w (L/day)	2.0	2.1	5.0	1.00	3.0	50	1.00
Absorption Fraction Water, AF_w (unitless)	0.50	0.525	5.0	1.00	0.75	50	1.00
Exposure Frequency, EF (days/yr)	350	365[2]	4.29	1.00	365[2]	4.29	1.00
Exposure Duration, ED (years)	7	7.35	0.00	0.00	10.5	0.00	0.00
Body Weight, BW (kg)	75	78.75	- 4.46	- 0.89	112.5	- 33.33	- 0.67

[1]Only input variables that represent variability are included. Concentrations are point estimates of uncertainty. Averaging time is a function of exposure duration. RfD is a fixed point estimate.
[2]Maximum EF of 365 days/yr represents a 4.29% change in the nominal RME value of 350 days/yr.

The following observations can be made from these results:

▸ In decreasing order of sensitivity:

  Local SR (Δ = 5%) rankings: I_w = AF_w = EF > BW > ED

  Range SR (Δ = 50%) rankings: I_w = AF_w = EF > BW > ED

▸ SR values for variables in the numerator (I_w, AF_w, and EF) are all equal to 1.0, so the SR approach suggests that they contribute equally to the HI estimate.

▸ BW values are the same as in Table A-4. They are negative, and the values change as a function of the percent change in the nominal RME value (Δ).

Tables A-4 and A-5 suggest that the SR approach provides essentially the same information about sensitivity as other Tier 1 methods. Specifically, inspection of the risk equation reveals that ED does not contribute to HI. In addition, for pathway-specific variables in the numerator, like ingestion rates and absorption fractions, SR values are equal to the percent contributions of the exposure pathways. This actually reflects the fact that each factor in the numerator of a multiplicative equation has an SR of 1.0.

The results of the SR approach applied to the example above can be generalized to all multiplicative and additive risk equations, as discussed below.

Generalizing the Limitations of the SR Approach

In many cases, the general equation for SR (Equation A-5) will give values that can be determined *a priori*, without doing many calculations. To understand why this is true, it is useful to simplify the algebraic expression given by Equation A-5. Let Δ equal the percentage change in the input variable, X_1. For *SR* calculations, Δ may be either positive or negative (e.g., ±5% for local SR; ±100% for range SR), and the new value for the input variable (i.e., X_2) is given by Equation A-6.

$$\begin{aligned} X_2 &= X_1 + (X_1 \times \Delta) \\ &= X_1 \times (1 + \Delta) \end{aligned}$$

<div align="right">Equation A-6</div>

Therefore, the denominator in Equation A-5 reduces to Δ:

$$\frac{X_2 - X_1}{X_1} = \frac{X_1(1 + \Delta) - X_1}{X_1} = \frac{(1 + \Delta) - 1}{1} = \Delta$$

and Equation A-5 reduces to Equation A-7:

$$SR = \frac{1}{\Delta} \times \left(\frac{Y_2 - Y_1}{Y_1} \right)$$

<div align="right">Equation A-7</div>

Equation A-7 can be used to evaluate SR for different types of exposure models in which the intake equation is generally expressed as a simple algebraic combination of input variables. Solutions to SR calculations for input variables in both multiplicative and additive equations are given in Table A-6. For any such risk equation, the solution will fall into one of the five categories given by Exhibit A-4.

EXHIBIT A-4

**CATEGORIES OF SOLUTIONS FOR SENSITIVITY RATIOS OF
MULTIPICATIVE OR ADDITIVE EQUATIONS**

Case 1   SR is a constant (e.g., 1.0). SR is independent of the choice of nominal (default) values for input variables and the choice of Δ.

Case 2   SR is a constant determined only by the nominal values for the input variables. SR is independent of the choice of Δ.

Case 3   SR is constant determined only by the choice of Δ. SR is independent of the nominal values for the input variables.

Case 4   SR is a function of both the nominal values for the input variables and the choice of Δ.

Case 5   SR is 0. The variable does not contribute to the risk estimate.

Table A-6. Examples of algebraic solutions to Sensitivity Ratio calculations for additive and multiplicative forms of risk equations.[1, 2]

Equation Type (Output = Y, Inputs = A, B, C, D)		$SR_A =$	$SR_B =$	$SR_C =$	$SR_D =$
1) Additive in Numerator	$Y = \dfrac{A + B}{C}$	$\dfrac{A}{A + B}$	$\dfrac{B}{A + B}$	$-\dfrac{1}{1 + \Delta}$	NA[3]
2) Additive in Denominator	$Y = \dfrac{A}{C + D}$	1.0	NA	$-\dfrac{C}{C\,(1 + \Delta) + D}$	$-\dfrac{D}{D\,(1 + \Delta) + C}$
3) Multiplicative in Numerator	$Y = \dfrac{A \times B}{C}$	1.0	1.0	$-\dfrac{1}{1 + \Delta}$	NA
4) Multiplicative in Denominator	$Y = \dfrac{A}{C \times D}$	1.0	NA	$-\dfrac{1}{1 + \Delta}$	$-\dfrac{1}{1 + \Delta}$

[1]Sensitivity Ratio for input variable A for an equation that is additive in the numerator: $SR_A = A / (A + B)$.
[2]Δ=% change in input variable. For example, Δ for $C=[(C_2 - C_1)/C_1] \times 100\%$, where C_1=the original point estimate and C_2=the modified point estimate. Similarly, $C_2 = C_1 (1 + \Delta)$.
[3]NA=not applicable because the variable is not in the equation.

The following observations can be made for the four forms of the risk equation, based on one of the five cases described in Exhibit A-4:

(1) Additive in Numerator

▸ **Case 2**: SR values for variables in the numerator depend exclusively on the nominal point estimates for all variables in the numerator. The values are independent of the choice of percent change in the inputs (Δ).

▸ **Case 3**: SR values for variables in the denominator depend exclusively on Δ, and are negative (i.e., inversely related to the output). Also, the lower the choice for Δ, the higher the resulting SR values. Therefore, SR is somewhat arbitrary, especially for the range SR approach since input variables may have different plausible minimum and maximum values.

(2) Additive in Denominator

▸ **Case 1**: SR values for variables in the numerator are always equal to 1.0. Since they are independent of the nominal values and Δ, there is no way to distinguish the relative contributions to the output.

▸ **Case 4**: SR values for variables in the denominator are a function of both the nominal values of variables in the denominator and Δ.

(3) Multiplicative in Numerator and *(4) Multiplicative in Denominator*

▸ **Case 1**: SR values for variables in the numerator are always equal to 1.0. Since they are independent of the nominal values and Δ, there is no way to distinguish the relative contributions to the output.

▸ **Case 3**: SR values for variables in the denominator depend exclusively on the Δ, and are negative (i.e., inversely related to the output). Also, the lower the choice for Δ, the higher the resulting SR values. Therefore, SR is somewhat arbitrary, especially for range SR since input variables may have different plausible minimum and maximum values.

These generalized results highlight a major limitation in the use of the SR approach for obtaining information from sensitivity analysis. For simple exposure models in which the relationship between exposure and risk is linear (e.g., multiplicative), the ratio offers little information regarding the relative contributions of each input variable to the risk estimate. In many cases, all of the input variables will have the same constant, either equal to 1.0 (in the case of a single exposure pathway) or equal to the relative contributions of the exposure pathways. For more complex models that combine additive, multiplicative, and nonlinear relationships between inputs and outputs (e.g., environmental fate and transport models, pharmacokinetic models), the ratio is likely to be an effective screening tool for identifying potentially influential input variables and assumptions.

Another difficulty with the SR approach is that it generally requires an assumption that the input variables are independent. Two variables may actually be positively correlated (e.g., high values of X_1 correspond with high values of X_2) or negatively correlated (e.g., high values of X_1 correspond with low values of X_2). If input variables are correlated, holding the value for one variable fixed while allowing the other to vary may produce misleading results, especially with the range sensitivity ratio approach. For example, it may not be realistic to hold body weight fixed at a central tendency while allowing skin surface area to vary from the minimum to maximum values. An improvement over the sensitivity ratio approach would be to allow correlated input variables to vary simultaneously.

A.2.1.4 SENSITIVITY SCORE

A variation on the sensitivity ratio approach may provide more information from a Tier 1 sensitivity analysis, but it requires that additional information be available for the input variables. The *sensitivity score* is the SR weighted by a normalized measure of the variability in an input variable (U.S. EPA, 1999). Examples of normalized measures of variability include the coefficient of variation (i.e., standard deviation divided by the mean) and the normalized range (i.e., range divided by the mean), as given by Equation A-8.

$$Sensitivity\ Score = SR \times \frac{\sigma}{\mu} \quad or \quad SR \times \frac{(max - min)}{\mu} \qquad \text{Equation A-8}$$

By normalizing the measure of variability (i.e., dividing by the mean), this method effectively weights the ratios in a manner that is independent of the units of the input variable, and provides a more robust method of ranking contributions to the risk estimates than the SR alone. This approach does require that the coefficient of variation or range can be calculated for each variable. Tables A-7 and A-8 present the results of the sensitivity scores based on the CV applied to the hypothetical example from Section A.2.0.

Table A-7. Calculation of coefficient of variation (CV = SD / Mean) for the hypothetical example of RME HI given in Section A.2.0.

Input Variable , X in Equation A-2[1]	Probability Distribution[2]	Mean[3]	SD[3]	CV = SD/Mean
Tap Water Ingestion Rate, I_w (L/day)	lognormal (1.3, 0.75)	1.3	0.75	0.58
Absorption Fraction, Water, AF_w (unitless)	beta (2.0, 3.0)	0.4	0.2	0.50
Soil Ingestion Rate, I_s (kg/day)	uniform (0, 0.13)	0.065	0.038	0.58[2]
Absorption Fraction, Soil, AF_s (unitless)	beta (1.22, 4.89)	0.20	0.15	0.75
Exposure Frequency, EF (days/yr)	triangular (180, 250, 350)	260	35	0.13[3]
Exposure Duration, ED (years)	empirical CDF (see Table A-2 for parameters)	1.75	3.86	2.21
Body Weight, BW (kg)	lognormal (74.6, 12.2)	74.6	12.2	0.16

[1]Only input variables that represent variability are included. Concentrations are point estimates of uncertainty. Averaging time is a function of exposure duration. RfD is a fixed point estimate.
[2]Beta (a, b): mean=a / (a+b) and SD = ((a x b) / [(a + b)^2 x (a+b+1)])^0.5)
Uniform (min, max): mean = (min + max)/2 and SD = ((1/12)^0.5) x (max - min) = 0.289 x (max - min)
Triangular (min, mode, max): mean = (min + mode + max)/3 and SD = (1/18) x (min^2 + mode^2 + max^2 - min x max - min x mode - mode x max)
Empirical CDF ({x}, {p}): mean and SD were estimated by Monte Carlo simulation.
[3]Mean=arithmetic mean; SD=arithmetic standard deviation

Table A-8. Results of the Sensitivity Score (Score) approach applied to the hypothetical example of RME HI given in Section A.2.0. Calculations for Sensitivity Ratio (SR) and Coefficient of Variation (CV) are given in Table A-4 and Table A-7, respectively.

Input Variable , X in Equation A-2[1]	Nominal RME value (X_1)	CV (Table A-7)	Local SR ($\Delta = +5\%$)		Range SR ($\Delta = +50\%$)	
			SR (Table A-4)	Score[2]	SR (Table A-4)	Score[2]
Tap Water Ingestion Rate, I_w (L/day)	2.0	0.58	0.87	**0.50**	0.87	**0.50**
Absorption Fraction, Water, AF_w (unitless)	0.50	0.50	0.87	**0.44**	0.87	**0.44**
Soil Ingestion Rate, I_s (kg/day)	0.100	0.58	0.13	**0.06**	0.13	**0.06**
Absorption Fraction, Soil, AF_s (unitless)	0.30	0.75	0.13	**0.10**	0.13	**0.10**
Exposure Frequency, EF (days/yr)	350	0.13	1.00	**0.13**	1.00	**0.13**
Exposure Duration, ED (years)	7	2.21	0.00	**0**	0.00	**0**
Body Weight, BW (kg)	75	0.16	- 0.89	**- 0.14**	- 0.67	**- 0.11**

[1]Only input variables that represent variability are included. Concentrations are point estimates of uncertainty. Averaging time is a function of exposure duration. RfD is a fixed point estimate.
[2]Score=SR x CV (see Equation A-8)

The following observations can be made from these results:
- ▸ In decreasing order of sensitivity:

  - Score based on local SR ($\Delta = 5\%$): I_w > AF_w > BW > EF > AF_s > IR_s > ED

  - Score based on range SR ($\Delta = 50\%$): I_w > AF_w > EF > BW > AF_s > IR_s > ED

- ▸ Compared with the SR approach alone in which sensitivity can only be expressed for exposure pathways, the sensitivity score approach provides a measure of sensitivity for exposure variables within each exposure pathway.

- ▸ Although ED has the highest CV, it continues to have no contribution to the HI.

  - ▸ If Tier 1 sensitivity analysis is based on the sensitivity score, the highest ranked variables are generally those with the highest CV in the exposure pathway that contributes the most to the total risk (HI). For this hypothetical example, I_w and AF_w are the two highest ranked variables.

A.2.2 TIER 2 APPROACHES

Approaches for sensitivity analysis in Tier 2 of a PRA utilize the results of Monte Carlo simulations, which allows multiple input variables to vary simultaneously. The methods are relatively simple to perform with spreadsheets or commercial statistical software. The results are generally easy to communicate, although the details of the methodology are more complex than Tier 1 approaches. As given by Table A-1, goals for the sensitivity analysis in Tier 2 are the same as Tier 1:quantifying the relative contributions of the exposure pathways, identifying potential nonlinear relationships that may exist between input variables and the risk estimate, and rank ordering the relative contribution of exposure variables to variability or uncertainty in the risk estimate. In addition, since the output is a distribution, Tier 2 sensitivity analysis methods can also utilize graphical techniques to observe nonlinear relationships, as well as evaluate potential changes in relative importance of variables and assumptions for risks in the RME risk range. Methods are applied to the hypothetical example presented in Section A.2.0 in order to demonstrate the advantages over the Tier 1 methods.

A.2.2.1 GRAPHICAL TECHNIQUES

Simple scatter plots of the simulated input and output (e.g., risk vs. exposure frequency, or risk vs. arithmetic mean soil concentration) can be used to qualitatively and quantitatively evaluate influential variables. A "tight" best-fit line through the scatter plot, as indicated by the magnitude of the r^2, suggests that a variable may significantly influence the variance in risk. Hypothetical scatter plots used to identify sensitive and insensitive variables are shown in Figure A-2. Another method for visualizing the relationship between all of the inputs and outputs is to generate a scatterplot matrix (Helsel and Hirsch, 1992). This graphic shows both histograms and scatter plots for all variables on the same page.

Figure A-3 illustrates scatter plots for the 1-D MCA simulations associated with the example from Section A.2.0. Based on the r^2 values (i.e., coefficient of determination for simple linear regression analysis), the relationship between HI and I_w is very strong ($r^2 = 0.47$) while the relationship between HI and I_s is very weak ($r^2 < 0.01$), suggesting that HI is more sensitive to variability in I_w than I_s.

A.2.2.2 CORRELATION COEFFICIENTS

The variance in a risk estimate from a Monte Carlo simulation is due to the variance in the probability distributions used in the risk equation. It is commonly said that a Monte Carlo model propagates sources of variability simultaneously in a risk equation. Numerous statistical techniques, known collectively as correlation analysis and regression analysis, can be applied to a linear equation to estimate the relative change in the output of a Monte Carlo simulation based on changes in the input variables. Examples of metrics of sensitivity include the simple correlation coefficient, the rank correlation coefficient, and a variety of coefficients from multiple regression techniques. The underlying assumptions associated with these approaches are discussed in greater detail in Section A.3. As explained in Section A.3.3.1, correlation coefficients and regression coefficients are based on different interpretations of the input variables, but they can be calculated with similar equations.

When the output distribution is compared with the distribution for one input variable at a time, two of the more common approaches are to calculate the Pearson product moment correlation and the Spearman rank correlation. Correlation analysis with one input variable will generally yield reasonable results when the input variables are sampled independently in a Monte Carlo simulation. Some statistical packages offer the correlation coefficient as an index of sensitivity, so it is important to identify which

coefficient is being calculated. *Crystal Ball*® and *@Risk* can be used to calculate the Spearman rank correlation, which tends to be more robust when the relationships between inputs and outputs are nonlinear. If the relationships are linear, such as with the product quotient models presented in this appendix, the two metrics of correlation will yield similar rankings of input variables. Rank correlation coefficients shown in *Crystal Ball*® and *@Risk* are calculated by the standard method provided in most statistics texts. *Crystal Ball*® also indicates that sensitivity can be determined as contribution to variance. This is not the relative partial sum of squares techniques discussed in Section A.3.3.2 (Equation A-19). Instead, *Crystal Ball*® calculates the contribution to the variance by squaring the rank correlation coefficients and normalizing them to 100%. Many other commonly used commercial software packages will perform Spearman rank correlation. Pearson product moment correlations (*r*) can be calculated in Microsoft Excel using the trendline feature in a scatter plot chart, or by using the function *Correl(X array, Y array)*, where *X array* corresponds with the Monte Carlo simulation of an input variable, and *Y array* corresponds with the output of the simulation.

Figure A-4 illustrates results of the correlation analysis for the 1-D MCA simulations associated with the example from Section A.2.0. The graphics were generated using *Crystal Ball*® 2000. The results are summarized in Table A-9. If the model output variable (e.g., HI) and input variable are highly correlated, it means that the output is sensitive to that input variable. By squaring the coefficient, the results can be expressed in terms of the percentage contribution to variance in the output (Figure A-4, top panel). To determine if the correlation is positive or negative, the correlation coefficient should not be squared (Figure A-4, bottom panel). For risk equations, in general, variables in the numerator of the equation (ingestion rate, absorption fraction, exposure frequency, etc.) will tend to be positively correlated with risk, while variables in the denominator (body weight) will tend to be negatively correlated with risk. The greater the absolute value of the correlation coefficient, the stronger the relationship.

Table A-9. Results of Tier 2 sensitivity analyses applied to hypothetical example in Section A.2.0: Pearson product moment correlations and Spearman rank correlations.[1]

Exposure Variable	Product Moment Correlation		Spearman Rank Correlation[2]		
	r	r^2 x 100%	r	r^2 x 100%	normalized r^2 x 100%
Tap Water Ingestion Rate, I_w (L/day)	0.644	41.4	0.603	36.3	39.5
Absorption Fraction Water, AF_w (unitless)	0.583	34.0	0.666	44.4	48.3
Body Weight, BW (kg)	- 0.216	4.7	- 0.229	5.2	5.7
Exposure Frequency, EF (days/yr)	0.174	3.0	0.167	2.8	3.0
Absorption Fraction Soil, AF_s (unitless)	0.109	1.2	0.149	2.2	2.4
Soil Ingestion Rate, I_s (g/day)	0.061	0.4	0.099	1.0	1.1
Exposure Duration, ED (years)	0.010	0.0	0.010	0.0	0.0

[1]Monte Carlo simulation using *Crystal Ball*® *2000*, Latin Hypercube sampling, and 5000 iterations.
[2]*Crystal Ball*® *2000* output includes Spearman rank correlations, *r*, and *normalized* r^2 values, calculated by dividing each r^2 value by the sum of all the r^2 values (i.e., 0.920 in this example). Figure A-4 illustrates the *r* and normalized r^2 values for the Spearman rank correlation analysis.
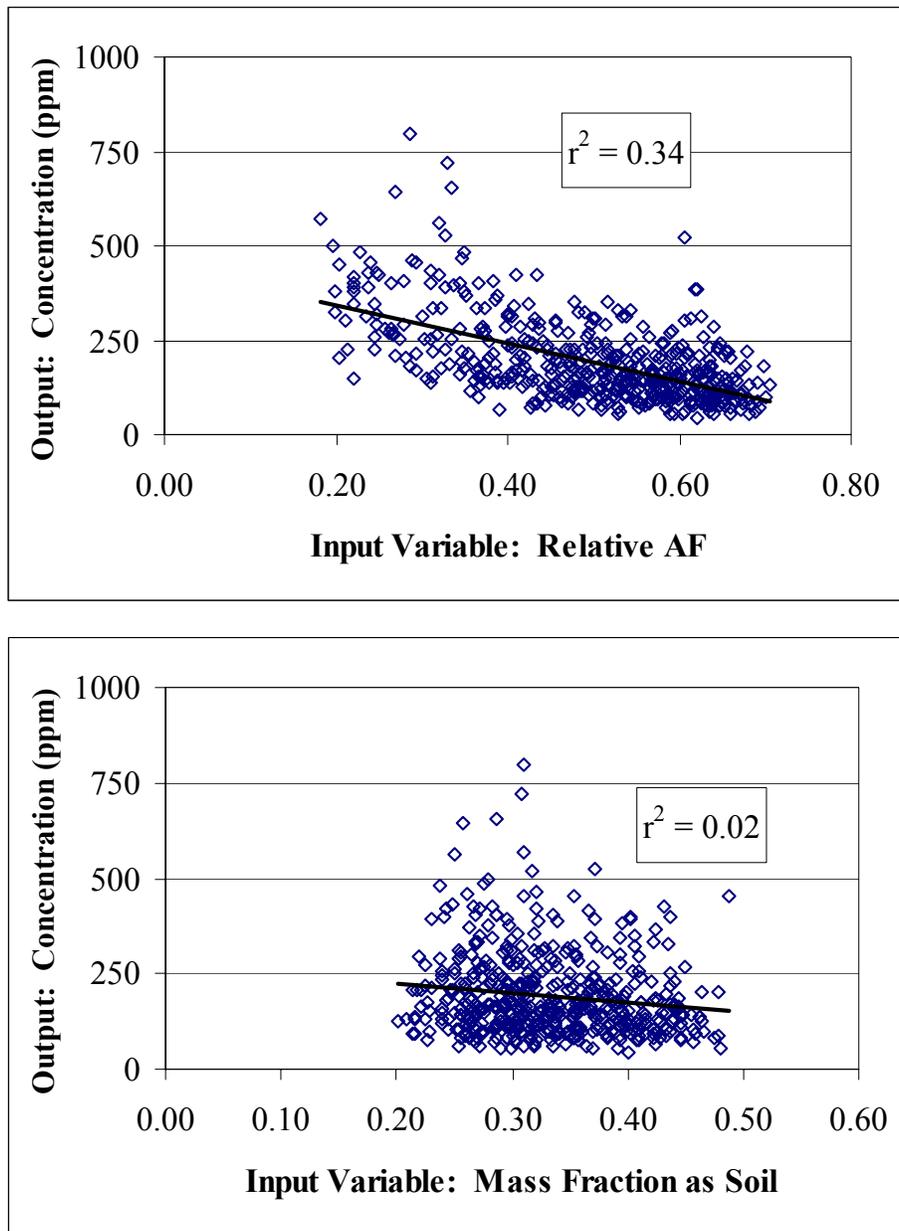
Figure A-2. Scatterplots of simulated random values from a 1-D MCA of variability. The output from the model is a contaminant concentration in soil (C) that corresponds with a prescribed (fixed) level of risk for a hypothetical population (based on Stern, 1994). For each iteration of a 1-D MCA simulation, random values were simultaneously selected for all model variables and the corresponding concentration (C) was calculated. Inputs were simulated as independent random variables. Scatterplots of 500 consecutive random values and estimates of C are shown for two input variables: relative absorption fraction, RAF (top graph); and mass fraction of dust as soil, F (bottom graph). There is a moderate, indirect relationship between C and RAF (r^2=0.34), compared with the weak relationship between C and F (r^2=0.02), suggesting that the model output (C) is more sensitive to variability in RAF than F.
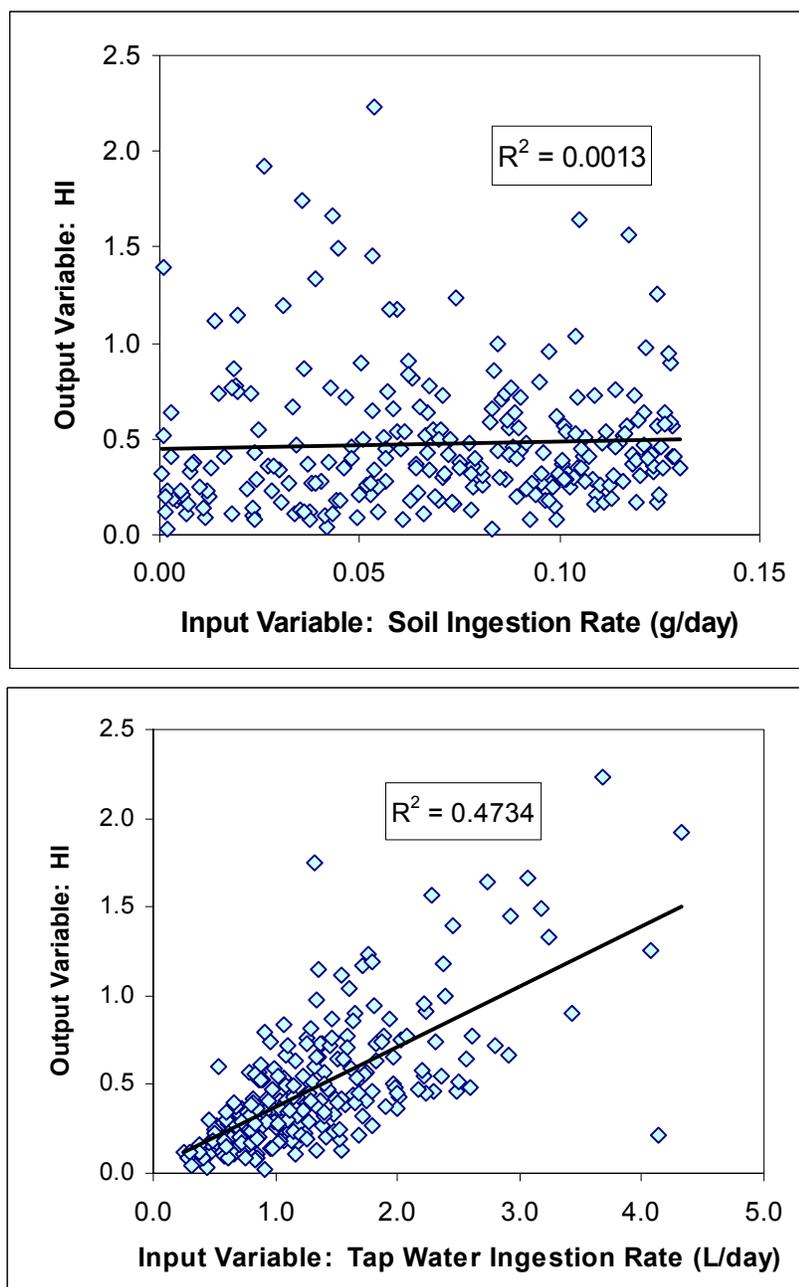
Figure A-3. Scatterplots of simulated random values from a 1-D MCA of variability for example in Section A.2.0. The output from the model is HI. For each iteration of a 1-D MCA simulation, random values were simultaneously selected for all model variables and the corresponding HI was calculated. Inputs were simulated as independent random variables. Scatterplots of 250 consecutive random values and estimates of HI are shown for two input variables: soil ingestion rate, I_s (top graph); and tap water ingestion rate, I_w (bottom graph). There is a negligible relationship between HI and I_s ($r^2 < 0.01$), compared with the strong relationship between HI and I_w ($r^2=0.47$), suggesting that the model output (HI) is more sensitive to variability in I_w than I_s. Best-fit lines were generated with the Simple Linear Regression in Microsoft Excel's trendline option for scatterplots; r^2 values represent the coefficient of determination (see Section A.3).
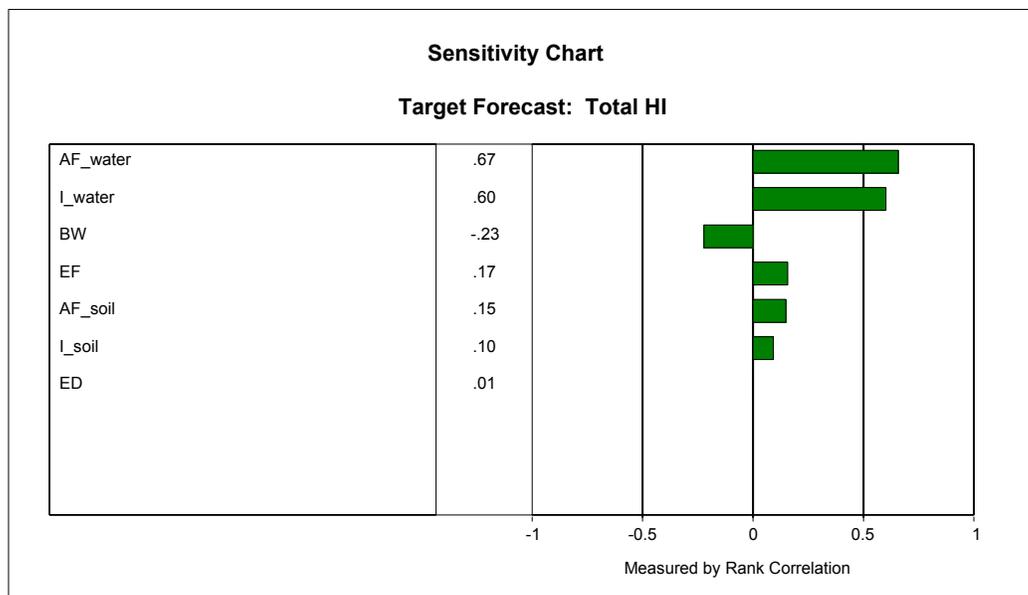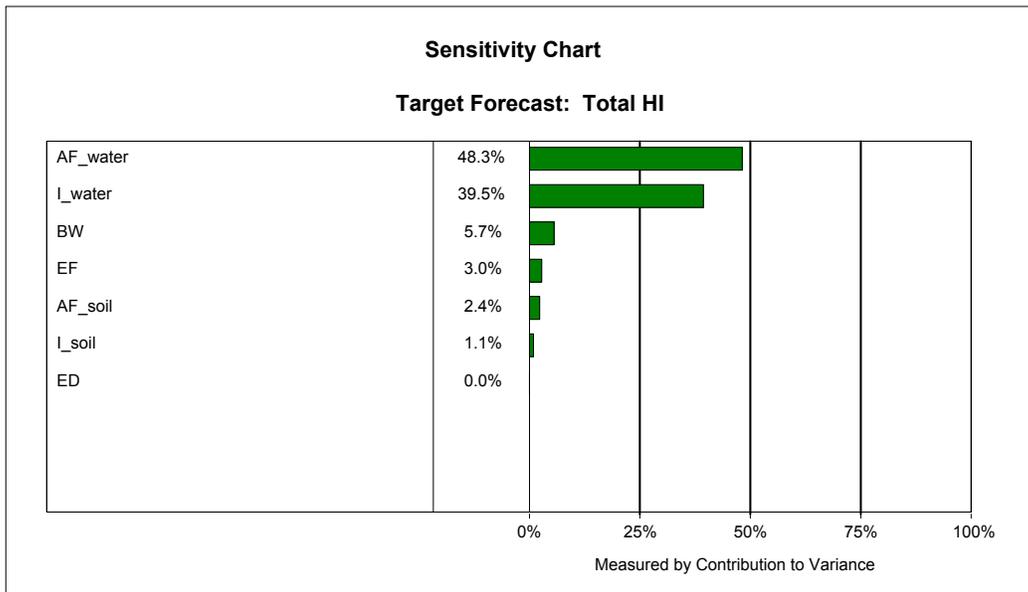
Sensitivity Chart

Target Forecast: Total HI

AF_water	48.3%
I_water	39.5%
BW	5.7%
EF	3.0%
AF_soil	2.4%
I_soil	1.1%
ED	0.0%

0%   25%   50%   75%   100%

Measured by Contribution to Variance

Sensitivity Chart

Target Forecast: Total HI

AF_water	.67
I_water	.60
BW	-.23
EF	.17
AF_soil	.15
I_soil	.10
ED	.01

-1   -0.5   0   0.5   1

Measured by Rank Correlation

Figure A-4. Top panel - bar graph showing the r^2 values (square of Spearman rank correlation coefficient), a metric for the dependence of HI on exposure factors based on 1-D MCA for variability. Bottom panel - bar graph, sometimes referred to as "tornado plot", showing rank correlation coefficient. This graph is effective for showing both the relative magnitude and direction of influence (positive or negative) for each variable. Abbreviations for input variables are given in Table A-4. In this example, the variable with the greatest effect on HI is the absorption fraction in water (AF_w), followed by the water ingestion rate (I_w). Concentration does not influence variability because, in this example, long-term average concentration is characterized by a point estimate (i.e., 95% UCL), rather than a probability distribution. Exposure duration does not influence variability because variability in ED is expressed in both the numerator (ED) and denominator (AT=ED x 365 for noncarcinogenic effects), and cancels out. Output was generated with *Crystal Ball®,* which calculates the contribution to variance by squaring the rank correlation coefficient and normalizing to 100%.

In this example, seven exposure variables are used to characterize variability in HI. The remaining variables in the risk equation (i.e., concentration terms, and RfD) are characterized by point estimates. Because point estimates do not vary in a Monte Carlo simulation, they do not contribute to the variance in the output. This result does not mean that concentration is an unimportant variable in the risk assessment. Concentration may still contribute greatly to the uncertainty in the risk estimate. A sensitivity analysis of parameter uncertainty in a risk equation can be explored using iterative simulations, such as with 2-D MCA.

Results of the Pearson correlation and Spearman rank correlation give similar rankings of the input variables, with absorption fraction of water (AF_w) and tap water ingestion rate (I_w) being the two dominant exposure variables. Pearson correlations suggest that I_w is the most sensitive variable (r =0.644), whereas the highest Spearman rank correlation is for AF_w (r = 0.603). This may reflect the fact that I_w is characterized by an untruncated lognormal distribution, whereas AF_w is bounded between 0 and 1.0. The effect on the correlations of the occasional high-end value for I_w generated from random sampling of the lognormal distribution will tend to be expressed by Pearson correlations, but muted by the Spearman rank correlations.

A comparison of the Tier 1 and Tier 2 results is given below:

> ‣ **Tier 1, Sensitivity Ratios:**

> > - Local SR (Δ = 5%) rankings: EF > BW > I_w = AF_w > I_s = AF_s > ED

> > - Range SR (Δ = 50%) rankings: EF > I_w = AF_w > BW > I_s = AF_s > ED

> ‣ **Tier 1, Sensitivity Scores:**

> > - Score based on local SR (Δ = 5%): I_w > AF_w > BW > EF > AF_s > IR_s > ED

> > - Score based on range SR (Δ = 50%): I_w > AF_w > EF > BW > AF_s > IR_s > ED

> ‣ **Tier 2, Correlation Coefficients:**

> > - Pearson: I_w > AF_w > BW > EF > AF_s > IR_s > ED

> > - Spearman Rank: AF_w > I_w > EF > BW > AF_s > IR_s > ED

The Tier 1 sensitivity scores and Tier 2 correlation coefficients yield similar results, suggesting that, if sufficient information is available to estimate the coefficient of variation in the input variables, a Tier 1 analysis can help to focus efforts on the variables that contribute most to the variance in risk. By contrast, the Tier 1 sensitivity ratio approach suggested that EF was the most influential variable, when in fact it contributes less than 5% to the variance in the HI. These results suggest that Tier 1 sensitivity ratios are best applied to identify dominant exposure pathways, rather than dominant exposure variables in the risk equation.

A.2.2.3 FOCUSING ON THE RME RANGE OF THE RISK DISTRIBUTION

Monte Carlo methods can also be used to determine the sensitivity over a subset of the output distribution, such as the RME range (i.e., 90th to 99.9th percentiles). For some exposure models, the relative contribution of exposure variables may be different for the high-end exposed individuals than for the entire range of exposures. The general strategy for exploring sensitivity over subsets of risk estimates is to first sort the distribution of simulated output values in ascending (or descending) order, and then apply a sensitivity analysis to the subset of interest (e.g., > 90th percentile). For the hypothetical example presented in this appendix, there was no difference in the relative rankings of inputs in the RME range.

A.2.2.4 INSPECTION

With Monte Carlo analysis, the probability distributions assumed for the various input variables are used to generate a sample of a large number of points. Statistical methods are applied to this sample to evaluate the influence of the inputs on the model output. A number of different "indices" of sensitivity can be derived from the simulated sample to quantify the influence of the inputs and identify the key contributors. Most of these are based on an assumption that the model output Y varies in a monotonic, linear fashion with respect to various input variables (X_1, X_2, etc.). For example, an estimate of average daily intake (mg/kg-day) from multiple exposure pathways is linear with respect to the intake from each pathway. Since most risk models are linear with respect to the input variables, the output distribution (particularly its upper percentiles) tends to be dictated by the input variables with the largest coefficient of variation (CV), or the ratio of the standard deviation to the mean. For example, Equation A-9 represents a simple expression for intake rate as a function of random variables X_1 and X_2 :

$$Y = X_1 + X_2$$

Equation A-9

where X_1 and X_2 may represent dietary intake associated with prey species 1 and 2, respectively. If the same probability distribution was used to characterize X_1 and X_2, such as a lognormal distribution with an arithmetic mean of 100 and standard deviation of 50 (i.e., CV=50/100=0.5), each variable would contribute equally to variance in Y. If, however, X_2 was characterized by a lognormal distribution with an arithmetic mean of 100 and standard deviation of 200 (i.e., CV=200/100=2.0), we would expect Y to be more sensitive to X_2. That is, X_2 would be a greater contributor to variance in Y.

While the coefficient of variation may be a useful screening tool to develop a sense of the relative contributions of the different input variables, a common exception is the case when X_1 and X_2 have different scales. For example, Equation A-10 is an extension of Equation A-9:

$$Y = a_1 X_1 + a_2 X_2$$

Equation A-10

where a_1 and a_2 are constants that may represent the algebraic combination of point estimates for other exposure variables. If the means of X_1 and X_2 are equal, but $a_1 \gg a_2$, then X_1 would tend to be the dominant contributor to variance, regardless of the CV for X_2. This concept was demonstrated by the sensitivity score calculations given in Table A-8. Water ingestion rate (I_w) and soil ingestion rate (I_s) had the same CV (0.58), but I_w was the dominant variable because tap water ingestion contributed approximately 90% to the HI.

The most influential random variables generally have the highest degrees of skewness or are related to the output according to a power function (Cullen and Frey, 1999). For example, Equation A-11 presents an extension of Equation A-10 in which there is a power relationship between X_2 and Y. In this

$$Y = a_1 X_1 + a_2 X_2^{\theta} \qquad \text{Equation A-11}$$

example, assume Y represents the total dietary intake rate of cadmium for muskrats, X_1 and X_2 represent the dietary intake rate associated with prey species 1 and 2, respectively, a_1 and a_2 represent additional point estimates in the equation, and θ is the power exponent. In general, for $\theta > 1$, the total dietary intake rate (Y) will be more sensitive to the intake rate associated with species 2 (X_2) than species 1. Assume (hypothetically) that the power relationship stems from the fact that there is a direct relationship between availability of prey species X_2 and chemical body burdens of prey species X_2 because individuals that are more accessible to the muskrat also happen to frequent areas of the site with higher concentrations.

A.3.0 ADVANCED CONCEPTS IN SENSITIVITY ANALYSIS

This section provides additional information on the underlying principles of sensitivity analysis, although it is not a comprehensive summary and is not intended to substitute for the numerous statistical texts and journal articles on sensitivity analysis. Section A.3.1 begins with a general framework for relating model output to model input. Section A.3.2 explains the sensitivity ratio approach and highlights some of its limitations. Section A.3.3 reviews some of the metrics reported by the commercial software that report results of sensitivity analysis following Monte Carlo simulations (e.g., *Crystal Ball*®, *@Risk*). While statistical software for MCA provides convenient metrics for quantifying and ranking these sources, it is strongly recommended that risk assessors and risk managers develop an understanding of the underlying principles associated with these metrics.

A.3.1 RELATING THE CHANGE IN RISK TO THE CHANGE IN INPUT VARIABLE X

For purposes of discussion, let *Y* denote a model output (e.g., risk) and suppose that it depends on the input variable *X*. In general, a risk assessment model may use any number of inputs; however, for purposes of illustrating concepts, it is convenient to restrict this discussion to one variable. The model relates the output *Y* to values of *X* (i.e., $x_0, x_1, \bullet\bullet\bullet, x_n$) based on the function expressed as *Y=F(x)*. The sensitivity of *Y* to *X* can be interpreted as the slope of the tangent to the response surface *F(X)* at any point x_i. This two-dimensional surface can be a simple straight line, or it may be very complex with changing slopes as shown in Figure A-5a. The sensitivity, therefore, may depend on both the value of *X* and the amount of the change Δx about that point. This concept can be extended to two input variables, X_1 and X_2, where the response is characterized by a three-dimensional surface. The shape may be a simple plane (Figure A-5c) or it may be very complex with many "hills" and "valleys" depending on the defining function $F(X_1, X_2)$. In a typical risk assessment with ten or more variables, the surface can be very complex, but the shape is likely to be dominated by a small subset of the input variables.

A sensitivity analysis based on a relatively small deviation about the point may be referred to as a local sensitivity analysis, while a large deviation may be referred to as range sensitivity analysis. In either case, the objective is to evaluate the sensitivity at some nominal point (X_1^*, X_2^*) such as the point defined by the mean or median of X_1 and X_2. At any point, the sensitivity of the model output, $Y^* = F(X_1^*, X_2^*)$, to one of the inputs (X_1 or X_2), is represented by the rate of change in *Y* per unit change in *X*. This is the slope of the surface at that nominal point in the direction of *X* and is expressed as $\partial Y/\partial X_i$, the *partial derivative* of *Y* with respect to *X*.

$$\textit{Partial Derivative} = \frac{\partial Y}{\partial X} \approx \frac{\Delta Y}{\Delta X}$$

If the function F(X_1, X_2) is known explicitly, it may be possible to determine the partial derivatives analytically. This is not a requirement, however, because an estimate can be obtained by incrementing X_i by a small amount, ΔX_i, while keeping the other inputs fixed and reevaluating the model output Y. The resulting change in Y divided by ΔX_i will approximate $\partial Y/\partial X_i$ at the nominal point. In practice, analytical solutions can be approximated using Monte Carlo techniques. This information is presented to highlight the fundamental concepts of sensitivity analysis. The partial derivative, *per se*, would typically not be one of the methods of sensitivity analysis used in a PRA. However, all of the approaches that are presented in this appendix are variations on this concept.

One drawback to using the partial derivative to quantify the influence of X_i is that the partial derivative is influenced by the units of measurement of X_i. For example, if the measurement scale for X_i is changed from grams to milligrams, the partial derivative $\partial Y/\partial X_i$ will change by a factor of 1,000. Therefore, it is necessary to **normalize the partial derivative** to remove the effects of units (see Section A.3.2).

If the relationship between Y and all of the inputs is linear, then the response surface is a flat plane and each of the partial derivatives at each point, (X_i, Y), will remain constant regardless of where the point is in the surface (Figure A-5b). In this case, it is a simple matter to determine the relative influence that the various inputs have on the model output. When the relationship is nonlinear, however, the situation is more complex because the influence of a particular input may vary depending on the value of that input.
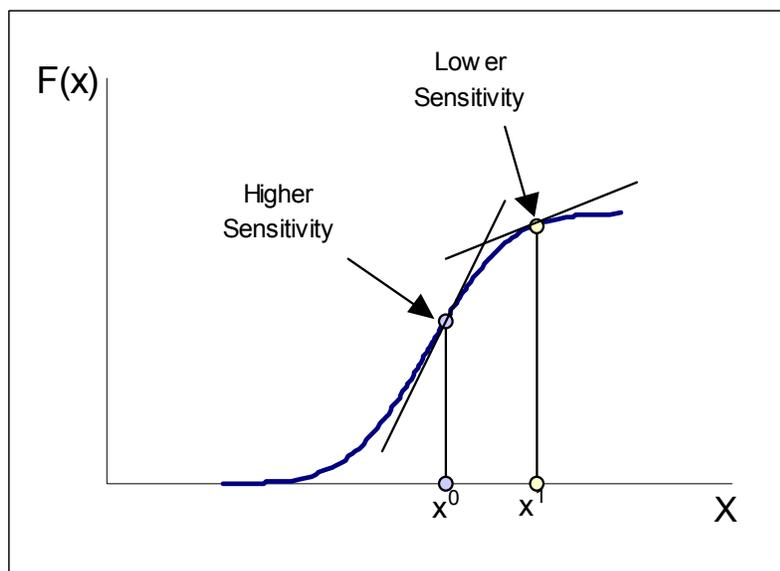

Figure A-5a. Hypothetical 2-D response surface for Y given one input variable: Y=$F(X)$. The sensitivity of Y with respect to X is calculated as the slope at a specific point on the surface (x^0, x^1), or the partial derivative, $\partial Y/\partial X_i$.
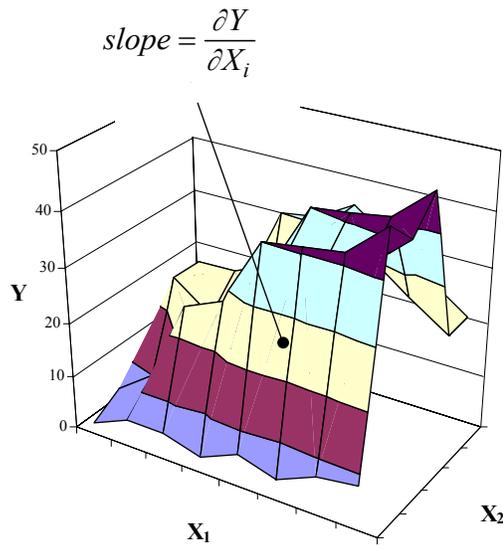
$$slope = \frac{\partial Y}{\partial X_i}$$

Figure A-5b. Hypothetical 3-D response surface for Y given two input variables: $Y = f(X_1, X_2)$. The sensitivity of Y with respect to X_i is calculated as the slope at a specific point on the surface, or the partial derivative, $\partial Y/\partial X_i$.
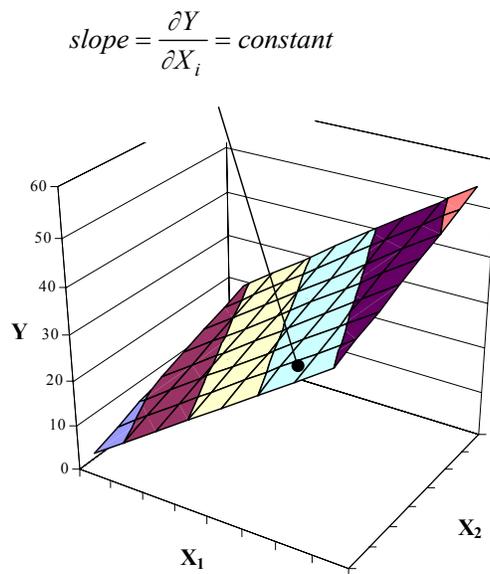
$$slope = \frac{\partial Y}{\partial X_i} = constant$$

Figure A-5c. Hypothetical 3-D response surface when Y is a linear function of two input variables: $Y=f(X_1, X_2)$. The slope (i.e., the partial derivative, $\partial Y/\partial X_i$) is constant for any point (X_i, Y) on the surface in the direction of X_i. In this case, $\partial Y/\partial X_1=5$ while $\partial Y/\partial X_2=2$.

A.3.2 NORMALIZED PARTIAL DERIVATIVE

Classical sensitivity analysis methods use estimates of the partial derivatives of the model output with respect to each variable. For the purpose of evaluating the relative influence of the various input variables on the model output at a single point, the **normalized partial derivative** provides a useful index.

If the input variables are all discrete and take on a small number of values, then it is possible to evaluate the influence of the various input variables at each of the points defined by considering all possible combinations of the inputs. Then the influence can be evaluated for each input by computing normalized partial derivatives at each point. This approach is limited to situations where the number of inputs as well as the number of possible values for each input is relatively small; otherwise, the number of combinations to be evaluated will be unmanageable. Furthermore, when evaluating the influence at different points on the input-output surface simultaneously, it is important to take into account the probability associated with each of those points. For example, the fact that a particular input has a large influence on the model output at a particular point would be discounted if the probability associated with that particular point is very low.

A similar approach may be used to analyze inputs that are continuous variables if a few points representing the range of values are selected. For example, low, medium (or nominal), and high values may be selected for each of the continuous input variables and then the relative influence of each of the input variables can be computed as in the case of discrete inputs. One limitation of this approach, however, is that the continuous nature of the inputs makes it impossible to calculate an exact probability for each of the points. Generally, in a PRA, many if not all of the inputs will be random variables described by probability distributions and it will be necessary to quantify the influence of each input, X_i, over the entire range of X_i.

An estimate of the partial derivative can be obtained by incrementing X_i by a small amount, say ΔX_i while keeping the other inputs fixed and reevaluating the model output Y. The resulting change in Y divided by ΔX_i will approximate $\partial Y/\partial X_i$ at the nominal point.

$$\text{Partial Derivative} = \frac{\partial Y}{\partial X} \approx \frac{\Delta Y}{\Delta X}$$

As previously noted, one complication to using the partial derivative to quantify the influence of X_i is that the partial derivative is influenced by the units of measurement of X_i. One way this is accomplished is to divide the partial derivative by the ratio of the nominal point estimates, Y^* / X_i^* (or equivalently multiply by X_i^* / Y^*). An approximation of the normalized partial derivative is given by Equation A-12.

$$\text{Normalized Partial Derivative} \approx \frac{\Delta Y}{\Delta X} \times \frac{X_1}{Y_1} = \frac{\left(\dfrac{Y_2 - Y_1}{Y_1}\right)}{\left(\dfrac{X_2 - X_1}{X_1}\right)} \quad \text{Equation A-12}$$

This is the same as the equation for calculating sensitivity ratios (Section A.2.1.3), or elasticity (see Equation A-5). As with the SR approach, the normalized partial derived can be weighted by characteristics of the input variable (Section A.2.1.4). One approach is to divide by the ratio of standard deviations (σ_Y / σ_X), where σ_Y is the standard deviation of Y and σ_X is the standard deviation of X. This method requires that the standard deviations be known, or that a suitable estimate can be obtained.

As previously noted, if the relationship between Y and all of the inputs is nonlinear, the influence of a particular input may vary depending on the value of that input. One approach to this problem is to consider a range of values for the input and to examine the influence over that range. If the input is considered to be a random variable following some specified probability distribution, then it may be desirable to look at the influence that the random input has on the model output across the distribution of input values. This can be accomplished with a Monte Carlo approach. Another technique that addresses nonlinearities is to calculate contributions to variance using input variables that are transformed (e.g., lognormal or power transformation).

A.3.3 REGRESSION ANALYSIS: R^2, PEARSON R, AND PARTIAL CORRELATION COEFFICIENTS

In order to understand R^2, it is necessary to first understand simple and multiple linear regression. In regression analysis, we are interested in obtaining an equation that relates a dependent variable (Y) to one or more independent variables (X):

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{Equation A-13}$$

where β_0 and β_1 are regression coefficients, and ε is called a random error. Equation A-13 is the general equation for a simple linear regression, because there is only one Y and one X variable, and their relationship can be described by a line with intercept β_0 and slope β_1.

Note that *linear* regression refers to the linear relationship between parameters (β_0, β_1), not X and Y. Thus, the equation $Y = \beta_0 + \beta_1 X_1^2 + \varepsilon$ is considered linear. *Multiple* linear regression involves more than one X related to one Y $\left[Y = \beta_0 + \beta_1 X_1 + \beta_2 X \cdots\right]$, while *multivariate* regression involves more than one Y to more than one X.

The random error, ε, represents the difference between an observed Y value (calculated from the observed input variables), and a Y value predicted by the regression line ($\hat{y}$). It is also called

EXHIBIT A-5

SIMPLIFYING ASSUMPTIONS IN REGRESSION ANALYSIS

- Y is a linear function of the unknown coefficients (β_i)
- Successive values of Y are uncorrelated
- Variance of Y is constant for all values of inputs (X_i)

the *residual* (i.e., $\varepsilon = y - \hat{y}$). The random error takes into account all unpredictable and unknown factors that are not included in the model. Exhibit A-5 gives some of the simplifying assumptions that apply to regression analysis. Assumptions about ε are that the random error has mean = 0 and constant variance, and is uncorrelated among observations. One method of finding the best regression line is to minimize the residual sum of squares (i.e., least-squares method), also called the sum of squares due to error (SSE).

In terms of sensitivity analysis, we are interested in how much of the variation in *Y* can be explained by the variation in *X*, and how much is unexplained (due to random error). If a scatter plot of paired observations (*x*, *y*) shows that our regression line intersects all of the observations exactly, then all of the variation in *Y* is explained by *X*. Another way of stating this is that the difference between the mean output ($\bar{y}$) and an observed *y* (y_i), or ($y_i - \bar{y}$), is equal to the difference between the mean output and a predicted *y* or ($\hat{y} - \bar{y}$).

In general, the total deviation of y_i from $\bar{y}$ is equal to the sum of the deviation due to the regression line plus the deviation due to random error:

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$
$$\Sigma\,(y_i - \bar{y})^2 = \Sigma\,(y_i - \hat{y}_i)^2 + \Sigma\,(\hat{y}_i - \bar{y})^2 \quad \text{Equation A-14}$$
$$SST = SSE + SSR$$

Thus, the total sum of squares (SST) equals the sum of squares due to error (SSE) plus the sum of squares due to regression (SSR).

A.3.3.1 CALCULATIONS OF R² AND ADJUSTED R²

The R^2 term is a measure of how well the regression line explains the variation in *Y*, or:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$
$$R = \sqrt{\frac{\text{variation explained by regression}}{\text{total variation in Y}}} \quad \text{Equation A-15}$$

where R^2 is called the *coefficient of multiple determination* and R is called the *multiple correlation coefficient*. If $R^2 = 0.90$ for a certain linear model, we could conclude that the input variables (X_1, X_2,...X_k) explain 90% of the variation in the output variable (*Y*). R^2 reduces to the *coefficient of determination r^2* for simple linear regression when one independent variable (*X*) is in the regression model. The *sample correlation coefficient*, *r*, is a measure of the association between *X* and Y, and calculated by Equation A-16. It is also referred to as the Pearson product moment correlation coefficient.

$$r = \frac{\displaystyle\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\left[\displaystyle\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2\right]^{0.5}} \qquad \text{Equation A-16}$$

In addition, r is an estimate of the unknown population parameter, ρ, defined by Equation A-17:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad \text{Equation A-17}$$

where σ_X and σ_Y denote the population standard deviations of the random variables X and Y, and where σ_{XY} is called the covariance between X and Y. The covariance σ_{XY} is a population parameter describing the average amount that two variables "covary". Thus, another way of thinking about a correlation coefficient (R) is that it reflects the ratio of the covariance between two variables divided by the product of their respective standard deviations; and the value always lies between -1 and +1. *@Risk* and *Crystal Ball*® provide both the R^2 for the entire model, as well as the correlation coefficients for each input variable (or regressor). The higher the value of R_i for X_i, the more sensitive the output variable is to that input variable.

Although the calculations are the same, there is a subtle conceptual difference between the coefficient of determination (r^2) from regression, and the square of the correlation coefficient. When evaluating two variables *(X, Y)*, the key is whether X is interpreted as a "fixed" quantity (i.e., an explanatory variable), or a random variable just like Y. In regression analysis, r^2 measures how well the regression line explains the variation in Y given a particular value for X (Equation A-15). Correlation requires that X be considered a random variable, typically having a bivariate normal distribution with Y (see Appendix B).

One artifact of regression analysis is that R^2 increases as you add more and more input variables to your model; however, the increased fit of the model due to one or more of the input variables may be insignificant. Sometimes an adjusted R^2 is calculated to take into account the number of input variables (called regressors) in the model (k) as well as the number of observations in the data set (n):

$$R^2_{adj} = \frac{(n-1)R^2 - 1}{n - k - 1} \qquad \text{Equation A-18}$$

While R^2 gives the proportion of the total *variation* of Y that is explained, R^2_{adj} (Equation A-18) takes into account the degrees of freedom (df), and gives the proportion of the total *variance* of Y that is explained (variance = variation /df); or stated simply, R^2_{adj} is the R^2 corrected for df, where df is described by [1 - k/(n-1)].

- If the relationship between an input variable and an output variable is strong, but nonlinear, the R^2 statistic will be misleadingly low.
- If the means of the sampling data are used rather than the individual observations for each variable, R^2 will be misleadingly high. This is because taking the mean of a sample reduces the fraction of the

total variation due to *random* variation (see discussion of random error above). This is an important consideration when trying to interpret the results of regression analyses that incorporate data averaged over different spatial scales (e.g., regression of PbB on soil lead concentrations taken at the city block level may give an inflated R^2 value if the sampling data are averaged over a larger spatial scale, such as the census tract level).

A multiple regression analysis can also be performed to estimate the **regression coefficients** (see Appendix A.3.3). Each coefficient essentially represents an "average" value of the partial derivative across the entire distribution of the input. The regression coefficient, like the partial derivative, depends on the units of measurement so, as in the case of the partial derivative, it must be normalized. This can be accomplished by multiplying the regression coefficient by the ratio of estimated standard deviations s_y/s_x.

A convenient way to carry out a sensitivity analysis is to perform a stepwise regression analysis. Some statistical software packages (e.g., SAS, SPSS) offer a variety of different approaches for this; however, in general, they can be classified into two general categories: forward selection and backward elimination. In the forward selection, the inputs are added to the model one by one in the order of their contribution. In the backward elimination, all of the inputs are used in the model initially and then they are dropped one by one, eliminating the least important input at each step. A true stepwise procedure is a variation on the forward selection approach where an input can drop out again once it has been selected into the model if at some point other inputs enter the model that account for the same information.

A.3.3.2 RELATIVE PARTIAL SUM OF SQUARES (RPSS)

The **relative partial sum of squares (RPSS)** measures the sensitivity of the model output to each of the input variables by partitioning the variance in the output attributable to each variable using multiple regression techniques (Rose et al., 1991). The RPSS is presented as a percentage reflecting the proportion of influence a given variable has on risk. The results of RPSS are intuitive and generally easy to understand.

Briefly, the RPSS represents the percentage of the total sum of squares attributable to each of the variables. To calculate RPSS for variable V_i, the difference between the regression sum of squares (RSS) for the full model and the regression sum of squares for the model with V_i missing (RSS$_{-i}$) is divided by the total sum of squares (TSS) and expressed as a percentage:

$$RPSS_i = \frac{100\,(RSS - RSS_{-1})}{TSS} \qquad \text{Equation A-19}$$

This procedure can be thought of as analogous to least squares linear regression, but performed in the *n*-dimensional parameter space of the risk equation. Since this approach depends on the adequacy of the linear regression model between the output variable (e.g., risk) and all the variables, an additional diagnostic is to check how close R^2 is to 1.0. For equations with more than three parameters (such as those used in Superfund risk assessments), the computational overhead of this process is large and requires specific computer programs. The software program *Crystal Ball*® does not perform this calculation, but it can be determined with most standard statistical software packages that perform multiple regression. *@Risk* performs a calculation similar to this called multivariate stepwise regression that yields correlation coefficients in lieu of percent contributions to output variance.

A.3.3.3 SPEARMAN'S RANK CORRELATION COEFFICIENT (RHO)

The validity of using indices such as regression coefficients, correlation coefficients, and partial correlation coefficients depends on the assumptions of the underlying <u>linear</u> model being met. If there is any doubt that a data set satisfies the model assumptions, a nonparametric measure of correlation based on the rank orders of the inputs and associated outputs can be used. The Spearman Rank correlation coefficient is a nonparametric statistic; it measures an association between variables that are either count data or data measured on an ordinal scale, as opposed to data measured on an interval or ratio scale. An example of an ordinal scale would be the ranking of sites based on their relative mean soil concentrations. For example, if there are four categories of soil contaminant concentrations, sites with the highest concentrations may receive a rank of 1 while sites with lowest concentrations may receive a rank of 4. Ordinal scales indicate relative positions in an ordered series, not "how much" of a difference exists between successive positions on a scale.

To calculate the Spearman rank correlation coefficient, assign a rank to each of the input variables (X_j) and output variables (Y_k). For each ranked pair (X_j, Y_k), calculate the difference, d, between the ranks. For example, if the first observation for variable X has a ranking of 5 (relative to all of the observations of X), and the corresponding value of Y has a ranking of 3 (relative to all of the observations of Y), the difference (d) is equal to 5–3=2. Spearman rho (r_s) is calculated as:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{(n^3 - n)} \qquad \text{Equation A-20}$$

Hence ($-1 \leq r_s \leq 1.0$), and r_s=-1 describes a perfect indirect or negative relationship between ranks in the sense that if an X element increases, the corresponding Y element decreases. Similarly, r_s=0 suggests that there is no relationship between X and Y.

The Pearson product moment correlation coefficient is equal to the Spearman rank correlation coefficient when interval/ratio values of the measured observations (X, Y) are replaced with their respective ranks.

REFERENCES FOR APPENDIX A

Cullen, A.C. and H.C. Frey. 1999. Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press.

Hamby, D.M. 1994. A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environ. Monit. and Assess.* 32:135–154.

Helsel, D.R. and R.M. Hirsch. 1992. Statistical Methods in Water Resources. Elsevier Science B.V.

Iman, R.L. and J.C. Helton. 1988. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Anal.* 8:71–90.

Iman, R.L. and J.C. Helton. 1991. The Repeatability of Uncertainty and Sensitivity Analyses for Complex Probabilistic Risk Assessments. *Risk Anal.* 11:591–606.

Merz, J., M.J. Small, and P. Fischbeck. 1992. Measuring Decision Sensitivity: A Combined Monte Carlo-Logistic Regression Approach. *Medical Decision Making*, 12: 189–196.

Morgan, M.G. and M. Henrion.1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press.

Palisade Corporation. 1994. *Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3.* Windows Version Release 3.0 User's Guide, Palisade Corporation, Newfield, NY.

Rose, K.A., E.P. Smith, R.H. Gardner, A.L. Brenkert, and S.M. Bartell. 1991. Parameter Sensitivities, Monte Carlo Filtering, and Model Forecasting Under Uncertainty. *J. Forecast* 10:117–133.

Saltelli, A and J. Marivort. 1990. Non-Parametric Statistics in Sensitivity analysis for Model Output: A Comparison of Selected Techniques. *Reliab. Engin. Syst. Saf.* 28:299–253.

Shevenell, L. and F.O. Hoffman. 1993. Necessity of Uncertainty Analyses in Risk Assessment. *J Hazard Mater*. 35:369–385.

Stern, A.H. 1994. Derivation of a Target Level of Lead in Soil at Residential Sites Corresponding to a *de minimis* Contribution to Blood Lead Concentration. *Risk Anal.* 14:1049–1056.

U.S. EPA. 1997. *Guiding Principles for Monte Carlo Analysis*. Risk Assessment Forum and National Center for Environmental Assessment. EPA/630/R-97/001.

U.S. EPA. 1999. *TRIM, Total Risk Integrated Methodology, TRIM FATE Technical Support Document Volume I: Description of Module*. Office of Air Quality Planning and Standards. EPA/43/D-99/002A.

APPENDIX B

SELECTION AND FITTING OF DISTRIBUTIONS

B.0   INTRODUCTION

An important step in Monte Carlo analysis (MCA) is to select the most appropriate distributions to represent the factors that have a strong influence on the risk estimates. This step in the development of a Monte Carlo model can be very challenging and resource intensive.

> ☞ *Specifying probability distributions for all of the input variables and parameters in a probabilistic risk assessment (PRA) will generally not be necessary.*

If the sensitivity analysis indicates that a particular input variable does not contribute significantly to the overall variability and uncertainty, then this variable may be represented as a point estimate. As discussed in Appendix A, however, different approaches to sensitivity analysis may be applied throughout the tiered approach (e.g., sensitivity ratios, correlation analysis), and the ability to reliably identify variables as being minor or major can vary. Sometimes it can be helpful to develop probability distributions based on preliminary information that is available from Tier 1 in order to explore alternative options for characterizing variability and uncertainty. Likewise, sometimes the important "risk drivers" are apparent, and resources can be allocated to fully characterize the variability and uncertainty in those input variables. Therefore, the process of selecting and fitting distributions may also be viewed as a tiered approach. This appendix reviews the methods available to select and fit distributions and provides guidance on the process for determining appropriate choices depending on the information needed from the assessment and the information available to define the input variables.

In PRA, there are some important distinctions in the terminology used to describe probability distributions. A probability density function (PDF), sometimes referred to as a probability model, characterizes the probability of each value occurring from a range of possible values. Probability distributions may be used to characterize variability (PDFv) or uncertainty (PDFu). One advantage of using a PDFv and PDFu is that distributions represent a large set of data values in a compact way (Law and Kelton, 1991). For example, a lognormal distribution provides a good fit to a large data set of tap water ingestion rates (n=5,600) among children ages 1 to 11 years (Roseberry and Burmaster, 1992). Therefore, the distribution type (lognormal) and associated parameters (mean and standard deviation) fully describes the PDFv for intake rates, from which other statistics of interest can be calculated (e.g., median, and 95th percentile). Reducing a complex exposure model to a series of representative and well-fitting distributions can facilitate both the quantitative analysis and the communication of the modeling methodology. Alternatively, a PDFu may be specified to characterize parameter uncertainty. For example, the sample mean ($\bar{x}$) is generally an uncertain estimate of the population mean (μ) due to measurement error, small sample sizes, and other issues regarding representativeness (see Section B.3.1). A PDFu can be used to represent the distribution of possible values for the true, but unknown parameter. Understanding whether uncertainty or variability is being represented by a PDF is critical to determining how the distribution and parameters should be specified and used in a PRA.

EXHIBIT B-1

DEFINITIONS FOR APPENDIX B

Bayesian Analysis - Statistical analysis that describes the probability of an event as the degree of belief or confidence that a person has, given some state of knowledge, that the event will occur. Bayesian Monte Carlo combines a prior probability distribution and a likelihood function to yield a posterior distribution (see Appendix D for examples). Also called subjective view of probability, in contrast to the frequentist view of probability.

Bin - Regarding a histogram or frequency distribution, an interval within the range of a random variable for which a count (or percentage) of the observations is made. The number of bins for a histogram is determined on a case-by-case basis. In general, equal interval widths are used for each bin; however, in some cases (e.g., Chi-square test), individual bin widths are calculated so as to divide the distribution into intervals of equal probability.

Countably Infinite - Used to describe some discrete random variables, this term refers to a set of numbers that can be counted with integers (e.g., one, two, three) and that has no upper limit. Examples include the number of tosses required for a coin to show a head—we can count each toss, but it is possible that at least one more toss is needed. The number of dust particles in a volume of air is another example. Countably finite implies there is an upper limit (e.g., days of work per year).

Cumulative Distribution Function (CDF) - Obtained by integrating the PDF, gives the cumulative probability of occurrence for a random independent variable. Each value c of the function is the probability that a random observation x will be less than or equal to c.

Empirical Distribution Function (EDF) -The EDF, also called the empirical CDF (ECDF), is based on the frequency distribution of observed values for a random variable. It is a stepwise distribution function calculated directly from the sample, in which each data point is assigned an equal probability.

Frequency Distribution or Histogram - A graphic (plot) summarizing the frequency of the values observed or measured from a population. It conveys the range of values and the count (or proportion of the sample) that was observed across that range.

Goodness-of-Fit (GoF) Test - A method for examining how well (or poorly) a sample of data can be described by a hypothesized probability distribution for the population. Generally involves an hypothesis test in which the null hypothesis H_0 is that a random variable X follows a specific probability distribution F_0. That is, $H_0: F=F_0$ and $H_a: F \neq F_0$.

Independence - Two events A and B are independent if whether or not A occurs does not change the probability that B occurs. Likewise, knowing the value of B does not affect the value of A. Input variables, X and Y, are independent if the probability of any paired values (X, Y) is equal to the probability of X multiplied by the probability of Y. In mathematical terms, X and Y are independent if f(X, Y)=f(X) x f(Y). Independence is not synonymous with correlation. If X and Y are independent, then their correlation is zero, Cor(X, Y)= 0. But, the converse is not always true. There may be a nonlinear relationship between X and Y that yields Cor(X, Y)=0, but the variables are highly dependent.

Nonparametric Method - Also called a *distribution-free* method, a procedure for making statistical inferences without assuming that the population distribution fits a theoretical distribution such as normal or lognormal. Common examples are the Spearman rank correlation, (see Appendix A) and the bootstrap-t approach..

Parameter - In PRA, a parameter is a quantity that characterizes the probability distribution of a random variable. For example, a normal probability distribution may be defined by two parameters (e.g., arithmetic mean and standard deviation).

Parametric Distribution - A theoretical distribution specified by a distribution type and one or more parameters. Examples include the normal, Poisson, and beta distributions.

EXHIBIT B-1 —*Continued*
DEFINITIONS FOR APPENDIX B

<u>Probability Density Function</u> (PDF) **-** A function representing the probability distribution of a continuous random variable. The density at a point refers to the probability that the variable will have a value in a narrow range about that point.

<u>Probability Distribution</u> - The mathematical description of a function that associates probabilities with specified intervals or values for a random variable. A probability distribution can be displayed in a graph (e.g., PDF or CDF), summarized in a table that gives the distribution name and parameters, or expressed as a mathematical equation. In PRA, the process of selecting or fitting a distribution that characterizes variability or uncertainty can also be referred to as applying a *probability model* to characterize variability or uncertainty. In this guidance, the probability model is considered to be one source of model uncertainty.

<u>Step Function</u> - A mathematical function that remains constant within an interval, but may change in value from one interval to the next. Cumulative distribution functions for discrete random variables are step functions.

<u>Z-score</u> - The value of a normally distributed random variable that has been standardized to have a mean of zero and a SD of one by the transformation $Z=(X–\mu)/\sigma$. Statistical tables typically give the area to the left of the z-score value. For example, the area to the left of z=1.645 is 0.95. Z-scores indicate the direction (+/-) and number of standard deviations away from the mean that a particular datum lies assuming X is normally distributed. Microsoft Excel's *NORMSDIST(z)* function gives the probability p such that p=Pr(Z $\leq$ z), while the *NORMSINV(p)* function gives the z-score z_p associated with probability p such that p=Pr(Z $\leq$ z$_p$).

B.1.0 CONCEPTUAL APPROACH FOR INCORPORATING A PROBABILITY DISTRIBUTION IN A PRA

Often, more than one probability distribution may appear to be suitable for characterizing a random variable. A step-wise, tiered approach is recommended for incorporating probability distributions in a PRA. This appendix provides guidance on selecting and fitting distributions for variability and parameter uncertainty based on the overall strategy given in Exhibit B-2. Many of the same principles of selecting and fitting distributions are also given in EPA's *Report of the Workshop on Selecting Input Distributions for Probabilistic Assessments* (U.S. EPA, 1999a).

EXHIBIT B-2

GENERAL STRATEGY FOR SELECTING AND FITTING DISTRIBUTIONS

(1)   Hypothesize a family of distributions
(2)   Assess quality of fit of distribution
(3)   Estimate distribution parameters
(4)   Assess quality of fit of parameters

Probability distributions may be developed to characterize variability or uncertainty. Example flow charts for specifying a PDFv and PDFu are given in Figures B-1 and B-2, respectively. Both approaches outline an iterative process that involves three general activities: (1) identify potentially important sources of variability or uncertainty to determine if a PDF may be needed; (2) apply the general strategy given in Exhibit B-1 and evaluate plausible alternatives for distributions and parameter estimates; and (3) document the decision process. The flowcharts provide a general outline of the process and contain terms which are explained in subsequent sections. Just as with the point estimate approach, different sites may require different probability distributions for input variables, depending on the unique risk management issues and sources of uncertainty.

B.2.0   PRELIMINARY SENSITIVITY ANALYSIS

Selecting and fitting probability distributions for *all* of the input variables can be resource intensive and is generally unnecessary. Ideally, a subset of variables could be identified that contribute to most of the variability and uncertainty in a risk estimate. Sensitivity analysis can play an important role in helping to identify and quantitatively rank the major exposure pathways and variables. Since the information obtained from a sensitivity analysis may vary, depending on the approach(es) used and the information available to characterize the input variables, risk assessors should understand inherent limitations of each approach. A variety of approaches that are common for Tier 1 and 2 analyses are described and applied to a hypothetical example in Appendix A.

In a Tier 1 assessment, sensitivity analysis is typically limited to exploring the effect of alternative point estimates on the risk estimate. These methods can be helpful if additional information regarding the variability in the input variables is incorporated into the analysis (i.e., sensitivity scores). Alternatively, a reasonable approach is to specify preliminary probability distributions for one or more inputs in order to maximize the advantages of probabilistic methods. The difference between a preliminary distribution and a subsequent distribution reflects the level of effort invested in characterizing variability and uncertainty. If a robust data set is available in Tier 1 to define point estimates, then a preliminary distribution may, in fact, fully characterize variability with very high confidence. For other variables, summary statistics, rather than sample data, may be available, allowing for estimates of central tendency or plausible ranges. The use of preliminary distributions reflects an effort to employ more robust sensitivity analysis techniques without expending the effort and resources that might otherwise be applied to a PRA in Tier 2. The goal of the preliminary analysis would not be necessarily to evaluate risks and/or develop a PRG; rather, the focus would be on identifying input variables that may be important to explore more fully. Preliminary sensitivity analysis can provide insight into the importance of selecting among alternative probability distributions and exposure scenarios.

One-dimensional Monte Carlo simulations with preliminary (or screening-level) distributions can be run prior to engaging in a more involved process of selecting and fitting distributions. The distributions can be selected based on knowledge regarding the mechanisms that result in variability, and information already available for determining point estimates (e.g., summary statistics, U.S. EPA guidance, etc.). Table B-1 provides examples of preliminary distributions that might be selected based on the type of information available, sometimes referred to as the *state of knowledge*. In many cases, the distribution is intended to estimate the plausible bounds of a variable, while requiring no additional data collection effort. For example, given estimates of a lower bound [min], upper bound [max], and the assumption that each value is equally likely, a uniform distribution would be used to represent variability (or parameter uncertainty). If no mechanistic basis for selecting a distribution exists, then the preliminary distribution would be chosen based on the available information. For example, given the estimates of the arithmetic mean [μ] and a percentile value [a] for a random variable, an exponential distribution might be recommended with $\lambda = 1/\mu$.

Guidance on matching the choice of the distribution to the state of knowledge is extended to a more diverse array of scenarios later in this appendix (see Table B-4).

Table B-1. Examples of Preliminary Distributions Based on Information Available[1], [2]

Information / Constraints	Distribution Shape
[a, b]	uniform
[a, m, b]	triangular
[a, b, α_1, α_2, β]	beta
[μ, σ]	normal
γ	exponential
[a, b, μ, σ]	Johnson Sb, Lognormal
[α, β]	gamma

a=minimum, b=maximum, m=mode, α=shape parameter, μ=mean,
σ=standard deviation, γ=average rate of occurrence of events, β=scale,

It may be informative to explore alternative choices for distributions applied to the same variable. For example, a simple yet informative approach is to run two 1-D MCA simulations for variability with an input variable characterized first by a Johnson Sb (i.e., a four-parameter lognormal distribution; Hahn and Shapiro, 1967) and then by a normal distribution. The difference in the risk distribution, especially at the percentile that is relevant to the risk management decision (e.g., 95th percentile), may offer insights regarding the importance of the shape of the PDFv.

B.3.0 WHAT DOES THE DISTRIBUTION REPRESENT?

Distributions may be specified to characterize variability or uncertainty. Often, a Monte Carlo simulation of variability will focus on describing differences between individuals in a population (i.e., inter-individual variability). In this case, the goal is to select a distribution that is representative of the *target* population—the set of all receptors that are potentially at risk. There may be uncertainty that the choice of PDFv reflects variability in the target population. In general, risk assessors should fully disclose uncertainties in the PDFv, especially because the use of a distribution instead of a point estimate may inappropriately suggest that there is a greater state of knowledge. Following the tiered process (see Chapter 2, Figure 2-1), there are multiple opportunities to consider consequences of alternative modeling approaches early in the process of developing a probabilistic model. The importance of relating the distribution to the *target* population, clearly distinguishing between variability and uncertainty, and evaluating data representativeness is emphasized in Sections B.3.1, B.3.2 and B.4.

---

[1]The preliminary distributions are based in part on maximum entropy concepts. Maximum entropy is a technique for determining the distribution that represents the maximum uncertainty allowed by the available information and data (Vose, 1996). Although the approach can be used to quickly define distributions that maximize uncertainty, the credibility of the distribution depends on the use of accurate, unbiased information.

[2]See Table B-2 for more detailed descriptions of selected distributions.

B.3.1   CONCEPTS OF POPULATION AND SAMPLING

The distinction between a *target* population, a *sampled* population, and a *statistical* population should be considered carefully when evaluating information for use in both Tier 1 and Tier 2 of a PRA. The *target* population is often considered to be the "population of concern". A risk assessor is often interested in quantifying specific attributes of the population (e.g., exposure duration, exposure frequency, etc.). A *sampled* population is the set of receptors available for selection and measurement. For purposes of this appendix/guidance, the *sampled* population may be the *target* population or it may be a different population that is thought to be representative of the *target* population. For purposes of this guidance, a *statistical* population is an approximation of the *target* population based on information obtained from the *sampled* population.

Distributions are generated from representative *sample* populations to make inferences about the *target* population. Ideally, a *sampled* population should be a subset of a *target* population and should be selected for measurement to provide accurate and representative information about the exposure factor being studied. However, defining representative samples is a matter of interpretation.
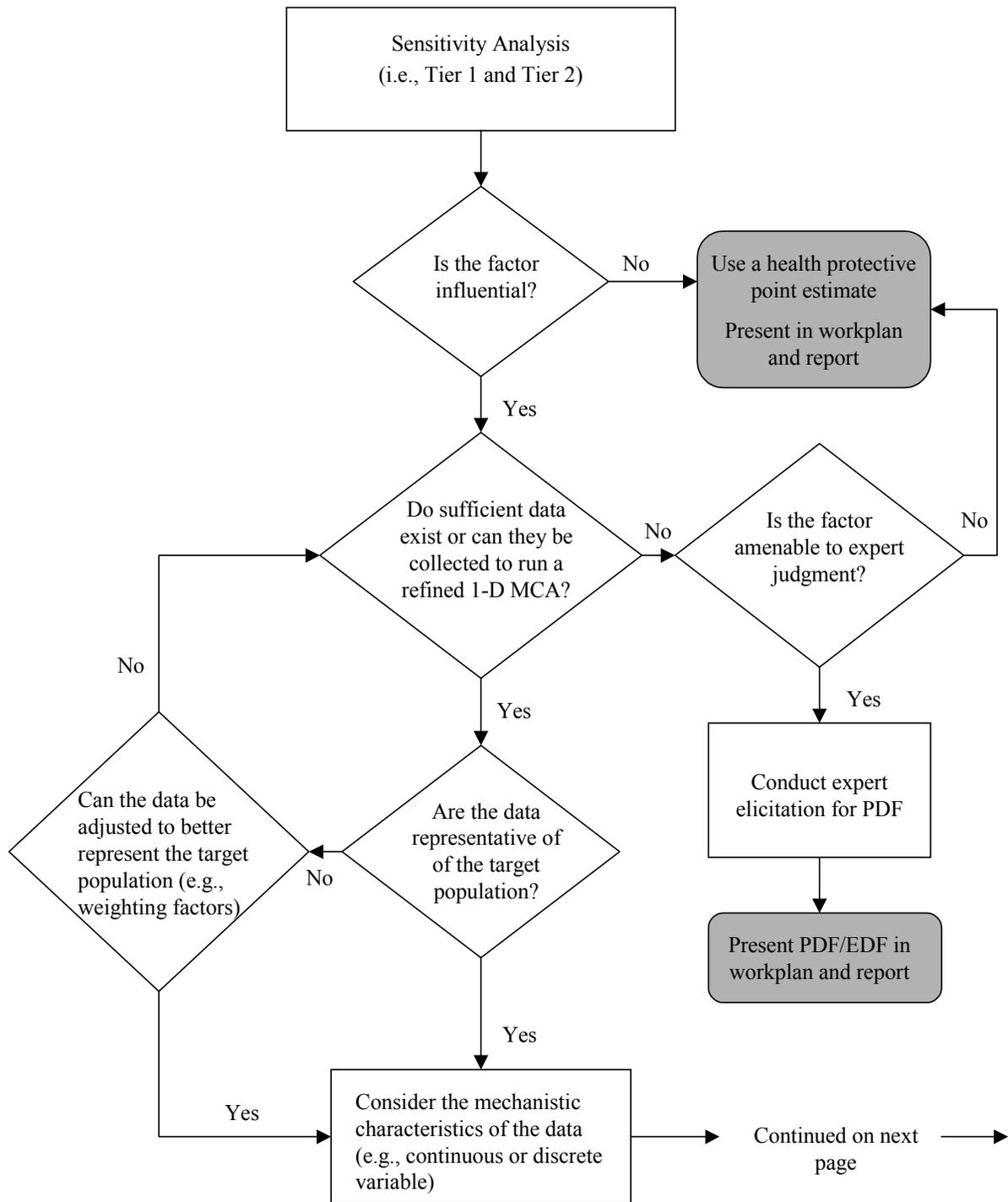
Figure B-1 (page 1 of 2). Conceptual approach for incorporating probability distributions for variability in PRA.
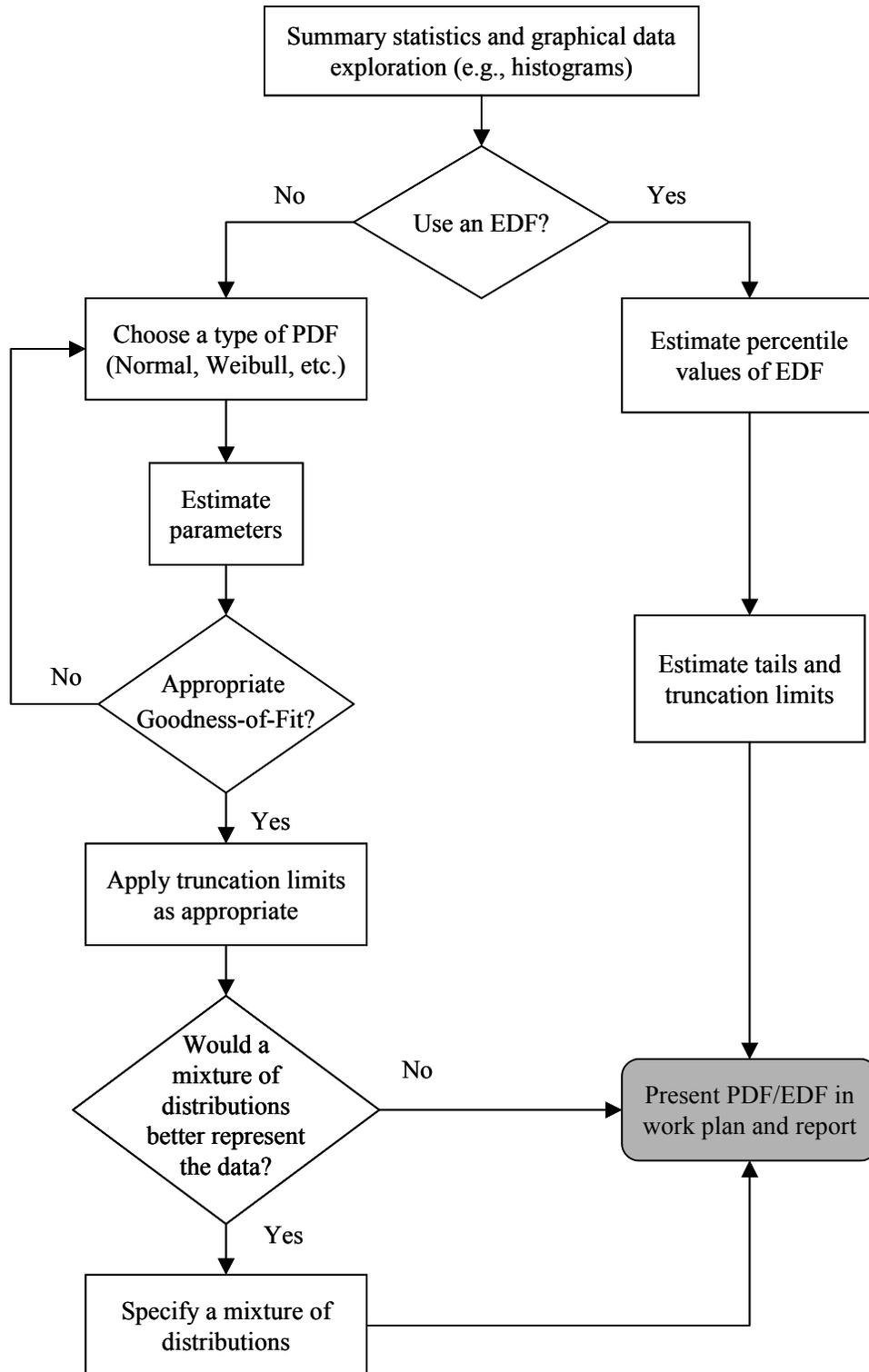
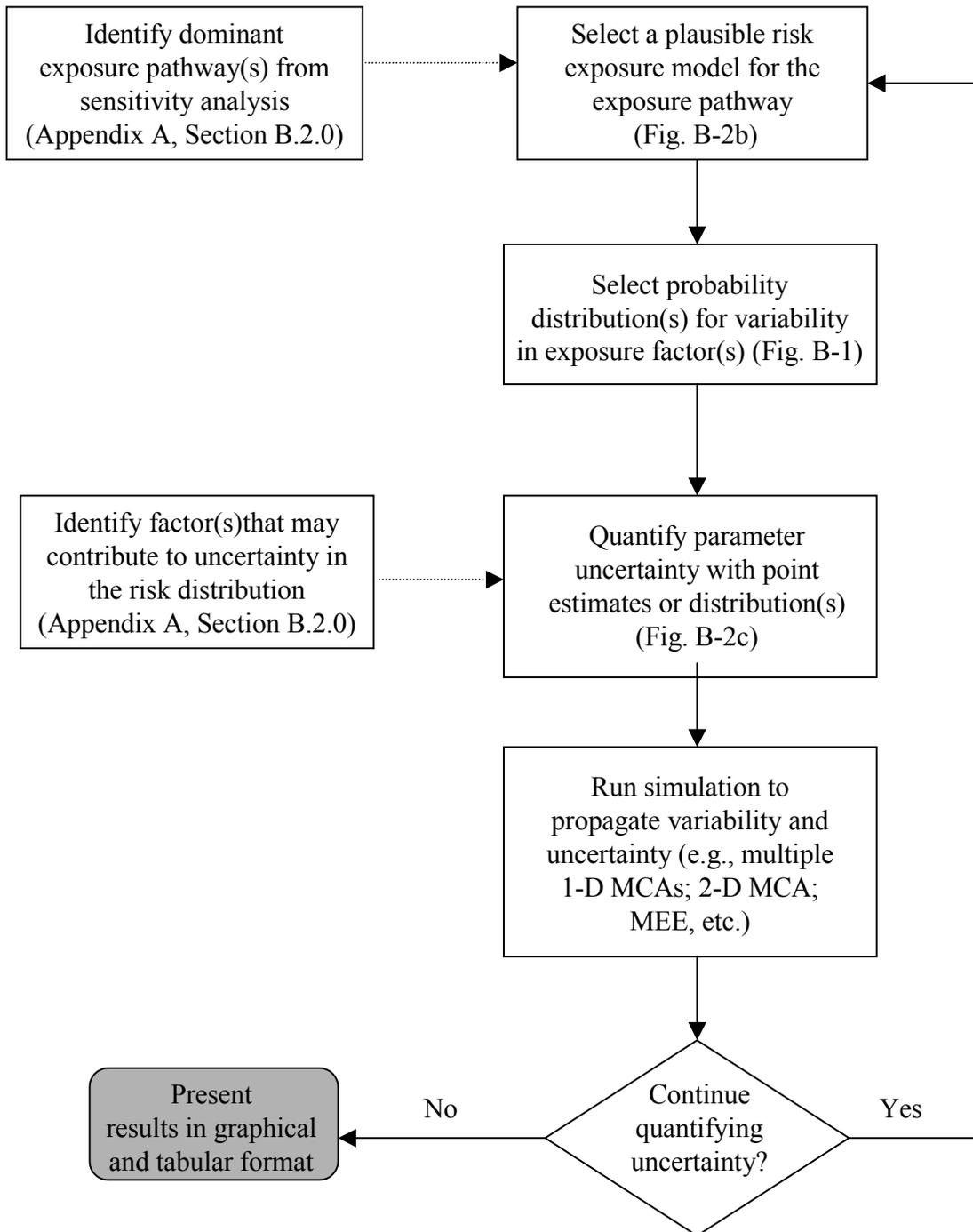(Continued from previous page)

```
                    ┌─────────────────────────────┐
                    │ Summary statistics and      │
                    │ graphical data exploration  │
                    │ (e.g., histograms)          │
                    └─────────────────────────────┘
                                  │
                                  ▼
              No          ◇ Use an EDF? ◇          Yes
        ┌────────────────                ────────────────┐
        ▼                                                ▼
┌──────────────────────┐                    ┌──────────────────────┐
│ Choose a type of PDF │                    │ Estimate percentile  │
│ (Normal, Weibull,    │                    │ values of EDF        │
│ etc.)                │                    └──────────────────────┘
└──────────────────────┘                              │
        │                                             ▼
        ▼                                   ┌──────────────────────┐
┌──────────────────────┐                    │ Estimate tails and   │
│ Estimate parameters  │                    │ truncation limits    │
└──────────────────────┘                    └──────────────────────┘
        │
        ▼
  ◇ Appropriate
    Goodness-of-Fit? ◇ ── No ──┐ (loops back to Choose a type of PDF)
        │
       Yes
        ▼
┌──────────────────────┐
│ Apply truncation     │
│ limits as appropriate│
└──────────────────────┘
        │
        ▼
  ◇ Would a mixture
    of distributions
    better represent
    the data? ◇ ─── No ───▶ ┌──────────────────────┐
        │                   │ Present PDF/EDF in   │
       Yes                  │ work plan and report │
        ▼                   └──────────────────────┘
┌──────────────────────┐
│ Specify a mixture of │
│ distributions        │
└──────────────────────┘
```

Figure B-1 (page 2 of 2). Conceptual approach for incorporating probability distributions for variability in PRA.

```
┌─────────────────────┐                    ┌─────────────────────┐
│  Identify dominant  │                    │  Select a plausible │
│ exposure pathway(s) │ ·················▶  │  risk exposure model│ ◀──┐
│ from sensitivity    │                    │  for the exposure   │    │
│ analysis            │                    │  pathway            │    │
│ (Appendix A,        │                    │  (Fig. B-2b)        │    │
│ Section B.2.0)      │                    └─────────────────────┘    │
└─────────────────────┘                              │               │
                                                     ▼               │
                                        ┌─────────────────────┐      │
                                        │  Select probability │      │
                                        │  distribution(s) for│      │
                                        │  variability in     │      │
                                        │  exposure factor(s) │      │
                                        │  (Fig. B-1)         │      │
                                        └─────────────────────┘      │
                                                     │               │
                                                     ▼               │
┌─────────────────────┐                 ┌─────────────────────┐      │
│ Identify factor(s)  │                 │  Quantify parameter │      │
│ that may contribute │ ·············▶  │  uncertainty with   │      │
│ to uncertainty in   │                 │  point estimates or │      │
│ the risk            │                 │  distribution(s)    │      │
│ distribution        │                 │  (Fig. B-2c)        │      │
│ (Appendix A,        │                 └─────────────────────┘      │
│ Section B.2.0)      │                           │                  │
└─────────────────────┘                           ▼                  │
                                       ┌─────────────────────┐       │
                                       │  Run simulation to  │       │
                                       │  propagate          │       │
                                       │  variability and    │       │
                                       │  uncertainty (e.g., │       │
                                       │  multiple 1-D MCAs; │       │
                                       │  2-D MCA; MEE, etc.)│       │
                                       └─────────────────────┘       │
                                                 │                   │
                                                 ▼                   │
┌─────────────────┐                          ◇─────────◇            │
│   Present       │         No             ╱  Continue   ╲   Yes    │
│ results in      │ ◀─────────────────────╱  quantifying  ╲────────┘
│ graphical and   │                       ╲  uncertainty? ╱
│ tabular format  │                        ╲─────────────╱
└─────────────────┘                          ◇─────────◇
```

Figure B-2a (page 1 of 3). Conceptual approach for quantifying model and parameter uncertainty in PRA.

Identify dominant
exposure pathway(s)
from
Sensitivity Analysis
(Appendix A and Section
B.2.0)

Is more than
one model
plausible for an
exposure pathway?

No

Yes

Identify and evaluate the exposure
factor(s) quantified by a
candidate exposure model

Purpose and Objectives
• regulatory context
• scientific questions addressed
• application niche (physical,
  chemical, biological system)
• status of agency and/or peer review

Defining and Limiting Components
• process(es) characterized
  (e.g., transport, diffusion,
  volatilization, bioavailability, etc.)
• temporal and spatial scales
• level of aggregation/simplification

Theoretical Basis
• mechanistic basis for algorithms
  • numerical or analytic solution

Select alternative
exposure model

No

Is the
exposure model
appropriate?

Yes

Run simulation
with candidate
exposure model

Figure B-2b (page 2 of 3). Detailed conceptual approach for incorporating model uncertainty in PRA.

Identify candidate probability distribution(s) for variability (Fig. B-1):
• mechanistic basis for variability
• exploratory data analysis
• expert judgment

Estimate parameters (e.g., MLE, method of moments, etc.)

Is information available to quantify parameter uncertainty?

No → Is the parameter amenable to expert elicitation?

No → Continue with process acknowledging limits of data

Yes

Yes → Conduct expert elicitation

Select distribution (or point estimate) for uncertainty

Run simulation to propagate variability and uncertainty

Run sensitivity analysis to identify important sources of uncertainty

Should an alternative probability model (i.e., PDFv or PDFu) be explored?

Yes

No → Present results in graphical and tabular format

Figure B-2c (page 3 of 3). Detailed conceptual approach for incorporating parameter uncertainty in PRA.

B.3.2 CONSIDERING VARIABILITY AND UNCERTAINTY IN SELECTING AND FITTING DISTRIBUTIONS

Multiple probability distributions may be used to describe variability and uncertainty in an input variable. For example, a normal probability distribution may be selected to characterize variability in body weight, whereas a uniform distribution may selected to characterize uncertainty in the estimate of the arithmetic mean of the normal distribution. The appropriate interpretation and analysis of data for an exposure variable will depend on whether one is specifying a PDFv or PDFu. Figure B-1 outlines one useful process for selecting distributions for variability, whereas Figure B-2 (three pages) outlines a useful process for quantifying both model and parameter uncertainty.

Variability generally refers to observed differences attributable to true heterogeneity or diversity in a population (U.S. EPA, 1997b). Variability results from natural random processes. Inter-individual variability may stem from environmental, lifestyle, and genetic differences. Examples include human physiological variation (e.g., natural variation in body weight, height, breathing rates, drinking water intake rates), changes in weather, variation in soil types, and differences in contaminant concentrations in the environment. Intra-individual variability may reflect age-specific changes (e.g., body weight and height). Variability is not reducible by further measurement or study. A PDF for variability can usually be obtained by fitting a distribution to the sample measurements.

Sources of Uncertainty

Uncertainty generally refers to the lack of knowledge about specific factors, parameters, or models (U.S. EPA, 1997b). Although uncertainty in exposure and risk assessment may be unavoidable due to the necessary simplification of real-world processes, it generally can be reduced by further measurement and study. Parameter uncertainty may stem in part from measurement errors, sampling errors, or other systematic errors in the collection and aggregation of data. Model uncertainty may reflect the simplification of a complex process, a mis-specification of the exposure model structure, a misuse or misapplication of an exposure model, use of the wrong distributional model, and the use of surrogate data or variables. Scenario uncertainty may reflect uncertainty in an exposure model, such as the relevance of specific exposure pathways to the target population. A conceptual exposure model can be used to provide direction in specifying a probability distribution for uncertainty. For example, the concentration term in a Superfund risk assessment typically represents the long-term average concentration to which a receptor is exposed (see Chapter 5). An uncertainty distribution for the concentration term could be developed in part from ideas about the statistical uncertainty of estimating the long-term average from a small sample, and the assumption of random movement of the receptors within a defined exposure unit.

Probability Distributions and Model Uncertainty

This appendix primarily focuses on methods for quantifying uncertainty associated with both the selection of a variability distribution, and estimating parameters of a distribution. A probability distribution can be referred to as a type of model in the sense that it is an approximation, and often a simplified representation of variability or uncertainty that combines both data and judgment. A broader use of the term model refers to a representation of a chemical, physical, or biological process. In risk assessment, many different models have been developed, with varying objectives, major defining and limiting components, and theoretical basis. Figure B-2b provides a general process for exploring model uncertainty of this type. This figure reflects the concepts and spirit of the *Agency Guidance for Conducting External Peer Review of Environmental Regulatory Modeling* (U.S. EPA, 1994). In general, EPA regional risk assessors should be consulted in order to determine the types of exposure and risk models that may be plausible for quantifying exposure at a particular site.

Parameter Uncertainty

Quantifying parameter uncertainty in a probabilistic model typically requires judgment (see Appendix C). When data are uncertain due to, for example, small sample sizes or questionable representativeness (Section B.3.1), Monte Carlo simulation can be a useful tool for demonstrating the effect of the uncertainty on the risk estimates. It is most important to model uncertainty when the sensitive input variables are uncertain. Uncertainty can be quantified in both the point estimate approach (e.g., a range of possible central tendency exposure values) or a probabilistic approach (e.g., a range of possible values for the arithmetic mean of a distribution). While a quantitative uncertainty analysis may complicate a risk management decision by suggesting that risk estimates are highly uncertain, this information can be helpful by focusing additional efforts towards collecting data and reducing uncertainty in the most sensitive input variables. Likewise, if an estimated risk is below a regulatory level of concern, even after quantifying highly uncertain inputs to the exposure model, the risk manager may be more confident in a decision. As emphasized in Figures B-2a, B-2b, and B-2c, risk assessors should generally refrain from setting *ad hoc* probabilities to different candidate distributions in a single Monte Carlo simulation. Instead, this guidance strongly recommends exploring model or parameter uncertainty by running a separate simulation with each candidate model. For example, rather than randomly assigning a beta distribution or a lognormal distribution to an exposure variable for each iteration of a simulation, separate simulations should be run with the candidate probability distributions. Similarly, if a range of temporal or spatial scales is plausible for quantifying exposure, multiple simulations should be designed to demonstrate the importance of these assumptions on the risk estimates.

Uncertainty in parameter estimates may be characterized using a variety of methods. Similar to a PDF for variability, a PDF for parameter uncertainty may be represented by a probability distribution with a unique set of parameters. Sometimes the distribution for uncertainty can be specified by knowing (or assuming) a distribution for variability. For example, if X is a normally distributed random variable, the Student's t distribution and the Chi-square (χ^2) distribution can be used to develop PDFu's for random measurement error uncertainty in the sample mean and variance, respectively. The PDFu for both the Student's *t* and Chi-square distributions is determined by the sample size (*n*). If a PDFu cannot be determined from the PDF for variability, or assumptions regarding the underlying distribution for variability are not supportable, nonparametric or "distribution free" techniques may be used (e.g., bootstrapping). Both parametric and nonparametric techniques may yield confidence intervals for estimates of population parameters.

B.4.0   DO DATA EXIST TO SELECT DISTRIBUTIONS?

Developing site-specific PDFs for every exposure assumption (or toxicity value, in the case of ecological risk) can be time and resource intensive, and in many cases, may not add value to the risk management decision. For those exposure variables that do exert a significant influence on risk, a PDF may be developed from site-specific data, data sets available in the open literature (e.g., EPA's *Exposure Factors Handbook*, U.S. EPA 1997a), or from existing PDFs in the literature (e.g., Oregon DEQ, 1998).

At Superfund sites, perhaps the most common exposure variable that will be described by site-specific data will be the media concentration term. The sample (i.e., collection of empirical measurements) will most often be used to estimate either a point estimate of uncertainty (e.g., an upper confidence limit for the arithmetic mean concentration—the 95% UCL), or a distribution that characterizes the full distribution of uncertainty in the mean. Exposure variables such as ingestion rates, exposure duration, and exposure frequency will most likely be derived from existing PDFs or data sets in the open literature. The Agency supports the development PDFs that may be generally applicable to

different sites (e.g., body weight, water intake, and exposure duration) (U.S. EPA, 1999b, 2001). Until final recommendations of PDFs are available for the more generic exposure variables, PDFs for exposure variables that lack adequate site-specific data will typically be selected from: (1) existing PDFs; (2) data on the entire U.S. population; or (3) data on subsets of the U.S. population that most closely represent the target population at a site. If risks to a sensitive subpopulation, such as young children, elderly adults, ethnic groups, or subsistence fishermen, are a concern at a site, then existing PDFs or data sets that best characterize these subpopulations would be preferable to national distributions based on the entire U.S. population. If adequate site-specific data are available to characterize any of the exposure variables, distributions can be fit to those data.

Uncertainty Associated with Sample Size

An appropriate question to consider when evaluating data sets for use in exposure and risk assessment is, "What sample size is sufficient?" Generally, the larger the sample size (n), the greater one's confidence in the choice of a probability distribution and the corresponding parameter estimates. Conversely, for small n, Goodness-of-fit (GoF) tests (see Section B.6.2) will often fail to reject many of the hypothesized PDFs. In general, there is no rule of thumb for the minimum sample size needed to specify a distribution for variability or uncertainty. Increasing a sample size may be an appropriate option to consider when evaluating risk management strategies to reduce uncertainty.

Statistical sampling, in general, is important to consider when estimating parameters of a probability distribution. One rule of thumb is that the parameters that reflect the central tendency of a distribution (e.g., arithmetic mean, median, mode) can be estimated with greater confidence than parameters that reflect the extremes of the distribution (e.g., 95th percentile). When deciding on appropriate truncation limits (minimum and maximum values), it is unlikely that the statistical sample actually includes the plausible bounds. See Section B.5.7 for more detailed guidance on specifying truncation limits for probability distributions.

B.4.1 WHAT ARE REPRESENTATIVE DATA?

The question, "What is a representative sample?", is important to address when selecting and fitting distributions to data. Many of the factors that may determine representativeness (e.g., sample size and the method of selecting the target, and sample population (Section B.3.1)) are relevant to both point estimate and PRA. EPA's *Guidance for Data Usability in Risk Assessment, Part A* (U.S. EPA, 1992) describes representativeness for risk assessment as the extent to which data define the true risk to human health and the environment.

The goal of representativeness is easy to understand. However, evaluating data to determine if they are representative is more difficult, especially if the problem and decision objectives have not been clearly defined.

The importance of representativeness also varies with the level of complexity of the assessment. If a screening level assessment is desired, for example, to determine if concentrations exceed a health protective exposure level, then representativeness may not be as important as health protectiveness. However, if a complete baseline risk assessment is planned, the risk assessor should generally consider the value added by more complex analyses (e.g., site-specific data collection, sensitivity analysis, and exposure modeling). A tiered approach for making these decisions for a PRA is presented in Chapter 2, and examples of more complex analyses are presented in Appendix D. In addition, the Agency (U.S.

EPA, 1999a) summarizes the advantages and weaknesses of proposed checklists for risk assessors to evaluate representativeness of exposure factors data.

For purposes of this guidance, a surrogate study is one conducted on a sampled population that is similar to, but not a subset of, the target population. When using surrogate data, the risk assessor should generally exercise judgment about the representativeness of the data to the target population. For example, the distribution of body weights of deer mice from two independent samples from similar ecosystems may differ depending on the age structure, proportion of males and females, and the time of year that the samples were obtained. When in doubt about which study results to use in defining a probability distribution, one option is to develop a distribution and calculate risks with each sample independently, and compare the results. This approach can be a simple, but effective type of uncertainty analysis. At a minimum, uncertainties associated with the use of surrogate studies should be discussed in the assessment.

In many cases, the surrogate population shares common attributes with the target population, but is not truly representative. The risk assessor should then determine the importance of the discrepancies and whether adjustments can be made to reduce those differences. There are a wide variety of methods that can be used to account for such discrepancies, depending on the available information. Summary statistics (e.g., as presented by the *Exposure Factors Handbook*, U.S. EPA, 1997a) can be used to estimate linear characteristics of the target population from the sample population. For example, if the mean, standard deviation, and various percentiles of the sample population are known, then the mean or proportion exceeding a fixed threshold can be calculated using a simple weighted average. Adjustment options are more numerous if the risk assessor has access to the raw data. Adjustments for raw data include: weighted averages, weighted proportions, transformations, and grouping of the data based on the available information (e.g., empirical data, and professional judgment).

In most cases, the evaluation of data representativeness will necessarily involve judgment. The workplan should generally include a description of the data, the basis for the selection of each distribution, and the method used to estimate parameters (see Chapter 2). Empirical data (i.e., observations) are typically used to select distributions and derive parameter estimates. However, it may be necessary to use expert judgment or elicitation in cases where the quality or quantity of available data are found to be inadequate.

B.4.2   THE ROLE OF EXPERT JUDGMENT

Expert judgment refers to inferential opinion of a specialist or group of specialists within an area of their expertise. When there is uncertainty associated with an input variable, such as a data gap, expert judgment may be appropriate for obtaining distributions. Note that distributions elicited from experts reflect individual or group inferences, rather than empirical evidence. Distributions based on expert judgment can serve as Bayesian priors in a decision-analytic framework. The distributions and Bayesian priors can be modified as new empirical data become available. There is a rich literature base regarding the protocol for conducting expert elicitations and using the results to support decisions (Morgan and Henrion, 1990). Elicitation of expert judgment has been used to obtain distributions for risk assessments (Morgan and Henrion, 1990; Hora, 1992; U.S. EPA, 1997b) and for developing air quality standards (U.S. EPA, 1982).

Bayesian analysis is a statistical approach that allows the current state of knowledge, expressed as a probability distribution, to be formally combined with new data to reach an updated information state. In PRA, Bayesian Monte Carlo analysis (Bayesian MCA) can be used to determine the reduction in

uncertainty arising from new information. When combined with techniques from decision analysis, Bayesian MCA can help to determine the type and quantity of data that generally should be collected to reduce uncertainty. The benefits and limitations of expert elicitation, Bayesian statistics, Bayesian MCA, and decision analysis (i.e., value of information [VOI]), as applied to PRA, are discussed in greater detail in Appendix D.

B.5.0 FITTING DISTRIBUTIONS TO DATA

Sometimes more than one probability distribution may adequately characterize variability or uncertainty. The choice of a distribution should be based on the available data and on knowledge of the mechanisms or processes that result in variability. In general, the preferred choice is the simplest probability model that adequately characterizes variability or uncertainty and is consistent with the mechanism underlying the data. For example, a log-logistic distribution would not necessarily be selected over a 2-parameter lognormal distribution simply because it was ranked higher in a GoF test by a

> **EXHIBIT B-3**
>
> **FACTORS TO CONSIDER IN SELECTING A PROBABILITY DISTRIBUTION***
>
> - *Is there a mechanistic basis for choosing a distributional family?*
> - *Is the shape of the distribution likely to be dictated by physical or biological properties or other mechanisms?*
> - *Is the variable discrete or continuous?*
> - *What are the bounds of the variable?*
> - *Is the distribution skewed or symmetric?*
> - *If the distribution is thought to be skewed, in which direction?*
> - *What other aspects of the shape of the distribution are known?*
> - *How well do the tails of the distribution represent the observations?*
>
> *Source: U.S. EPA, 1997b

statistical software package. Some distributions (e.g., normal, lognormal) are well known among risk assessors. The statistical properties for these distributions are well understood and the formal descriptions can often be brief.

Important factors to consider in selecting a PDF are described in Exhibit B-3. An initial step in selecting a distribution should be to determine if the random variable is discrete or continuous. Continuous variables take any value over one or more intervals and generally represent measurements (e.g., height, weight, concentration). For a continuous variable, a mathematical function generally describes the probability for each value across an interval. Discrete variables take either a finite or *countably infinite* number of values. Unique probabilities are assigned to each value of a discrete variable. The number of rainfall events in a month is an example of a discrete random variable, whereas the amount of rainfall is a continuous variable. Similarly, the number of fish meals per month is a discrete variable, whereas the average size (mass) of a fish meal is continuous.

Another important consideration is whether there are plausible bounds or limits for a variable. For example, it is highly unlikely that an American adult will weigh less than 30 kg or more than 180 kg. Most exposure variables may assume any nonnegative value within a plausible range. Therefore, distributions will generally be truncated at a minimum of zero (or higher), or a probability distribution that is theoretically bounded at a nonzero value may be specified (see Table B-3). A more detailed discussion of factors to consider in selecting a PDF and specifying parameter values is provided below.

B.5.1   CONSIDERING THE UNDERLYING MECHANISM

There may be mechanistic reasons depending on known physical or biological processes that dictate the shape of the distribution. For example, normal distributions result from processes that sum random variables whereas lognormal distributions result from multiplication of random variables. A Poisson distribution is used to characterize the number of independent and randomly distributed events in a unit of time or space. An exponential distribution would describe the inter-arrival times of independent and randomly distributed events occurring at a constant rate. If, instead, the elapsed time until arrival of the kth event is of interest, then the appropriate probability distribution would be the gamma distribution (Morgan and Henrion, 1990).

> ☞ *In all cases, it is incumbent on the risk assessor to explain clearly and fully the reasoning underlying the choice of a distribution for a given exposure variable—primarily from a mechanistic standpoint if possible.*

Table B-2 lists some of the probability distributions that may commonly be used in PRA. This is not an exhaustive list, and the scientific literature contains numerous examples with alternative distributions. Where practicable, a mechanistic basis is presented for the choice of the distribution. For some distributions, such as beta, triangular, and uniform, a mechanistic basis is not offered because it is unlikely that a chemical or biological process will yield a random variable with that particular shape. Nevertheless, such distributions may be appropriate for use in PRA because they reflect the extent of information that is available to characterize a specific random variable. Preliminary distributions are discussed in Section B.2.0 and Table B-4. Because many of the distributions given in Table B-2 can assume flexible shapes, they offer practical choices for characterizing variability.

Table B-2 also illustrates probability distributions (both PDFs and CDFs) commonly used in PRA. While intuitively appealing, identifying a mechanistic basis for a distribution can be difficult for many exposure variables; however, it may be relatively apparent that the variable is bounded by a minimum (e.g., ingestion rate $\geq$ 0 mg/day) and a maximum (e.g., absorption fraction $\leq$ 100%), or that the relevant chance mechanism results in a discrete distribution rather than a continuous distribution, as described above.

For each distribution, one or more examples with different parameter estimates are given to demonstrate the flexibility in the shape of the PDF. In addition to the descriptions of the distributions in Tables B-2, Table B-3 provides a summary of the parameters and theoretical bounds that define the PDFs. For a further discussion of characteristics of PDFs see Thompson, 1999. Figures (a-h) immediately following Table B-2 present examples of PDFs and the corresponding CDFs for distributions commonly used in PRA.

Table B-2. Examples of Selected Probability Distributions for PRA.

Distribution	Mechanistic Basis	Example(s)
Beta Figure (e)	Describes a continuous random variable with finite upper and lower bounds. This distribution can take on very flexible shapes, but generally does not have a mechanistic basis.	Absorption fraction bounded by 0 and 100%; fraction of time an individual spends indoors.
Binomial	Describes a discrete random variable produced by processes that: (1) occur in a fixed number n of repeated independent "trials"; (2) yield only one of two possible outcomes (e.g., "success" or "failure") at each trial; and (3) have constant probability p of "success". A binomial distribution is characterized by parameters n, p, and x, representing the number of trials, the probability of success of each trial, and the number of successes, respectively.	The number of animals with tumors (or some other quantitative outcome) in a chronic animal bioassay.
Exponential Figure (h)	If instead of counting the number of events in the Poisson process (below), one measures the time (or distance) between any two successive, random, independent events.	The length of time between two radiation counts; length of time between major storm events; distance between impact points of two artillery shells.
Gamma Figure (g)	Similar to exponential except that time until occurrence of the kth event in the Poisson process is measured (rather than time between successive events). Reduces to exponential when k=1.	Time until k^{th} radiation count; elapsed time until k^{th} major storm event.
Lognormal Figure (b)	Multiplication of a large number of random variables, or equivalently adding the logarithms of those numbers, will tend to yield a distribution with a lognormal shape.	Chemical concentrations in environmental media; media contact rates; rates and flows in both fate and transport models. Because the basic risk equation is multiplicative, distributions of risk are generally lognormal. In practice, lognormal distributions often provide good fits to data on chemical concentrations in a variety of media (Gilbert, 1987; Ott, 1990).
Normal Figure (a)	Addition of independent random variables, with no one variable contributing substantially to the total variation of the sum, will tend to yield a distribution with a normal shape. This result is established by the central limit theorem.	The "Gaussian Plume Model" for the dispersion of air pollutants is based on the idea that, at a micro level, individual parcels of air, or molecules of pollutants, are subject to many random collisions from other molecules that act together as if a large number of random numbers were being added/subtracted from an initial 3-dimensional description of a position.
Poisson	Observed when counting the frequency of discrete events, where the events are independent of one another, and randomly distributed in space or time. Approximates the binomial distribution when sample size, n, is large and probability, p, is small.	The number of counts of radiation that occur in a particular time interval; the release of synaptic transmitter from nerve cells; the number of artillery shells falling within a fixed radius; the occurrence of major storm events in a month; number of leaks in average length of pipe.

Table B-2. Examples of Selected Probability Distributions for PRA.

Distribution	Mechanistic Basis	Example(s)
Triangular Figure ©)	The PDF is shaped like a triangle, with parameters representing plausible bounds and a most likely value (i.e., mode). This is a "rough" probability model that generally describes a random variable based on limited information rather than mechanistic basis.	Variability in shower droplet diameter. Uncertainty in the mean air exchange rate in a shower.
Uniform Figure (d)	The PDF is shaped like a rectangle, with parameters representing plausible bounds. This is a "rough" probability model that generally describes a random variable based on limited information rather than a mechanistic basis.	Variability in the air ventilation rate in a house.
Weibull Figure (f)	Originated in reliability and (product) life testing as a model for time to failure or life length of a component when the failure rate changes with time. A very flexible model taking a wide range of shapes. If the failure rate is constant with time, the Weibull reduces to the exponential distribution.	Examples for exponential and gamma would also be appropriate for Weibull.

Normal

(a1)

(a2)

Lognormal

(b1)

(b2)

Triangular

(c1)

(c2)

Uniform

(d1)
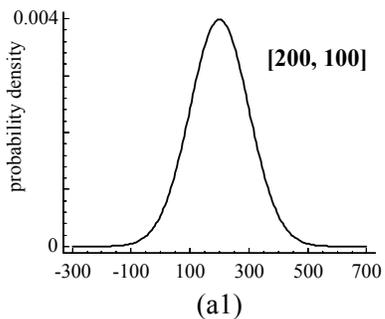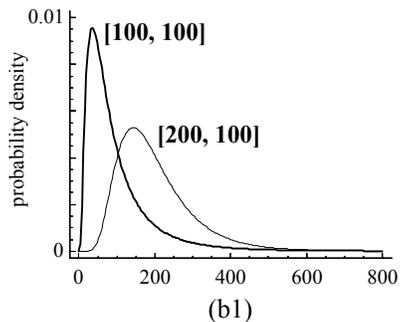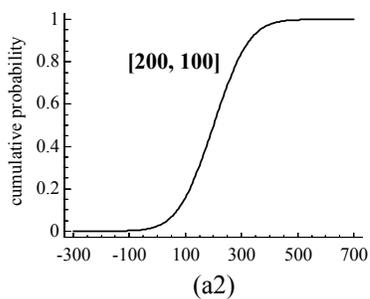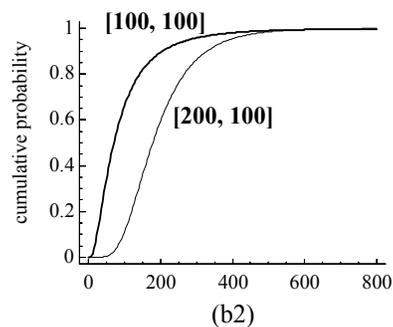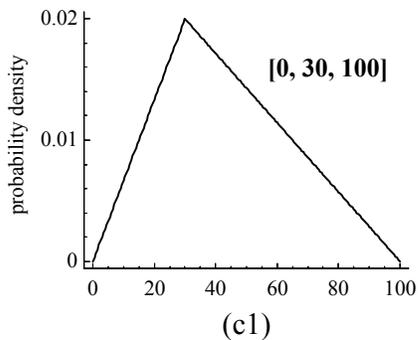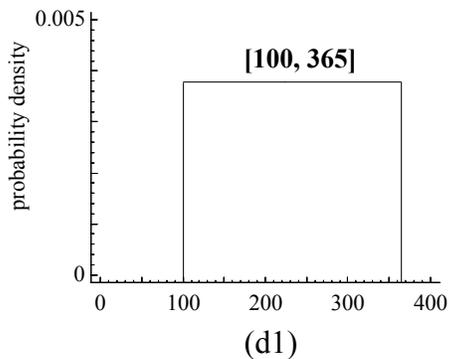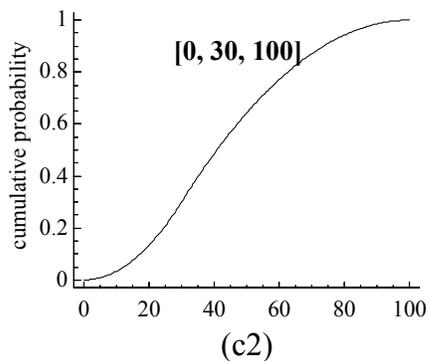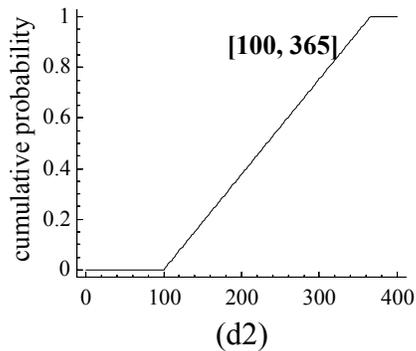
(d2)

Beta
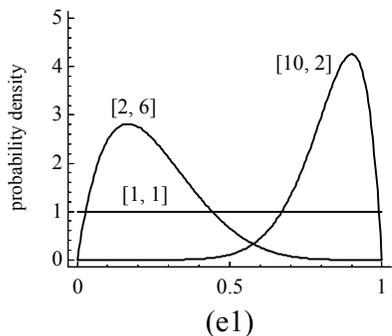
(e1)
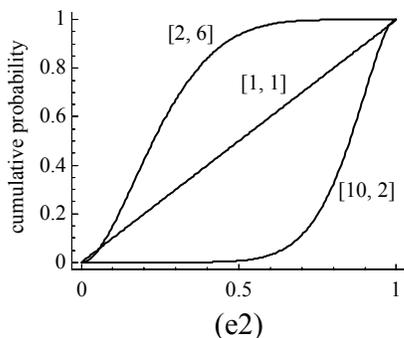
(e2)

Weibull
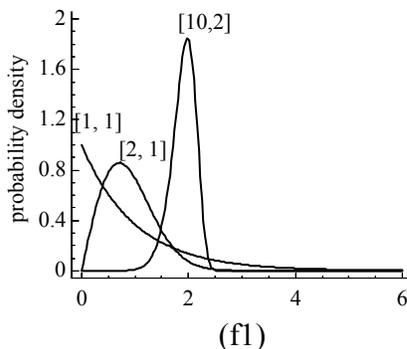
(f1)
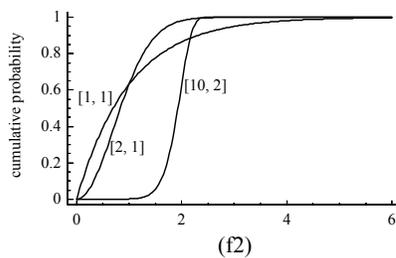
(f2)

Gamma
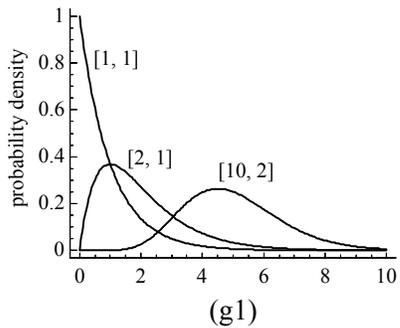
(g1)
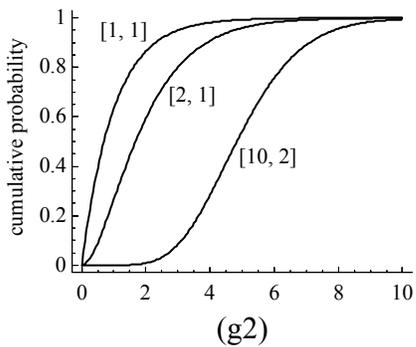
(g2)

Exponential
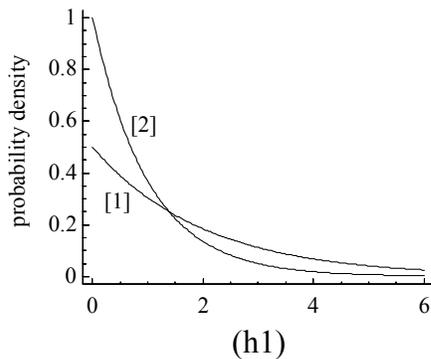
(h1)
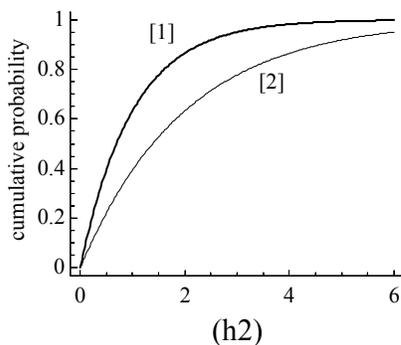
(h2)

B. 5.2 EMPIRICAL DISTRIBUTION FUNCTIONS (EDFs)

In some cases, an empirical distribution function (EDF) may be preferred over fitting the data set to a hypothesized distribution. EDFs, also called empirical cumulative distribution functions (ECDF), provide a way to use the data itself to define the distribution of the relevant variable. Briefly, an EDF for a random variable is described by a step function based on the frequency distribution of observed values. An EDF for a continuous random variable may be linearized by interpolating between levels of the various bins in a frequency distribution. The CDF for a linearized EDF appears as a line, rather than steps. Example B-3 at the end of this Appendix illustrates an EDF, linearized EDF, and beta distribution (α_1=0.63, α_2=2.85, rescaled to min=0, max=364) fit to percentile data for soil ingestion rates in children (Stanek and Calabrese, 1995). A plausible range (i.e., minimum and maximum values) was imposed on the data set for this example.

EDFs provide a complete representation of the data with no loss of information. They do not depend on the assumptions associated with estimating parameters for theoretical probability models. EDFs are designed to provide direct information about the shape of the distribution, which reveals skewness, multimodality, and other features of the data set. However, EDFs may not adequately represent the tails of a distribution due to limitations in data acquisition. In the simplest case, an EDF is constrained to the extremes of the data set. This may be an unreasonable restriction if limiting the EDF to the smallest and largest sample values is likely to greatly underestimate the distributional tails. If this is an important source of uncertainty, the risk assessor may choose to extend the tails of the distribution to plausible bounds or to describe the tails with another distribution (see Exhibit B-4). For example, an exponential distribution may be used to extend the tails based on the last 5% of the data. This method is based on extreme value theory, and the observation that extreme values for many continuous, unbounded distributions follow an exponential distribution (Bratley et al., 1987). As with other probability models, uncertainty in the plausible bounds of an EDF may be reduced by obtaining additional information.

EXHIBIT B-4

VARIATIONS OF THE EDF

Linearized - Linearly interpolates between two observations, yielding a linearized cumulative distribution pattern.

Extended - In addition to linearizing (see above), adds lower and upper bounds based on expert judgment.

Mixed Exponential - Adds an exponential upper and/or lower tail to the EDF.

Advantages and disadvantages of using EDFs in PRA are discussed in detail in the *Report of the Workshop on Selecting Input Distributions for Probabilistic Assessments* (U.S. EPA, 1999a).

B.5.3 GRAPHICAL METHODS FOR SELECTING PROBABILITY DISTRIBUTIONS

Graphical methods can provide valuable insights and generally should be used in conjunction with exploratory data analysis. Examples of graphical methods are frequency distributions (i.e., histograms), stem-and-leaf plots, dot plots, line plots for discrete distributions, box-and-whisker plots, and scatter plots (Tukey, 1977; Conover, 1980; Morgan and Henrion, 1990).

☞ *Graphical methods are invaluable for exploring a data set to understand the characteristics of the underlying population.*

Together with statistical summaries, graphical data summaries can reveal important characteristics of a data set, including skewness (asymmetry), number of peaks (multi-modality), behavior in the tails, and data outliers.

Frequency Distribution or Histogram

The frequency distribution, or histogram, is a graphical approximation of the empirical PDF. Frequency distributions can be plotted on both linear and log scales. The general strategy for selecting the number of bins to partition the data is to avoid too much smoothing and too much jaggedness. Equation B-1 (U.S. EPA, 1999a) provides a starting point for estimating the number of bins based on the sample size (n).

$$Number\, of\, Bins = 1 + 3.322 \log_{10} n \qquad \text{Equation B-1}$$

Probability Plotting

Another method that may be used to visualize distributions and estimate parameters is probability plotting, also referred to as linear least squares regression or regression on ordered statistics. This technique involves finding a probability and data scale that plots the CDF of a hypothesized distribution as a straight line. The corresponding linearity of the CDF for the sample data provides a measure of the GoF of the hypothesized distribution. The general approach involves sorting the sample data in ascending order and converting the ranks to percentiles. The percentile value for the ith rank is calculated according to Gilbert (1987) as:

$$Percentile = 100 \times \frac{i - 0.5}{n} \qquad \text{Equation B-2}$$

An alternative formula is provided by Ott (1995):

$$Percentile = 100 \times \frac{i}{n + 1} \qquad \text{Equation B-3}$$

Plotting positions given by Equations B-2 and B-3 are special cases of the more general formula given by Equation B-4 (Helsel and Hirsch, 1992):

$$Percentile = 100 \times \frac{i - a}{n + 1 - 2a} \qquad \text{Equation B-4}$$

where a is a constant that varies from 0 (Equation B-3) to 0.5 (Equation B-2).

The percentiles are used to calculate the *z-scores*, which represent the number of standard deviations away from the mean that a particular datum lies assuming the data are normally distributed. For normal distributions, the data are plotted against the z-scores; for lognormal distributions, the data are log-transformed and plotted against the z-scores. In both cases, parameters of the distribution can be estimated from the least-squares regression line. When the hypothesized distribution is a poor fit to the data, p-plots can yield misleadingly low estimates of the standard deviation (Cullen and Frey, 1999). Both Gilbert (1987) and Ott (1995) provide excellent descriptions of the use of probability plotting to derive parameter estimates for a given distribution. Probability plotting techniques with best-fit lines have been used to estimate parameters for a wide variety of distributions, including beta, Weibull, and gamma.

Cullen and Frey (1999) point out that probability plotting may not be a primary choice for selecting a fitting distributions because the method violates an important assumption of least squares regression—independence of the observations (see Appendix A, Exhibit A-5). This is because the rank-ordered data are no longer independent. Nevertheless, this approach may yield good results when the fit is good and the choice of distributions is somewhat subjective.

B.5.4   PARAMETER ESTIMATION METHODS

As a rule, there are often a number of different methods available for estimating a given parameter. The most appropriate method to apply may require judgment, depending on the relative difficulty in applying a method for a particular parameter, as well as the desired statistical properties of the method. The following simple example provides a useful analogy. Suppose that the parameter of interest, A, is the total area of an approximately square exposure unit. If the exposure unit is a perfect square, and the length of one side (L_1) is known, the area would be equal to L_1^2 (i.e., for a square, $A=L_i^2$). Suppose L is unknown, but two independent measurements, X_1 and X_2, are available to estimate the length (see Exhibit B-5). If it is assumed that the random variable, L, has a probability distribution with mean μ, then the area of the square piece of property is $A=\mu^2$. What is a reasonable estimate of the area (i.e., $\hat{A} = \hat{\mu}^2$) based on X_1 and X_2? Three plausible methods for calculating $\hat{\mu}^2$ are given below.

1. $\hat{\mu}_a^2 = \left(\dfrac{X_1 + X_2}{2} \right)^2$

2. $\hat{\mu}_b^2 = \dfrac{X_1^2 + X_2^2}{2}$

3. $\mu_c^2 = X_1 \times X_2$

EXHIBIT B-5

ESTIMATING THE AREA OF A HYPOTHETICAL EXPOSURE UNIT
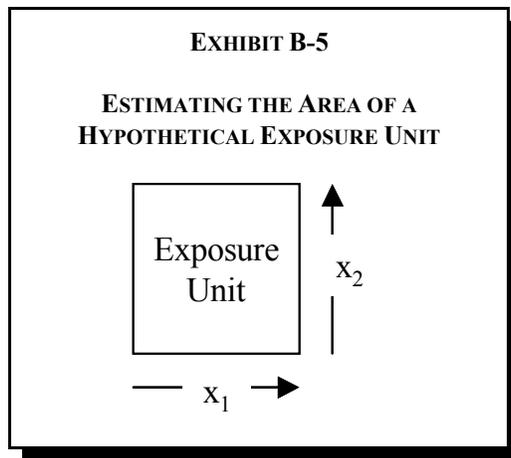
Because these three estimators will, as a rule, give different answers, it may be useful to set criteria for selecting which one gives the "best" answer. Some of the statistical criteria that are used for this purpose are *consistency*, *efficiency, robustness, sufficiency*, and *unbiasedness* (see Exhibit B-6). It turns out, each method is relatively easy to implement, but the third method is preferred because it is a more efficient estimator.

In many cases, particularly if a model is complex, potential estimators of the unknown parameters are not readily apparent. To assist in developing estimators, several general methods have been developed. Exhibit B-7 lists some of the more common parameter estimation methods.

Perhaps the simplest method is the method of matching moments (MoMM), also called the method of moments. MoMM is appropriately named, as it involves expressing the unknown parameters in terms of population moments and then "matching", or equating the sample moments to the population

moments. For example, the sample mean ($\bar{x}$) and standard deviation (*s*) are estimators for the corresponding population parameters (μ and σ).

Maximum Likelihood Estimation (MLE) is a commonly applied method, that is often thought of as a parameter estimate for which the observed data are most "likely". The likelihood function is defined for independent continuous random variables as follows:

$$L(\theta_1, \theta_2,...\theta_k) = \prod_{I=1}^{n} f(x_1|\theta_1, \theta_2, ..., \theta_k)$$

The likelihood function is evaluated based on the product of the PDF for each value of x. The parameters of the probability model, (θ_k), are chosen to maximize the likelihood function value and thereby are most likely to produce the sample data set (Cullen and Frey, 1999).

It has also been demonstrated that MLE yields estimators that generally have good properties when evaluated by the criteria listed above. In some cases (e.g., for smaller sample sizes), these estimators are not *unbiased*; however, this can often be accounted for by "adjusting" the estimator. A familiar example of this adjustment is in estimation of the variance of a normal distribution. The MLE for the variance is biased by a factor of ((n-1)/n), but this is easily corrected by multiplying the MLE by (n/(–1)). For some distributions, calculations of the MLE are straightforward. For example, MLE for parameters of a normal distribution are given by the mean and standard deviation of the sample data, the same as MoMM. MLE for parameters of a lognormal distribution are given by the mean and standard deviation of the log-transformed data, which is different from MoMM. In general, MLE calculations are complex, and commercial software such as *@Risk* and *Crystal Ball*® may be used. A more detailed discussion of the derivation and properties of MoMM and MLE can be found in the statistics literature (e.g., Chapter 5 of Mood and Graybill, 1963; Chapter 9 of Mendenhall and Scheaffer, 1973; Section 6.5 of Law and Kelton, 1991; Section 5.6 of Cullen and Frey, 1999).

EXHIBIT B-6

CRITERIA FOR EVALUATING PARAMETER ESTIMATION METHODS*

Consistency — A consistent estimator converges to the "true" value of the parameter as the number of samples increases.

Efficiency — An efficient estimator has minimal variance in the sampling distribution of the estimate.

Robustness — A robust estimator is one that works well even if there are departures from the assumed underlying distribution.

Sufficiency — A sufficient estimator is one that makes maximum use of information contained in a data set.

Unbiasedness — An unbiased estimator yields an average value of the parameter estimate that is equal to that of the population value.

*Source: Cullen and Frey, 1999

EXHIBIT B-7

PARAMETER ESTIMATION METHODS

- Method of Matching Moments
- Maximum Likelihood
- Minimum Chi-Square
- Weighted Least-Squares

B.5.5 DEALING WITH CORRELATIONS AMONG VARIABLES OR PARAMETERS

Correlations between exposure variables or between parameters of the probability distribution may be important components of a probabilistic model. Correlation is a measure of association between two quantitative random variables. Two random variables may either be positively or negatively correlated. A positive correlation exists between two variables if the value of X_1 increases as the value of X_2 increases. For example, higher hand dust lead levels have been associated with higher pediatric blood lead levels (Charney et al., 1980). A negative correlation exists between two variables if the value of X_1 increases as the value of X_2 decreases. For example, studies suggest the ingestion of soil and dust particles increases as particle size decreases (Calabrese et al., 1996).

A first step in identifying correlations is to assess the possible physical and statistical relationships that exist between variables. In an ecological risk assessment (ERA), for example, the largest surf scoter (diving duck) does not consume the least amount of food, nor does the smallest surf scoter consume the greatest amount of food. Random sampling of body weight and ingestion rate as separate parameters, however, allows for these two possibilities. Neglecting a correlation between two variables may restrict (underestimate) the tails of the ecological Hazard Quotient (HQ) for each chemical of concern (COC), which are frequently the areas of the distribution of most interest.

The degree to which correlations affect the output of a risk model depends on: (1) the strength of correlations between the two variables, and (2) the contribution of the correlated variables to overall variance in the output (Cullen and Frey, 1999). Therefore, it is useful to conduct a preliminary sensitivity analysis to assess the impact of alternative correlation assumptions on the model output. If the impact is significant, correlations should be identified and accounted for in the PRA.

There are several approaches to account for dependencies in MCA including: (1) modifying the model to include the correlation; and (2) simulating dependence between variables for sample generation (Cullen and Frey, 1999). Modifying the model is preferred as simulation techniques cannot capture the full complexity between model inputs. However, when this is not possible, dependencies between variables can be simulated and approximated by correlation coefficients and bivariate normal distributions.

Correlation coefficients are a numerical measure of the strength and direction of the relationship between two variables. Sample correlation coefficients measure the linear relationship between variables. However, if two variables are from different probability distributions, it is unlikely that they are linearly related. Consequently, simulation software programs such as *Crystal Ball*® and *@Risk* can be used to calculate and employ the nonparametric statistic, Spearman's Rank Correlation Coefficients (Rho) in simulating correlation between inputs. Rank Correlation Coefficients measure the linear dependence not of the data values themselves, but of the rank value of the data. The ranks indicate relative positions in an ordered series, not the quantitative differences between the positions. The disadvantage of losing information by using the rank values (rather than the actual values) is offset by the ability to correlate random variables from different distribution types (See Appendix A).

Exhibit B-8 gives an example of a straightforward approach to specifying a rank correlation between two input variables in a one-dimensional Monte Carlo analysis (1-D MCA) for variability. A range of correlations is explored as a form of uncertainty analysis on the distribution of intakes given a fish advisory of 7.0 µg/day for a chemical.

EXHIBIT B-8

CORRELATION OF INPUT VARIABLES FOR 1-D MCA OF VARIABILITY

Intake Equation   Intake = (CF x IR x FI x EF x ED)/(BW x AT)

Variables	Description and Units	Units	Point Estimate or PDFv
CF	concentration in fish	ug/kg	25
IR	fish ingestion rate	kg/meal	lognormal (0.16, 0.07)[1]
FI	fraction ingestion from source	unitless	1.0
EF	exposure frequency	meals/yr	lognormal (35.5, 25.0)[1]
ED	exposure duration	years	30
BW	body weight	kg	70
AT	averaging time	days	10950

[1]Lognormal PDF parameters: arithmetic mean, standard deviation

▸ Correlation between IR and EF is suggested by Burger et al. (1999) study of 250 anglers on the Savannah River, South Carolina. Moderate correlation (Kendall's tau=0.17, p=0.04)

▸ Uncertainty Analysis: 1-D MCA simulations of variability correlating IR and EF using *Crystal Ball® 2000* (5,000 iterations, Latin Hypercube sampling). Spearman rank correlations: 0.10, 0.50, 0.90

Statistics of PDFv for Intake (ug/day) compared to Fish Advisory of 7.0 ug/day

Rank Correlation (r)	0.10	0.50	0.90
Intake Statistics (ug/day)			
mean	1.6	1.8	2.0
50th percentile	1.1	1.1	1.1
95th percentile	4.4	5.4	6.5
97.5th percentile	5.7	7.0	9.0

▸ For this example, only IR and EF are characterized by PDFs. They contribute approximately equally to the distribution of intakes. Positive rank correlations have little effect on the median (50th percentile) of the output distribution, but tend to widen the tails of the distribution. Increasing the correlation from 0.10 to 0.90 increases the 90th percentile from 4.4 to 6.5 ug/day, and the 97.5th percentile from 5.7 to 9.0 ug/day.

▸ If the fish advisory is 7.0 ug/day, uncertainty in the correlation coefficient may have important consequences for the risk management decision.

Correlations may also be specified for parameters of a probability distribution. This is an important concept when designing a two-dimensional Monte Carlo analysis (2-D MCA) in which parameters of the same PDFv might be otherwise be described by independent PDFu's. A common approach for correlating two parameters is to specify a bivariate normal distribution (Nelsen, 1986, 1987; Brainard and Burmaster, 1992). A bivariate normal distribution allows for the distribution of one variable to be sampled conditional on the other. This is a special case of a joint distribution in which both x and y are random variables and normally distributed (as the conditional distribution of x or of y is always normal) (Wonnacott and Wonnacott, 1981). Example B-4 further explains bivariate normal distributions and demonstrates this approach applied to coefficients of a simple linear regression model that relates contaminant concentrations in soil and dust.

The results of correlation analysis should be interpreted with caution. Two variables may be associated due to: (1) a dependency between the two variables; (2) chance (two independent variables appear dependent due to chance in the sampling procedure); and (3) variables not included in the analysis (lurking variables) are affecting the two variables being analyzed. Likewise, a low correlation measure does not necessarily mean the two variables are independent. As a lurking variable may cause the appearance of an association between the two independent variables, it may also mask the association between two dependent variables.

☞ *Correlation describes a degree of mathematical association, not a causal relationship between the two variables.*

Efforts to extrapolate or predict correlations outside the range of observed values should also be done with caution. For example, there may be a strong linear relationship between age and height in children; however, it would be inappropriate to apply this correlation to adults. Additional caution is needed when correlating more than two factors at a time. In general, because of the complexity of specifying a valid covariance matrix when correlating more than two factors at a time, risk assessors may need to consult a statistician to avoid generating misleading risk estimates.

B.5.6 CENSORED DATA

In order to define the exposure point concentration, estimates of summary statistics representative of the entire distribution of data are needed (Helsel and Hirsch, 1992). Censored data complicate the process of selecting and fitting PDFs and estimating parameter estimates. A censored data set is a data set for which measurements above or below a certain threshold are not available. Left censored data occurs frequently at Superfund sites, where samples for a number of chemicals are often below the reporting limit. A censored datum (often denoted by ND) commonly represents a value of half of the laboratory reporting limit.

Three general methods for estimating summary statistics for left censored data sets include: (1) simple substitution; (2) distributional methods; and (3) robust methods (Helsel and Hirsch, 1992). These methods may be evaluated based on the root mean squared error (RMSE) estimate, a measure of the difference between the sample statistic (e.g., the sample mean, $\bar{x}$) and the true population parameter (e.g., population mean, μ).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \frac{(\overline{x} - \mu)^2}{\mu}}{N}}$$

Methods which yield estimates closer to the true parameter value have lower bias, higher precision, and lower RMSEs.

Simple Substitution Methods

Simple substitution methods entail substituting values equal to or lower than the reporting limit in the data set. These surrogate values are then included in the calculation of the summary statistics and in determining the distributional shape of the data set. Although this method is frequently used, it is important to understand its limitations; depending on the surrogate value used (e.g., half the reporting limit) the simple substitution method may yield biased parameter estimates (e.g., low estimates of the mean) and may yield misleading distributional shapes. Studies such as those reported by Gilliom and Helsel (1986) have determined, in terms of the RMSE, that simple substitution methods perform more poorly than the distributional and robust methods described below.

Distributional Methods

With distributional methods, the entire data set is assumed to follow a theoretical distribution (e.g., normal distribution). Assuming a theoretical distribution, MLE and probability plotting (p-plot) methods provide summary statistics that best match the reported values of the data and the percentage of samples below the threshold value. If the data fit the theoretical distribution exactly, or if the sample size is large, both MLE and p-plots are unbiased methods. Often, however, the sample size is small and the distribution deviates from a theoretical distribution. In this case, the MLE and p-plot methods may yield biased and imprecise methods (Hesel and Hirsch, 1992).

Robust Methods

With robust methods, a theoretical distribution is needed. A theoretical distribution is fit to the data above the detection limit by MLE or p-plot methods. Based on this assumed PDF, the value of the data points below the detection limit are extrapolated and used in the summary statistics calculation. Unlike the simple substitution method, these extrapolated values are not estimates for the data points; rather, they are only used jointly to calculate summary statistics (Hesel and Hirsch, 1992). The method is considered robust as it uses the actual values of the sample data, rather than the distribution above the detection limit.

B.5.7   TRUNCATION

Truncation refers to imposing a minimum and/or maximum value on a probability distribution. The main purpose of truncation is to constrain the sample space to a set of "plausible values". For example, a probability distribution for adult body weight might be truncated at a minimum value of 30 kg and a maximum value of 180 kg in order to avoid the occasional selection of an unlikely value (e.g., 5 or 500 kg). Given the subjectiveness involved in selecting truncation limits, such choices should clearly be made with caution, and involvement of stakeholders who may be aware of site-specific circumstances.

For example, there may well be individuals who weigh more than 180 kg and less than 30 kg. The purpose for truncating the tails of a distribution is to confine each risk estimate of a Monte Carlo simulation to a combination of plausible input values. The advantage of truncating unbounded probability distributions in PRA is that central tendency and high-end risk estimates will not be biased by unrealistic values. The disadvantage is that the original parameter estimates of the nontruncated distribution are altered by constraining the sample space. The bias in the parameter estimates increases as the interval between the minimum and maximum truncation limit is reduced. For example, a normal distribution with an arithmetic mean of 100 may be fit to a data set; imposing a truncation limit of 300 may result in a truncated normal distribution with an arithmetic mean of 85. The relationship between the truncated and nontruncated parameter estimates can be determined analytically (Johnson et al., 1995) or approximated using Monte Carlo simulations under both truncated and nontruncated scenarios.

Table B-3. Theoretical bounds and parameter values for selected distributions.

Probability Distribution	Parameters[1]	Theoretical Bounds
Normal	(μ, σ)	$(-\infty, +\infty)$
Lognormal	(μ, σ)	$[0, +\infty)$
Weibull	(α, β)	$[0, +\infty)$
Exponential	(β)	$[0, +\infty)$
Gamma	(α, β)	$[0, +\infty)$
Beta	$(\alpha_1, \alpha_2, a, b)$	$[a, b]$
Uniform	(a, b)	$[a, b]$
Triangular	(a, m, b)	$[a, b]$
Empirical (bounded EDF)	$(a, b, \{x\}, \{p\})$	$[a, b]$

[1] a=minimum, b=maximum, μ=mean, σ=standard deviation, m=mode,
α=shape parameter, β=scale parameter, x=value, p=probability

Truncation is typically considered when using unbounded probability distributions (e.g., normal, lognormal, gamma, Weibull) to characterize variability. Table B-3 gives the theoretical bounds for selected probability distributions that may be more commonly used in PRA. Truncating the minimum value may also be appropriate for distributions whose minimum is defined as zero (e.g., lognormal, gamma, Weibull). Truncation is generally less important when a PDF is used to characterize uncertainty in a parameter estimate (e.g., arithmetic mean), since distributions for uncertainty are often bounded by definition (e.g., triangular, uniform). Bounded continuous distributions, such as the beta distribution or empirical distribution (see Section B.5.2) are not subject to the parameter bias of truncation, although plausible minimum and maximum values must still be identified.

Identifying appropriate truncation limits that reflect "plausible bounds" for an exposure variable will often require judgment. Given that most data sets represent statistical samples of the target population, it is unlikely that the minimum and maximum observed values represent the true minimum and maximum values for the population. However, there may be physiological or physical factors that can aid in setting plausible truncation limits. For example, the maximum bioavailability of chemicals in the gastrointestinal (GI) tract is 100%. Similarly, the solubility of chemicals in aquatic environments

(accounting for effects of temperature) will generally be less than the chemical solubility in water free of particulates.

In general, sensitivity analysis can be used to determine if truncation limits are an important source of parameter uncertainty in risk estimates. For exposure variables in the numerator of the risk equation, the maximum truncation limit is of greatest concern. For exposure variables in the denominator of the risk equation, the minimum truncation limit is of greatest concern. Details regarding the fit of the tails of the probability distribution and the effect of truncation on the parameter estimates should generally be included in the workplan.

B.6.0   ASSESSING QUALITY OF THE FIT

The quality of the fit of a distribution may be evaluated in several ways. Standard statistical approaches are available to test the fit of a theoretical distribution to a data set (i.e., GoF tests). In addition, alternative choices for distribution shapes and plausible bounds might be explored as a form of sensitivity analysis. Together with graphical exploration (Section B.5.3), this information may be useful when deciding whether or not to incorporate a specific type of distribution for an exposure variable into a PRA.

☞   *GoF tests are one tool among several to assess the quality of a distribution.*

Although GoF testing is a necessary part of distribution fitting, and tests are readily available with commercial software, it is less important than mechanistic considerations or graphical data exploration for choosing a candidate distribution. Examples of GoF tests are discussed below, and cautions regarding GoF are outlined in Section B.6.3.

B.6.1   WHAT IS A GOODNESS-OF-FIT TEST?

Goodness-of-fit (GoF) tests are formal statistical tests of the hypothesis that the data represent an independent sample from an assumed distribution. These tests involve a comparison between the actual data and the theoretical distribution under consideration.

In statistical hypothesis testing the null hypothesis (H_0) is assumed to be true unless it can be proven otherwise. The "evidence" upon which we base a decision to reject or not to reject H_0 is a random sample. Typically, we seek to reject H_0 in favor of H_a. For example, with the two sample *t*-test, the null hypothesis is that the means of two populations are equal (not different) and the alternative is that they are different. This is expressed as:

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

Most often, the hypothesis test is used to show that the means are not equal (i.e., reject H_0 in favor of H_a) in order to state that there is a significant difference between the two populations at a specified significance level (e.g., $\alpha = 0.05$). Thus, the hypothesis test is often referred to as a significance test.

The *p-value* in a statistical test is calculated from a sample and represents the probability of obtaining a value of the test statistic as extreme or more extreme as the one observed if H_0 is in fact true. When the *p-value* is small it means either the null hypothesis is not true, or that we have witnessed an

unusual or rare event (by chance we drew an unusual sample that resulted in the extreme value of the test statistic). Often a value of 0.05 or 0.01 is designated as a cutoff, or significance level α. If the *p-value* is (e.g., p < 0.05), the null hypothesis is rejected in favor of the alternative, and we state that the test result is statistically significant at level α. This does not mean that we have proven H_a is true. Rather, we are saying that based on our sample results, it is unlikely that H_0 is true.

In a GoF test, the hypothesis test is set up the same way as a "traditional" hypothesis test, but the outcome is viewed a little differently. In GoF tests, we generally seek to *fail* to reject H_0 because the null hypothesis states that the data were obtained from a population described by the specified distribution (F_0). The alternative hypothesis is that the data were obtained from a population described by a different distribution. In most applications of GoF techniques, the alternative hypothesis is composite—it gives little or no information on the distribution of the data, and simply states that H_0 is false (d'Agostino and Stephens, 1986). This can be expressed as:

$$H_0 : F = F_0$$

$$H_a : F \neq F_0$$

where F_0 is a specific continuous distribution function, such as the CDF for a normal distribution.

☞ *GoF tests do not prove that the population is described by the specified distribution, but rather that this assumption could not be rejected.*

In general, *p-values* provide one metric of evaluating the fit of the distribution. For example, a *p-value* of 0.06 indicates that the null hypothesis (i.e., the assumption of a specified distribution) cannot be rejected at α=0.05. Larger *p-values* indicate a better fit and stronger evidence that the distribution specified by the null hypothesis may be appropriate. This guidance does not recommend an arbitrary cutoff for the *p-value*. A risk assessor performing a GoF test generally should report the *p-value* and whether the fit is considered "good" or "poor".

B.6.2 WHAT ARE SOME COMMON GOODNESS-OF-FIT TECHNIQUES?

The following GoF tests can also be found in most general statistical and spreadsheet software. Both *Crystal Ball*® and *@Risk* software present the results of chi-square, K-S, and Anderson-Darling tests in their fitting routines.

Shapiro-Wilk Test

The most widely used GoF test in risk assessment is the Shapiro-Wilk test for normality (Gilbert, 1987). This simple hypothesis test can determine whether or not a small data set ($n \leq 50$) is normally distributed. The test can also be run on log-transformed data to assess whether the data are lognormally distributed. D'Agostino's test may be used for samples sizes larger than those accommodated by the Shapiro-Wilk test (i.e., $n > 50$) (d'Agostino and Stephens, 1986). In addition, Royston (1982) developed an extension of the Shapiro-Wilk test for n as large as 2000 (Gilbert, 1987).

Probability Plot Correlation Coefficient Test

The correlation coefficient r (or the coefficient of determination, r^2) between the data and the z-scores of a normal probability plot (Filliben, 1975; Helsel and Hirsch, 1992) is similar to the W statistic

of the Shapiro-Wilk test. A detailed comparison of the Shapiro-Wilk test and the product correlation coefficient test is given by Filliben (1975) and d'Agostino and Stephens (1986). Helsel and Hirsch (1992) summarize critical r* values derived by Looney and Gulledge (1985) for the probability plot correlation coefficient test.

Chi-Square Test

The chi-square test is a general test that may be used to test any distribution (continuous or discrete), and for data that are ordinal (e.g., categories such as high/medium/low). Chi-square is a measure of the normalized difference between the square of the observed and expected frequencies. For example, by constructing a frequency distribution of the data with k adjacent bins, $j=1...k$, the number of data points in the jth bin can be compared with the expected number of data points according to the hypothesized distribution. Note that in the case of continuous, unbounded distributions (e.g., normal), the first and last intervals may include $[-\infty, a_1]$ or $[a_k, +\infty]$ (Law and Kelton, 1991). The chi-square test is very sensitive to the chosen number and interval width of bins—different conclusions can be reached depending on how the intervals are specified. Strategies for selecting bins (e.g., setting interval widths such that there are no fewer than 5 data points expected per bin) are given in the statistical literature (d'Agostino and Stephens, 1986; Law and Kelton, 1991). The test statistic is compared with a value of the chi-square distribution with $(k - r - 1)$ degrees of freedom, where k is the number of sample values and r is the number of parameters of the hypothesized distribution. As described in Section B.6.1, in general, higher *p-values* suggest better fits.

Kolmogorov-Smirnov (K-S) Test

The K-S test is a nonparametric test that compares the maximum absolute difference between the step-wise empirical CDF and the theoretical CDF. Because the maximum discrepancy is compared with the test statistic, K-S is sometimes referred to as a *supremum* test (Cullen and Frey, 1999). In general, lower values of the test statistic indicate a closer fit. The K-S test is most sensitive around the median of a distribution, and, hence, it is of little use for regulatory purposes when the tails of distributions are most generally of concern (U. S. EPA, 1999a). Although it does not require grouping data into bins like the chi-square test, critical values for the K-S test depend on whether or not the parameters of the hypothesized distribution are estimated from the data set (Gilbert, 1987; Law and Kelton, 1991). The Lilliefors test was developed to surmount this problem when the hypothesized distribution is normal or lognormal (Gilbert, 1987).

Anderson Darling Test

The Anderson-Darling test assesses GoF in the tails (rather than the mid-ranges) of a PDF using a weighted average of the squared differences between the observed cumulative densities. The Anderson-Darling test is sometimes referred to as the *quadratic* test (Cullen and Frey, 1999). The test statistic should be modified based on sample size prior to comparison with the critical value. Like the K-S test, in general, lower values of the test statistic indicate a closer fit (i.e., if the adjusted test statistic is greater than the modified critical value for a specified α, the hypothesized distribution is rejected). The Anderson-Darling test may be particularly useful because it places more emphasis on fitting the tails of the distribution.

B.6.3   CAUTIONS REGARDING GOODNESS-OF-FIT TESTS

There are many statistical software programs that will run GoF tests against a long list of candidate distributions. It is tempting to use the computer to make the choice of distribution based on a test statistic. However, GoF tests have low statistical power and often provide acceptable fits to multiple distributions. Thus, GoF tests are better used for rejecting poorly fitting distributions than for ranking good fits. In addition, for many distributions, GoF statistics lack critical values when the parameters are unknown (i.e., estimated from the data). In practice, this limitation is often discounted and the critical values are interpreted as a semi-quantitative measure of the fit. It is most appropriate to form an idea of the candidate distributions based on some well reasoned assumptions about the nature of the process that led to the distribution, and then to apply a GoF test to ascertain the fit (U.S. EPA, 1999a). Whenever possible, mechanistic and process (i.e., phenomenologic) considerations should inform the risk assessor's choice of a particular distribution rather than the results of a comparison of GoF tests (Ott, 1995). In addition, the value of graphical evaluations of the fit cannot be overstated.

B.6.4   ACCURACY OF THE TAILS OF THE DISTRIBUTION

The tails of a distribution (e.g., $< 5^{th}$ and $> 95^{th}$ percentiles) for an input variable are often of greatest interest when characterizing variability in risk. Distributions fit to data may not characterize the tails of the distribution in a way that represents the target population. In general, the importance of uncertainty in the fit of the tails of particular distributions should be determined on a site-specific basis. For exposure variables in the numerator of the risk equation, the upper tail is of greatest concern. For exposure variables in the denominator of the risk equation, the lower tail is of greatest concern.

The tails of the input PDFs generally have a significant influence on the tails of the risk distribution, especially for those variables that are ranked highest in a sensitivity analysis. Different distributions may share the same mean and variance, but assume very different shapes. Experiments with Monte Carlo simulations have demonstrated that the shape of the input PDFs may have a minimal effect on the risk estimates in the tails of the probability distribution when the mean and variance of the input PDFs are held constant (Hoffman and Hammonds, 1992; Finley and Paustenbach, 1994). Nevertheless, it is generally a good practice in PRA to demonstrate that alternative choices of PDFs do not have a significant effect on percentiles in the RME risk range.

A common question when developing and evaluating Monte Carlo models is, "How many iterations is enough?". Since Monte Carlo sampling is approximately random, no two simulations will yield the same results (unless the same starting point, or seed, of the random number generator is used). A rule of thumb is that the stability of the output distribution improves with increasing numbers of iterations, although there will always remain some stochastic variability. The stability is generally better at the central tendency region of the output distribution than at the tails; therefore, more iterations may be needed when the risk management decision is associated with the higher percentiles (e.g., $> 95^{th}$ percentile). Risk assessors are encouraged to run multiple simulations (with the same inputs) using different numbers of iterations in order to evaluate the stability of the risk estimate of concern. The results of such an exercise should generally be reported to the Agency when submitting a PRA for review. Note that while the speed of modern computers has essentially eliminated the issue for 1-D MCA (e.g., 10,000 iterations of most 1-D MCA models can be run in less than 1 minute), it may still be an important issue for more complex modeling approaches such as Microexposure Event analysis (MEE) and 2-D MCA (see Appendix D).

B.7.0   SELECTING PROBABILITY DISTRIBUTIONS BASED ON STATE OF KNOWLEDGE

Table B-4 summarizes preliminary strategies for proceeding with a PRA based on the amount of available information. Recommended starting points for each of the three steps in the general process are provided. This table provides guidance on candidate distributions that are consistent with the available information, however, it is not intended to discourage the use or exploration of alternative choices.

☞ *Table B-4 provides recommended preliminary strategies, not steadfast rules. As an analyst works through the PRA, alternative distributions, estimation methods, consideration of mechanism, and GoF tests may better guide the selection process.*

Case 1 represents the best scenario, in which the analyst has access to the raw data and a sufficiently large sample size (or $\geq$ 6 percentiles). In this case, the analyst has a variety of choices for distribution fitting and estimating parameters. However, frequently raw data are inaccessible to the analyst. Cases 2 and 3 have limited information available (i.e., mean and upper percentile) and, therefore, have a narrower set of starting points. Case 4 is the most extreme scenario of data availability requiring expert judgment on selecting and fitting distributions.

Table B-4. Strategies for conducting PRA based on available information. Preferred methods in Case 1 (most information) are identified by an asterisk (*).

Evaluation Step	Case 1	Case 2	Case 3	Case 4
	Decreasing Information ⟶			
Data Availability	raw data of sufficiently large sample size *or* six or more percentiles	three to five statistics	two statistics	one statistic
Selection of Distribution Type	**Nonnegative Continuous** any in this category **Bounded** beta, Johnson's SB	**Nonnegative Continuous** lognormal, gamma, Weibull **Bounded** beta, Johnson's SB		case-by-case basis using expert judgment
Selection of Parameter Estimation / Fitting Method	maximum likelihood* regression methods matching moments	minimize average absolute percent error (MAAPE) for available statistics	exact agreement between 2-parameter PDF and available statistics	
Assessment of Quality of Fit	**Graphical Assessment** P-log Q plot*, P-Q plot* residual % error plot* P-P plot, Q-Q plot **GoF Tests** Anderson-Darling* K-S Chi-square	**Graphical Assessment** P-log Q plot, P-Q plot **GoF Test** Chi-square, Estimate p-value for MAAPE using parametric bootstrap (if sample size is known)	**Graphical Assessment** judgment based on comparative analysis of PDFs and CDFs	
Estimation of Parameter Uncertainty	**Large Sample** asymptotic normality assumption **Medium Sample** nonparametric bootstrap **Small Sample** parametric bootstrap	**Parametric bootstrap** generate random samples using the fitted distribution (if sample size is known)		

EXAMPLES OF FITTING DISTRIBUTIONS USING
GRAPHICAL METHODS, GOODNESS-OF-FIT, AND PARAMETER ESTIMATION

Example B-1. Empirical Distribution Function (EDF) for Soil Ingestion Rates

This hypothetical example illustrates how graphical methods can be used to select probability distributions for variability based on percentile data reported in the literature. Table B-5 gives the summary statistics that are reported by Stanek and Calabrese (1995) for average daily soil ingestion rates among young children. Three options are explored for selecting a distribution: (1) empirical distribution function (EDF) represented by a step function; (2) linearized and extended EDF; and (3) continuous parametric distributions (beta and lognormal).

In order to specify an EDF, a plausible range (minimum and maximum) must be inferred using judgment. Exposure factors such as ingestion rate are nonnegative variables (i.e., minimum ≥ 0); given the relatively low value for the 25th percentile (10 mg/day), it is assumed that 0 mg/day is a reasonable minimum value for this example. If children with pica for soil are excluded from the population of concern, the maximum value may be inferred from the relatively shallow slope at the high-end of the distribution. That is, the 90th percentile is reported as 186 mg/day while the 99th percentile is 225 mg/day, an increase of only 39 mg/day; it is assumed that 300 mg/day is a plausible maximum value for this example. Commercial software such as *Crystal Ball*® and *@Risk* can be used to input EDFs. Figure B-3 illustrates the basic step-wise EDF represented by the reported percentile values, as well as the "linearized, extended EDF" (i.e., linear interpolation between reported values and extended lower and upper tails).

An alternative to relying on a linear interpolation between the percentile values is to fit a continuous probability distribution to the reported percentiles. Since the original data are unavailable, standard GoF tests for the EDF, such as K-S and Anderson-Darling (d'Agostino and Stephens, 1986), cannot be applied. Note that computer software (e.g., *Crystal Ball*®, *@Risk*) will provide test statistics and corresponding *p-values*, however, these results will (inappropriately) reflect the number of percentile values reported rather than the sample size of the original data. Nevertheless, graphical methods may be employed to assess the adequacy of the fit of various PDFs. In this example, a beta distribution and lognormal distribution were fit to the EDF using *Crystal Ball*®. Figure B-4 illustrates the selected statistics for both distributions.

The beta distribution appears to more closely match the reported percentile values, especially at the upper tail of the distribution. The lognormal distribution has an unbounded maximum that, for this example, results in an extreme overestimate of the 95th and 99th percentiles. The beta distribution, by definition, is bounded at 0 and 1, and rescaled in this example to a maximum of 364 mg/day. This analysis would support the use of a beta distribution in a Monte Carlo simulation.

Table B-5. Selected statistics for reported and fitted distributions for ingestion rate (mg/day).

Summary Statistic	Reported Values	Linearized, Extended EDF	Beta Distribution[1]	Lognormal Distribution[2]
minimum	--	0	0	0
25th percentile	10	10	13	11
50th percentile	45	45	44	31
75th percentile	88	88	100	86
90th percentile	186	186	165	216
95th percentile	208	208	205	375
99th percentile	225	225	322	3346
maximum	--	300	364	$+\infty$

[1]Parameters of best-fit beta distribution: α_1=0.63, α_2=2.85, min=0, max=364.
[2]Parameters of best-fit lognormal distribution: μ=97.6, σ=291.8.
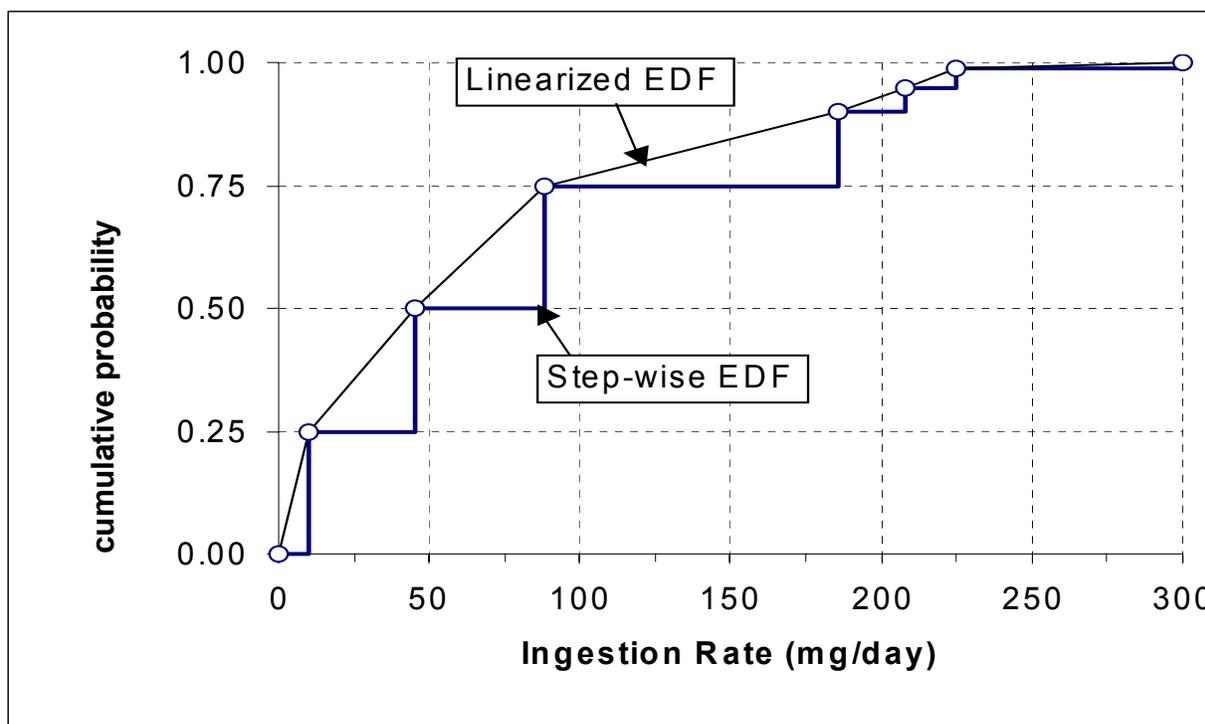
Figure B-3. Comparison of step-wise EDF and linearized EDF for ingestion rate. The upper and lower tails of both distributions are extended to a plausible range of [0, 300] mg/day.
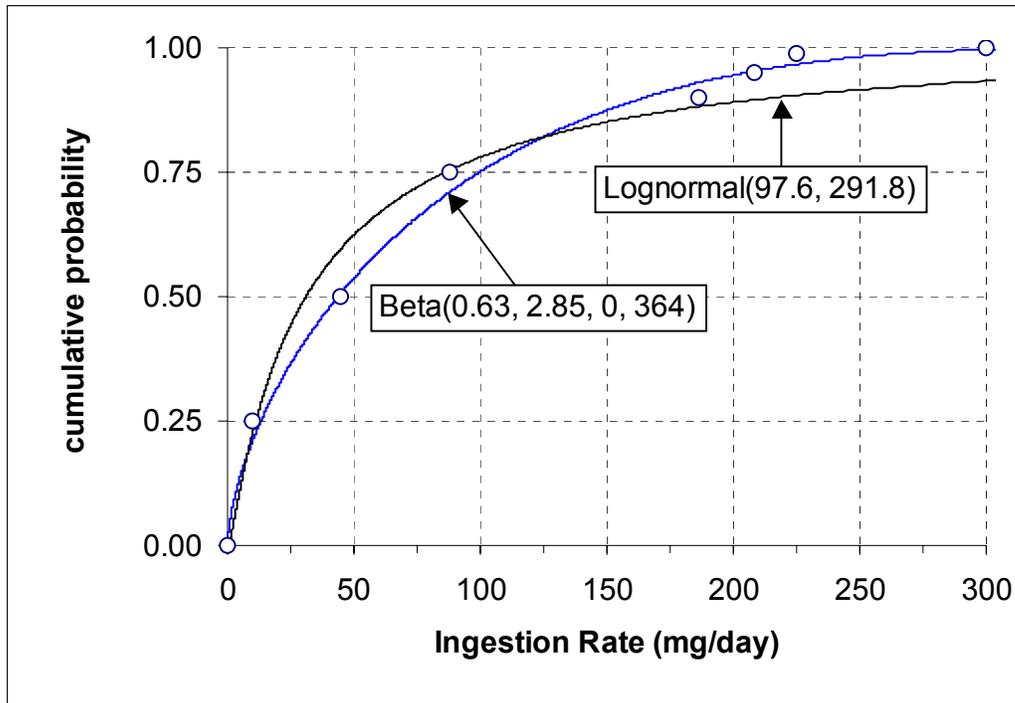
Figure B-4. Graphical assessment of beta and lognormal distributions fit to the cumulative distribution reported in the literature (circles). The beta distribution provides a closer fit to the percentile values in this example.

Example B-2. Variability in Lead Concentrations in Quail Breast Tissue

This hypothetical example demonstrates how the combination of graphical methods, GoF tests, and parameter estimation techniques provides strong evidence for selecting and fitting a lognormal distribution. Assume lead concentration in quail is an important variable for a food web model. Site-specific data (n=62) are used to estimate inter-individual variability in concentration (Table B-6). The histograms in Figure B-5 show lead concentrations in quail breast tissue collected near a settling pond at a plating works. Equation B-1 indicated that 7 bins is an appropriate starting point. The result (top left panel, Figure B-5) suggests that approximately 80% of the values are < 200 ppm and that the probability distribution for variability may be described by a nonnegative, right-skewed distribution (e.g., exponential, Weibull, lognormal, etc.). However, additional bins are needed to better understand the low-end of the distribution. After increasing the number of bins from 7 to 16 (top right panel, Figure B-5), graphical evaluation continues to suggest that the distribution is unimodal right skewed. The bottom panel of Figure B-5 illustrates that increasing the number of bins would not provide better resolution of the low-end of the distribution. For these data, 16 bins appear to provide a reasonable balance between too much smoothing and too much jaggedness.

Probability plots can be used to visually inspect the GoF of a specified distribution to the data, and, because the hypothesized distribution yields a straight line, the plots are particularly useful for evaluating deviations at the tails. In addition, parameter estimates can be obtained from the regression lines fit to the data, as discussed below. For this example, two lognormal probability plots are explored to evaluate how well the data can be described by a lognormal distribution (Figure B-6). The top panel gives the *z-score* on the abscissa (the "x" axis) and ln[concentration] on the ordinate (the "y" axis), while the bottom panel gives ln[concentration] on the abscissa and *z-score* on the ordinate. Plotting positions for both methods were calculated using Equation B-2. Equally plausible parameter estimates can be obtained from regression lines using either plotting method; however, the approach shown in the top panel may be easier to implement and interpret.

Despite the relatively large sample size of n=62, GoF tests generally fail to reject lognormality (i.e., normality of the log-transformed data) in this example. For the probability plot correlation coefficient test (Filliben, 1975; Looney and Gulledge, 1985), if r < r* (the value for r at a specified α), normality is rejected. For this example, *r* is 0.988, and r* is between 0.988 and 0.989 for n=62 and α=0.25; therefore, the *p-value* for the concentrations is approximately 0.25 and one fails to reject lognormality at $\alpha \leq 0.25$. D'Agostino's test yields essentially the same conclusion, with a calculated *Y* value of -1.9166. For this data set, with n=62 and α=0.10, one rejects normality if $Y < -2.17$ or $Y > 0.997$ (see Table 9.7 in d'Agostino and Stephens, 1986); therefore, since Y is within this interval, one fails to reject the normal distribution. However, for α=0.20, the rejection criteria is [$Y < -1.64$ or $Y > 0.812$], Y falls outside the low-end of the interval, resulting in a rejection of the normal distribution. For this data set, the *p-value* associated with d'Agostino's test is slightly less than 0.20 and one fails to reject normality at $\alpha < 0.20$.
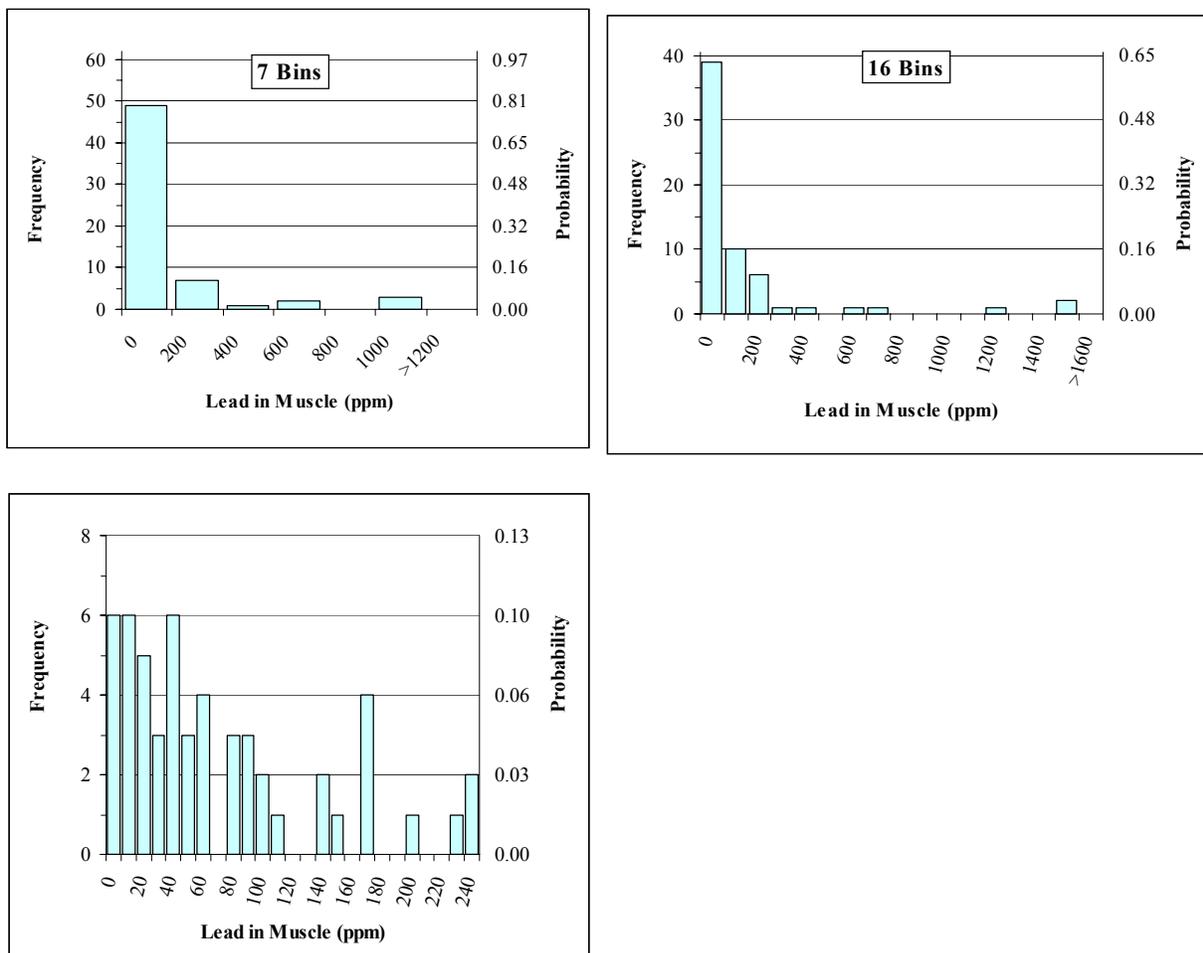
Figure B-5. Histograms of lead concentrations in quail breast muscle (*n*=62). The top left panel shows the result with seven bins; the top right panel shows the result with sixteen bins; the bottom panel uses bin widths of 10 ppm to highlight the lower tail (< 250 ppm) of the distribution.

Table B-6. Sample values of lead concentration (ppm) in quail breast muscle (*n*=62).

0.45	15.8	36.6	57	91	173	265
2.1	16	40	59.6	94.2	175.6	322
5.4	16.7	40.1	61.4	99	176	490
7.8	21	42.8	62	107	177	663.4
7.8	23	44	64	109	205	703
8.8	24	46	64	111	239	1231
11.8	24.8	47	84.6	149	241	1609
12	29.2	49	86.6	149	245	1634
15	35.5	53	86.8	154	264	

Different methods for obtaining the parameter estimates for the lognormal distribution can be explored in this example. For the lognormal distribution, MLE and MoMM simply require calculating the mean and standard deviation of the log-transformed sample data. For the lognormal probability plot method, the parameters can be obtained directly from the least squares regression line expressed as follows:

$$\ln(x) = [slope]z + [intercept]$$
<div align="right">Equation B-5</div>

such that exponentiating the intercept will give the geometric mean (GM) and exponentiating the slope will give the geometric standard deviation (GSD) (see Footnote 3 of Table B-7). Both the MLE and MoMM estimates will generally match the arithmetic mean of the log-transformed data (i.e., intercept) determined from lognormal probability plots; however, estimates of the standard deviation (i.e., slope) will vary (Cullen and Frey, 1999). In general, the probability plot method yields estimates of the standard deviation that are less than or equal to that of MoMM and MLE, and the results yield closer estimates as the correlation coefficient of the probability plot increases (Cullen and Frey, 1999). Table B-7 summarizes the parameter estimates using MLE, MoMM, and the two lognormal probability plotting techniques described above. The corresponding parameter estimates for the untransformed data are also presented.

In this example, the strong linearity of the probability plots (r^2=0.98) shown in Figure B-6 is an indication that a lognormal distribution is a reasonable model for describing variability in concentrations. The tails of the distributions fit the data fairly well, although the bottom panel suggests that the lognormal distribution slightly overestimates the lower tail. Furthermore, the parameter estimates of the lognormal distribution using probability plotting closely match the estimates using MLE and MoMM.

Table B-7. Parameter estimates for lognormal distribution of lead concentrations (ppm).

Parameter Estimation Method	Log-transformed Data		Untransformed Data[3]	
	Arithmetic mean [$\hat{\mu}$]	Arithmetic stdev [$\hat{\sigma}$]	Arithmetic mean [$\hat{\mu}$]	Arithmetic stdev [$\hat{\sigma}$]
Maximum Likelihood Estimate (MLE)	4.175	1.522	207	626
Method of Matching Moments (MoMM)	4.175	1.522	207	626
Log Probability Plot[1]	4.175	1.507	203	597
Log Probability Plot[2]	4.175	1.543	214	670

[1]Least squares regression line for Figure B-6, top panel.
[2]Least squares regression line for Figure B-6, bottom panel.
[3]For a lognormal distribution, the following equations can be used to convert parameters of the normal distribution of log-transformed data to corresponding parameters of the lognormal distribution of untransformed data. Assume μ* and σ* are the arithmetic mean and standard deviation, respectively, for the normal distribution of log-transformed data.

$$geometric\ mean = \exp[\mu^*]$$

$$geometric\ standard\ deviation = \exp[\sigma^*]$$

$$arithmetic\ mean = \exp[\mu^* + 0.5\sigma^{*2}]$$

$$standard\ deviation = \exp[\mu^*]\left(\exp[\sigma^{*2}]\exp[\sigma^{*2} - 1]\right)^{0.5}$$
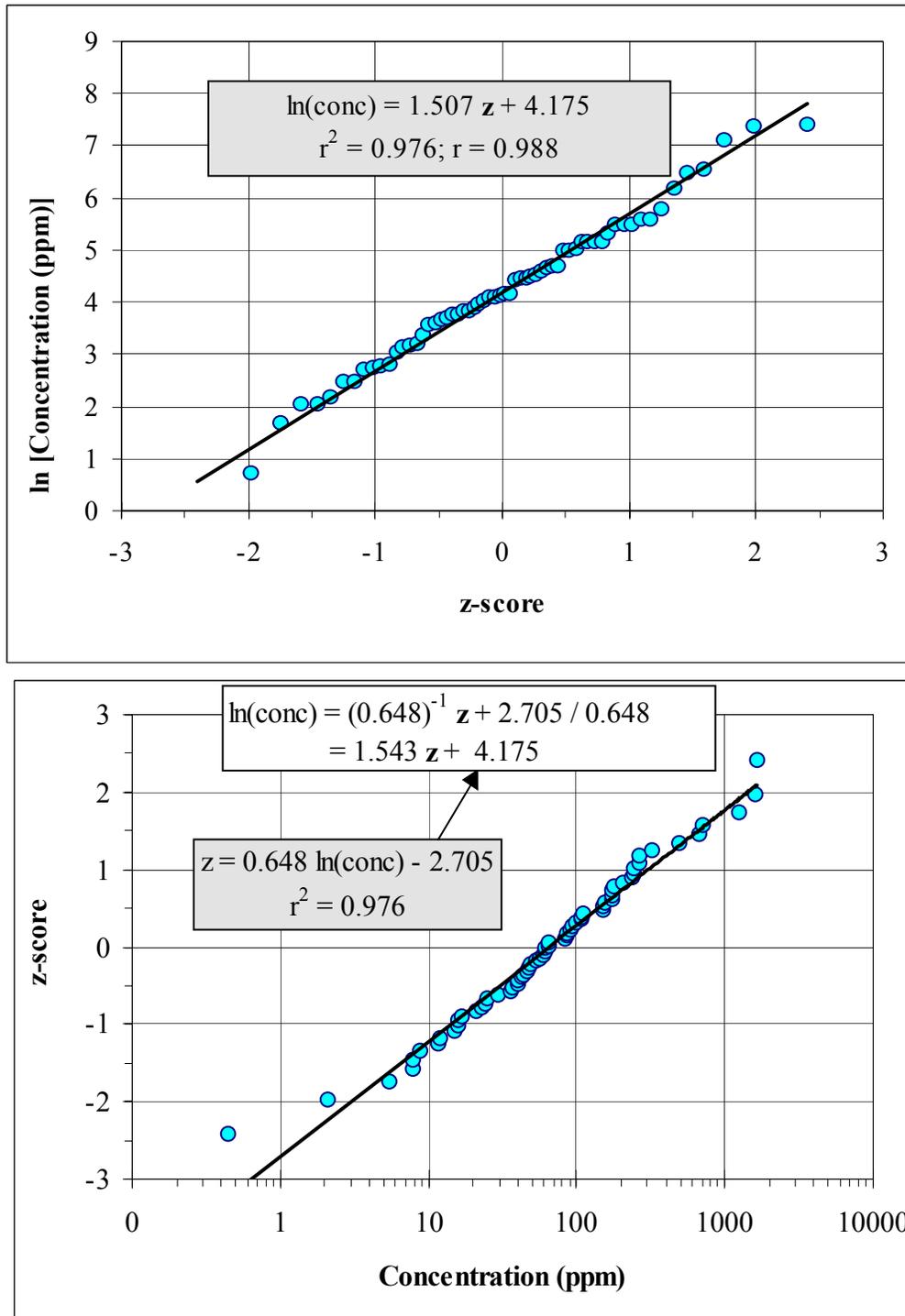
Figure B-6. Lognormal probability plots of lead in quail breast tissue. Top panel gives z on the abscissa and ln[concentration] on the ordinate. Bottom panel gives concentration (log scale) on the abscissa and z on the ordinate. Equally plausible parameter estimates can be obtained from regression lines using either plotting method. Bottom panel requires an additional step to express the equation that yields parameter estimates $[\ln(x)=(\text{slope})\, z + (\text{y-intercept})]$, where the slope estimates the standard deviation of $\ln(x)$ and the y-intercept (at $z=0$) estimates the arithmetic mean of $\ln(x)$.

Example B-3. Variability in Meal Sizes Among Consuming Anglers

A creel survey of anglers consuming contaminated fish was performed to estimate variability in fish meal sizes. The anglers were asked how many people would eat their fish. The lengths of the fish were measured and a regression equation was used to calculate the corresponding weights. The portion of the fish mass that is consumed was assumed to be 40% (e.g., fillets). Results given in Table B-8 are expressed in units of grams of fish per meal.

The appearance of the histograms (Figure B-7) suggests that the sample (n=52) may have been selected from a single distribution.

A normal probability plot of the meal sizes (Figure B-8) shows a departure from linearity. Specifically, there appears to be a "kink" in the probability plot at about 400 g/meal, suggesting that the sample may have been obtained from two unique distributions. Both the Filliben test and Shapiro-Wilk test indicated a significant departure from normality at α=0.01. Parameters may be read directly from the equations of the regression lines on the right hand panel of the graph. MoMM and MLE gave similar estimates.

Table B-8. Meal size (g/meal) (n=52).

65	182	310	405
74	208	314	415
74	221	318	416
77	226	318	477
90	241	327	531
110	248	332	572
111	253	336	608
133	260	337	745
143	261	350	831
150	281	351	907
163	303	360	1053
163	305	365	1189
174	305	390	1208
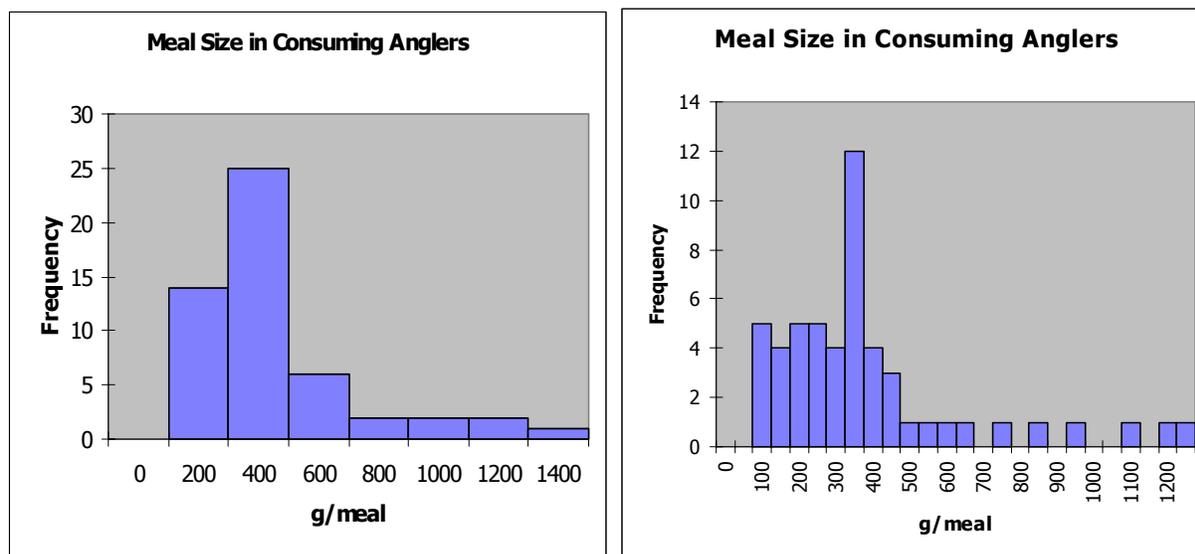
Figure B-7. Histograms of meal size (n=52) among consuming anglers. Left panel uses 7 bins, while the right panel uses 14 bins.
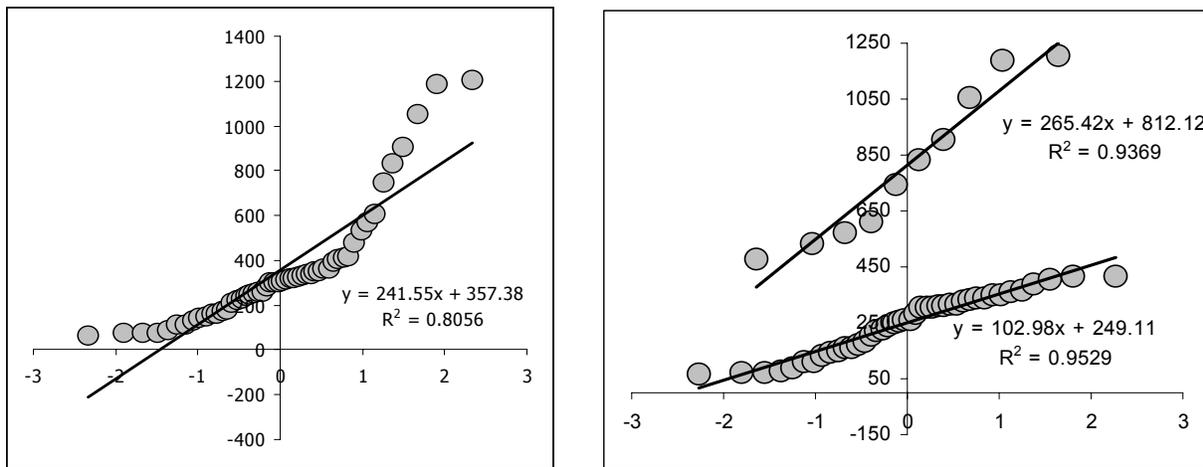
Figure B-8. Probability plot of meal size data from consuming anglers. The left panel shows the combined data, with a departure from linearity at ~ 400 g/meal. The right panel shows the data split between high consumers (top line) and low consumers (bottom line); note that separate lognormal probability plots were reconstructed for both subsets of the data. The point at which to "split" the distribution in the left panel is somewhat subjective. The break would be more obvious if the two distributions did not overlap.

Example B-4. Bivariate Normal Distributions

This example introduces the bivariate normal distribution to illustrate two concepts: (1) use of information on correlations in a Monte Carlo simulation; and (2) specifying distributions for uncertainty in parameter estimates. A brief explanation of the bivariate distribution is presented followed by an example comparing assumptions of no correlation and perfect correlation. A less complex example of a method for addressing correlations in PRA is given in Exhibit B-8.

> **THIS EXAMPLE PRESENTS...**
>
> • Description of the assumptions associated with the bivariate normal distribution
>
> • Guidance on simulating the bivariate normal distribution for two random variables
>
> • Application of bivariate normal to a simple linear regression equation relating contaminant concentrations in soil and dust (see Figure B-9). Results are compared to the assumption of no correlation and perfect correlation

Properties of a Bivariate Normal Distribution

One approach that can be used to correlate two random variables is to specify a bivariate normal distribution, which allows for the distribution of one variable to be sampled conditional on the other. A bivariate normal distribution is a special case of a joint distribution in which both x and y are random independent normally distributed variables. A bivariate normal distribution can be specified for all correlation coefficients including $\rho=0$, $\rho=1$, and $\rho=-1$. The bivariate distribution has a three dimensional shape and for $\rho=0$, from a bird's-eye view, is perfectly circular. As correlation increases (i.e. moves towards -1 or 1) this circle narrows and flattens to an elliptical shape, and finally for perfect correlation ($\rightarrow=1$ and $\rho=-1$) becomes a straight regression line with a $r^2=1$. In three dimensional space the probability of obtaining measurement pairs (x, y) in the region is equal to the volume under the surface in that region. To completely specify the bivariate normal, estimates of the arithmetic mean and variance of the two parameters, as well as the correlation coefficient (μ_X and μ_Y, variances σ^2_X and σ^2_Y , and correlation coefficient ρ) are needed.

In a bivariate normal distribution, values of y corresponding to each value of x follow a normal distribution (Snedecor and Cochran, 1989). Analogously, the values of x corresponding to each value of y follow a normal distribution. Furthermore, if two random variables, X and Y, jointly follow a bivariate normal distribution, the marginal distribution of X is normal with mean μ_X and variance σ^2_X, and the marginal distribution of Y is normal with mean μ_Y and variance σ^2_Y.

Conditional Distributions

Assume we are interested in the conditional distribution of X given a certain value for Y. For example, if X and Y are positively correlated, we would expect that relatively high values of X tend to correspond with relatively high values of Y. The conditional distribution of X given that Y=y, where y represents a specific value for the random variable Y, is a normal distribution with:

$$mean = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \quad and$$

$$variance = \sigma_X^2(1 - \rho^2)$$

Equation B-6

Likewise, the conditional distribution of Y given that X=x, is also normal with:

$$mean = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \quad and$$

$$variance = \sigma_Y^2(1 - \rho^2)$$

Equation B-7

These general equations can be used to generate a correlated pair (X, Y), as described below.

*Note that the mean of the conditional distribution of X is a function of the given value of Y but the variance depends only on the degree of correlation.

General Approach for Correlating X and Y

To generate a correlated pair (X, Y), first generate X using a random value Z_1 from the standard normal distribution:

$$X = \mu_X + \sigma_X \times Z_1$$

Equation B-8

Next, express Y as a function of the conditional mean and variance of Y given X and a second standard normal variate Z_2:

$$Y = \mu_Y + \sigma_Y \times Z_2$$

Equation B-9

and generate a correlated *Y* by plugging Equation B-7 into Equation B-9. Using algebra, the combined equations yield the following simplified expression for generating *Y*:

$$Y = \mu_Y + \sigma_Y \left[(\rho \times Z_1) + \sqrt{1 - \rho^2} \times Z_2 \right]$$
Equation B-10

The important component of this equation is that two random variates are needed (Z_1 and Z_2).

An alternative, but less general approach would be to obtain *Y* by first generating a normal variate *X* (Equation B-8) and then plugging that value into the regression equation of *Y* on *X* to obtain the associated value of *Y*. While this method maintains a correlation between *X* and *Y*, it will underestimate parameter uncertainty. The results are equal only for the special case of perfect correlation (ρ=1.0) between *X* and *Y*. Therefore, the more general bivariate normal distribution approach (given by Equations B-8 to B-10) is recommended for correctly correlating *X* and *Y* because it provides a more robust estimate of parameter uncertainty.

Application of Bivariate Normal Distribution to Correlate Concentrations of Zinc in Soil and Dust

---

EXHIBIT B-9

STEPS FOR SIMULATING UNCERTAINTY IN LINEAR REGRESSION EQUATION USING A BIVARIATE NORMAL DISTRIBUTION TO CORRELATE PARAMETERS (B_0, B_1)

(1) Select Z_1 from a standard normal distribution $Z \sim N(0, 1)$

(2) Calculate β_0 using Equation B-8, where X=β_0, μ_x=μ_{b0}, and σ^2_x=σ^2_{b0}

(3) Select Z_2 from a standard normal distribution $Z \sim N(0, 1)$

(4) Calculate β_1 using Equation B-10, where Y=β_1, μ_y=μ_{b1}, σ^2_y=σ^2_{b1}, ρ=correlation between β_0 and β_1

Assume random sampling of soil and dust zinc concentrations occurs in a residential area. Composite samples of soil and dust are collected from 21 locations such that samples are paired (i.e., each soil sample is co-located with a dust sample) (Table B-9). First the relationship between the zinc concentration in soil and dust is evaluated using simple least-squares regression. Next, the bivariate normal distribution for the slope (β_1) and intercept (β_0) is determined, yielding an arithmetic mean and standard deviation for each parameter (μ_{b0}, σ^2_{b0}, μ_{b1}, and σ^2_{b1}), and correlation coefficient ρ between β_1 and β_0. In this context, the bivariate normal distribution may be considered a distribution for uncertainty in the parameter estimates.

Three simulation methods are employed to demonstrate the effect of assuming a bivariate normal distribution for parameters vs. perfect correlation, or independent parameters. Specifically:

(1)   The slope and intercept of the regression line are described by a specific form of the bivariate normal distribution (i.e., follow *Steps 1, 2* in Exhibit B-9, and use Equation B-10 instead of *Step 4*).

(2)   The slope and intercept of the regression line are described by a general form of the bivariate normal distribution (i.e., follow *Steps 1 to 4* in Exhibit B-9).

(3)   The slope and intercept of the regression line are described by independent normal distributions (i.e., follow *Steps 1–4* in Exhibit B-9, but omit the correlation coefficient ρ in *Steps 2 and 4*). For each approach, Monte Carlo simulations with *I*=5,000 iterations were run to determine the set

of parameter values (β_0, β_1) for a simple linear regression equation. Typically, the uncertainty in the parameter estimates is not accounted for when simple linear regression equations are used to relate to exposure variables in a model. Such an approach may fail to account for important sources of parameter uncertainty. Figure B-10 (middle panel) illustrates the preferred approach for characterizing parameter uncertainty based on the bivariate normal distribution. (Note that the correlation coefficient relating the intercepts and slopes generated from the simulation is consistent with the correlation coefficient that describes the bivariate normal distribution; this is a good check that the simulation was set up correctly and run for a sufficient number of iterations). These results are contrasted with results using a form of the bivariate normal (Equation B-10) that underestimates uncertainty (top panel) unless parameters are perfectly correlated. In addition, the simplistic approach of sampling from independent normal distributions (bottom panel), yields a "shot gun" scatter plot. Sampling from independent normal distributions results in unlikely extreme combinations of the slope and intercept more often than the correct bivariate normal approach; propagating this bias through a risk model may severely bias estimates of uncertainty in risk.

Table B-9. Zinc concentrations in paired (i.e., co-located) soil and dust samples (ppm) for n=21 locations.

Sample	Soil (X_i)	Dust (Y_i)	Sample	Soil (X_i)	Dust (Y_i)
1	120	216	12	560	200
2	190	149	13	560	256
3	270	83	14	720	496
4	285	508	15	800	239
5	310	215	16	880	203
6	340	219	17	910	757
7	350	203	18	1035	676
8	380	101	19	1445	426
9	440	178	20	1600	522
10	480	232	21	1800	276
11	560	199			

Bivariate Normal Distribution for Parameters of the Regression Equation		
B_0	mean	173.9
	variance	4162.2
B_1	mean	0.193
	variance	0.0063
s^2		27857.4
Cov (B_0, B_1)		-4.2428
r		-0.8254

$y = 0.193x + 174$
$r^2 = 0.24$
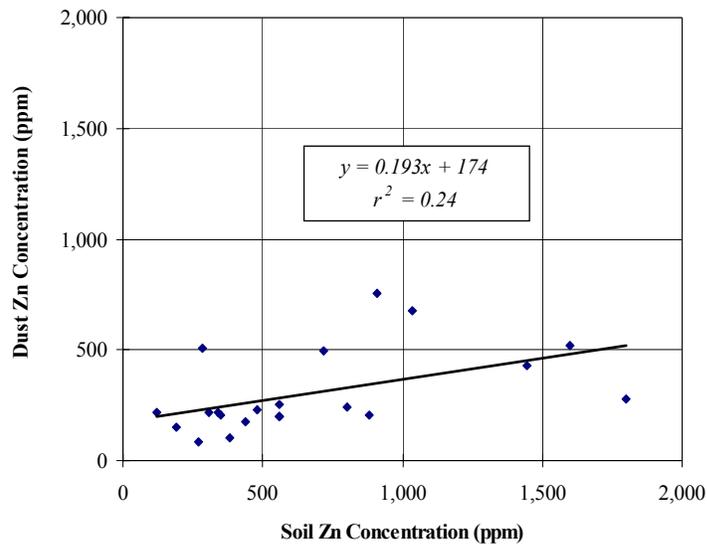
Figure B-9. Simple linear regression of zinc concentrations in soil and dust.
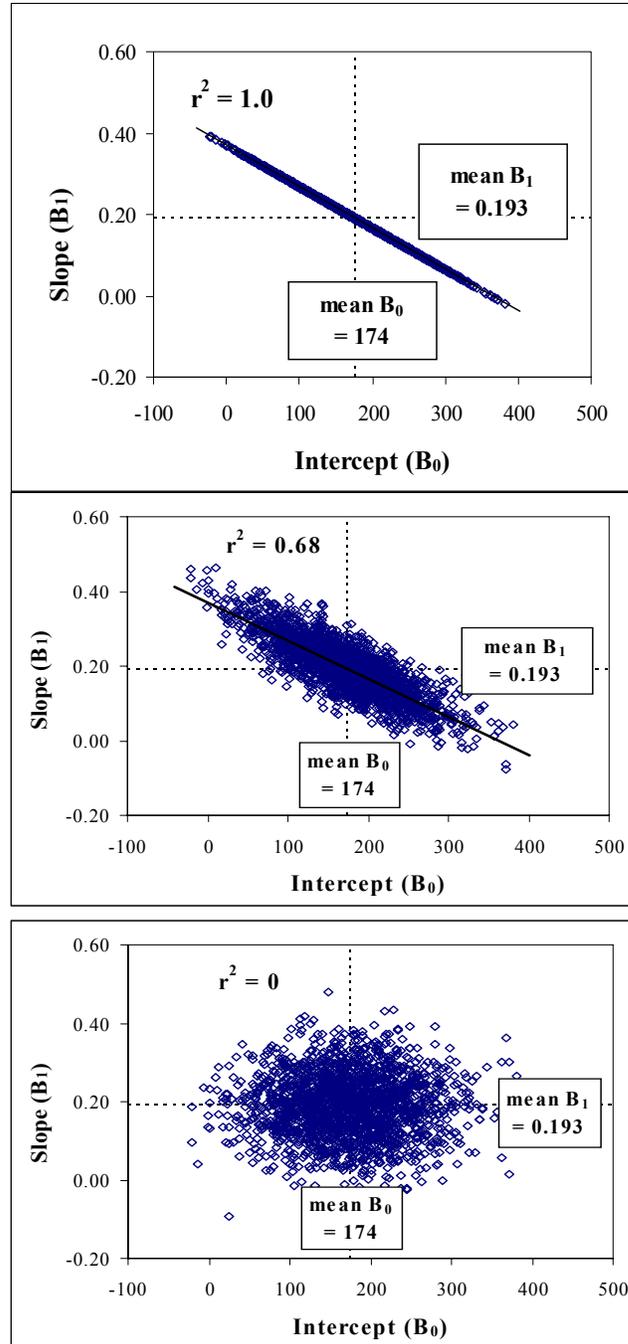
Figure B-10. Results of Monte Carlo simulation (*n*=5000 iterations) to estimate the slope and intercept of a regression equation. Top panel reflects the bivariate normal distribution for the special case that fails to capture the parameter uncertainty; middle panel reflects the preferred bivariate normal distribution with ρ=-0.825 based on empirical paired data; bottom panel reflects sampling from independent normal distributions.

REFERENCES FOR APPENDIX B

Brainard, J. and D.E. Burmaster. 1992. Bivariate Distributions for Height and Weight of Men and Women in the United States. *Risk Anal* 12(2):267–275.

Brately, P., B.L. Fox, and L.E. Schrage. 1987. *A Guide to Simulation*. Springer-Verlag, NY.

Burger, J., W. L. Stephens, Jr., C. S. Boring, M. Kuklinski, J.W. Gibbons, and M. Gochfeld. 1999. Factors in Exposure Assessment: Ethnic and Socioeconomic Differences in Fishing and Consumption of Fish Caught along the Savannah River. *Risk Anal*. 19(3):427–438.

Calabrese, E.J., Stanek, E.J., and Barnes R. 1996. Methodology to Estimate the Amount and Particle Size of Soil Ingested by Children: Implications for Exposure Assessment at Waste Sites. *Regul. Toxicol. Pharmacol*. 24:264–268.

Charney, E., J. Sayre, and M. Coulter. 1980. Increased Lead Absorption in Inner City Children: Where Does the Lead Come From? *Pediatrics* 65:226–231.

Conover, W.J. 1980. *Practical Nonparametric Statistics*. John Wiley & Sons, NY.

Cullen, A.C. and H.C. Frey. 1999. Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press.

d'Agostino, R.B. and M.A. Stephens. 1986. *Goodness-of-fit techniques*. Marcel Dekker, Inc, NY.

Filliben, J.J. 1975. The Probability Plot Correlation Coefficient Test for Normality. *Technometrics* 17(1):111–117.

Finley, B.L. and D.J. Paustenbach. 1994. The Benefits of Probabilistic Exposure Assessment: Three Case Studies Involving Contaminated Air, Water and Soil. *Risk Anal* 14(1):53–73.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Hostrand Reinhold, NY.

Gilliom, R.J. D.R. Helsel. 1986. Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 1. Estimation Techniques. *Water Resources Research*. 22:135–146..

Hahn, G.J. and S.S. Shapiro. 1967. *Statistical Models in Engineering*. John Wiley & Sons, NY.

Helsel, D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*. Elsevier. Amsterdam.

Hoffman, F.O. and J.S. Hammonds. 1992. *An Introductory Guide to Uncertainty Analysis in Environmental and Health Risk Assessment.* ES/ER/TM–35. Martin Marietta.

Hora, S.C. 1992. Acquisition of Expert Judgment: Examples From Risk Assessment. *J. Energy Eng*. 118(2):136–148.

Johnson, N.L., S. Kotz, and N. Balakrishnan. 1995. *Continuous Univariate Distributions*. Volume 2, Second Ed. John Wiley & Sons, NY.

Law, A.M. and W.D. Kelton. 1991. *Simulation Modeling and Analysis*. McGraw-Hill, NY.

Looney, S.W. and T.R. Gulledge. 1985. Use of the Correlation Coefficient with Normal Probability Plots. *American Statist.* 39:297–303.

Mendenhall, W. and R.L. Scheaffer. 1973. *Mathematical Statistics with Applications*. Duxbury Press.

Mood, A.M. and F.A. Graybill. 1963. *Introduction to the Theory of Statistics*. Second Edition. McGraw-Hill, Inc.

Morgan, G.M. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, NY.

Nelsen, R.B. 1986. Properties of a One-Parameter Family of Bivariate Distributions with Specified Marginals. *Comm. Stat. (Theory and Methods)* 15:3277–3285.

Nelsen, R.B. 1987. Discrete Bivariate Distributions with Given Marginals and Correlation. *Comm. Stat. (Simulation and Computation)* B16:199–208.

Oregon DEQ. 1998. *Guidance for the Use of Probabilistic Analysis in Human Health Exposure Assessments*. Waste Management and Cleanup Division. Interim Final. November.

Ott, W.R. 1990. A Physical Explanation of the Lognormality of Pollutant Concentrations. *J. Air Waste Manage Assoc.* 40(10):1378–1383.

Ott, W.R. 1995. *Environmental Statistics and Data Analysis*. CRC Press, Boca Raton.

Palisade Corporation. 1994. *Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3*. Windows Version Release 3.0 User's Guide, Palisade Corporation, Newfield, NY.

Roseberry, A.M. and D.E. Burmaster. 1992. Lognormal Distributions for Water Intake by Children and Adults. *Risk Anal.* 12(1):99–104.

Royston, J.P. 1982. An Extension of Shapiro and Wilk's *W* test for Normality to Large Samples. *Appl. Stat.* 31:115–124.

Snedecor, G.W. and W.G. Cochran. 1989. *Statistical Methods*. Eighth Edition. Iowa State University Press, Iowa.

Stanek, E.J. and Calabrese, E.J. 1995. Daily Estimates of Soil Ingestion in Children. *Environ. Health Perspect.* 103:176–285.

Thompson, K. 1999. Developing Univariate Distributions from Data for Risk Analysis. *Hum. Eco. Risk Assess.* 5(4):755–783.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley, Boston.

U.S. EPA. 1982. *Air Quality Criteria for Particulate Matter and Sulfur Oxides*. ECAO, ORD. EPA 600/8–82-029.

U.S. EPA. 1992. *Guidance for Data Useability in Risk Assessment, Part* A. Office of Emergency and Remedial Response, Washington, DC. OSWER Directive No. 9285.7-09A.

U.S. EPA. 1994. *Guidance for Conducting External Peer Review of Environmental Regulatory Models.* Office of the Administrator, Washington, DC. EPA/100/B-94-001. July.

U.S. EPA. 1997a. *Exposure Factors Handbook.* Office of Research and Development, Washington, DC. EPA/600/P-95/002Fa, Fb, and Fc.

U.S. EPA. 1997b. *Use of Probabilistic Techniques (Including Monte Carlo Analysis) in Risk Assessment*, Memorandum from Deputy Administrator Hansen and *Guiding Principles for Monte Carlo Analysis*. EPA/630/R-97-001.

U.S. EPA. 1999a. *Report of the Workshop on Selecting Input Distributions for Probabilistic Assessments.* Risk Assessment Forum. EPA/630/R-98/004. January.

U.S. EPA. 1999b. *Options for Development of Parametric Probability Distributions for Exposure Factors.* Office of Research and Development. Research Triangle Institute Final Report. April 6.

U.S. EPA. 2001. *Development and Evaluation of Probability Density Functions for a Set of Human Exposure Factors.* Office of Emergency and Remedial Response. University of California Draft Report. May.

Vose, D. 1996. *Quantitative Risk Analysis: A Guide to Monte Carlo Modeling*. John Wiley & Sons, NY.

Wonnacott and Wonnacott. 1981. *Regression: A Second Course in Statistics*. John Wiley & Sons, NY.

APPENDIX C

CHARACTERIZING VARIABILITY AND UNCERTAINTY IN THE CONCENTRATION TERM

C.0   THE CONCENTRATION TERM AND THE EXPOSURE UNIT

Incomplete knowledge of the concentration of one or more chemicals in various exposure media is often the major source of uncertainty in Superfund risk assessments. In any risk assessment, the derivation of the concentration term will reflect assumptions about: (1) properties of the contaminant, (2) the spatial and temporal variability in contamination, (3) the behavior of the receptor, and (4) the time scale of the toxicity of the chemical(s). This appendix expands upon concepts introduced in Chapter 5. This appendix does not provide detailed equations for performing calculations, but instead refers the reader to other Environmental Protection Agency (EPA) guidance documents in which both the recommended approaches and calculations are provided.

The concentration term is linked to the concept of an exposure unit (EU). For Superfund risk assessments, an EU is the geographical area in which a receptor is randomly exposed to a contaminated medium for a relevant exposure duration. Environmental sampling provides information about the contamination within and around an EU. Multiple EUs may be defined at a site based on the choice of a receptor, the exposure medium, and the nature of contact with the medium. For example, residential exposures to children may involve exposures via soil and dust ingestion both at the primary residence and recreational areas at a day care facility. Site-specific information regarding the activities of receptors should guide assumptions about the receptor's contact with exposure media.

> ☞ *Defining the EU is critical to the success of the remedial strategy, as it affects the calculation of the concentration to which receptors are exposed.*

C.1.0   VARIABILITY IN PRA

In general, variability and uncertainty should be kept separate to the extent possible in any probabilistic risk assessment (PRA). For example, assume a one-dimensional Monte Carlo Analysis (1-D MCA) was developed to characterize variability in risk, but it combined a distribution for uncertainty in mean concentration with distributions for variability in exposure variables. The result would yield a single distribution for risk, however, each risk estimate would reflect both uncertainty and variability and distinguishing between the two would not be possible. Therefore, EPA's *Guiding Principles for Monte Carlo analysis* recommends against mixing distributions of variability and uncertainty in a 1-D MCA (U.S. EPA, 1997b) to avoid such ambiguities.

A fundamental concept in Monte Carlo analysis is that there is variability in exposure between receptors (inter-individual variability) as well as day-to-day variability for each individual (intra-individual variability). In most Tier 2 analyses (see Chapter 2), the goal of a 1-D MCA is to characterize inter-individual variability in exposure and risk. Typically, probability distributions for exposure represent variability (PDFv's) between individuals in the average value over the entire exposure duration. In this case, the exposure point concentration (EPC) should represent the average exposure concentration over the entire exposure duration. Because an EPC is calculated from a sample, there is uncertainty that

the sample mean equals the true mean concentration within the EU; therefore, to account for associated uncertainty, the 95% upper confidence limit for the mean (95% UCL) is generally used for Superfund risk assessments (U.S. EPA, 1992).

In a 1-D MCA, a point estimate for the EPC is combined with PDFv's for other variables to yield a probability distribution for risk. An alternative approach is to simulate long-term average exposures as a series of consecutive short-term exposure events. This approach is referred to as MicroExposure Event (MEE) Monte Carlo modeling, and is discussed in detail in Appendix D. In MEE modeling, the goal is to develop PDFv's for exposure variables that capture the event-to-event variability in exposures at the individual level. The concept of an averaging time still applies, but generally to a shorter time frame. For example, seasonal variability in exposure frequency might be expected among outdoor occupational workers so that different PDFv's are representative of inter-individual for each season. In this case, the EPC continues to represent an average concentration within the EU, but it would be linked to season-specific activity patterns. It may be important to develop two different weighted averages to reflect season-specific activity patterns and locations that are more frequently contacted in the summer compared with the winter, for example. As the time frame for the exposure scenario is shortened from the entire exposure duration, to a season, to a day, to an individual event, the concentration term should be reevaluated to assess the relevance of the assumption that concentrations contacted by the receptor are represented by the mean of the measured sample.

The following discussion introduces concepts of temporal and spatial variability as they apply to the estimate of the EPC for different exposure media and exposure scenarios. While the general rule of thumb applies to all Monte Carlo models—use a measure of the average concentration within the EU over the time frame of exposure—it is important to apply the site sampling data in a way that is consistent with the exposure scenario.

C.1.1 TEMPORAL VARIABILITY

Temporal variability in chemical concentrations may be an important consideration when developing a preliminary remediation goal (PRG) for any exposure medium (refer to Chapter 5 for a comprehensive discussion of using PRA to evaluate PRGs). For example, wind erosion may change chemical concentrations in surface soil over time; leaching may change concentrations in both subsurface soil and groundwater; and bioaccumulation may result in increasing concentrations in predatory fish with time. If possible, such factors should be considered early in the risk assessment process and included in the conceptual site model.

Development of the EPC normally will depend on the averaging time relevant to the exposure scenario and health endpoint of concern. In the shorter term, it may be unlikely that receptors are exposed throughout the entire EU due to temporal (and spatial) variability in the contaminant and inter-individual variability in activity patterns. Therefore, inter-individual variability in the EPC might be expected, and a distribution of EPCs may be developed to represent differences in exposure among the population. Variability in short-term exposure may be an important factor for assessing variability in acute toxicity. However, over time, short-term variability in the EPC will tend to smooth out and approach a long-term average concentration. A single estimate of the long-term average EPC may be reasonable to use in assessing risks to the receptor population. This is true regardless of the underlying distribution of the environmental sampling data (e.g., lognormal, normal, beta, etc.).

While most chemicals regulated by the Superfund program are based on concerns for chronic toxicity (e.g., lifetime cancer risk from exposure to a carcinogen for ten or more years), for some

chemicals, toxic effects occur with shorter exposure durations (e.g., nitrate in drinking water and methemoglobinemia in infants). Differences between acute and chronic health endpoints are important to consider for ecological receptors such as transient migratory species. Superfund guidance distinguishes between acute and chronic exposure to provide risk assessors the option of evaluating risk under different time frames. The EPC should be estimated within an EU during a period of time that has toxicological relevance for the exposed population.

☞ *The time scale of the concentration term should match the time scale of the toxicity criterion and exposure duration.*

C.1.2 SPATIAL VARIABILITY

Spatial variability in chemical concentrations is also an important property to consider when developing a PRG. Spatial variability arises from many factors, including the mechanism of contamination, physical and chemical dilution and transformation processes, and physical characteristics of the site (Cullen and Frey, 1999). Similarly, receptors may exhibit spatial variability in their contact with an exposure medium. In general, receptors are assumed to have equal access to all areas within an EU so that the concept of a long-term average concentration is applicable.

Often, the EPC is estimated without regard to the spatial patterns in contamination. The sampling design yields a measure of the variability in concentrations that is assumed to be representative of the receptor's contact with the exposure medium. However, even when the sampling design is representative (e.g., both are simple random samples within the EU), the concentrations may exhibit clear spatial patterns that could be used to reduce uncertainty in the EPC. Geostatistics (see Section C.5.2 and Appendix D) offers a wide range of techniques for incorporating spatial information into estimates of the EPC. These techniques are particularly useful when there is uncertainty in the representativeness of site sampling, due to a difference in scale between site sampling and the size of the EU, or the use of targeted sampling designs that oversample areas within an EU believed to contain the highest levels of contamination.

In point estimate risk assessments (Tier 1 of the PRA), the EPC is most often characterized by a point estimate of the mean concentration, typically given by the 95% UCL for the mean to account for uncertainty in the site characterization (U.S. EPA, 1992). Variability in concentrations is an important consideration for determining appropriate statistical methods used to estimate the 95% UCL. In addition, for some Monte Carlo models, a PDFv may be developed to determine the EPC for the exposure model. A PDFv for the EPC may be warranted in short-term exposure scenarios, particularly when the sampling density is relatively sparse in relation to the size of the EU (i.e., poor site characterization). For example, a risk assessment may include a future use residential scenario (e.g., currently the site is undeveloped) in which the EPC that is relevant to a potentially exposed population of children is the average concentration within a 0.5 acre lot. If the soil sampling yields 100 measurements, but a small subset of the samples (e.g., less than three) are available for any 0.5 acre area, the most appropriate measure of the average concentration for a hypothetical residence may be the maximum detected concentration or a single value from the PDFv in concentration among hypothetical receptors. In general, for any of the EU's that define a randomly located residence, the poor site characterization would be a source of uncertainty in both a point estimate and probabilistic risk assessment.

At the vast majority of sites, concentration data is the easiest data to obtain of all the exposure variables. In cases of poor site characterization, risk managers may opt to perform a point estimate risk assessment only using the maximum detected concentration and highly protective exposure assumptions.

In the scenario described above for 0.5 acre residential lots, it is possible that a residence would be located in an area in which the average concentration is represented by the maximum detected concentration in the sample. Should the risk manager opt for a Tier 1 point estimate risk assessment, the use of the maximum detected concentration of a chemical on the site should ensure the performance of a health-protective risk assessment within a smaller EU.

Consideration of variability is also warranted in short-term scenarios for ecological risk assessment (ERA) when the EU is much smaller than the site (see Section C.3.1.1). For example, the home range of the receptor populations may be relatively small in comparison to the spatial distribution of sampling locations (e.g., benthic invertebrates living in the sediment at the bottom of a river or soil invertebrates in a terrestrial habitat). In these cases, the receptor would be exposed to an area smaller than the sampling grid or measure of areal sampling density. A value from the PDFv that characterizes variability in the concentrations across a relatively large spatial scale may be used to define the EPC for a receptor population at a smaller scale. Again, risk assessors should take care in designing a 1-D Monte Carlo model when using a PDFv for the concentration term. It is unadvisable to mix a PDFv for the concentration term with PDFv's for other exposure scenarios when estimating risks within one EU. Use of the PDFv in this manner would incorrectly suggest that the mean concentration varied for each individual within the same EU according to the variability in concentration measured across a much larger area. A preferred approach is to use a PDFv to obtain a point estimate that represents the EPC, and then combine this point estimate with PDFv's for other variables in the Monte Carlo simulation to estimate risks in the small EU. If there are many EU's at a site, or if the boundaries of EUs are undefined, more advanced modeling approaches can be developed to efficiently run multiple scenarios. Methods for characterizing exposure point concentrations for ecological receptors are further discussed in Sections C.2 and C.3.

C.1.3   EXAMPLE OF TEMPORAL AND SPATIAL VARIABILITY

Exposure scenarios often require consideration of both temporal and spatial variability. The MEE might be used to assess temporal variability by simulating long-term intake as the sum of individual exposure events. The time step for MEE is an important consideration and will depend on the rate of change of the most rapidly changing exposure variable. In addition, there should be a correspondence between the time periods over which data were obtained and the time step used in the MEE model. For example, when a MEE is used for the risk assessment, the concentration term selected at each time period should match the "average" concentration within the EU appropriate for that particular time period. Assume that the receptor is a residential child, and the time period is a single day, and the child may contact only 1,000 square feet within the 0.5 acre (20,000 square feet) residential EU. The specific 1,000 square foot area may change with each day as the child chooses different areas in the yard to frequent. Hence, the variability in the sample may be a more appropriate measure of the concentration contacted by residential child receptor on a day-to-day basis than the long-term average within the 0.5 acre EU. Over the long-term, this receptor will be exposed to the entire EU and hence the average contaminant concentration within the 0.5 acre EU. Note that the day-to-day variability in concentration undergoes the familiar phenomenon of "regression to the mean" when considered over the long-term.

C.1.4   SPATIAL AND TEMPORAL VARIABILITY FOR DIFFERENT EXPOSURE MEDIA

C.1.4.1 VARIABILITY OF CONCENTRATIONS IN SOIL

Surface soil is subject to erosion by wind and surface water runoff. Over time, concentrations in surface soil may change, but generally at a slow rate relative to other media. The spatial variability of chemical contamination is most often due to the mechanism by which the contamination occurred. For example, particulate stack emissions will tend to fall in an even pattern downwind of the stack whereas over-application of pesticides and chemical spills can result in a patchy pattern of contamination.

Subsurface soil is not subject to wind erosion, so concentrations change mostly due to degradation processes or leaching of the contaminant to groundwater. At most Superfund sites, concentrations of chemicals in subsurface soil will remain relatively constant.

C.1.4.2 VARIABILITY OF CONCENTRATIONS IN GROUNDWATER

Exposure to groundwater contamination mostly occurs at a fixed point in space (e.g., the wellhead). Groundwater is subject to a variety of influences that can alter chemical concentrations within this medium such as aerobic and anaerobic biodegredation, volatization, and absorption. Due to these influences, monitored natural attenuation is an appropriate remedy under certain site conditions. If a risk assessor wishes to use a measure of the long-term average of a concentration in groundwater, a hydrogeologist should be consulted.

C.1.4.3 VARIABILITY OF CONCENTRATIONS IN SURFACE WATER

Concentrations in surface water can be very dynamic. Streams are constantly flowing and the effects of mixing, dilution and evaporation can change the chemical concentrations in surface water over relative short time periods. Any sampling of surface water is truly a "snapshot" in time. The sampling methods used to characterize spatial and temporal variability of concentrations in surface water will have a direct effect on the uncertainty in estimates of the average concentration over both short and long time frames.

C.1.4.4 VARIABILITY OF CONCENTRATIONS IN SEDIMENT

In some situations, sediment may be considered a relatively stable medium, similar to soil. Alternatively, sediment may be physically moved by currents, tides, the movement of ships and other events. Trend analysis may be used to establish the long-term average sediment transport at a site. This information could provide the basis for choosing a representative "average" concentration in the sediment available to ecological receptors (Piest and Miller, 1975; Van Sickel and Beschta, 1983; Walling, 1983; Meade et al., 1990).

C.1.4.5 VARIABILITY OF CONCENTRATIONS IN FISH

Concentrations in fish may vary due to a change in the availability of food and environmental conditions. Factors that may be used to model population dynamics may include intensity of angler harvest, death/attrition of the population, and the introduction of a predator species or a more adaptive species. In risk assessments that include a fish ingestion exposure pathway, the activities of the angler may be a more important factor in determining the EPC than the changes in concentrations in fish over time. For example, an avid recreational angler may harvest fish from different locations within a lake and

consume fish of different sizes and species. In this way, with the consumption of contaminated fish, both the contaminated medium and the exposure point change throughout the exposure duration.

Unless, samples of fish are collected over time, knowledge of these factors will generally be unknown. Concentrations of bioaccumulative chemicals in territorial fish (e.g., largemouth bass) obtained in different locations will generally reflect the concentrations in the sediment in the individual's home territory. Concentrations of bioaccumulative chemicals in migratory fish will be more difficult to predict as the fish will contact areas with varying sediment and surface water concentrations.

C.1.4.6 EXAMPLES OF TEMPORAL AND SPATIAL VARIABILITY IN THE CONCENTRATION TERM FOR SELECTED EXPOSURE MEDIA

Whatever medium is considered in the development of EPCs, the risk assessor should be aware that the EPC embodies aspects of both the spatial distribution of contamination, the movement of the receptor, and possibly the contaminated medium within the EU. Table C-1 presents examples of sources of temporal and spatial variability in the concentration term based on both the contamination in selected exposure media and the receptor.

Table C-1. Examples of temporal and spatial variability in selected media for the concentration term in common exposure scenarios.

Factor		Soil	Groundwater	Fish
Temporal Variability	Contaminant	• none, if contaminant source is inactive • aerial deposition from ongoing source emissions affected by wind patterns • degradation over time • volatilization • migration to groundwater • radioactive growth and decay	• seasonal fluctuation in groundwater table • migration of contaminant plume • natural attenuation	• seasonal changes in species availability • bioconcentration • long-term changes in population dynamics • fish tissue concentrations linked to temporal variability in water and sediment concentrations • physical and chemical processes
	Receptor	• changes in activity patterns and behaviors over time (e.g., with age)	• none, fixed location at specific wellhead • changes in well location over time	• dietary preferences for fish species • cooking practices
Spatial Variability	Contaminant	• heterogeneity in concentrations over a small area and with depth, including presence of hotspots • heterogeneity in soil properties that influence bioavailability	• migration of contaminant plume, based on hydrogeology and source emissions (e.g., bulk flow or continuous source)	• migration of fish • changes in fish population structure
	Receptor	• daily activity patterns involve contact with different areas of the EU	• none, fixed location at specific wellhead • changes in well location over time	• change in recreational habits, and areas fished

C.2.0   NONRANDOM EXPOSURES

As discussed in Section C.1.2, in the long-term it is generally assumed receptors exhibit random movement, such that there is an equal probability of contacting any area within the entire EU. Therefore, the long-term exposure concentration will most likely be the arithmetic mean of the concentration within the EU. However, in many situations, the assumption of random exposures in space may clearly be an oversimplification. People's behavior and preferences will cause them to access specific areas within an EU with greater frequency than others. The same is true in terms of ecological receptors with specific habitat preferences.
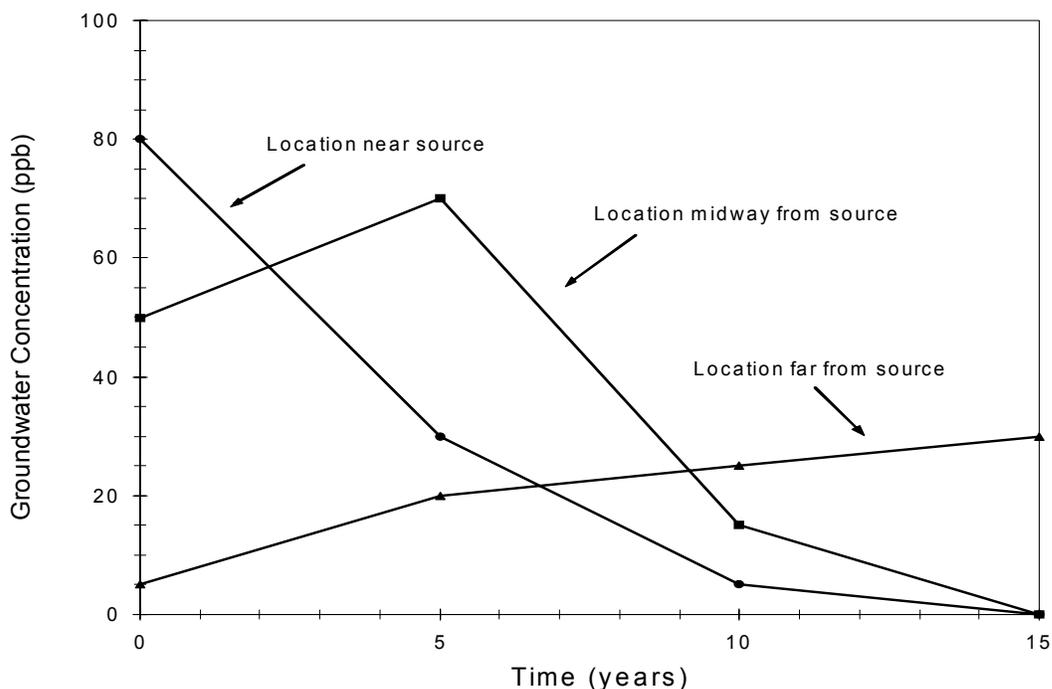
Figure C-1. Spatial and temporal variability in contaminant concentrations in groundwater.

For example, groundwater concentrations may show a large variation when sampled from wells in different locations (Figure C-1). Typically, residential receptors do not sample randomly from different wells, but draw chronically from individual wells. In such a case, the EU is a single wellhead. Fluctuations in the groundwater plume will depend on the hydrogeology of the site as well as the seasonal fluctuations in the water table. In this hypothetical example, concentrations are declining over time at distances nearest to the source, and concentrations are increasing as the plume moves farther from the source.

Incomplete information regarding the behavior patterns of people and environmental systems can be a large source of uncertainty in a risk assessment. Because of this, methods are being developed to model spatial relationships (between the contaminant and receptor) and nonrandom exposures. Recently, a quantitative technique to model nonrandom exposure has been proposed for ERA (Hope, 2000, 2001). Briefly, this technique divides the EU into smaller subunits and uses information about the attractiveness

of each subunit to assign a probability of the receptor occupying a given subunit for a period of time. Receptor movements are modeled stochastically and a time-weighted average of all the subunits provides a measure of the EPC. In some ecological risk assessments, telemetry data can be used to better characterize the areas of contamination that overlap with habitats of selected species. Hoff (1998) demonstrates an approach for American badgers (*Taxidea taxus*) in which telemetry data and geostatistical modeling provide an improved relationship between contaminant concentrations, tissue residues, and effects.

C.3.0   SOURCES OF UNCERTAINTY IN THE CONCENTRATION TERM

There are numerous potential sources of uncertainty in the estimate of the true mean concentration within an EU. As discussed in Chapter 5 (Section 5.1.1), sources of uncertainty can be grouped into four broad categories: sample data, location of the EU, behavior of the receptor, and from miscellaneous sources (e.g., physical and chemical processes). Development of an uncertainty distribution for the average concentration requires knowledge of the variability in chemical concentrations within the EU (unless distribution-free approaches are used), the toxicity of the chemicals, and the receptor's behavior. These distributions should be developed by risk assessors with the concept of the EU in mind. Differences in scale (e.g., small home range of an ecological receptor population relative to the site sampling design) can be a major source of uncertainty in ecological risk assessments. Methods for addressing such uncertainties in the concentration term are presented below. By incorporating these methods into the quantitative uncertainty analysis, risk managers may more effectively evaluate the importance of data-gaps and design subsequent rounds of site sampling to reduce the uncertainty in the EPC.

C.3.1   QUANTIFICATION OF UNCERTAINTY BASED ON THE SIZE OF THE EXPOSURE UNIT

Site characterization sometimes occurs before an EU has been defined. Therefore, an EU may be smaller than an entire site, equal to the site itself, or larger than the site. These three conditions lead to different conclusions and methods about the determination of the EPC. The most complex situation is when the EU is smaller than the site and the site can contain multiple EUs. For future scenarios in which the land use differs from the current land use, the difficulty in predicting the exact size and location of EUs necessitates accounting for the uncertainty in the EU.

Composite sampling is often used to maximize site information. However, it is important to note that the use of composite sampling influences the concentration term. If composite sampling is used exclusively at a site, the actual maximum concentration present or the best estimate of this maximum concentration will not be available. Depending on the time scale of the toxic effect or whether acute toxicity should be considered, this lack of knowledge of the maximum concentration present may be a large data gap. Risk assessors are urged to consider composite sampling and its ramifications for the concentration term.

C.3.1.1 WHEN THE EXPOSURE UNIT IS SMALLER THAN THE SITE

The size of the EU will be different depending on the length of exposure. A receptor can access a greater area if given more time. In almost all cases, the size of the EU for short-term exposure will be smaller than the EU for long-term exposure. Therefore, in addition to the uncertainty associated with sampling and analysis (which can be quantified with existing methods for calculating confidence intervals), there is uncertainty about the location of the EU within the site.

If contamination is evenly spread across the site, the location of the EU may not have any bearing on the EPC. In such a case, uncertainty may depend on the sample size or density of measurements within the EU relative to the entire site. In point estimate risk assessments, the concentrations of chemicals at the sampling location that poses the greatest risk may be considered as estimates of the EPC for this small EU. Using this "riskiest" sampling location as an estimate of the mean within an EU of unknown location accounts for both the uncertainty associated with limited sampling within a single EU and the uncertainty of the location within the site of the EU.

To express the uncertainty in location of the EU as a distribution, methods have been developed to place an EU of a given size randomly about a site (Burmaster and Thompson, 1997). A concentration term is developed for each of a large number of randomly located EUs. The distribution of these concentration terms will express the uncertainty in the location of the EU.

Risk assessors are cautioned to consider whether the statistical method used to estimate the EPC in an EU accounts for all sources of uncertainty in the concentration term. If only a few samples are used to characterize the average concentration within an EU, then the uncertainty in the EPC is large and should be presented in the risk characterization. These conditions may warrant additional sampling or the use of analytical methods that account for spatial variability within the entire site.

At some sites, geostatistical methods, pattern recognition, and geographical information systems (GIS) methods may provide additional insight and will aid in the development of the concentration term (see Section C.5.2). Although Table 3-1 shows several statistical methods for estimating both point estimates and distributions that encode uncertainty in the concentration term, a risk assessor's understanding of these uncertainties should be conceptual as opposed to purely statistical.

C.3.1.2 WHEN THE EXPOSURE UNIT IS THE SAME SIZE AS THE SITE

In this case, the entire environmental data set within the site boundaries can be used for the determination of the concentration term. Assuming the EU occupies the entire site, then the source of uncertainty associated with knowing the average concentration within the EU is the sampling and analytical uncertainty.

C.3.1.3 WHEN THE EXPOSURE UNIT IS LARGER THAN THE SITE

In this case, the EU extends beyond the site boundaries. Therefore, the entire environmental data set within the site boundaries can be used for determination of the concentration term. However, an additional term in the exposure assessment may be needed to account for the fraction of the exposures that are expected to occur off site. Essentially, the contribution of the chemical concentrations measured on and off site are weighted by the fraction ingested or contacted in each area. Similarly, the term "area use factor" is used in ecological risk assessments to refer to the percentage of time or area an animal inhabits a contaminated area. An exposure scenario in which the EU is defined by the multiple locations that may be visited would be a common extension of this concept. One reasonable assumption regarding off site exposures is that the concentrations would be equal to the "background" concentrations. If this assumption is made, a site risk assessor should be consulted to determine appropriate methods for incorporating background concentrations into the risk assessment. Alternatively, additional sampling at off site locations would be needed to estimate the concentrations.

C.4.0 SUMMARY OF RECOMMENDATIONS FOR THE CONCENTRATION TERM

Table C-2 presents general guidelines for establishing a concentration term in various media based on exposure time and the size of the EU. These general guidelines along with site-specific exposure conditions are the driving factors in risk assessment decision making for establishing the concentration term.

Table C-2. Summary of factors that may be considered in developing an EPC.

Medium	Exposure Time	Random	Non-Random	Size of EU relative to the site/sampling density	Recommendation (Human Health and Ecological)
Soil	Short-term		X	small	HH - consider variability in concentration relative to the time scale of toxicity. ECO - time weighted average of smaller subunits.
Soil	Long-term	X		variable	HH, ECO - consider uncertainty in the average concentration within an EU.
Fish	Short-term		X	variable	HH, ECO - consider variability in sample concentrations relative to the exposure time.
Fish	Long-term	X		variable	HH - consider uncertainty in the average concentration in consumed portion of fish. ECO - consider uncertainty in average concentration of whole fish.
Ground-water	Short-term		X	small - single well head	HH - consider either the highest detected concentration or uncertainty around the concentration at the center of the plume as a measure of a single well and relate to the time scale of the toxic effect. ECO - not applicable
Ground-water	Long-term		X	small - single well head	HH - consider variability among the higher concentration samples as a protective EPC. Alternatively, hydrogeologic modeling may be used to obtain a long-term average concentration in the most contaminated area. ECO - not applicable

C.5.0 METHODS FOR ESTIMATING UNCERTAINTY IN THE MEAN CONCENTRATION

Confidence intervals (CIs) and UCLs are computed to characterize uncertainty in a parameter estimate. CIs can be computed for any parameter. The general method for estimating confidence intervals is presented in equation C-1.

$$CI = \text{parameter estimate} \pm (\text{critical value}) \times SE \qquad \text{Equation C-1}$$

The parameter estimate is the estimated value for the unknown population parameter. The critical value is the number, z, with probability, p, lying to its right (for an upper critical value) or left (for a lower critical value). For a standard normal distribution (i.e., arithmetic mean=0, standard deviation=1), critical values are referred to as the *z-score* or *z-statistic*. These values are commonly given in statistics texts, and

may also be calculated using the Microsoft Excel function *Normsinv(p)*, where *p* corresponds to the probability lying to the right of the value. Distributions that characterize parameter uncertainty are sometimes referred to as sampling distributions. The standard error (SE) is the standard deviation of the sampling distribution for the parameter estimate. The confidence interval conveys two concepts: (1) an upper and lower confidence limit (for a 2-sided CI), and (2) a confidence level $(1-\alpha)$, which gives the probability that the method yields an interval that encloses the parameter (Moore and McCabe, 1993). Methods for estimating SE vary for specific parameters. For example, the SE of a mean concentration may be calculated based on the sample variance and the sample size (due to Central Limit Theorem). Methods for calculating the SE for other parameters, such as the 95th percentile, are more complex, and may be estimated from a series of nested bootstrap simulations (Efron and Tibshirani, 1993; U.S. EPA, 2001a).

When comparing alternative approaches for quantifying parameter uncertainty, criteria that are important to consider include the variance of the original data set, and the bias and coverage of the CIs generated by each method. In statistics, a method is unbiased if the mean of the sampling distribution is equal to the true value of the parameter. Similarly, a method has accurate coverage if the probability *p* that a CI does not cover the true parameter is equal to the probability level used to construct the CI. For risk assessment, the most desirable method is one that deals well with high variance, yields CIs that are sufficiently wide (i.e., the CI does not underestimate the probability of enclosing the population parameter), and, more specifically, yields upper confidence limits that are not biased low. The choice of the most appropriate method will depend on the characteristics of the data set and a balance between two objectives: (1) the desire to be health protective and, therefore, have a low probability of underestimating the mean, and (2) a desire to be accurate, in the sense of choosing a method whose expected coverage equals the true coverage. As a general principle for quantitative uncertainty analysis, if alternative methods yield very different answers, it is helpful to explore the reasons for the differences. The objective is to explain why the estimates of the 95% UCL differ, and to determine if the differences are sufficiently great that they could alter the risk management decision or PRG. This information should be presented as part of the risk communication process associated with the scientific management decision points of the tiered process for PRA (see Chapter 2).

As discussed in Chapter 5, in Superfund risk assessment, the EPC is usually calculated as the 95% UCL for the mean to account for the uncertainty in estimating the average concentration within an EU. The 95% UCL is defined as a value that, when repeatedly calculated for randomly drawn subsets of size (*n*), equals or exceeds the true population mean 95% of the time. In other words, it is calculated and applied as a 1-sided confidence limit. The 95% UCL is one percentile on the probability distribution that characterizes uncertainty in the mean (i.e., the PDFu for the mean). It is equal to the 95th percentile of the sampling distribution for the mean. EPA's guidance on calculating the concentration term describes the rationale and methodology for selecting the 95% UCL as the point estimate for the concentration term (U.S. EPA, 1992).

Common methodologies for characterizing the 95% UCL for the arithmetic mean concentration include the following: (1) application of Equation C-1 using Student's t-statistic (for normal distributions), (2) Land method using H-statistic (for lognormal distributions) (Land 1971, 1975), and (3) bootstrap and Jacknife resampling techniques (Efron and Tibshirani, 1993). Details on these methods and on choosing an appropriate method are provided in the ORD/OSWER guidance bulletin, *Lognormal Distribution in Environmental Applications* (U.S. EPA, 1997a), and the more recent OSWER guidance bulletin, *Guidance on Calculation of UCLs at Superfund Sites* (U.S. EPA, 2001a). An overview of methods that may be used when data are not normal or lognormal is also provided by Schulz and Griffin (1999). It is the responsibility of the regional risk assessor to ensure that an appropriate method for

calculating a UCL or for developing an uncertainty distribution is chosen. Chapter 3 (Table 3-1) provides an overview of approaches for characterizing uncertainty in the concentration term in both 1-D MCA and 2-D MCA.

C.5.1 QUANTIFYING UNCERTAINTY WITHOUT INFORMATION ABOUT LOCATIONS OF SAMPLES AND RECEPTORS

Knowledge of both the sampling locations and the receptor's activity patterns with the EU can be used to derive a more representative estimate of the 95% UCL. If a risk assessor has access to an environmental data set without information about the sample locations, the risk assessor is forced to assume that the sample consists of a number of independent observations. The validity of this assumption depends on the unknown spatial variability of contamination at the site. The size and location of an EU, as well as the choice of a statistical method for estimating the distribution of uncertainty around the mean concentration will require often implicit (and possibly incorrect) assumptions about the spatial distribution of contamination. Similarly, if information regarding receptor activity patterns is unavailable, one must assume that any area within the EU is equally representative of potential exposures. The risk assessor is urged to explore the effects of these various assumptions and to make choices that are protective of human health and the environment.

C.5.2 QUANTIFYING UNCERTAINTY WITH INFORMATION ABOUT LOCATIONS OF SAMPLES AND RECEPTORS

In classical statistics, observations are assumed to be independent. This assumption is often invalid at contaminated sites where the method by which a chemical is released into the environment (e.g., deposition from airborne emissions; migration of contaminant plume from a point source) results in positive spatial autocorrelation. In other words, observations located next to each other tend to contain similar levels of contamination (i.e., redundant information) (Griffith and Layne, 1999). For example, the higher the spatial autocorrelation, the less incremental information is provided by adding observations in close proximity to existing observations. This decrease in the information content of a site sample is exacerbated by the tendency to choose sampling locations in the most contaminated areas rather than distributed at regular spatial intervals or specified using random sampling methodology.

At many hazardous waste sites, environmental sampling plans are designed with remedial actions rather than risk assessment in mind. Therefore, the risk assessor must establish a correspondence between the actual sampling locations and the locations a receptor would be expected to frequent. Geostatistics may provide information to establish this correspondence.

Geostatistics is a branch of spatial statistics that can be used to model spatial variability and parameter uncertainty. Geostatistics offers two fundamental contributions to risk assessment: (1) a group of methods to describe the spatial distribution of a contaminant in a quantitative fashion, and (2) the ability to maximize the information available in the data set (Deutsch and Journel, 1988; Isaacs and Srivastava, 1989).

Geostatistics is capable of using the information revealed by a correlation analysis of the data to estimate concentrations at unsampled locations. For example, geostatistics is able to use the spatial information contained in the data to model uncertainty in contaminant concentrations for areas where data are sparse, a situation commonly encountered in site assessment work. Using geostatistics, information from samples collected from outside an EU can be used to model the uncertainty in the mean concentration within an EU. Approaches that do not consider the geospatial information present in the

data are limited to the subset of samples within an EU. However, this ability to model uncertainty in areas where data are sparse is also limited, and a well characterized site is still the best path to understanding the risk at that site.

Geostatistical methods may be used to calculate a distribution of uncertainty in the mean of the concentration term for use in PRAs. In the past, geostatistics has not been widely applied to risk assessment, even though uncertainty in the exposure concentration is often a major source of uncertainty in risk estimates. Most risk assessors quantify uncertainty in the long-term average concentration without explicitly considering the spatial information present in data obtained from environmental sampling or knowledge of the receptor's movement and activities within the EU. When spatial information does not exist, the inherent assumption is that environmental sampling yields a data set that is representative of the spatial variability in concentrations encountered by a receptor. This assumption represents one source of uncertainty in the EPC. In addition, data collected outside an EU are often ignored in the analysis, even though they can provide a more comprehensive view of patterns of contamination across the site, including the EU of interest. Ignoring site-wide information may result in less informed estimates of risk and, therefore, less effective remedial designs (i.e., too little or too much remediation). In the past five years, with rapidly expanding software and hardware capabilities, some examples of the application of geostatistics can be found in exposure assessment and remedial design (e.g., Gomez-Hernandez, 1996; Goovaerts, 1996, 1997; Kriakidis, 1996; Ginevan and Splitstone, 1997; McKenna, 1997, 1998) as well as site assessment guidance (e.g., U.S. EPA, 2000).

A limit to applying geostatistics at hazardous waste sites is that the method is resource intensive and requires personnel experienced with the software and techniques. Risk assessors and risk managers should ensure that contractors and other personnel have the necessary capabilities before applying geostatistical methods to risk assessment or site cleanup. Geostatistics is a powerful tool, but it cannot incorporate quantitative knowledge regarding all sources of uncertainty. The risk assessor is cautioned to consider all possible sources of uncertainty as described in Chapter 5. As indicated previously, a full discussion of geostatistics is beyond the scope of this guidance, and interested readers are urged to consult the OSWER guidance document, *Guidance on Strategy for Surface Soil Cleanup at Superfund Sites* (U.S. EPA, 2001b).

EPA has produced several software packages used for geostatistical estimation. Among these are GEO-EAS and GEO-PACK. Expertise in geostatistics can be obtained from ORD/Las Vegas.

REFERENCES FOR APPENDIX C

Burmaster, D.E. and K.M. Thompson. 1997. Estimating Exposure Point Concentrations for Surface Soils for use in Deterministic and Probabilistic Risk Assessments. *Human Eco. Risk Assess*. 3(3): 363-84.

Cullen, A.C. and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum Press.

Deutsch, C.V. and Journel, A.G. 1998. *Geostatistical Software Library and User's Guide, 2nd Ed.*, Oxford University Press, NY.

Efron, B. and Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, CRC Press.

Ginevan, M.E. and D.E. Splitstone. 1997. Improving Remediation Decisions at Hazardous Waste Sites with Risk-Based Geostatistical Analysis. *Environ. Science Tech.* 31(2):92A–96A.

Gomez-Hernandez, J.J. 1996. Issues on Environmental Risk Assessment. In: *Proceedings of the Fifth International Geostatistics Congress*, Vol. 1. Baafi, E.Y. and N.A. Schofield, eds.). Kluwer Academic Publishers, Dordrecht, 15-26.

Goovaerts, P. 1996. Accounting for Local Uncertainty in Environmental Decision-Making Processes. In: *Proceedings of the Fifth International Geostatistics Congress*, Vol. 2. Baafi, E.Y and N.A. Schofield, eds.). Kluwer Academic Publishers, Dordrecht, 929-940.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.

Griffith, D.A. and L.J. Layne. 1999. *A Casebook for Spatial Statistical Analysis*. Oxford University Press, NY.

Hoff, D.J. 1998. Integrated Laboratory and Field Investigations Assessing Contaminant Risk to American Badgers (*Taxidea taxus*) on the Rocky Mountain Arsenal National Wildlife Refuge. Ph.D. Dissertation, Clemson University, Clemson, S.C.

Hope, B.K. 2000. Generating Probabilistic Spatially-Explicity Individual and Population Exposure Estimates for Ecological Risk Assessment. *Risk Anal.* 20(5):575–590.

Hope, B.K. 2001. A Case Study Comparing Static and Spatially Explicit Ecological Exposure Analysis Methods. *Risk Anal.* 21(6):1001–1010.

Isaacs, E.H. and R.M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, NY.

Kriakidis, P.C. 1996. Selecting Panels for Remediation in Contaminated Soils via Stochastic Imaging. In: *Proceedings of the Fifth International Geostatistics Congress*, Vol. 2. Baafi, E.Y. and N.A. Schofield, eds.). Kluwer Academic Publishers, Dordrecht, 973-983.

Land, C.E. 1971. Confidence Intervals for Linear Functions of the Normal Mean and Variance. *Ann. Math. Stat*. 42:1197–1205.

Land, C.E. 1975. Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance. In: *Selected Tables in Mathematical Statistics*, Vol. 3. American Mathematical Society, Providence, RI.

McKenna, S.A. 1997. *Geostatistical Analysis of Pu-238 Contamination in Release Block D, Mound Plant, Miamisburg, Ohi*o. SAND97-0270, Sandia National Laboratories, Albuquerque, NM.

McKenna, S.A. 1998. Geostatistical Approach for Managing Uncertainty in Environmental Remediation of Contaminated Soils: Case Study. *Environmental and Engineering Geoscienc*e, 4(2):175-184.

Meade R.H., T.R. Yuzyk, and T.J. Day. 1990. *Movement and Storage of Sediment in Rivers of the United States and Canada*. In: Wolman et al. (eds) Surface Water Hydrology. The Geology of North America. Geological Society of America, Boulder, CO.

Moore, D.S. and G.P. McCabe. 1993. *Introduction to the Practice of Statistics*. W.H. Freeman and Company, NY.

Piest R.F. and C.R. Miller. 1975. *Sediment Yields and Sediment Sources*. In: Vanoni V.A. (ed.) Sedimentation Engineering, American Society of Civil Engineers, NY.

Schulz, T.W. and S. Griffin. 1999. Estimating Risk Assessment Exposure Point Concentrations when the Data are not Normal or Lognormal. *Risk Anal*. 19(4): 577– 584.

U.S. EPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. Office of Solid Waste and Emergency Response. Washington, DC. OWSER Directive No. 9285.7-081.

U.S. EPA. 1997a. *Lognormal Distribution in Environmental Applications*. Office of Research and Development and Office of Solid Waste and Emergency Response. Washington, DC. EPA/600/R-97/006.

U.S. EPA. 1997b. *Memorandum from Deputy Administrator Fred Hansen on the Use of Probabilistic Techniques (including Monte Carlo Analysis) in Risk Assessment, and Guiding Principles for Monte Carlo Analysis*. Office of Research and Development, Washington, DC. EPA/630/R-97/001. May 15.

U.S. EPA. 2000. *Statistical Estimation and Visualization of Ground-water Contamination Dat*a. Office of Research and Development, Washington, DC. EPA/600/R-00/034.

U.S. EPA. 2001a. *Draft Guidance on Calculation of Upper Confidence Limits for Exposure Point Concentrations at Superfund Sites*. Office of Emergency and Remedial Response, Washington, DC.

U.S. EPA. 2001b. *Supplemental Guidance for Developing Soil Screening Levels for Superfund Sites*. Peer review draft. Office of Solid Waste and Emergency Response. Washington, DC. OSWER No. 9355.4-24. March.

Van Sickel J. and R.L. Beschta. 1983. Supply-Based Models of Suspended Sediment Transport in Streams. *Water Resour. Res*. 19:768–78.

Walling D.E. 1983. The Sediment Delivery Problem. *J. Hydrol*. 65:209–37.

APPENDIX D

ADVANCED MODELING APPROACHES FOR CHARACTERIZING VARIABILITY AND UNCERTAINTY

D.0   INTRODUCTION

This appendix briefly describes the following advanced modeling approaches that can be used in probabilistic risk assessment (PRA) to characterize variability and uncertainty: two-dimensional MCA (2-D MCA), microexposure event analysis (MEE), geospatial statistics, and Bayesian analysis. Except for 2-D MCA, these approaches can also be applied to point estimate risk assessment. The application of many of these approaches will require access to expertise in specialized areas of statistics and, in some cases, specialized or even custom-designed computer software. The intent here is to introduce some of the basic concepts and terminology, as well as to provide references where the reader can find more exhaustive coverage of these topics.

D.1.0   EXPRESSING VARIABILITY AND UNCERTAINTY SIMULTANEOUSLY

A Monte Carlo analysis that characterizes either uncertainty or variability in each input variable (see Chapter 1) can be described as a one-dimensional Monte Carlo analysis (1-D MCA). A 2-D MCA is a term used to describe a model that simulates both uncertainty and variability in one or more input variables. All probability distributions that are used to describe variability in a PRA model have a certain degree of associated uncertainty. For example, suppose variability in soil concentration (ppm) is estimated using a normal probability density function (PDF) defined by a mean (μ_{soil}=5) and standard deviation (σ_{soil}=1), and subjectively truncated (min, max) at (0, 50). Uncertainty in the parameter estimates can be represented in a PRA model by assuming both parameters are also random variables. To illustrate this concept, assume normal PDFs for *uncertainty* can be specified for both parameters. Uncertainty in the mean is described by the normal PDF with parameters (μ_{mean}=5, σ_{mean}=0.5); similarly, uncertainty in the standard deviation is described by the normal PDF with parameters (μ_{SD}=1, σ_{SD}=0.5). Model variables are represented in this manner when there is a compelling reason to believe that a unique probability distribution does not adequately describe one's knowledge of each variable in the model. A variable described in this way is called a second order random variable. Figure D-1 (Panel A) shows a collection of *n*=20 cumulative probability distributions (CDFs), each curve representing a unique set of (mean, SD) parameter estimates for the normal PDF for variability. Panel B shows the 90% *confidence interval*[1] based on 2,500 simulated CDFs. The 95% lower and upper bounds correspond to the distribution of 5th percentiles and 95th percentiles, respectively (i.e., CDF for 2,500 5th percentiles and CDF for 2,500 95th percentiles). The 90% credible interval (CI) for the 50th percentile is (3.4, 6.7).

---

[1]Note that the term "credible interval" may be more appropriate than "confidence interval" given that the range is based on subjective as well as statistical considerations. Brattin, Barry, and Chiu (1996) provide additional examples of uncertain PDFs that illustrate this concept.
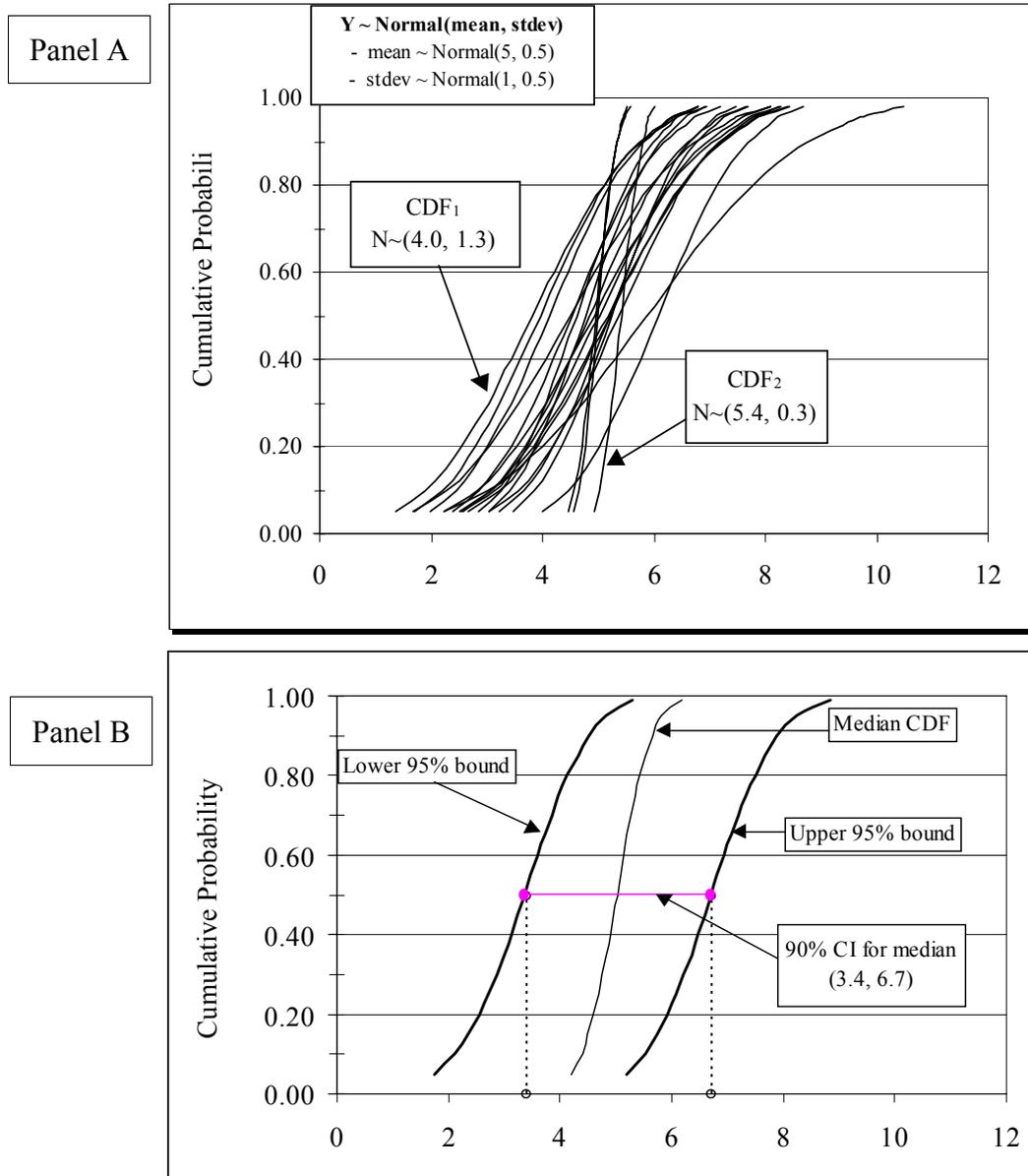
Figure D-1. Panel A shows a family of 20 CDFs for a hypothetical random variable, *Y* (e.g., concentration in units of ppm), characterized by a normal PDF where both the mean and SD are also random variables representing uncertainty in the parameter estimates: Mean~ Normal(5, 0.5), SD~ Normal(1, 0.5). Each CDF represents a single simulation of *n*=2500 iterations using a unique set of parameters. For example, CDF_1 represents N~(4.0, 1.3) while CDF_2 represents N~(5.4, 0.3). **Panel B** shows the "90% credible interval" for the CDF based on 2,500 *simulations*, each simulation using *n* = 2500 iterations (i.e., a 2-D MCA with 2,500 outer loop iterations and 2,500 inner loop iterations). Lower, median, and upper bounds represent the simulated 5th, 50th, and 95th percentiles, respectively. The 90% confidence interval for the estimate of the 50th percentile is: {3.4, 6.7}.

EXHIBIT D-1

DEFINITIONS FOR APPENDIX D

Bayesian Statistics - A specialized branch of statistics that views the probability of an event occurring as the degree of belief or confidence in that occurrence.

Geospatial Statistics - A specialized branch of statistics that explicitly takes into account the georeferenced context of data and the information (i.e., attributes) it contains.

Frequentist - A term referring to classical statistics in which the probability of an event occurring is defined as the frequency of occurrence measured in an observed series of repeated trials.

Image Analysis - A technique in geostatistics used to restore a degraded image or interpret images that have been contaminated by noise or possibly some nonlinear transformation.

Kriging - A geostatistical method of spatial statistics for predicting values at unobserved locations.

Likelihood Function - A Bayesian term referring to a probability distribution expressing the probability of observing a piece of new information given that a particular prior belief is true.

Location Tag - The spatial coordinates of a sampling location (e.g., longitude, latitude).

Microexposure Event Analysis (MEE) - An approach to modeling exposure in which long-term exposure of an individual is simulated as the sum of separate short-term exposure events.

Point Pattern Analysis - A technique in geostatistics of restricting the analysis to location information, ignoring attribute information, addresses two location problems: (1) describing points according to spacing, and (2) describing points according to density.

Posterior Distribution - A Bayesian term referring to a probability distribution that has been updated with new information.

Prior Distribution - A Bayesian term referring to the hypothesized, expected, or calculated probability distribution for an event prior to the collection of new information.

Spatial Autocorrelation - The tendency of data from locations that are relatively close together to be geographically correlated.

Thiessen (Voronoi) Polygon Analysis - A method of spatial statistics in which an area is subdivided into subregions, or polygons, in order to predict values at unobserved locations.

Time Step - A modeling term used to describe the time interval within which variable values do not change.

Two-Dimensional Monte Carlo analysis (2-D MCA) - Separate representation of variability and uncertainty in an MCA, usually accomplished using nested computation loops.

In the example shown in Figure D-1, the mean and standard deviation for soil concentration were allowed to vary independently. Thus, a distribution could be defined by a combination of a low mean and a high standard deviation, high mean and low standard deviation, or any other combination in between. The assumption of independence of variable parameters may not be valid in all cases. It may be unreasonable to assume that a high mean soil concentration would occur with a low standard deviation. An alternative assumption would be that the standard deviation of the mean is a constant proportion of the mean (i.e., a constant coefficient of variation). Correlations between parameters should be considered in the design of the PRA. One approach that is especially useful for characterizing relationships between the slope and intercept of a simple linear regression is to specify the bivariate normal distribution for the parameter estimates.

D.2.0 TWO-DIMENSIONAL MONTE CARLO ANALYSIS (2-D MCA)

Two-dimensional MCA is an approach for computing risk (or hazard) when combining distributions that represent variability and uncertainty. In 2-D MCA, distributions representing variability and uncertainty are sampled using nested computational loops (Figure D-2). The inner loop simulates variability by repeatedly sampling values for each variable from their defined probability distributions. With each circuit of the outer loop, new parameter values for each variable are selected, and the inner loop sampling is repeated. The result is a collection of inner loop simulations, one for each parameter value selected. If the inner loop samples 5,000 times, and the outer loop samples 1,000 times, then each

---

variable is sampled 5,000,000 times and 1,000 simulated probability distributions of risk are generated from the PRA model. These probability distributions can be analyzed to estimate the distributions for specific risk estimates. For example, confidence limits on the estimate of specific risk percentiles can be simulated using 2-D MCA (Figure D-3).

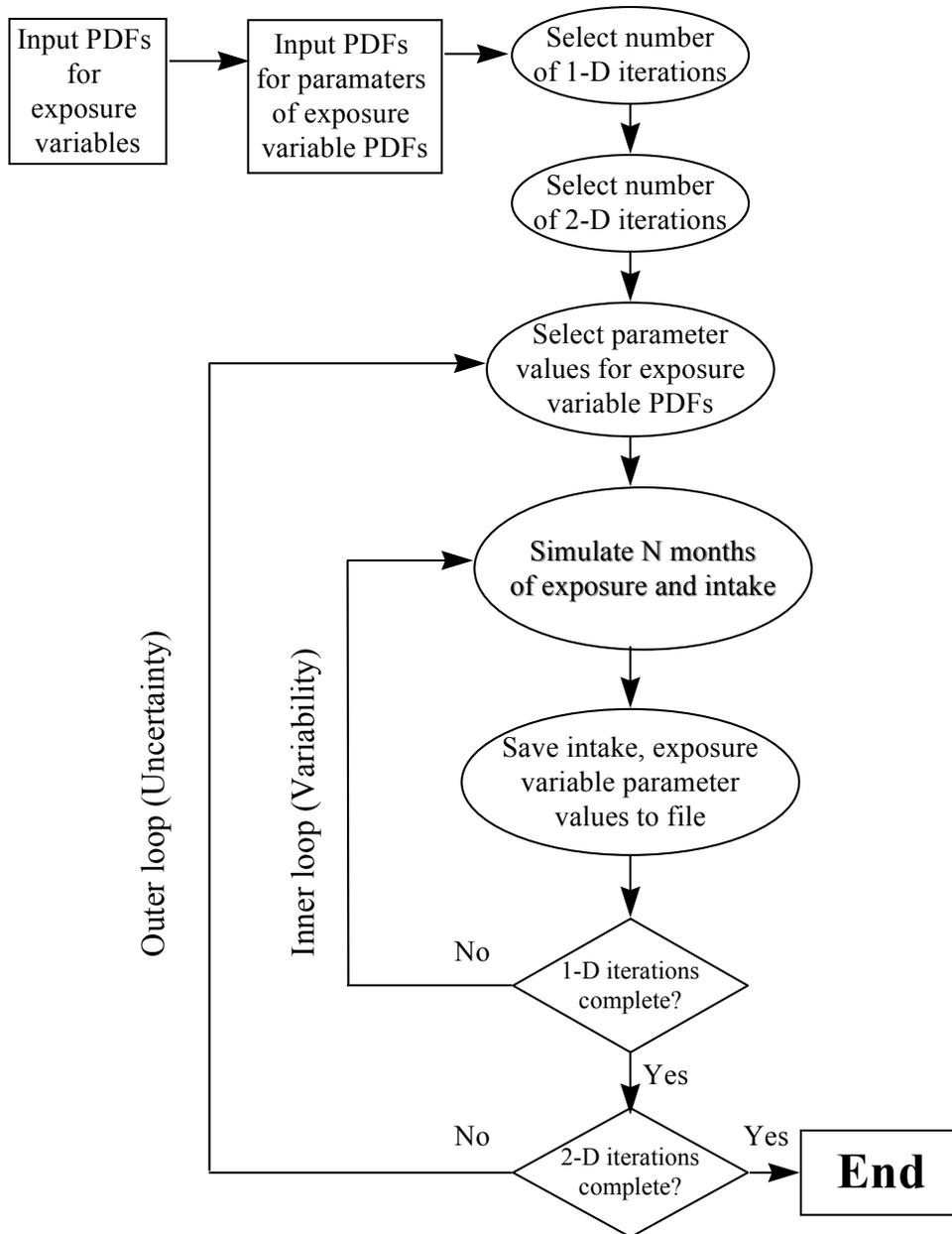## *Simulation Logic for 2-Dimensional MCA*

Figure D-2. Diagram showing of a 2-D Monte Carlo model in which the variability and uncertainty dimensions are computed in nested loops. In this example, values for exposure variables in the inner loop represent monthly averages.
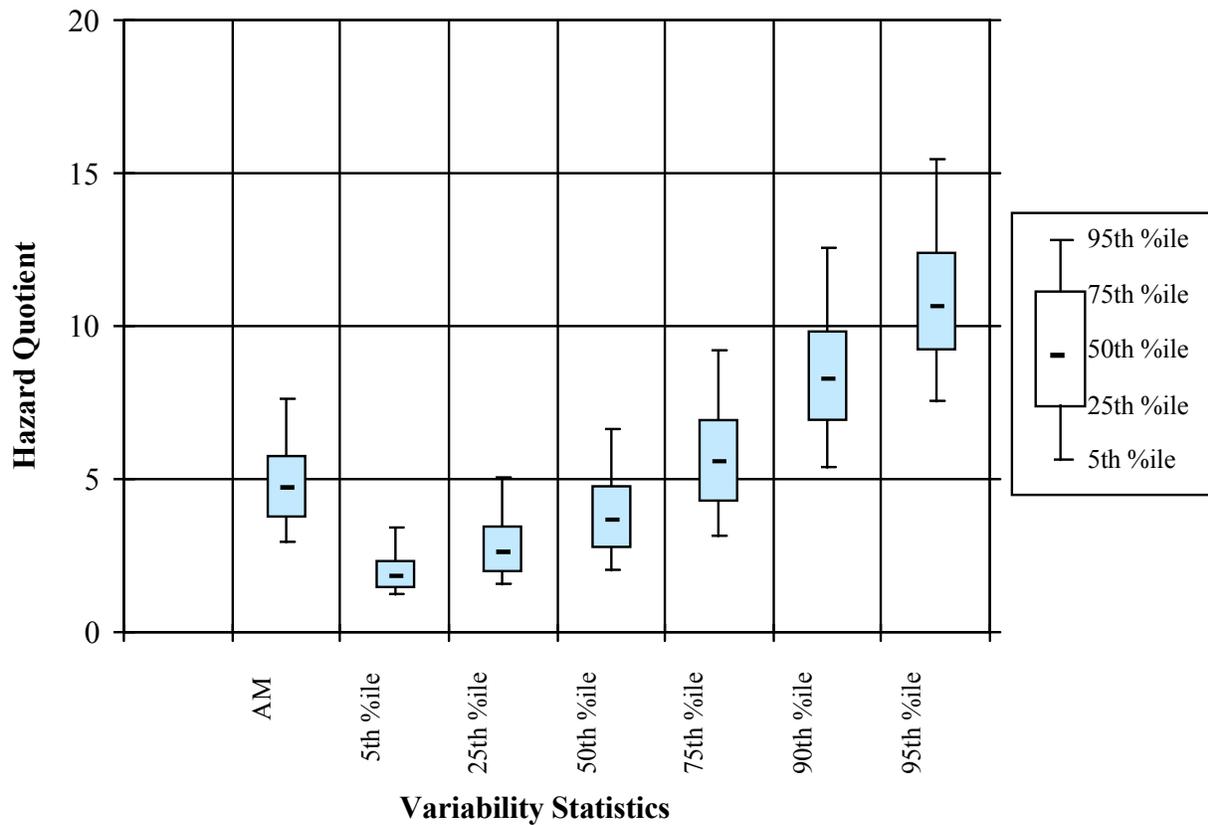
Figure D-3. Output from a 2-D MCA showing the estimated mean Hazard Quotient (HQ) and the 90% confidence interval for the arithmetic mean (AM) and selected percentiles of the HQ distribution. The 95th %ile HQ would be the reasonable maximum exposure (RME) risk estimate. The simulation suggests that there is a 95% probability that the RME HQ (95th percentile) is below 16.

D.3.0 MICROEXPOSURE EVENT ANALYSIS

The standard dose equation generally used in Superfund site risk assessments represents exposures averaged over a specified time period that is relevant to the health endpoint of concern (Equation D-1). If the risk assessment is directed at assessing life-time risk to humans, the averaging time used in Equation D-1 would generally be 70 years (i.e., estimated average human lifetime), and the calculated chemical intake would generally represent the life-time average daily dose (LADD). Where information is available to characterize variability on a smaller time scale than life-time, an alternative expression of dose that accommodates such variability may be desirable.

Standard Time-Averaging

$$\text{DOSE} = \frac{C \times IR \times EF \times ED}{BW \times AT} \qquad \text{Equation D-1}$$

Microexposure Event Modeling

$$\text{DOSE} = \frac{1}{AT} \sum_{j=1}^{ED} \frac{1}{BW_j} \sum_{i=1}^{Events_j} C_{ij} \cdot IR_{ij} \qquad \text{Equation D-2}$$

C   = Concentration; I = exposure event; j = year of life
IR   = Intake Rate
EF   = Exposure Frequency
ED   = Exposure Duration
BW = Body Weight
AT   = Averaging Time

Concentrations in various environmental media can be expected to vary over time. For example, wind erosion may change chemical concentrations in surface soil. Leaching may change concentrations in both subsurface soil and groundwater. The change in the concentration term is most readily apparent when considering anglers harvesting fish. If an angler consumes a large amount of fish from a single location (e.g., a specific lake, pond, or river), then the average chemical concentration in the fish consumed by that angler can be expected to be similar to the average of the chemical concentration of fish in the population. However, if an angler consumes fish only occasionally, or harvests fish from different locations, there will be considerably more uncertainty in the concentration term. In addition, a harvesting angler may consume varying amounts of fish over the period of the exposure duration due to changing tastes, changes in the fish population size or other factors.

Daily activity patterns, food intake, soil ingestion and other behavioral factors are measured in a time period of less than a year. The extrapolation of these short term results to the chronic exposure situation is a source of uncertainty. Exposure events are real but unknowable, whereas data regarding the nature and magnitude of these events is known but its application to a real world situation is uncertain. Microexposure event analysis (MEE) attempts to explicitly quantify this uncertainty. Figure D-5 presents the general approach for MEE analysis. (Price et al., 1996, 2000). MEE modeling provides an alternative to the standard time-averaging approach represented by Equation D-1. In the MEE approach, long term intake is viewed as the sum of individual exposure events (Equation D-2). Implementing the MEE approach in a PRA requires dividing the exposure duration into short epochs, or time steps, within which the values assigned to exposure variables remain constant, but are allowed to vary from one time step to the next. In a PRA model, exposure variables are adjusted at each time step by selecting values from the probability distributions representing each variable (Figure D-4). Discussion of the implementation of

MEE analysis in risk assessment and its merits and limits can be found in Wallace et al. (1994), Price et al. (1996), Slob (1996), and Buck et al. (1997).

In MEE modeling, the time step becomes an important variable, with associated uncertainty. The time step should be selected based on information available to describe how exposures change over time. For example, a model of a moving plume of solvents in groundwater might suggest that chemical concentrations in a given location are dropping by between 16 and 25% quarterly. Several rounds of sampling may support this prediction. This rapid decline in concentrations suggests that an appropriate time step might be one quarter (i.e., three months).

On the other hand, where risk is being assessed for metals, dioxin, or PAHs in soil, the concentrations might be expected

Figure D-4. Time Step for MEE.

to change much more slowly, if at all, and the basis of the time step might be the increase in age and corresponding changes in behavior of the receptor. The time step may be global; that is, one time step may apply to all variables in the model. In this case, the same number of random values would be selected for each exposure variable in a Monte Carlo simulation. A more complex model may use different time steps for different variables, requiring some probability distributions to be sampled more often than others. The selection of a value for a time step implies that the value represents the average value for that variable during the time step.

Figure D-5. Flowchart showing general approach for Microexposure Event (MEE) analysis.

Two important issues related to time step should be considered in implementing the MEE approach in PRA models. The first is the relationship between the length of the time step and the number of times random values are generated from a defined probability distribution. As the time step decreases, more time steps are needed to simulate exposures over a specified duration. For example, given a time step of one year and an exposure duration of 30 years, each random variable will be sampled 30 times (once per year); for a time step of one month and an exposure duration of 30 years, each random variable would be sampled 360 times (i.e., 12 months/year x 30 years). The Central Limit Theorem indicates that as n increases, the distribution of sample means is approximately normal, and the standard deviation of the sample distribution is inversely proportional to the square root of n. Thus a highly skewed input distribution (e.g., lognormal) may tend to become less skewed with increasing n (Figure D-6). A biased estimate of the RME risk in a PRA model may result if an inappropriately small or large time step is used in the model. This emphasizes the importance of having an empirical basis for selecting the time step and of exploring the time step as a variable in a sensitivity analysis of the model.

The second issue related to the time step concerns temporal correlations. Is it reasonable to assume that random values selected for consecutive time steps are completely independent? For example, consider body weight. The body weights of an individual measured at different times would be expected to show positive temporal autocorrelation; that is, body weight is likely to be similar (but not constant) from one time step to the next. For example, if an individual weighs 60 kg during one month, it is unlikely that they will weigh 80 kg the next month. If this scenario is accepted, then body weight should not be allowed to vary independently from one monthly time step to the next in the model. At shorter time steps, temporal correlation becomes more likely as a result of temporal autocorrelation. For example, one can expect a higher correlation between body weights on an individual measured on two successive days (one-day time step) than between weights measured at the midpoint of two successive years. Approaches to simulating temporal correlations in probabilistic models might include fixing an individual within a percentile range of a distribution (e.g., randomly assigned quartile) or using randomly assigned fluctuations (e.g., $BW_t = BW_{t-1} \pm x$).
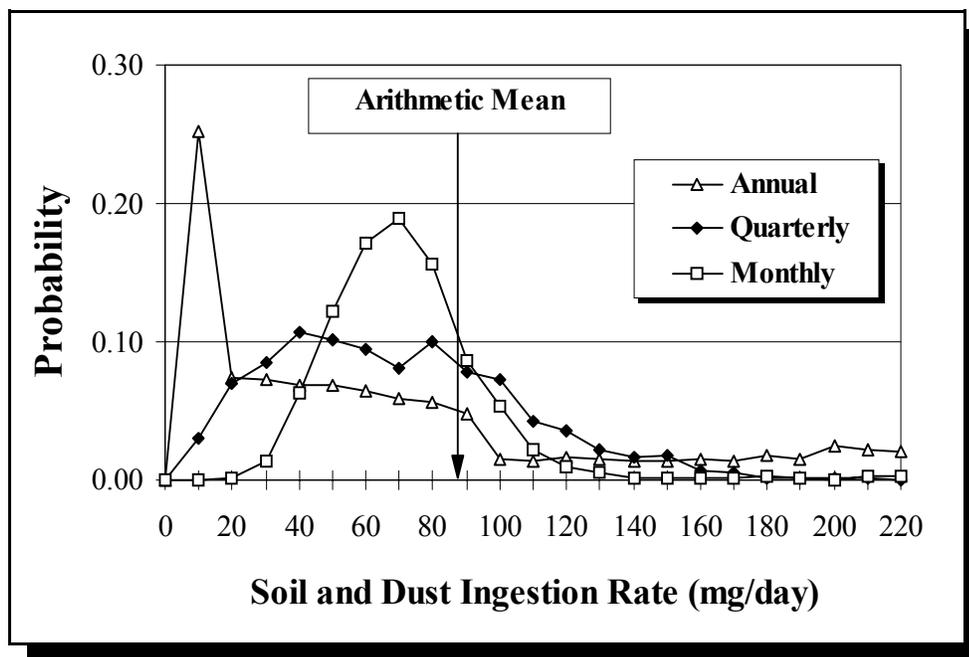

Figure D-6. Hypothetical example showing the effect of model time step on the probability distribution for soil and dust ingestion rate in children over a 1-year period. Number of samples (n) needed to simulate exposures: Annual (1), Quarterly (4), Monthly (12).

D.4.0 GEOSPATIAL STATISTICS

Spatial statistics is a specialized branch of statistics, falling under the heading of multivariate statistics, that explicitly takes into account the georeferenced or locational tagged context of data. Generally, environmental samples collected at Superfund sites have this geolocational information By acknowledging the geography of site chemicals, information about the spatial distribution of contamination can be incorporated into an exposure assessment. In addition, knowledge about a receptors home range or patterns of movement may also be incorporated into the definition of the exposure unit (see Appendix C, Section C.2.0). Explicitly accounting for spatial relationships may lead to a more accurate estimate of the confidence limits for the arithmetic mean concentration. Geospatial statistics quantifies the spatial autocorrelation (Exhibit D-2) of sample measurements and allows for the exploration of the spatial distribution of exposure and risk using techniques of map generalization. By recording locational tags for each sample, information about spatial patterns within an exposure unit (EU) can be exploited to estimate both pre- and post-remediation exposure and risk.

> **EXHIBIT D-2**
>
> **POSITIVE SPATIAL AUTOCORRELATION**
>
> • Locations with a high value of Y tend to be surrounded by nearby high values of Y.
> • Locations with a medium value of Y tend to be surrounded by nearby medium values of Y.
> • Locations with a low value of Y tend to be surrounded by nearby low values of Y.

In the past five years, with rapidly expanding software and hardware capabilities, some examples of the application of geostatistics can be found in exposure assessment and remedial design (e.g., Gomez-Hernandez, 1996; Goovaerts, 1996, 1997; Kriakidis, 1996; Ginevan and Splitstone, 1997; McKenna, 1998; Hope, 2000; 2001) as well as site assessment guidance (e.g., U.S. EPA, 2000).

Several important risk assessment issues are closely linked to geospatial statistics, as described in Exhibit D-3. Geospatial statistics comprises:

- *spatial autoregression*
- *geostatistics*
- *point pattern analysis*
- *image analysis*

The first three of these subjects can contribute to spatial statistical support of site risk assessments. The key concept linking all three is spatial autocorrelation, which refers to covariation among samples for a single chemical, or the tendency of data from locations that are relatively close together to be geographically correlated. By analogy, classical statistics treats soil samples as though they are balls, each having a battery of attributes, that can be placed into an urn for statistical analysis; geospatial statistics treats soil samples as though they are clusters of grapes,

> **EXHIBIT D-3**
>
> **EXAMPLES OF RISK ASSESSMENT ISSUES LINKED TO GEOSPATIAL STATISTICS**
>
> • Sampling tends to disproportionately represent "hot spots" (i.e., a relatively large portion of a data set with a small sample size (n) tends to be concentrated at "hot spots").
> • The upper confidence limit (UCL) for the arithmetic mean exposure concentration (e.g., chemical concentrations in soil) depends on the sample size.
> • Additional sampling may be needed, especially to better define the spatial patterns or the extent of contamination.
> • There is uncertainty about locations not sampled at a site, as well as uncertainty regarding the representativeness of neighboring samples in nearby EUs.

with the branchy stems representing locational tags. Concentrations located on the same "branch" will be more strongly correlated than concentrations on different branches.

How is Geostatistics Different from Classical Statistics?

In general, geostatistics provides information beyond that provided by classical statistical techniques for at least two reasons. First, in classical statistics, observations are assumed to be independent. This assumption is often invalid at contaminated sites where the method by which a chemical is released into the environment (e.g., deposition form airborne emissions; migration of contaminant plume from a point source) often results in positive spatial autocorrelation (see Section D.4.1). In other words, observations located next to each other tend to contain similar levels of contamination (i.e., redundant information). For example, the higher the spatial autocorrelation, the less incremental information is provided by adding observations in close proximity to existing observations. This issue is compounded when the sample locations have been preferentially determined (e.g., "hot spot" sampling) rather than distributed at regular intervals or specified using random sampling methodology.

Second, geostatistics is able to use the geospatial information contained in the data to model uncertainty in contaminant concentrations for areas where data are scarce, a situation commonly encountered in site assessment work. Using geostatistics, information from samples collected from outside an EU can be used to model the uncertainty in the mean concentration within an EU. Approaches that do not consider the geospatial information present in the data are limited to the subset of samples within an EU.

D.4.1 CORRELATION AND SPATIAL AUTOCORRELATION

Several simple bivariate statistical approaches may be used to introduce the concept of spatial autocorrelation. Consider two variables, X and Y. For positive correlation there is a tendency for high values of X to be paired with the high values of Y, medium values of X to be with the medium values of Y, and low values of X with the low values of Y. The tendency is in the opposite direction for negative correlation; high values of X tend to be paired with low values of Y, and so on. Spatial autocorrelation, which virtually always is positive, directly parallels these definitions, but is written in terms of a single variable as shown in Exhibit D-2.

Just as the bivariate relationship between two variables, X and Y, can be portrayed by a scatter plot (Y versus X), the spatial autocorrelation relationship can be portrayed for a single variable, Y, (e.g., Y versus Y). A good example is the Moran scatterplot, which plots the sum or average of nearby values of Y versus Y. This plot is most effective when Y has been converted to z-scores. As shown in Figure D-7 and Section D.4.2, scatter plots can be used to illustrate some important issues related to sample size.
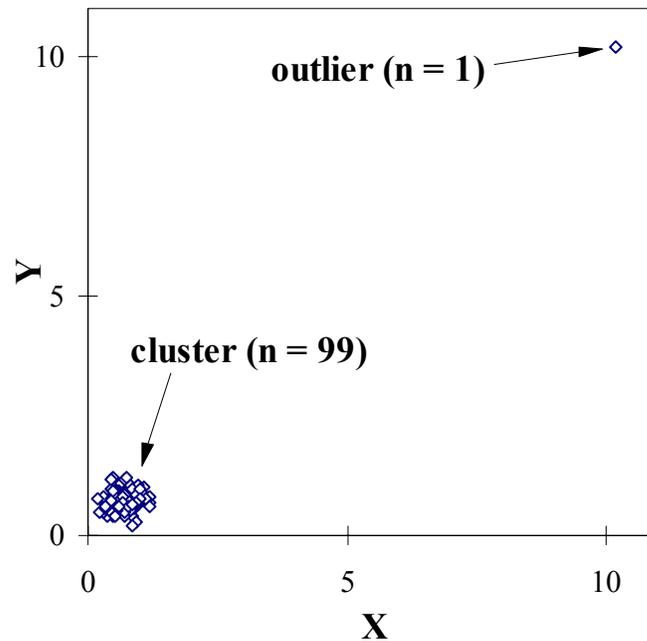
Figure D-7. Effect of an outlier on measured correlation: r=0.956
with outlier (*n*=100), whereas r=0.086 excluding outlier (*n*=99
clustered points).

If no soil samples were collected at a site (*n*=0), there is no information about the chemical
concentrations in soil, and any guess may be considered an estimate. However, if the chemical
concentration of a single sample (*n*=1) is measured, some information is obtained that partly restricts this
estimate. As each additional independent sample is taken, more information is obtained, and the
restriction on the estimate becomes more binding. If the same location is selected repeatedly for
sampling, then the repeated measures, which may vary through time, will tend to be highly positively
correlated; part of the information obtained from each sample is the same, and should not be counted
more than once in estimating the site-wide soil concentration. Similarly, if immediately adjacent
locations are sampled, the measures will often tend to be highly positively correlated (spatial
autocorrelation). Once the first sample is taken, each additional sample provides only a fractional
increment of new information about the site in its entirety.

D.4.2   EFFECTIVE SAMPLE SIZE (N*) AND DEGREES OF FREEDOM

Repeated measures can result in data clustering, which can be illustrated in a scatter diagram.
Because two points determine a straight line, if (*n–1*) points cluster together on a scatter diagram while a
single additional point occurs far away from this cluster (i.e., an outlier), then the resulting bivariate
correlation will be very high (see Figure D-7). This situation alludes to the notion of effective sample
size (*n**): the *n** is no longer equal to the number of observations (*n*), but rather is dramatically reduced
by the presence of inter-observational correlation. For the example shown in Figure D-7, *n** is slightly
greater than 2 rather than 100 (i.e., *n*).

Spatial autocorrelation plays an analogous role in georeferenced data. If a sampling network is arranged as a 25-by-25 square grid (one sample point per grid cell), and superimposed over a large site so that a very large distance separates nearby sample locations, then essentially zero spatial autocorrelation should be present in the geographic distribution of the concentrations of any given chemical. Concentrations will appear to be haphazard across the site, rendering the effective sample size as $n^*=625$. If the distance between nearby locations on the sampling mesh is decreased so that the spatial correlation is only r=0.050, then the effective sample size decreases to $n^*=514$. The effect of reducing the inter-sample distance on spatial autocorrelation and n^* for a 25-by-25 grid is shown in Exhibit D-4. If r increases to 1, then n^* reduces to 1. Therefore, obtaining a measure of latent spatial autocorrelation is essential to estimating n^*; this in turn is critical to determining confidence limits for estimates of mean concentrations, which are sensitive to sample size. The UCL for the mean will be biased *only* when very high levels of spatial autocorrelation are present; this is because the Student-t statistic used to estimate the UCL (assuming a normal distribution) changes very little as the degrees of freedom (related to sample size) increases above 10; part of the difference between n and n^* is offset by an inflation of the variance.

EXHIBIT D-4 — EFFECT OF SPATIAL AUTOCORRELATION (r) ON EFFECTIVE SAMPLE SIZE (n^*)	
r	**n^***
0.000	625
0.050	514
0.539	64
0.957	3
1.000	1

The concept of effective degrees of freedom is important in exposure assessment because high positive spatial autocorrelation can bias the estimate of the UCL concentration if geospatial statistics are not considered. This should be of particular concern when specific locations at a site are intensively sampled (e.g., suspected "hot spots"), and other locations are relatively undersampled. Accordingly, the design of the sampling network itself can be evaluated from the perspective of geospatial statistics in order to ascertain the quality of sample information. The ideal sampling network should provide geographic representativeness, should be roughly uniformly distributed over a site, and is best implemented as a stratified random sampling design; that is, the site is partitioned into geographic stratum (e.g., EUs), and then a random sampling of points is selected within each strata. In practice, sample designs may need to focus on objectives that are in conflict with the above ideals. For example, intense sampling of suspected "hotspots" may be necessary at some sites, at the expense of a more representative spatial coverage of the site. In such cases, several statistical techniques are available for assessing the statistical benefit (in terms of reducing uncertainty) of additional sampling at undersampled locations.

D.4.3   ASSESSMENT OF ADDITIONAL SITE SAMPLING

Thiessen Polygons. In addition to calculating nearest neighbor statistics, the adequacy of a sampling network can be assessed by Voronoi (i.e., Thiessen polygon) surface partitioning, a popular approach used in mapping intra-site geographic distributions. This procedure divides a site into a mutually exclusive set of polygons, each polygon containing a single measured concentration. Each polygon has the unique property that any location within the polygon is closer to the polygon's sample location than to any other sample point (Clifford et al., 1995). The concentration measured at the sample point in the polygon is assigned to the entire area of the polygon. The intensity of sample points on a surface can be measured by Equation D-3 mean inverse polygon areas:

$$SI = \frac{1}{m} \sum_{i=1}^{m} A_i^{-1}$$

Equation D-3

where *SI* is a measure of the sampling intensity, A_i is the area of the ith polygon, and *m* is the number of interior polygons (those not along the edge of the site); $m < n$. The variance of the sampling intensity can be expressed by Equation D-4:

$$SI_{Variance} = \frac{1}{m-1} \left[\sum_{i=1}^{m} A_i^{-2} - \frac{1}{m} \left(\sum_{i=1}^{m} A_i^{-1} \right)^2 \right]$$

Equation D-4

If the sampling network is uniform (i.e., polygon areas are equal), the variance will be essentially zero. The variance will increase as the network deviates from uniform. This measure can be used to assess whether or not additional samples will improve the spatial coverage.

☞ *Sampling locations that would yield a dramatic reduction in the variance should be given priority for future sampling efforts.*

Thiessen polygons can be used to develop area-weighted estimates of the arithmetic mean concentration ($C_{soil,w}$) according to the following general equation:

$$C_{soil,w} = \sum_{i=1}^{n} C_i \frac{A_i}{A_T}$$

Equation D-5

where C_i is the concentration in the ith polygon, A_i is the area of the ith polygon in the EU, and A_T is the total area of the EU. The weight for each measurement is essentially the ratio of the area of each polygon to the total area of the site. Clifford et al. (1995) applied this approach to an ecological risk assessment of the burrowing owl with the following simplifying assumptions: habitat range is circular, size of EU is constant (75 ha) although location may vary, and organisms spend equal time in all portions of their habitat. Given these assumptions, a nonparametric bootstrap method can be used to determine the approximate 95% UCL for the mean concentration (see Appendix C). Using Monte Carlo analysis, $C_{soil,w}$ can be estimated for different locations of the EU according to Equation D-5, and confidence limits can be generated from the multiple bootstrap estimates. Burmaster and Thompson (1997) demonstrate a similar approach in which the EU (with constant area but random rectangular dimensions) is overlaid on the Theissen polygon surface and 95% UCL for the mean is calculated from the bootstrap sample.

Linear Regression. Another diagnostic is found in the linear regression literature. The locational tag coordinates (e.g., longitude, latitude) can be converted to z-scores (say z_u and z_v) for the following calculation:

$$ Y \; = \; \frac{1}{n} + \frac{z_u^2 + z_v^2 - 2r_{u,v} z_u z_v}{(n-1)(1-r_{u,v}^2)} \qquad\qquad \text{Equation D-6} $$

where Y is a measure of the sampling network, r_{uv} is the correlation between the coordinate axes, and n is the number of samples. Any sampling location (z_u, z_v) in which $Y > 9/n$ may be considered too isolated in the sampling network. Additional sampling locations would be positioned closer to it to improve the overall coverage of the sampling network.

D.4.4 MAP GENERALIZATION

Another important application of geospatial statistics to risk assessment is that of map generalization, which draws on the subjects of geostatistics and spatial autoregression. Techniques developed for both topics exploit spatial autocorrelation in order to produce a map.

Kriging and Semivariograms. Geostatistics may employ kriging, which yields statistical guesses at values of a chemical at unsampled locations based on information obtained from sampled locations. Kriging assumes that the underlying geographic distribution is continuous, evaluates spatial autocorrelation in terms of distance separating sample points, and employs a scatter diagram similar to the Moran scatter plot to portray this relationship (i.e., the semivariogram plot: half the squared difference between measured concentrations for two sampled locations versus distance separating these two locations). The best-fit line to this scatter of points is described by one of about a dozen equations (semivariogram models).

Many different kriging approaches can be applied to quantify the spatial relationships among geographic attributes within an exposure unit. For example, site-specific chemical concentrations may be correlated with geologic information, such as glacial deposits, soil characteristics of core samples, and attributes that represent favorable habitats for ecological receptors. This information can be used to expand the available data and improve estimates of chemical concentrations at unsampled locations by employing a technique called co-kriging.

Thiessen Polygons and Spatial Autoregression. Spatial autoregression assumes a discretized surface, uses the Thiessen polygon surface partitioning to construct a Moran scatter plot, and can be used to estimate values at selected points with a regression-type equation. Theoretically, the exponential semivariogram model relates to the conditional autoregressive model, and the Bessel function semivariogram model relates to the simultaneous autoregressive model; in practice, though, the spherical semivariogram model often provides the best description of a semivariogram plot. Regardless of which approach is taken to map generalization, one relevant contribution of these two subjects is the following observation:

☞ *Including positive spatial autocorrelation results in more accurate variance estimates; this in turn yields more accurate estimates of the 95% UCL for the mean concentration.*

D.4.5 **IMPLEMENTATION ISSUES RELATED TO GEOREFERENCED DATA**

Estimation of parameters, for either geostatistical or spatial autoregressive models, cannot be achieved with ordinary least squares (OLS) techniques; nonlinear least squares must be used. While OLS provides unbiased regression coefficients, these estimates are not necessarily sufficient (i.e., they do not summarize all of the information in a sample pertaining to the population), efficient (i.e., the standard errors often are incorrect), and consistent (i.e., the asymptotic sampling distribution concentration will not be at the parameter value). In other words, OLS essentially uses the wrong degrees of freedom in its calculations, as described in Section D.4.2. Two additional complications of georeferenced data that do not appear in other types of data are (1) spatial autocorrelation might be directional (i.e., directional dependency); and (2) variance might be nonconstant over space as well as over the magnitude of the dependent variable, Y (e.g., chemical concentration). Several statistical approaches, which are beyond the scope of this guidance, are available for analyzing these potential sources of bias in the exposure concentration estimates (Isaaks and Srivastava, 1989; Cressie, 1991; Griffith, 1993; Ginevan and Splitstone, 1997).

D.5.0 **EXPERT JUDGMENT AND BAYESIAN ANALYSIS**

Up to this point in RAGS Volume 3: Part A, risk has been characterized as having a population probability distribution with parameters (e.g., mean, standard deviation) that can, theoretically, be estimated from observation. In theory, risk estimates could be derived by repeatedly measuring risk in subsets of the population of interest (e.g., repeated measurements of site-related cancer risk). The unstated expectation, or goal, is that the PRA model will accurately simulate this *real* risk distribution. This approach derives from a *classical* view of probability. The *classical* or *frequentist* view defines the probability of an event as the frequency with which it occurs in a long sequence of similar trials. From the *frequentist* perspective, the probability of having a flipped coin land *heads-up* is given by the frequency distribution of heads-up results derived from repeated similar trials of coin flips. For real-world decisions such as those informed by Superfund risk assessments, there is uncertainty that the sample data are representative of the population (see Chapter 1, Section 1.2.4).

Bayesian View of Probability. A Bayesian perspective on probability allows distributions to be constructed based on the judgment of an expert in the field. The subjectivist or Bayesian view is that the probability of an event occurring is the degree of belief a person has in the occurrence. Probabilities can be assessed by experts using scientific knowledge, judgment, data, past experience, and intuition. Different people may assign different probabilities to an event, and a single individual may assign different probabilities to the same event when considered at different times. The consequence is that probabilities become conditional and the conditions must be explicitly stated (Howson and Urbach, 1989; Morgan and Henrion, 1990; Ott, 1995; Sivia, 1996). These conditional probabilities can, of course, be updated with new information.

Using the coin flip analogy above, a Bayesian perspective might be that, based on experience with coins, assuming that most coins are *fair*, and that a fair coin would be expected to land heads-up half the time, the expected probability of the tossed coin landing heads-up is 0.5. If the outcome of repeated trials was different from the expected, the Bayesian approach would be to update the probability based on the new data. In the coin flip example, both the Bayesian and frequentist approaches will arrive at the same conclusions, because the outcome is amenable to rigorous experimentation. Where the two approaches can be expected to differ is in the assignment of probabilities to events that cannot be rigorously measured; for example, the probability of a site-related cancer risk, or the probability of a child ingesting a specific amount of soil.

The subjective judgment of experts is, therefore, an important tool in the Bayesian approach to risk assessment. For example, the input distributions for a PRA may be based upon the judgment of one or more experts who rely upon estimates from the literature, data from experimental studies, and any other information they consider relevant. Even when formal elicitations of expert opinion are not done, the final selection of the form and parameters of the input distributions usually involves some subjective judgment by the analyst. One of the challenges of incorporating judgments from experts or lay people is that there can be overconfidence bias (i.e., people tend to underestimate their uncertainty). There is a rich literature about the protocol for conducting expert elicitations and using the results to support decisions (Lichtenstein and Fischoff, 1977; Morgan and Henrion, 1990; Shlyakhter and Kammen, 1992). Elicitation of expert judgment has been used to obtain distributions for use in risk assessments (Morgan and Henrion, 1990; Hora, 1992; U.S. EPA, 1997;) and in developing air quality standards (U.S. EPA, 1982).

In addition to providing input distributions for PRAs, Bayesian analysis allows the current state of knowledge, expressed as a probability distribution, to be formally combined with new data to reach an updated information state. The distribution expressing the current knowledge is the *prior distribution* and may be the output of a PRA (Figure D-8). An appropriate *likelihood function* for the data must also be formulated. The likelihood function is based upon an understanding of the data gathering process and is used to determine the probability of observing a new set of data given that a particular risk estimate is true.

EXHIBIT D-5

COMPONENTS OF BAYES THEOREM IN PRA

- Input probability distributions for exposure (or toxicity) based on available data or expert judgment
- Prior probability distribution for risk based on input probability distributions (output from PRA)
- New data
- Likelihood function, expressing the probability of observing the new data conditional on prior risk estimates
- Posterior (updated) probability distribution for risk

Once the prior distribution is determined, the new data values are collected, and the likelihood function is assumed, Bayes theorem (Exhibit D-5) provides a systematic procedure for updating the probabilistic assessment of risk. The updated information state is called the *posterior distribution* and reflects the reduction in uncertainty arising from the new information.
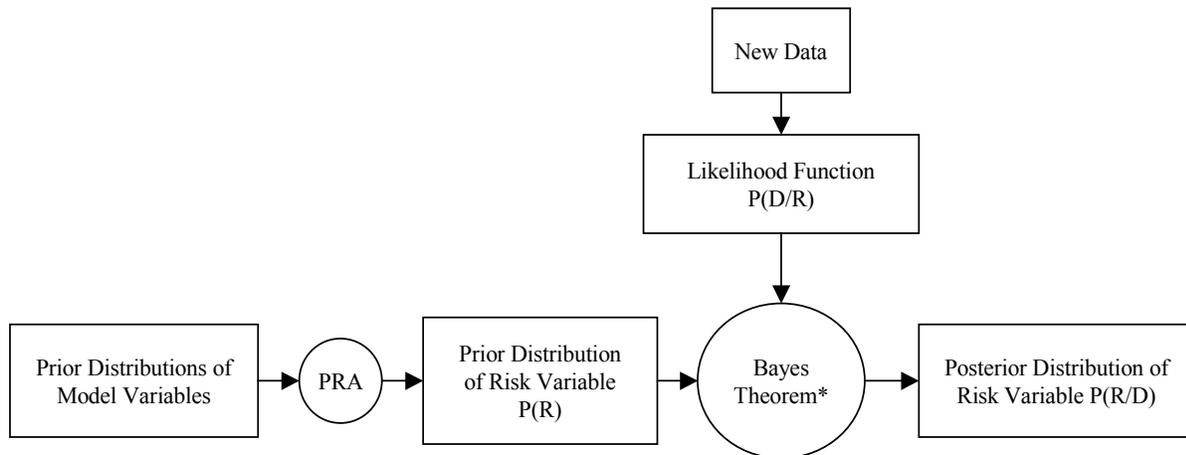
```
                                    ┌─────────────┐
                                    │  New Data   │
                                    └─────────────┘
                                           │
                                           ▼
                               ┌──────────────────────┐
                               │  Likelihood Function  │
                               │        P(D/R)         │
                               └──────────────────────┘
                                           │
                                           ▼
┌────────────────┐   ┌─────┐  ┌──────────────────┐   ╭─────────╮  ┌──────────────────────┐
│ Prior          │   │     │  │ Prior Distribution│   │  Bayes  │  │ Posterior Distribution│
│ Distributions of│→ │ PRA │→ │ of Risk Variable  │ → │Theorem* │→ │  of Risk Variable P(R/D)│
│ Model Variables │   │     │  │      P(R)         │   ╰─────────╯  └──────────────────────┘
└────────────────┘   └─────┘  └──────────────────┘
```

Figure D-8. Conceptual model of Bayesian Monte Carlo analysis. A PRA simulation yields a prior distribution of risk based on probability distributions for input variables. Given new data for an input variable, and a likelihood function for risk, Bayes Theorem (Eq. D-7) can be used to generate a posterior distribution of risk. The expression P(D/R) refers to a conditional probability, *"the probability of D, given R".* Conditional probabilities can be thought of as relative frequencies, where R is the information given, and D is the event being computed when a particular value of R occurs.

$$\textit{Bayes Theorem *}: \quad P(R_i/D) \;=\; \frac{P(D/R_i)\,P(R_i)}{\displaystyle\sum_{j=1}^{N} P(D/R_j)\,P(R_j)} \qquad \text{Equation D-7}$$

D = new data
R_i = i^{th} risk prediction associated with new data
Rj = j^{th} risk estimate simulated from PRA model
N = number of risk estimates from the PRA model

For example, suppose a model is available to relate soil tetrachlorodibenzodioxin (TCDD) concentrations at a site with serum concentrations of TCDD. A probability distribution of soil concentrations is created based upon expert judgment and a limited amount of site specific data. Using the model, the soil concentrations can be associated with a distribution of serum TCDD concentrations (P®), the prior distribution). New site-specific data (D) are subsequently collected on serum TCDD concentrations in order to reduce uncertainty in the risk estimate. Assume that it is known that serum TCDD concentrations generally follow a lognormal distribution and that the best estimate of the parameters of this distribution come from the prior distribution on serum TCDD. This creates the likelihood function (P(D|R)). Using Bayes Theorem, the new data are used to form a revised distribution of serum TCDD. This is the posterior distribution (P(R|D)).

Bayesian Monte Carlo analysis. In the past, the use of Bayesian analysis was limited by the degree of mathematical complexity involved. Using Monte Carlo analysis to carry out the PRA, rather than mathematical equations to describe the distributions, allows the calculations to be done much more easily. This variation on traditional Bayesian methods is called Bayesian Monte Carlo analysis (Patwardan and Small, 1992; Dakins et al., 1996). In the TCDD example discussed above and illustrated in Figure D-7, the required calculations are carried out for each of the N iterations of the Monte Carlo analysis (I and j go from 1 to N).

Bayesian Monte Carlo analysis is appropriate in several situations. If a model has been created and a distribution developed using PRA, new information may be incorporated without the need to repeat the entire analysis. This information could be on one of the uncertain parameters of the model or on the model output variable. Similarly, a generalized risk model with generic parameter distributions may be used for a Superfund risk assessment with the model predictions fine-tuned using data from a particular site of interest. Finally, after a distribution is developed, the amount of uncertainty that exists may be too large for the risk manager to make a decision. In this case, the risk manager might seek out new information that would refine the analysis and decrease the uncertainty.

Bayesian Monte Carlo analysis can also be combined with techniques from decision analysis to help determine the type and quantity of data that should be collected to reduce uncertainty. Decision analysis is a technique used to help organize and structure the decision maker's thought process and identify a best strategy for action. To determine the appropriate action, one defines the range of possible decisions, evaluates the expected value of the utility or loss function associated with each decision, and selects the decision that maximizes the expected utility or minimizes the expected loss.

☞ *Decision analysis provides a quantitative approach for evaluating the benefits of including an expanded assessment of uncertainty and the subsequent benefits of reducing this uncertainty.*

Value of Information. Value of information (VOI) analysis involves estimating the value that new information can have to a risk manager before that information is actually obtained (Clemen, 1996). It's a measure of the importance of uncertainty in terms of the expected improvement in a risk management decision that might come from better information. Examples of VOI quantities are the expected value of including uncertainty (EVIU), the expected value of sample information (EVSI), the expected value of perfect information (EVPI). Calculation of these quantities can be done using mathematical methods, numerical integration (Finkel and Evans, 1987), or Monte Carlo techniques (Dakins, 1999)

Value of information calculations require the specification of either a utility or a loss function. A loss function states the losses associated with making different types of decision errors including both direct monetary costs and losses associated with other consequences. Loss functions take various forms depending on the risk management situation (Morgan and Henrion, 1990).

Expected Value of Including Uncertainty. The expected value of including uncertainty, EVIU, is a measure of the value of carrying out a PRA. It's the difference between the expected loss of a decision based on a point estimate risk assessment and the expected loss of the decision that considers uncertainty (Figure D-9). If uncertainty in a risk assessment has been estimated using Monte Carlo techniques and a loss function has been specified, the EVIU can be easily calculated. First, the management decision from the point estimate assessment is determined. The loss from making this decision is calculated for each iteration of the Monte Carlo, each time assuming that the risk estimate from that iteration is true. The expected loss is the average of these individual losses. The expected loss for the PRA is determined by calculating the expected loss for a full range of management decisions and selecting the decision with the lowest expected loss. The EVIU is calculated by subtracting the loss associated with the PRA from that associated with the point estimate risk assessment.

Expected Value of Sample Information. The expected value of sample information is the difference between the expected loss of the decision based on the PRA and the expected loss of the decision from an improved information state. As such, the EVSI is a measure of the value that may result from the collection and use of new information (Figure D-9). Calculation of the EVSI involves a technique called preposterior analysis and is somewhat more complicated.

This type of analysis is termed "preposterior" because it involves the possible posterior distributions resulting from potential samples that have not yet been taken. For each replication from the Monte Carlo simulation, the predicted value from the model is used to randomly generate a set of K data points. Each set of data points is then used to calculate the posterior probabilities for the N Monte Carlo simulated values. These posterior probabilities are then used to obtain the optimal answer to the management question at this new level of uncertainty by selecting the decision that minimizes the expected loss over all possible management decisions.

This procedure is repeated for each of the N replications of the Monte Carlo analysis resulting in N posterior distributions, N management decisions, and N associated expected losses. Because each of these outcomes is equally weighted, the expected loss associated with the state of uncertainty expected to exist after the data collection program is carried out is simply the average of the N expected losses. The EVSI is the difference between the expected loss based on the results of the PRA and the expected loss from the updated information state.

Expected Value of Perfect Information. The EVPI is the difference between the expected loss of the decision based on the results of the PRA and the expected loss of the optimal management decision if all uncertainty were eliminated. In actual application, no research plan or data collection program can completely eliminate uncertainty, only reduce it. The EVPI is an upper bound for the expected value of efforts to reduce uncertainty and so provides the ultimate bound on what should be spent on research and data collection efforts.
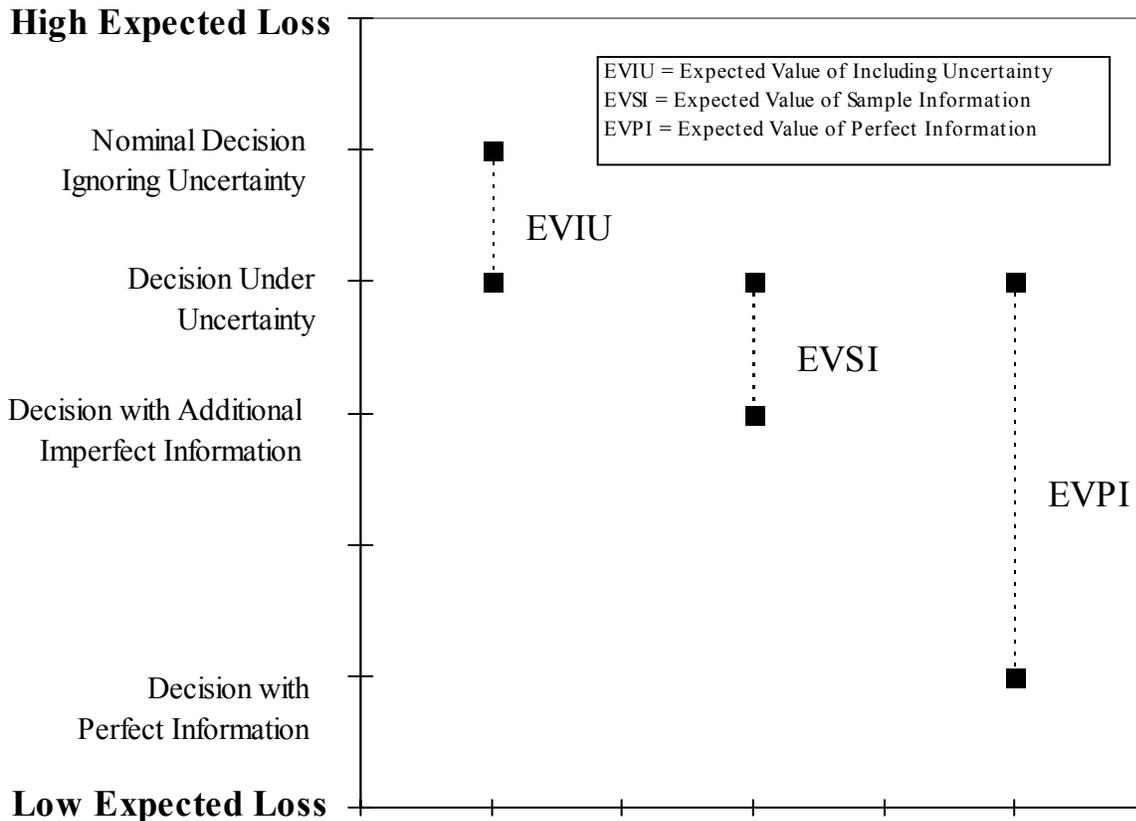
Figure D-9. Expected Loss associated with various types of information incorporated into a generic uncertainty analysis. The x-axis reflects different categories of value of information (VOI) quantities. The y-axis reflects the increasing Expected Loss with increasing uncertainty.

When a PRA has been carried out using Monte Carlo techniques, the expected loss associated with perfect information is calculated by determining the expected loss for each iteration of the Monte Carlo, assuming that the correct management decision, if that iteration were true, is made. As always, the expected loss is the average of these losses, and the EVPI is calculated by subtraction.

Uses of Value of Information in Risk Assessment. VOI analysis has many benefits for risk managers. First, VOI analysis makes the losses associated with decision errors explicit, balances competing probabilities and costs, and helps identify the decision alternative that minimizes the expected loss. VOI analysis can help a decision maker overcome a fear of uncertainty by developing a method to handle it. If the losses associated with making a poor decision are unclear, small uncertainties can take on major importance. Conversely, if the losses associated with different risk management decisions are similar, little additional effort need be expended to continue to consider the alternatives.

In addition, VOI analysis helps prioritize spending on research. It provides insights into how resources could be spent to achieve the most cost-effective reduction in uncertainty by identifying which sources of uncertainty should be reduced, what type of data should be obtained, and how much data is

needed. Finally, VOI analysis may help decision makers explain the rationale for their decisions to the public and help the public understand the multiple objectives considered in managing risks.

Expected Loss is usually greatest when uncertainty in risk estimates is ignored. For example, by quantifying uncertainty in risk (e.g., 2-D MCA, Bayesian Monte Carlo analysis) a risk manager may determine that the cleanup level associated with the 90th percentile of the risk distribution (rather than the 95th percentile) is adequately protective. Quantifying uncertainty may also result in lower expected loss when more soil remediation is required due to the losses associated with possible under-remediation, e.g., cost of additional sampling or lost revenue due to failure to meet land use requirements. The expected loss may be further reduced by collecting additional soil samples, which would presumably reduce uncertainty in estimates of mean exposure point concentrations. The expected loss may be minimized by obtaining "perfect" information (i.e., no uncertainty); however, as shown in Figure D-9, EVPI spans a wide range of expected loss because the value associated with reducing uncertainty may be tempered by costs associated with additional sampling and analysis. In practice, risk assessors consider this issue when deciding to obtain additional samples for site characterization.

The decision to obtain additional information in order to reduce uncertainty should be made on a site-specific basis, taking into account the potential impact that reducing uncertainty may have on the overall remedial decision. Important questions to consider include: (1) Are the risk estimates sufficiently sensitive to an exposure variable that collecting further data will reduce uncertainty? and (2) Are the confidence limits on the 95th percentile risk estimate sufficiently wide that reducing uncertainty may alter the cleanup goal? An example of decision framework applicable to PRA is presented in Figure D-10. The framework has three tiers. Tier 1 includes the point estimate approach and an assessment of the need for PRA. In Tier 2, the EVIU is calculated and, if warranted, a PRA is conducted. In Tier 3, the value of additional information is assessed and Bayes Theorem would be used to incorporate the new information and update probability distributions.

Limitations of These Techniques. Figure D-10 illustrates situations where Bayesian analysis and value of information quantities may not be helpful. For example, if point estimate risk assessment is selected as the appropriate method, these techniques do not apply. In addition, as site-specific data become available that are increasingly comprehensive and representative of the population of interest, Bayesian Monte Carlo analysis and the Monte Carlo analysis using the classical (frequentist) methods will approach the same result. This is because the site-specific data are incorporated into both approaches. To be representative and comprehensive, the data set must be sufficiently large, randomly selected, and represent the full range of variability that exists in the population (e.g., temporal, spatial, inter-individual). However, data sets are rarely perfect, often too small, suffer from relatively high sampling and/or measurement errors, or don't represent the entire population variability over time, space, age, gender, or other important variables. If the data cannot be assumed to describe the population distribution sufficiently well, then PRA will help to more fully develop the entire range of the population distribution and the Bayesian Monte Carlo analysis will act to refine the model estimates.
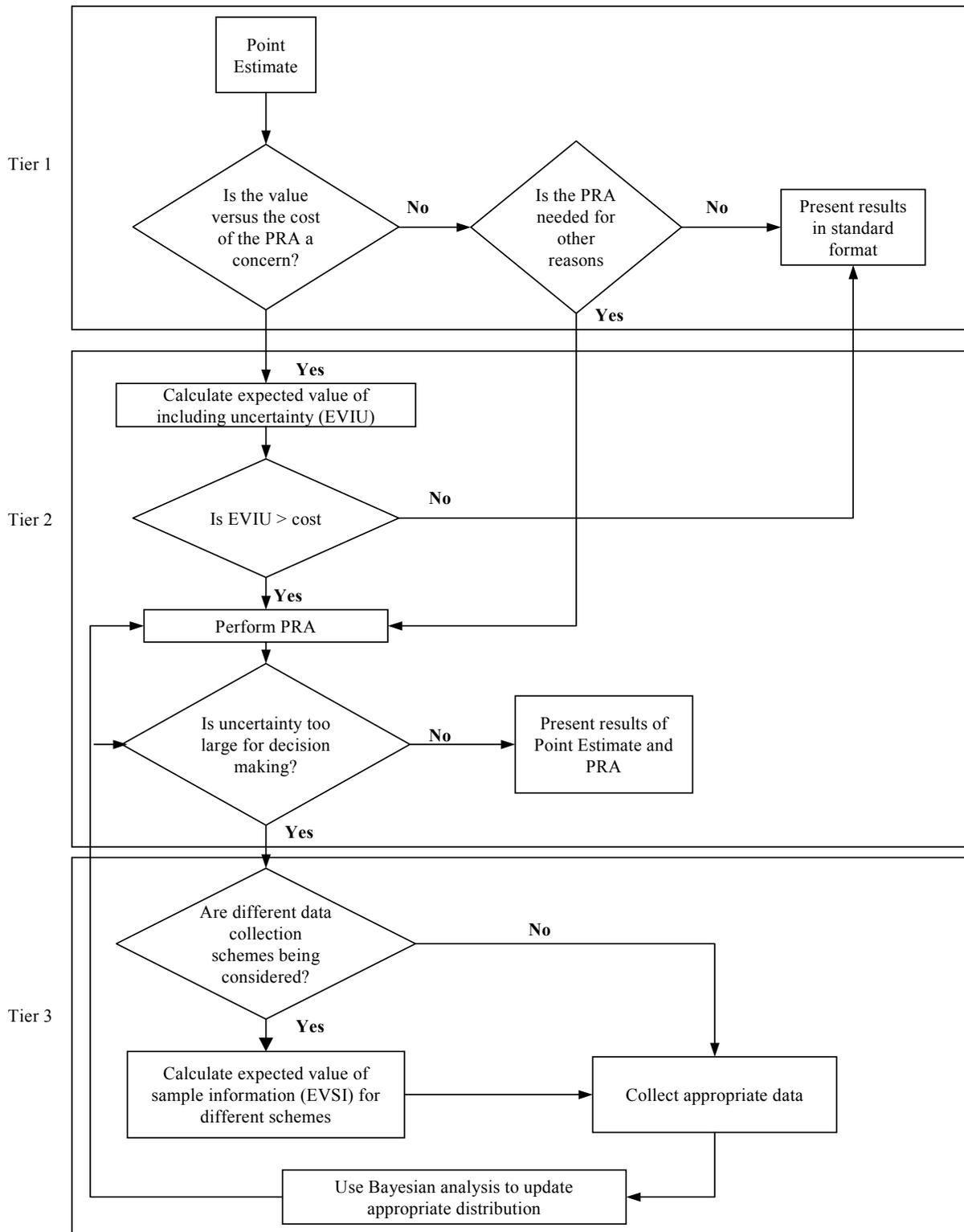
Tier 1

Point Estimate

Is the value versus the cost of the PRA a concern?

No → Is the PRA needed for other reasons

No → Present results in standard format

Yes

Tier 2

Yes

Calculate expected value of including uncertainty (EVIU)

Is EVIU > cost

No

Yes

Perform PRA

Is uncertainty too large for decision making?

No → Present results of Point Estimate and PRA

Yes

Tier 3

Are different data collection schemes being considered?

No

Yes

Calculate expected value of sample information (EVSI) for different schemes → Collect appropriate data

Use Bayesian analysis to update appropriate distribution

Figure D-10. Conceptual model for evaluating the expected value of including uncertainty in a Bayesian Monte Carlo anaylsis.

In order to carry out VOI calculations, a loss function must be assumed. Definition of the loss function may be complex due to multiple decision goals and/or multiple decision makers and may be difficult to capture in an equation. Finally, for Bayesian analysis and the calculation of the EVSI to be helpful, one or more sources of new data must exist. In addition, some information must be available about these data since a likelihood function describing its probability distribution must be assumed.

REFERENCES FOR APPENDIX D

Brattin, W.J., T.M. Barry, and N. Chiu. 1996. Monte Carlo Modeling with Uncertainty Probability Density Functions. *Hum. Eco. Risk Assess*. 2(4):820–840.

Buck, R.J., K.A. Hammerstrom, and P.B. Ryan. 1997. Bias in Population Estimates of Long-Term Exposure From Short-Term Measurements of Individual Exposure. *Risk Anal.* 17:455–466.

Burmaster, D.E. and K.M. Thompson. 1997. Estimating Exposure Point Concentrations for Surface Soils for Use in Deterministic and Probabilistic Risk Assessments. *Hum. Eco. Risk Assess.* 3(3):363–384.

Clemen, R.T. 1996. *Making Hard Decisions: An Introduction to Decision Analysis*. Duxbury Press, Pacific Grove, CA.

Clifford, P.A., D.E. Barchers, D.F. Ludwig, R.L. Sielken, J.S. Klingensmith, R.V. Graham, and M.I. Banton, 1995. An Approach to Quantifying Spatial Components of Exposure for Ecological Risk Assessment. *Environ. Toxicol. Chem*. 14(5):895–906.

Cressie, N. 1991. *Statistics for Spatial Data.* Wiley, New York, NY.

Dakins, M.E., J.E. Toll, M.J. Small, and K.P. Brand. 1996. Risk-based Environmental Remediation: Bayesian Monte Carlo Analysis and the Expected Value of Sample Information. *Risk Anal.* 16:67–79.

Dakins, M.E. 1999. The Value of the Value of Information. *Human Eco. Risk Assess*. 5(2):281–289.

Finkel, A.M. and J.S. Evans. 1987. Evaluating the Benefits of Uncertainty Reduction in Environmental Health Risk Management. *J. Air Pollut. Control Assoc*. 37:1164–1171.

Ginevan, M.E. and D.E. Splitstone. 1997. Improving Remediation Decisions at Hazardous Waste Sites with Risk-Based Geostatistical Analysis. *Environ. Sci. Technol*. 31(2):92A–96A.

Gomez-Hernandez, J.J. 1996. Issues on Environmental Risk Assessment. In: *Proceedings of the Fifth International Geostatistics Congress*, Vol. 1. (Baafi, E.Y. and N.A. Schofield, eds.). Kluwer Academic Publishers, Dordrecht, 15–26.

Goovaerts, P. 1996. Accounting for Local Uncertainty in Environmental Decision-Making Processes. In: *Proceedings of the Fifth International Geostatistics Congress*, Vol. 2. (Baafi, E.Y. and N.A. Schofield, eds.). Kluwer Academic Publishers, Dordrecht, 929–940.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation.* Oxford University Press, NY.

Griffith, D.A. 1993. *Spatial Regression Analysis on the PC: Spatial Statistics Using SAS*. Association of American Geographers. Washington, DC.

Hope, B.K. 2000. Generating Probabilistic Spatially-Explicity Individual and Population Exposure Estimates for Ecological Risk Assessment. *Risk Anal.* 20(5):575–590.

Hope, B.K. 2001. A Case Study Comparing Static and Spatially Explicit Ecological Exposure Analysis Methods. *Risk Anal.* 21(6):1001–1010.

Hora, S.C. 1992. Acquisition of Expert Judgment: Examples From Risk Assessment. *J. Energy Eng*. 118:136–148.

Howson, C. and P. Urbach. 1989. *Scientific Reasoning: The Bayesian Approach*. Open Court, LaSalle, IL.

Isaaks, E. and R. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford.

Kriakidis, P.C. 1996. Selecting Panels for Remediation in Contaminated Soils via Stochastic Imaging. In: *Proceedings of the Fifth International Geostatistics Congress*, Vol. 2. (Baafi, E.Y. and N.A. Schofield, eds.). Kluwer Academic Publishers, Dordrecht, 973–983.

Lichtenstein, S. and B. Fischoff, 1977. Do Those Who Know More Also Know More About How Much They Know? *Organizational Behavior and Human Performance* 20:159.

McKenna, S.A. 1998. Geostatistical Approach for Managing Uncertainty in Environmental Remediation of Contaminated Soils: Case Study. *Environ. Engin. Geosci.* 4(2), Summer, 175–184.

Morgan, G.M. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, NY.

Ott, W.R. 1995. *Environmental Statistics and Data Analysis*. CRC Press. Boca Raton.

Patwardhan, A. and M.J. Small. 1992. Bayesian Methods for Model Ancertainty Analysis with Application to Future Sea Level Rise. *Risk Anal.* 12:513–523.

Price, P.S., C.L. Curry, P.E. Goodrum, M.N. Gray, J.I. McCrodden, N.H. Harrington, H. Carlson-Lynch, and R.E. Keenan. 1996. Monte Carlo Modeling of Time-Dependent Exposures Using a Microexposure Event Approach. *Risk Anal.* 16:339–348.

Price, P.S., J.Y. Young, C.F. Chaisson. 2000. *Assessing Aggregate and Cumulative Pesticide Risks Using LifeLine™* Version 1.0. A Report Submitted to U.S. EPA Science Advisory Panel. August 31.

Shlyakhter, A.I. and D.M. Kammen. 1992. Sea-level Rise or Fall? *Nature* 253:25.

Sivia, D.S. 1996. *Data Analysis: A Bayesian Tutorial*. Clarendon Press. Oxford.

Slob, W. 1996. A Comparison of Two Statistical Approaches to Estimate Long-Term Exposure Distributions from Short-Term Measurements. *Risk Anal.* 16: 195–200.

U.S. EPA. 1982. *Air Quality Criteria for Particulate Matter and Sulfur Oxides*. ECAO, Office of Research and Development. EPA/600/8-82/029.

U.S. EPA. 1997. *Exposure Factors Handbook. Update to Exposure Factors Handbook.* Office of Research and Development, NCEA. EPA/600/8-89/043, May 1989. August.

U.S. EPA. 2000. *Statistical Estimation and Visualization of Ground-water Contamination Data.* Office of Research and Development, Washington, DC. EPA/600/R-00/034.

Wallace, L.A., N, Duan, and R. Ziegenfus. 1994. Can Long-term Exposure Distributions be Predicted From Short-Term Measurements? *Risk Anal.* 14:75–85.

APPENDIX E

DEFINITIONS OF TERMS RELEVANT TO PRA
AND REFERENCES FOR FURTHER READING

E.0   DEFINITIONS OF TERMS

Definitions for the specialized terms pertaining to probabilistic analysis are presented in this appendix. Some of the same terms are also defined at the beginning of each chapter, sometimes with additional examples that are relevant to concepts presented in the chapter. The definitions in this guidance are intended to be consistent with definitions used in the National Contingency Plan (NCP) and other Environmental Protection Agency (EPA) guidance, including the definitions of variability, uncertainty, and Monte Carlo simulation found in EPA's *Guiding Principles for Monte Carlo Analysis* (U.S. EPA, 1997a). Note that if a definition uses a term that is defined elsewhere in the appendix, it is highlighted in bold text.

Definitions of Terms Used in PRA	
50th percentile	The number in a distribution such that half the values in the distribution are greater than the number and half the values are less. The 50th **percentile** is equivalent to the **median**.
95th percentile	The number in a distribution such that 95% of the values in the distribution are less than or equal to the number and 5% of the values are greater than the number.
95% Upper Confidence Limit for a Mean	The 95 percent upper **confidence limit** (95% UCL) for a **mean** is defined as a value that, when repeatedly calculated for randomly drawn subsets of size n, equals or exceeds the true population **mean** 95% of the time. The 95% UCL provides a measure of u**ncertainty** in the **mean**; it is not a measure of **variability** and should not be confused with a 95th **percentile**. As sample size increases, the difference between the UCL for the **mean** and the true **mean** decreases, while the **95th percentile** of the distribution remains relatively unchanged, at the upper end of the distribution. EPA's Superfund program has traditionally used the 1-sided 95% UCL for the **mean** as the **concentration term** in **point estimates** of reasonable maximum exposure (**RME)** for human health **risk assessment** (U.S. EPA, 1992, 1997b).
Applicable or Relevant and Appropriate Requirements (ARARs)	Federal or state environmental standards; the NCP states that ARARs should be considered in determining **remediation goals**. ARARs may be selected as site-specific **cleanup levels**.
Arithmetic Mean (AM)	A number equal to the average value of a population or sample. Usually obtained by summing all the values in the sample and dividing by the number of values (i.e., sample size).
Assessment Endpoint	A term usually associated with **ecological risk assessment**; an explicit expression of an environmental value (ecological resource) that is to be protected, operationally defined by risk managers and risk assessors as valuable attributes of an ecological entity. Examples include 1) sustained aquatic community structure, including species composition and relative abundance and trophic structure; 2) reductions in populations of fish-eating birds; and 3) reductions in survival, reproduction or species diversity of indigenous benthic communities (U.S. EPA, 1997c, 1999a).

Definitions of Terms Used in PRA

Backcalculation	A method of calculating a **preliminary remediation goal (PRG)** that involves algebraic rearrangement of the risk equation to solve for concentration as a function of risk, exposure, and toxicity.
Background Exposure	Exposures that are not related to the site. For example, exposure to chemicals at a different time or from locations other than the **exposure unit (EU)** of concern. Background sources may be either naturally occurring or anthropogenic (man-made).
Bayesian Analysis	Statistical analysis that describes the probability of an event as the degree of belief or confidence that a person has, given some state of knowledge, that the event will occur. Bayesian **Monte Carlo** combines a prior **probability distribution** and a **likelihood function** to yield a posterior distribution (see Appendix D for examples). Also called subjective view of probability, in contrast to the frequentist view of probability.
Bootstrap Methods	A method of sampling actual data at random, with replacement, to derive an estimate of a population **parameter** such as the **arithmetic mean** or the standard error of the **mean**. The sample size of each bootstrap sample is equal to the sample size of the original data set. Both parametric and nonparametric bootstrap methods have been developed.
Boxplot	Graphical representation showing the center and spread of a distribution, sometimes with a display of outliers (e.g., Figure 7-3). This guidance uses boxplots to represent the following **percentiles**: 5th, 25th, 50th, 75th, and 95th.
Cancer Slope Factor (CSF)	A plausible upper-bound estimate of the probability of a response per unit dose of a chemical over a lifetime. The CSF is used to estimate an upper-bound probability of an individual developing cancer as a result of a lifetime of exposure to a particular level of a potential carcinogen.
Central Limit Theorem	If random samples of size n are repeatedly drawn from a population of any distribution, the distribution of sample **means** converges to the normal distribution. The approximation improves as n increases.
Central Tendency Exposure (CTE)	A **risk descriptor** representing the average or typical individual in the population, usually considered to be the **arithmetic mean** or **median** of the risk distribution.
CTE Risk	The estimated risk corresponding to the **central tendency exposure.**
Cleanup Level	A chemical concentration chosen by the risk manager after considering both RGs and the nine selection-of-remedy criteria of the NCP (U.S. EPA, 1990; 40CFR 300.430(e)(9)(iii)). Also referred to as Final Remediation Levels (U.S. EPA, 1991), chemical-specific cleanup levels are documented in the Record of Decision (ROD). A cleanup level may differ from a **PRG** for several reasons, including various uncertainties in the risk estimate, the technical feasibility of achieving the **PRG**, and application of the nine criteria outlined in the NCP.
Coefficient of Variation	Ratio of the **standard deviation** (SD) to the **arithmetic mean** (AM) (CV=SD/AM). Dimensionless measure of the spread of a distribution, therefore, useful for comparing **probability density functions (PDFs)** for different **random variables**.

Definitions of Terms Used in PRA

Community Advisory Group (CAG)	A group formed to provide a public forum for community members to present and discuss their needs and concerns related to the Superfund decision-making process. A CAG serves as the focal point for the exchange of information among the local community, EPA, State regulatory agency, and other pertinent Federal agencies involved in the cleanup of a Superfund site.
Community Involvement Coordinator (CIC)	As a member of the **CAG** and site team, the CIC coordinates communication plans (i.e., the **Communicty Involvement Plan (CIP)** and addresses site-specific **CAG** organizational issues.
Community Involvement Plan (CIP)	A plan that identifies community concerns and the preferences of the community for the communication of site-related issues.
Concentration Term	The concentration **variable** used in **exposure assessment**. Concentration terms are expressed in units applicable to the media of concern (e.g., mg/L for water, $\mu g/m^3$ for air; mg/kg for soil and dust.
Confidence Interval	A range of values that are likely to include a population **parameter**. **Confidence intervals** may describe a **parameter** of an input **variable** (e.g., **mean** ingestion rate) or output **variable** (e.g., **95th percentile** risk). When used to characterize **uncertainty** in a risk estimate, it is assumed that methods used to quantify **uncertainty** in the model inputs are based on statistical principles such as sampling distributions or **Bayesian** approaches. For example, given a randomly sampled data set, a 95% confidence interval for the **mean** can be estimated by deriving a sampling distribution from a Student's t distribution.
Confidence Limit	The upper or lower value of a **confidence interval**.
Continuous Variable	A **random variable** that can assume any value within an interval of real numbers (e.g., concentration).
Countably Infinite	Used to describe some discrete **random variables**, this term refers to a set of numbers that can be counted with integers (e.g., one, two, three) and that has no upper limit. Examples include the number of tosses required for a coin to show a head—we can count each toss, but it is possible that at least one more toss is needed. The number of dust particles in a volume of air is another example. Countably finite implies there is an upper limit (e.g., days of work per year).
Correlation	A quantitative relationship between two or more input **variables** of a model (e.g., body weight, inhalation rate, skin surface area). In analyses involving time-dependent **variables**, a change in one **variable** is accompanied by a change in another time-dependent, correlated **variable**. Ignoring correlations in **probabilistic risk assessment (PRA)** may lead to unrealistic combinations of values in a risk calculation. Correlations can also be defined as relationships between inputs and outputs.

Definitions of Terms Used in PRA

Coverage	**Confidence intervals** are expected to enclose a true but unknown **parameter** according to a specified probability, such as 90% or 95%. This is the expected coverage of the **confidence interval**, given a specified significance level (alpha). The difference between the expected coverage and the actual coverage is one metric for evaluating statistical methods that yield different **confidence intervals**.
Credible Interval	A range of values that represent plausible bounds on a population **parameter**. Credible intervals may describe a **parameter** of an input **variable** (e.g., **mean** ingestion rate) or output **variable** (e.g., **95th percentile** risk). The term is introduced as an alternative to the term **confidence interval** when the methods used to quantify **uncertainty** are not based entirely on statistical principles such as sampling distributions or **Bayesian** approaches. For example, multiple estimates of an **arithmetic mean** may be available from different studies reported in the literature—using professional judgment, these estimates may support a decision to describe a range of possible values for the **arithmetic mean.**
Cumulative Distribution Function (CDF)	A graph that shows the cumulative probability of occurrence for a **random** independent **variable** (e.g., Fig. 6-1). The cumulative probability is typically given as the y-axis, ranging from 0 to 1.0. Each value c of the function is the probability that a random observation x will be less than or equal to c. Mathematically, the function that defines the CDF is obtained from the **PDF** by integration (in the case of a **continuous** random **variable**) or by summation (for **discrete random variables**).
Discrete Variable	A **random variable** that can assume any value within a finite set of values (e.g., number of rainfall events in one month) or at most a **countably infinite** set of values.
Empirical Distribution	A distribution obtained from actual data and possibly smoothed with interpolation techniques. Data are not fit to a particular **parametric distribution** (e.g., normal, lognormal), but are described by the **percentile** values.
Expected Value of Information (EVOI)	The expected increase in the value (or decrease in the loss) associated with obtaining more information about quantities relevant to the decision process. EVOI is a measure of the importance of **uncertainty** in risk and the potential for changing a **risk management** decision if **uncertainty** is reduced (see Appendix D).
Expert Judgment	An inferential opinion of a specialist or group of specialists within an area of their expertise. Expert judgment (alternatively referred to as professional judgment) may be based on an assessment of data, assumptions, criteria, models, and **parameters** in response to questions posed in the relevant area of expertise (see Appendix D).
Exposure Assessment	The qualitative or quantitative estimate (or measurement) of the magnitude, frequency, duration, and route of exposure. A process that integrates information on chemical fate and transport, environmental measurements, human behavior, and human physiology to estimate the average doses of chemicals received by individual receptors. For simplicity in this guidance, exposure encompasses concepts of absorbed dose (i.e., uptake and bioavailability).
Exposure Point Concentration (EPC)	The contaminant concentration within an **exposure unit** to which receptors are exposed. Estimates of the EPC represent the **concentration term** used in **exposure assessment**.

Definitions of Terms Used in PRA

Exposure Unit (EU)	A geographic area where exposures occur to the receptor of concern during the time of interest. Receptors may be human or ecological (e.g., plants, birds, fish, mammals). For purposes of **PRA**, **probability distributions** for exposure and toxicity **variables** apply equally to all members of a population at a given exposure unit. Ecological exposure units often consider habitat and seasonality factors that enhance exposure in a spatial area usually related to home ranges.
Forward Calculations	A method of calculating a risk estimate that involves the standard arrangement of the risk equation to solve for risk as a function of concentration, exposure, and toxicity.
Frequency Distribution	A graph or plot that shows the number of observations that occur within a given interval; usually presented as a **histogram** showing the relative probabilities for each value. It conveys the range of values and the count (or proportion of the sample) that was observed across that range.
Frequentist	A term referring to classical statistics in which the probability of an event occurring is defined as the frequency of occurrence measured in an observed series of repeated trials.
Geometric Mean (GM)	The nth root of the product of n observations. For lognormal distributions, the GM is equal to the median and is less than the **arithmetic mean**. For normal distributions, all three measures of central tendency (GM, **AM**, **median**) are equal.
Geostatistics	Branch of statistics that focuses on data that have a spatial or geographic components. In risk assessment, geostatistics is a general term for a variety of techniques that are typically applied to chemical concentrations in soil or groundwater in which the sampling locations are considered in quantifying the **exposure point concentration**.
Goodness-of-Fit (GoF) Test	A method for examining how well (or poorly) a sample of data can be described by a hypothesized **probability distribution** for the population. Generally involves an **hypothesis test** in which the null **hypothesis** H_0 is that a **random variable** X follows a specific **probability distribution** F_0. That is, $H_0: F = F_0$ and $H_a: F \neq F_0$.
Hazard Index (HI)	The sum of more than one **hazard quotient** for multiple substances and/or multiple exposure pathways. The HI is calculated separately for chronic, subchronic, and shorter-duration exposures.
Hazard Quotient (HQ)	The ratio of estimated site-specific exposure to a single chemical from a site over a specified period to the estimated daily exposure level, at which no adverse health effects are likely to occur.
Hazardous Substance Research Centers (HSRC)	Research centers providing free technical assistance to communities with environmental contamination programs through two distinct outreach programs: **Technical Outreach Services for Communities (TOSC)** and Technical Assistance to Brownfields Community (TAB).
High-end Risk	A **risk descriptor** representing the high-end, or upper tail of the risk distribution, usually considered to be equal to or greater than the 90th **percentile**.

Definitions of Terms Used in PRA

Histogram	A graphing technique which groups the data into intervals and displays the count of the observations within each interval. It conveys the range of values and the relative frequency (or proportion of the sample) that was observed across that range.
Hypothesis Testing	Statistical test of an assumption about a characteristic of a population. The goal of the statistical inference is to decide which of two complementary hypotheses is likely to be true.
Image Analysis	A technique in **geostatistics** used to restore a degraded image or interpret images that have been contaminated by noise or possibly some nonlinear transformation.
Independence	Two events *A* and *B* are independent if knowing whether or not *A* occurs does not change the probability that *B* occurs. Two **random variables** *X* and *Y* are independent if the joint **probability distribution** of *X* and *Y* can be expressed as the product of the individual marginal **probability distributions**. That is, f(X, Y) = f(X) · f(Y). Independence of *X* and *Y* is *not* synonymous with zero **correlation** (i.e., Cor(X, Y) = 0). If *X* and *Y* are independent, then Cor(*X, Y*) = 0; however, the converse is not necessarily true because *X* and *Y* may be related in a nonlinear fashion but still maintain zero **correlation** (Law and Kelton, 1991).
Independent and Identically Distributed (IID)	**Random variables** that are independent and have the same **probability distribution** of occurrence.
Individual-Level Effect	An **assessment endpoint** that focuses on protecting a hypothetical or real individual in a population. Individual-based models may account for unique exposure and toxicological response to chemicals among individual receptors.
Iterative Reduction (IR)	A method of calculating a **PRG** that involves successively lowering the **concentration term** until the calculated risk is acceptable. This method can be applied to any medium.
Iterative Truncation	A method of calculating a **PRG** that involves developing an expression for the **concentration term** in which high-end values are "truncated" to reduce the maximum concentration, and calculating risks associated with the reduced concentration. The method may be repeated with consecutively lower **truncation** limits until risk is acceptable. Iterative **truncation** methods avoid difficulties associated with applying **Monte Carlo analysis** to a **backcalculation**.
Kriging	A statistical interpolation method that selects the best linear unbiased estimate of the **parameter** in question. Often used as a **geostatistical** method of spatial statistics for predicting values at unobserved locations based on data from the surrounding area. Information on fate and transport of chemicals within the area lacking data can be incorporated into kriged estimates.
Kurtosis	The measure of peakedness of a distribution. A uniform distribution has a lower kurtosis than a peaked distribution such as the normal and lognormal distribution. Kurtosis is referred to as the 4th central **moment of a distribution**.

Definitions of Terms Used in PRA

Land Method	The conventional method for calculating **uncertainty** in the **mean** concentration (e.g., **95% UCL**) when the sample data are obtained from a lognormal distribution (U.S. EPA, 1992).
Latin Hypercube Sampling (LHS)	A variant of the **Monte Carlo sampling** method that ensures selection of equal numbers of values from all segments of the distribution. LHS divides the distribution into regions of equal sampling **coverage**. Hence, the values obtained will be forced to cover the entire distribution. It is more efficient than simple random sampling, i.e., it requires fewer iterations to generate the distribution sufficiently.
Likelihood Function	A term from **Bayesian** statistics referring to a **probability distribution** that expresses the probability of observing new information given that a particular belief is true.
Local Sensitivity Analysis	Evaluation of the model sensitivity at some nominal points within the range of values of input variable(s).
Location Tag	The spatial coordinates of a sampling location (e.g., longitude, latitude).
Low-end Risk	A **risk descriptor** representing the low-end, or lower tail of the risk distribution, such as the 5th or 25th **percentile**.
Maximum Detected Concentration (MDC)	The maximum concentration detected in a sample.
Mean	**Arithmetic mean** or average; the sum of all observations divided by the number of observations. Referred to as the first central **moment of a distribution**.
Microexposure Event (MEE) Analysis	A method of assessing risk based on an aggregate sum of a receptor's contact with a contaminated medium. MEE analysis simulates lifetime exposure as the sum of many short-term, or "micro" exposures (see Appendix D). MEE approaches can be used to explore **uncertainty** associated with the model **time step** in **PRA** (e.g., use of a single value to represent a long-term average phenomenon, seasonal patterns in exposure, or intra-individual **variability**).
Mode	The most probable value of a **random variable**; a value with the largest probability or highest probability density (or mass for discrete **random variable**). The second **parameter** of a triangular distribution.
Moments of a Distribution	Similar to a **parameter**; constant that represents a mathematical description of a **random variable**. Central moments are defined with respect to the **mean**. **Mean**, **variance**, **skewness**, and **kurtosis** are the first, second, third, and fourth central moments of a **probability distribution**.

Definitions of Terms Used in PRA

Monte Carlo Analysis (MCA) or Simulation	A technique for characterizing the **uncertainty** and **variability** in risk estimates by repeatedly sampling the **probability distributions** of the risk equation inputs and using these inputs to calculate a distribution of risk values. A set of iterations or calculations from Monte Carlo sampling is a simulation. For example, a single iteration for risk from ingestion of water may represent a hypothetical individual who drinks 2 L/day and weighs 65 kg; another iteration may represent a hypothetical individual who drinks 1 L/day and weighs 72 kg.
Monte Carlo Sampling	A method of simple random sampling used to obtain a distribution of values which may serve as an input to a **PRA**. The probability of obtaining any given sample is similar to the probability of a sample occurring within the distribution. Hence, for a given sample size, simple random sampling tends to produce values clustered around the **mean** of the distribution.
Multiple Regression Analysis	A statistical method that describes the extent, direction, and strength of the relationship between several (usually continuous) independent **variables** (e.g., exposure duration, ingestion rate) and a single continuous dependent **variable** (e.g., risk).
Nonparametric Method	A procedure for making statistical inferences without assuming that the population distribution has any specific form such as normal or lognormal. Sometimes referred to as *distribution-free* methods. Common examples are the sign test, **Spearman rank correlation**, and the **bootstrap**-t approach.
Numerical Stability	The property of a probabilistic simulation such that the a **parameter** value of the output distribution (e.g., **percentile**, **mean**, **variance**, etc.) remains sufficiently constant for a specified number of **Monte Carlo** iterations. Numerical **stability** is a measure of the precision of the output from a simulation; the tails of the distribution are typically less stable than the center. Sufficient precision is determined by professional judgment.
One-dimensional Monte Carlo Analysis (1-D MCA)	A method of simulating a distribution for an endpoint of concern as a function of **probability distributions** that characterize **variability** or **uncertainty**. In this guidance, distributions used to characterize variability may be abbreviated **PDF**v, whereas distributions used to characterize uncertainty may be abbreviated **PDF**u. It is good practice *not* to combine **PDF**s for **variability** and **uncertainty** in 1-D MCA.
Parameter	A value that characterizes the **probability distribution** of a **random variable**. For example, a normal **probability distribution** may be defined by two parameters (e.g., **AM** and SD). It is important to distinguish between this definition, and a second popular use of the term parameter when referring to an input variable in a mathematical equation or model. For this guidance, the term *variable* will be used to describe inputs to a model. For example, if body weight is a variable in the exposure assessment that we define with a probability distribution (e.g., normal) we would state that the variable is body weight and the parameters are the arithmetic **mean** and **standard deviation** values that characterize the normal distribution
Parametric Distribution	A theoretical distribution defined by one or more **parameters.** Examples are the normal distribution, the lognormal distribution, the triangular distribution, and the beta distribution.

Definitions of Terms Used in PRA

Percentile	The *p[th] percentile* of the distribution is the value such that *p* percent of the observations fall at or below it. Also called *quantiles* or *fractiles*; percentiles are expressed as a percent, ranging from 0 to 100, whereas quantiles or fractiles range from 0 to 1.
Point Estimate	A quantity calculated from values in a sample to represent an unknown population **parameter**. Point estimates typically represent central tendency or upper bound estimate of **variability**.
Point Estimate Risk Assessment	The familiar **risk assessment** methodology in which a single estimate of risk is calculated from a set of **point estimates**. The results provide **point estimates** of risk for the **CTE** and **RME** exposed individuals. **Variability** and **uncertainty** are discussed in a qualitative manner.
Point Pattern Analysis	A technique in **geostatistics** of restricting the analysis to location information, ignoring attribute information, addresses two location problems: (1) describing points according to spacing, and (2) describing points according to density.
Population-Level Effect	An ecological term for an **assessment endpoint** that focuses on protecting a group of individuals within a specified **exposure unit** and time that have similar exposures and toxicological responses to chemicals.
Posterior Distribution	A term from **Bayesian** statistics referring to a **probability distribution** that has been updated with new information.
Potentially Responsible Party (PRP)	Individuals, companies, or any other party that is potentially liable for Superfund cleanup costs.
Power	The probability that a test procedure detects a false null **hypothesis**; Power equals (1-β), where β is the probability of a **Type II error** (i.e., accepting H_0 when H_a is true). Power curves are a function of a fixed significance level (α), sample size, and **variability (SD)**.
Preliminary Remediation Goal (PRG)	A chemical concentration in an environmental medium associated with a particular exposure scenario that is expected to be protective of human health and ecosystems. PRGs may be developed based on (**ARARs**), or exposure scenarios evaluated prior to a **risk assessment** (e.g., generic PRG) or as a result of the baseline **risk assessment** (site-specific PRG). Exhibit 5-1 provides further detail on generic and site-specific PRGs.
Prior Distribution	A **Bayesian** term referring to the hypothesized, expected, or calculated **probability distribution** for an event prior to the collection of new information.
Probabilistic Risk Assessment (PRA)	A **risk assessment** that uses probabilistic methods to derive a distribution of risk or hazard based on multiple sets of values sampled for **random variables**.

Definitions of Terms Used in PRA

Probability Density Function (PDF)	A graph that shows the probability of occurrence of an unknown or **variable** quantity. A PDF is used to characterize a **continuous random variable;** the integral of all possible values is equal to 1.0 (i.e., the area under the curve). In PRA, PDFs can be used to display the shape of the distribution for an input variable (e.g., normal distribution for ingestion rate) as well as the output from a **Monte Carlo simulation** (e.g., risk distribution).
Probability Distribution	A function that associates probabilities with the values taken by a **random variable**. A probability distribution can be displayed in a graph (e.g., **PDF** or **CDF**), summarized in a table that gives the distribution name and parameters, or expressed as a mathematical equation. In PRA, the process of selecting or fitting a distribution that characterizes **variability** or **uncertainty** can also be referred to as applying a *probability model* to characterize **variability** or **uncertainty**. In this guidance, the probability model is considered to be one source of model uncertainty.
Probability Mass Function (PMF)	A **histogram** that shows the probability of occurrence of an unknown or **variable** quantity. A PMF is used to characterize a **discrete random variable**; similar to the **PDF**, the sum of all possible values of a PMF is equal to 1.0. The mass at a point refers to the probability that the **variable** will have a value at that point.
Random Variable	A **variable** that may assume any value from a set of values according to chance. Discrete random **variables** can assume only a finite or **countably infinite** number of values (e.g., number of rainfall events per year). A random value is continuous if its set of possible values is an entire interval of numbers (e.g., quantity of rain in a year) **variable** that may assume any of a set of values. The likelihood of each value is described by a **probability distribution**.
Range Sensitivity Analysis	Evaluation of the model sensitivity across the entire range of values of the input variable(s).
Rank	If a set of values is sorted in ascending order (smallest to largest), the rank corresponds to the relative position of a number in the sequence. For example, the set {7, 5, 9, 12} when sorted gives the following sequence {5, 7, 9, 12} with ranks ranging from 1 to 4 (i.e., rank of 5 is 1, rank of 7 is 2, rank of 9 is 3, and rank of 12 is 4).
Rank Correlation (Spearman Rank Order Correlation Coefficient)	A "distribution free" or nonparametric statistic *r* that measures the strength and direction of association between the ranks of the values (not the values themselves) of two quantitative **variables**.
Remedial Investigation/Feasibility Study (RI/FS)	Studies undertaken by EPA to delineate the nature and extent of contamination, to evaluate potential risk, and to develop alternatives for cleanup.
Reasonable Maximum Exposure (RME)	The highest exposure that is reasonably expected to occur at a site (U.S. EPA, 1989, 1990). The intent of the RME is to estimate a conservative exposure case (i.e., well above the average case) that is still within the range of possible exposures.
RME Risk	The estimated risk corresponding to the **reasonable maximum exposure**.

Definitions of Terms Used in PRA

Reference Dose (RfD)	An estimate of an exposure level for the human population, including sensitive subpopulations, that is likely to be without an appreciable risk of deleterious effects during a lifetime. Chronic RfDs are specifically developed to be protective for a long-term exposure to a chemical (e.g., >7 years) and account for **uncertainty** spanning perhaps an order of magnitude or greater.
Remediation Action Level (RAL)	Generally, a concentration such that remediation of all concentrations above this level in an **exposure unit** will result in the **95% UCL** being reduced to a level that does not pose an unacceptable risk to an individual experiencing random exposures. The RAL will depend on the **mean**, **variance**, and sample size of the concentrations within an **exposure unit** as well as considerations of acute toxicity of the chemicals of concern.
Remediation Goal	Generally, a health-based chemical concentration in an environmental medium chosen by the risk manager as appropriate for a likely land use scenario.
Risk Assessment	The use of available information to make inferences about the health effects associated with exposure of individuals or populations to hazardous materials or situations. Components of risk assessment include: hazard identification, dose-response assessment, **exposure assessment**, and **risk characterization** (NRC, 1983).
Risk Characterization	A component of **risk assessment** that describes the nature and magnitude of risk, including **uncertainty**. In assessments of Superfund sites, it includes the summary and interpretation of information gathered from previous steps in the site **risk assessment** (e.g., data evaluation, **exposure assessment**, toxicity assessment), including the results of a probabilistic analysis.
Risk Descriptor	A statistic (e.g., arithmetic **mean**, 95th **percentile**) that describes the risk to the **assessment endpoint**.
Risk Management	The process by which regulatory decisions are made using all available **risk assessment** information (including, but not limited to, the results of the **PRA**). The NCP provides nine criteria for remedial decisions (e.g., protection of human health, compliance with **ARARs**, etc.). Risk managers may include the Remedial Project Manager (RPM), section and branch chiefs, etc.
RME Range	The 90th to 99.9th **percentiles** of the risk distribution generated from a **PRA**, within which an **RME** risk value may be identified. The **95th percentile** is generally recommended as the starting point for specifying the **RME** risk in a Superfund **PRA**.
Scientific/Management Decision Point (SMDP)	A point during the **risk assessment** process when the risk assessor communicates results of the assessment at that stage to the risk manager. At this point, the risk manager determines whether the information is sufficient to arrive at a decision regarding **risk management** strategies and/or if additional information is needed to characterize risk.

Definitions of Terms Used in PRA

Sensitivity Analysis	Process for identifying the important sources of variability and uncertainty in a model's output. Different techniques can be used in each of the 3 tiers of the tiered process for PRA (see Chapter 2). In Tier 1, sensitivity ratios are used to quantify the effects of changes in one or more model inputs on the model output. In Tiers 2 and 3, correlation analysis can be used to rank inputs based on their relative contribution to **variance** in risk. Local sensitivity refers to nominal changes in inputs within a plausible range, whereas range sensitivity refers to changes in inputs across the minimum and maximum values of the plausible range. Further explanations of the different methods for sensitivity analysis are given in Appendix A.
Sensitivity Ratio	Ratio of the change in model output per unit change in an input **variable**; also called *elasticity*.
Skewness	The measure of asymmetry of a distribution. Coefficients of skewness are zero for symmetric distributions (e.g., normal), positive for right-skewed distributions (e.g., lognormal), and negative for left-skewed distributions (e.g., specific forms of beta) . Referred to as the third central **moment of a distribution**.
Spatial Autocorrelation	The tendency of data from locations that are relatively close together to be geographically correlated.
Stakeholder	Any individual or group who has an interest in or may be affected by EPA's site decision-making process.
Stability	**Stochastic variability**, or "wobble" associated with random sampling, calculated as the average percent change in the model output after rerunning **Monte Carlo simulations** with the same set of input assumptions. Used as a metric for evaluating the adequacy of the number of iterations in a **MCA**.
Standard Deviation, Arithmetic and Geometric	Standard deviation (or arithmetic standard deviation, SD) is a common measure of the spread of a distribution. Calculated as the square root of the **variance.** The geometric standard deviation (GSD) is the anti-log of the standard deviation of the logarithms of each value. The GSD is a unitless quantity that gives a measure of the ratio of the variance to the mean, similar in concept to the **coefficient of variation**.
Step Function	A mathematical function that remains constant within each of a series of adjacent intervals but changes in value from one interval to the next. **Cumulative distribution functions** for **discrete random variables** are step functions.
Stochastic Dominance	Implies no intersection between the **CDFs**; distribution A stochastically dominates distribution B if, for every **percentile** of the **CDF**, A > B. This characteristic may not be apparent from the **PDFs** of the distributions, which may overlap.
Stochastic Process	A process involving **random variables**, and characterized by **variability** in space or time.
Target Population	The set of all receptors that are potentially at risk. Sometimes referred to as the "population of concern". A sample population is selected for statistical sampling in order to make inferences regarding the target population (see Appendix B, Section B.3.1, Concepts of Populations and Sampling).

Definitions of Terms Used in PRA

Technical Assistance Grant (TAG)	A federal grant that is intended to provide a community with the opportunity to hire independent experts to help evaluate and explain the results of a **risk assessment**
Technical Outreach Services for Communities (TOSC)	A service of the **HSRC** with the aim to provide independent technical information and assistance to help communities with hazardous substance pollution problems.
Thiessen (Voronoi) Polygon Analysis	A method of spatial statistics in which an area is subdivided into subregions, or polygons, in order to predict values at unobserved locations.
Time Step	A **variable** in all exposure models that refers to the unit of time for which a **random value** is considered representative of intra-individual **variability** (e.g., average daily ingestion rates for an individual from one year to the next). A time step may be equal to an entire exposure duration (e.g., 30 years), or a fraction of the exposure duration during which changes in input **variables** may be expected (e.g., one year). Time steps need not be identical for all exposure **variables**, and should address the most rapidly changing **variable** in the risk equation. Time step can be an important consideration for **MEE analysis**.
Toxicity Reference Value (TRV)	A numerical expression of a chemical's dose-response relationship that is used in **ecological risk assessment**.
True Mean Concentration	The actual average concentration in an **exposure unit**. Even with extensive sampling, the true **mean** cannot be known. Only an estimate of the true **mean** is possible. A greater number of representative samples increases confidence that the estimate of the **mean** more closely represents the true **mean**.
Truncation	The process of setting lower and upper limits on the range of a distribution, in order to avoid unrealistic values for exposure **variables** (e.g., > 100% bioavailability). Most often used for continuous, unbounded **probability distributions** (e.g., normal).
Two-dimensional Monte Carlo Analysis (2-D MCA)	An advanced modeling technique that uses two stages of random sampling, also called nested loops, to distinguish between **variability** and **uncertainty** in exposure and toxicity **variables**. The first stage, often called the inner loop, involves a complete **1-D MCA** simulation of **variability** in risk. In the second stage, often called the outer loop, **parameters** of the **probability distributions** are redefined to reflect **uncertainty**. These loops are repeated many times resulting in multiple risk distributions, from which **confidence intervals** are calculated to represent **uncertainty** in the population distribution of risk.
Type I Errors	False positive; the error made when the null **hypothesis** is rejected in favor of the alternative, when in fact the null **hypothesis** is true.
Type II Errors	False negative; the error made when the null **hypothesis** is accepted when in fact the alternative **hypothesis** is true.
Uncertainty	Lack of knowledge about specific **variables**, **parameters**, models, or other factors. Examples include limited data regarding the concentration of a contaminant in an environmental medium and lack of information on local fish consumption practices. Uncertainty may be reduced through further study.

Definitions of Terms Used in PRA	
Variability	True heterogeneity or diversity in characteristics among members of a population (i.e., inter-individual variability) or for one individual over time (intra-individual variability). For example, body weights of a study population at one point in time will exhibit variability, and body weight will change as an individual ages. Further study (e.g., increasing sample size, *n*) will not reduce variability, but it can provide greater confidence in quantitative characterizations of variability.
Variable	A quantity that can assume many values.
Variance	Measure of the spread of a distribution, equal to the square of the **standard deviation** (SD). Calculated as the average of the squares of the deviations of the observations from their **mean**. Variance is referred to as the second central **moment of a distribution**.
Z-score	The value of a normally distributed **random variable** that has been standardized to have a **mean** of zero and a SD of one by the transformation $Z=(X–\mu)/\sigma$. Statistical tables typically give the area to the left of the z-score value. For example, the area to the left of $z =1.645$ is 0.95. Z-scores indicate the direction (+/-) and number of **standard deviations** away from the **mean** that a particular datum lies assuming X is normally distributed. Microsoft Excel's *NORMSDIST(z)* function gives the probability *p* such that $p=Pr(Z \le z)$, while the *NORMSINV(p)* function gives the z-score z_p associated with probability *p* such that $p=Pr(Z \le z_p)$.

E.1.0 ADDITIONAL INFORMATION

Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis (Morgan and Henrion, 1990) and *Probabilistic Techniques in Exposure Assessment* (Cullen and Frey, 1999) provide excellent philosophical and practical treatises on probabilistic risk assessment. These works are highly recommended to risk assessors who wish to know more about probabilistic risk assessment. The *Summary Report for the Workshop on Monte Carlo Analysis* (U.S. EPA, 1996) and the *Summary Report for the Workshop on Selecting Input Distributions for Probabilistic Assessments* (U.S. EPA, 1999b) are other sources of information to learn more about PRA. Other additional references for reading are listed in this Appendix.

REFERENCES FOR APPENDIX E

Cullen, A.C. and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment.* A Handbook for
    Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press.

Law, A.M. and W.D. Kelton. 1991. Simulation Modeling and Analysis. McGraw-Hill, Inc., NY.

Morgan, G.M. and M. Henrion, 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative*
    Risk and Policy Analysis. Cambridge University Press, NY.

National Research Council (NRC). 1983. *Risk Assessment in the Federal Government: Managing the*
    Process. National Academy Press. Washington, DC.

U.S. EPA. 1989. *Risk Assessment Guidance for Superfund* (RAGS): *Volume I. Human Health Evaluation*
    Manual (HHEM) *(Part A, Baseline Risk Assessment).* Interim Final. Office of Emergency and
    Remedial Response, Washington, DC. EPA/540/1-89/002. NTIS PB90-155581.

U.S. EPA. 1990. National Oil and Hazardous Substances Pollution Contingency Plan. Final Rule. 40
    CFR 300: *55 Federal Register*, 8666-8865, March 8.

U.S. EPA. 1991. *Risk Assessment Guidance for Superfund* (RAGS)*, Volume I: Human Health Evaluation*
    Manual (HHEM)*, Part B, Development of Risk-Based Preliminary Remediation Goals.* Office of
    Emergency and Remedial Response, Washington, DC. EPA/540/R-92/003. NTIS PB92-963333.

U.S. EPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. Office of Solid
    Waste and Emergency Response, Washington, DC. OSWER Directive No. 9285.7-081.

U.S. EPA. 1997a. *Memorandum from Deputy Administrator Fred Hansen on the Use of Probabilistic*
    Techniques (including Monte Carlo Analysis) in Risk Assessment, and Guiding Principles for
    Monte Carlo Analysis. Office of Research and Development, Washington, DC.
    EPA/630/R-97/001. May 15.

U.S. EPA. 1997b. *Lognormal Distribution in Environmental Applications*. Office of Research and
    Development, and Office of Solid Waste and Emergency Response, Washington, DC.
    EPA/600/R-97/006. December.

U.S. EPA. 1997c. *Ecological Risk Assessment Guidance for Superfund: Process for Designing*
    and Conducting Ecological Risk Assessments. Interim Final. U.S. Environmental
    Protection Agency, Environmental Response Team (Edison, NJ). June 5.

U.S. EPA. 1999a. Issuance of Final Guidance: Ecological Risk Assessment and Risk Management
    Principles for Superfund Sites. U.S. Environmental Protection Agency, Office of Solid Waste
    and Emergency Response. OSWER Directive 9285.7-28 P. Stephen D. Luftig for Larry D.
    Reed. October 7.

U.S. EPA. 1999b. *Summary Report for the Workshop on Selecting Input Distributions for Probabilistic*
    Risk Assessment. Risk Assessment Forum. EPA/630/R-98/004.

REFERENCES FOR FURTHER READING

Baird, BF. 1989. *Managerial Decisions Under Uncertainty*. John Wiley & Sons, Inc., NY.

Bevington, P.R. 1969. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, NY.

Bratley, P., B.L. Fox, and L.E. Schrage. 1987. *A Guide to Simulation*. Springer-Verlag, NY.

Burmaster, D.E. and P.D. Anderson. 1994. Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessment. *Risk Anal.* 14(4):477–481.

Clemen, R, 1990. *Making Hard Decisions*. Duxbury Press.

Conover, W.I. 1971. *Practical Nonparametric Statistics*. John Wiley & Sons, NY.

Cox, D.C. and P. Baybutt. 1981. *Methods for Uncertainty Analysis: A Comparative Survey. Risk Anal* 1(4):251–258.

Cullen, A.C. and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment*. Plenum Press, NY.

D'Agostino, R. and M.A. Stephens, eds. 1986. *Goodness-of-Fit Techniques*. Marcel Dekker, Inc., NY.

Devroye, L. 1986. *Non-Uniform Random Deviate Generation*. Springer-Verlag, NY.

Evans, M., N. Hastings, and B. Peacock. 1993. *Statistical Distributions*. John Wiley & Sons, NY.

Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers*. Resources for the Future, Washington, DC.

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

Hamby, D.M. 1994. A review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environ. Monit. and Assess*. 32:135–154.

Hammersley, J.M. and D.C. Handscomb. 1964. *Monte Carlo Methods*. John Wiley & Sons, NY.

Hertz, D.B. and H. Thomas. 1983. *Risk Analysis and Its Applications*. John Wiley & Sons, NY.

Hertz, D.B. and H. Thomas. 1984. *Practical Risk Analysis - An Approach Through Case Studies*. John Wiley & Sons, NY.

Hoffman, F.O. and J.S. Hammonds. 1992. *An Introductory Guide to Uncertainty Analysis in Environmental and Health Risk Assessment*. ES/ER/TM-35, Martin Marietta.

Hoffman, F.O. and J.S. Hammonds. 1994. *Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability*. *Risk Anal* 14(5):707–712.

Iman, R.L. and W.J. Conover. 1982. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Commun. Stat*, Part B 11:311–331.

Iman, R.L. and J.C. Helton. 1988. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Anal*. 8(1):71–90.

Iman, R.L., J.M. Davenport, and D.K. Zeigler. 1980. *Latin Hypercube Sampling (A Program Users Guide)*. Technical Report SAND 79:1473, Sandia Laboratories, Albuquerque.

Johnson, M.E. 1987. *Multivariate Statistical Simulation*. John Wiley & Sons, NY.

Johnson, N.L. and S. Kotz. 1970. *Continuous Univariate Distributions*. Vols. 1 & 2. John Wiley & Sons, NY.

Johnson, N.L., S. Kotz, and A.W. Kemp. 1992. *Univariate Discrete Distributions*. John Wiley & Sons, NY.

Kendall, M. and A. Stuart. 1979. *Advanced Theory of Statistics, Volume I - Distribution Theory, Volume II - Inference and Relationship*. MacMillan, Inc., NY.

Kennedy, W.J. and E. Gentle. 1980. *Statistical Computing*. Marcel Dekker, Inc., NY.

LePage, R. and L. Billard. 1992. *Exploring the Limits of Bootstrap*. John Wiley & Sons, NY.

Lipton, J., W.D. Shaw, J. Holmes, and A. Patterson. 1995. Short Communication: Selecting Input Distributions for use in Monte Carlo Analysis. *Regul. Toxicol. Pharmacol.* 21:192–198.

McKone, T.E. and K.T. Bogen. 1992. Uncertainties in Health Risk Assessment: An Integrated Case Based on Tetrachloroethylene in California Groundwater. *Regul. Toxicol. Pharmacol.* 15:86–103.

Megill, R.E., ed. 1985. *Evaluating and Managing Risk*. Penn Well Books, Tulsa, OK.

Morgan, G.M. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, NY.

NCRP. 1996. Commentary No. 14. *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*. National Committee on Radiation Programs, Scientific Committee 64-17. Washington, DC.

Palisade Corporation. 1994. *Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3*. Windows Version Release 3.0 User's Guide, Palisade Corporation, Newfield, NY.

Press, W.H., B.P. Flannery, S.A. Teulolsky, and W.T. Vetterling. 1989. *Numerical Recipes in Pascal: the Art of Scientific Computing*. Cambridge University Press, NY.

Press, W.H., S.A. Teulolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in FORTRAN: the Art of Scientific Computing*. Cambridge University Press, NY.

Press, W.H., S.A. Teulolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, NY.

Read, T. and N. Cressie. 1988. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, NY.

Rohatgi, V.K. 1984. *Statistical Inference*. John Wiley & Sons, NY.

Rubenstein, R.Y. 1981. *Simulation and the Monte Carlo Method*. John Wiley & Sons, NY.

Sachs, L. 1984. *Applied Statistics - A Handbook of Techniques*. Springer-Verlag, NY.

Saltelli, A and J. Marivort. 1990. Non-Parametric Statistics in Sensitivity Analysis for Model Output: A Comparison of Selected Techniques. *Reliab. Eng. Syst. Saf.* 28:299–253.

Schneider, H. 1986. *Truncated and Censored Distributions from Normal Populations*. Marcel Dekker, Inc., NY.

Seiler, F.A. 1987. Error Propagation for Large Errors. *Risk Anal*. 7(4):509–518.

Seiler, F.A. and J.L. Alvarez. 1996. On the Selection of Distributions for Stochastic Variables. *Risk Anal.* 16(1):5–18.

Slob, W. 1994. Uncertainty Analysis in Multiplicative Models. *Risk Anal*. 14(4):571–576.

Smith, A.E., P.B. Ryan, and J.S. Evans. 1992. The Effect of Neglecting Correlations when Propagating Uncertainty and Estimating the Population Distribution of Risk. *Risk Anal*. 12(4):467474.

Smith, R.L. 1994. Uses of Monte Carlo Simulation for Human Exposure Assessment at a Superfund Site. *Risk Anal* 14(4):433–439.

Sokal, R. and R. Rohlf. 1981. *Biometry: The Principles and Practice of Statistics in Biological Research*. Second Edition. W.H. Freeman & Co., NY.

U.S. EPA. 1978. *Source Assessment: Analysis of Uncertainty - Principles and Applications*. EPA/600/2-79-004.

U.S. EPA. 1992a. *Guidelines for Exposure Assessment*. *Federal Register*. 57(104):22888-22938. May 29.

U.S. EPA. 1992b. *Guidelines for Carcinogenic Risk Assessment*. *Federal Register*. 51(185):33992-34003. May 29.

U.S. EPA. 1996. *Summary Report for the Workshop on Monte Carlo Analysis*. Office of Research and Development, Washington, DC. EPA/630/R-96/010.

U.S. EPA. 1999. *Guidelines for Carcinogenic Risk Assessment*. Review Draft. Risk Assessment Forum. Washington, DC. NCEA-F-0644.

U.S. EPA. 2001. *Guidelines for Carcinogenic Risk Assessment*. *Federal Register*. 66(230):59593-59594. November 29.

Wilks, D.S. 1995. *Statistical Methods in the Atmospheric Sciences, An Introduction*. Academic Press, San Diego.

APPENDIX F

WORKPLAN AND CHECKLIST FOR PRA

F.0   INTRODUCTION

This appendix provides guidance on developing a workplan prior to the initiation of a probabilistic risk assessment (PRA), and using a checklist when reviewing a PRA. Like the quality assurance project plan (QAPP), the workplan for PRA generally should document the combined decisions or positions of the remedial project manager (RPM), risk assessor, and stakeholders involved in the risk assessment. Often there are many stakeholders in a risk assessment, and it is important to involve and engage all stakeholders early in the decision-making process. These are important steps that should save time and effort.

F.1.0   WORKPLAN

In general, PRAs may be developed by Environmental Protection Agency (EPA), EPA contractors, or a potentially responsible party (PRP) with appropriate EPA oversight. In each case, it is important to develop a workplan early in the risk assessment process. PRAs to be submitted by a contractor or PRP should generally be submitted for EPA review before commencing the analysis. The workplan should describe the software to be used, the exposure routes and models, and input probability distributions and their basis (e.g., relevance to the site-specific contamination and pathways), including appropriate literature references. Examples of the elements of a workplan are given in Exhibit F-1, as well as Exhibit 4-8 in Chapter 4 (Example Elements of a Workplan for Ecological PRA). It is important that the risk assessor and risk manager discuss the scope of the probabilistic analysis and the potential impact on the Remedial Investigation/Feasibility Study (RI/FS).

EXHIBIT F-1

EXAMPLES OF ELEMENTS
OF THE WORKPLAN FOR PRA

1. Statement of the ecological assessment endpoints and/or human risk

2. Summary of the point estimate risk assessment

3. Potential value added for risk management by conducting a PRA and proceeding to the subsequent tiers (quantify variability, uncertainty, or both)

4. Discussion of adequacy of environmental sampling for PRA (e.g., data quality issues)

5. Description of the methods and models to be used (e.g., model and parameter selection criteria)

6. Proposal and basis for probability distributions and point estimates

7. Methods for deriving the concentration term

8. Proposal for probabilistic sensitivity analysis

9. Method for dealing with correlations

10. Bibliography of relevant literature

11. Software (i.e., date and version of product, random number generator)

12. Simulation approach (e.g., iterations, Monte Carlo or Latin Hypercube sampling, time step)

13. Proposed schedule and expertise needed

☞   *Given the time and effort that can be expected to be invested in conducting a PRA, it is important that a workplan undergo review and approval by EPA, prior to proceeding with the assessment.*

The EPA generally will not accept probabilistic analysis where a workplan for the analysis has not been initially submitted to the Agency and approved by the Regional risk assessor and RPM. Exceptions to this process may be considered on a case-by-case basis.

Conducting a PRA is an iterative process. In general, as new information becomes available, it should be used to evaluate the need to move to a higher tier. The decision to move an assessment to a higher tier of complexity should result in a revised workplan and consultation with the Agency. The previous PRA, and its sensitivity analysis, should be included in the revised workplan, along with a point estimate risk assessment based on any data collected as part of a lower tier. The assessment will often be restricted to the chemicals and pathways of concern that contribute the greatest risk.

Throughout the process of developing the PRA, the EPA risk assessor and the personnel involved in developing the assessment should have a continuing dialogue to discuss the many Agency decisions and their potential impact on the assessment. This dialogue, along with interim deliverables, will help to ensure that the risk assessment report will meet the needs of the Agency and that any problems are identified and corrected early in the process.

F.2.0   FOCAL POINTS FOR PRA REVIEW

In reviewing a PRA, it is recommended that a systematic approach be adopted to ensure that all key technical elements of the PRA are evaluated and potential weaknesses are identified. A review check list can facilitate this process and promote consistency in the reviews of PRAs. Such a list can be developed from EPA's guiding principles (U.S. EPA, 1997) and other reviews on the subject of PRA quality review (e.g., Burmaster and Anderson, 1994).

In general, the review of a PRA can be organized into four focal points listed in Exhibit F-2. PRAs can vary in complexity, from relatively simple to very complicated; thus, the review strategy may need to be customized for specific sites.

> **EXHIBIT F-2**
>
> **KEY FOCAL POINTS FOR PRA REVIEW**
>
> 1. Clarity of and conformation to objectives.
>
> 2. Scientific basis and documentation of input distributions and assumptions.
>
> 3. Model structure and computational mechanics.
>
> 4. Results, including, limitations, reasonableness, and clarity of documentation.

F.3.0   CHECKLIST FOR REVIEWERS

The exposure pathways and chemicals considered in a PRA should be clearly stated and related to the assessment endpoint. Often, the simplest way of doing this is to use the site conceptual model.

Table F-1 provides a list of major points that may be used to evaluate the quality of a probabilistic assessment. This is not an exhaustive list. The ultimate judgment of the acceptability of a PRA is the responsibility of the regional EPA personnel.

The issues that a reviewer should focus on may be different for each assessment. The workplan and the assessment should address each of the items on the checklist, but the workplan may include

additional items. The reviewer is responsible for ensuring that the workplan and the assessment are complete and of sufficient quality to help support a risk management decision under the National Contingency Plan (NCP).

The report should include a discussion of the results of assessment and how they relate to the point estimate of risk and hazard. A clear and concise description of what the results mean is an important part of each report.

F.4.0 INTERNAL AND EXTERNAL REVIEW

There are two levels of review that may be appropriate for a PRA. If an EPA reviewer feels the need for help with a review, other EPA personnel may be contacted formally or informally to provide additional review capabilities. The EPA personnel should also review the draft workplan for PRA to evaluate the appropriateness and consistency with Agency guidance. If EPA personnel are contacted early in the risk assessment process, the review can occur in a more productive and timely manner.

When the issues at a particular site are complex or contentious, EPA reviewers may also wish to obtain the services of outside experts for peer review (U.S. EPA, 2000). According to EPA's Peer-Review Policy Statement dated June 7, 1994 (U.S. EPA, 1994), "Major scientifically and technically based work products related to Agency decisions normally should be peer-reviewed." External peer review should be considered when allocating resources for a PRA. The EPA reviewers generally should select external peer reviewers who possess no bias or agenda regarding the process or methods of PRA.

Table F-1. Example of a Generic Checklist for Reviewers [2 pages]

Focal Point	✔	Evaluation Criterion
Objectives and Purpose		
Assessment Endpoints	✔	Are the human health and/or ecological assessment endpoints clearly stated and consistent with the workplan?
Benefits	✔	Are the rationales for, and benefits of, performing the PRA clearly stated and consistent with the workplan?
Site Conceptual Model	✔	Is there a description or graphic representation of the receptors and pathways considered in the assessment? Has the PRA addressed each of the pathways for completeness (e.g., sources, release mechanisms, transport media, route of entry, receptor)?
Separation of Variability and Uncertainty	✔	What is the modeling strategy for separating variability and uncertainty in the PRA? Is this strategy consistent with the assessment endpoints?
Model Structure and Computational Mechanics		
Flow Chart	✔	Is a diagram of the computational sequence provided so that the pathways of inputs and outputs and data capture can be understood and easily communicated?
1-D MCA / 2-D MCA	✔	Is a 1-D MCA or 2-D MCA being implemented in the PRA? What is represented by either or both dimensions?
Algorithms	✔	Are all algorithms used in the model documented in adequate detail to recreate the analysis?
Integration	✔	Are the algorithms used in numerical integration identified and documented?
Dimensional Analysis	✔	Has a unit analysis been conducted to ensure that all equations balance dimensionally?
Random Number Generation	✔	What random number generator is used in model computations? Is it robust enough? What reseeding approach is used to minimize repeated sequences?
Input Distributions and Assumptions		
Variability and Uncertainty	✔	Is there a clear distinction and segregation of distributions intended to represent variability from distributions intended to represent uncertainty?
Data sources	✔	Are the data or analysis sources used in developing or selecting the input distributions documented and appropriate for the site?
Distribution Forms	✔	Are the analyses used in selecting the form of the distribution adequately documented (i.e., understandable and repeatable by a third party?)
Distribution Parameters	✔	Are the analyses used to estimate the distribution parameters adequately documented?
Distribution Tails	✔	Do the estimation methods precisely depict the tails of the input distributions; how was this evaluated? Is there sufficient information to depict tails for empirical distributions? Are these estimated as exponential tails with bounding values?
Truncations	✔	Are any input distributions truncated? Do these truncations make sense? Should truncations be applied to any of the distributions?
Concentration Term	✔	Is the derivation of a point estimate or distribution for the concentration term adequately documented? Is sufficient information provided to enable the reviewer to recreate the concentration term?
Variable Correlations	✔	Have variable independence and correlations been addressed? Has the methodology for representing variable correlations in the model been documented and is it reasonable in terms of the variables, the site, and the statistical approach?

Focal Point	✔	Evaluation Criterion
Time Step	✔	Has the basis for the time step used in the model been documented? Is a single time step used, or do variables have different time steps? Are the time steps conceptually reasonable for the variables; for the site? Has the time step been evaluated in the sensitivity analysis?
Sensitivity Analysis	✔	Has a sensitivity analysis been conducted? Are the methods used in the analysis statistically valid? What did the analysis reveal about uncertainties in the assessment and the relative contributions of input variables to uncertainty?
Results of Modeling		
Completeness	✔	Are all the exposure routes identified in the site conceptual model and workplan addressed in the model results? Has the PRA fulfilled the objectives and satisfied the purpose stated in the workplan?
Point Estimate Calculation	✔	Has a point estimate calculation, using mean or median values of the input distributions, been performed? How do these results compare with the central tendencies calculated with the probabilistic model? How do the reasonable maximum exposure (RME) estimates compare? Have the similarities or differences between risk estimates from the point estimate and probabilistic approaches been adequately addressed?
Stability of Output Tails	✔	Has the stability of the high-end tail of the risk distribution been adequately evaluated? How stable are the estimated tails (in quantitative terms?) Is this level of stability adequate to support the risk management decisions that the model is intended to support?
Significant Figures	✔	Is the number of significant figures used in the output reasonable and consistent with model uncertainty?
Limitations	✔	Are the strengths and weaknesses of the PRA methodology and limitations of the results for decision making clearly presented?
Clarity	✔	Are the results and conclusions clearly presented and consistent with model output (e.g., central tendency exposure (CTE) and RME identified in the Executive Summary along with discussion of uncertainty)?
Graphics	✔	Are there graphics included that show both the risk distribution and PRA results (e.g., CTE and RME risk)?

REFERENCES FOR APPENDIX F

Burmaster, D.E. and P.D. Anderson. 1994. Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessment. *Risk Anal.* 14(4):477–481.

U.S. EPA. 1994. Memorandum from Deputy Administrator Carol Browner on *Peer Review and Peer Involvement at the U.S. Environmental Protection Agency*. June 7.

U.S. EPA. 1997. Memorandum from Deputy Administrator Fred Hansen on the *Use of Probabilistic Techniques (including Monte Carlo Analysis) in Risk Assessment*, and *Guiding Principles for Monte Carlo Analysis*. Office of Research and Development, Washington, DC. EPA/630/R–97/001. May 15.

U.S. EPA. 2000. *Peer Review Handbook: 2nd Edition.* Science Policy Council. Washington, DC. EPA/100/B-00/001. December.

APPENDIX G

FREQUENTLY ASKED QUESTIONS FOR PRA

INTRODUCTION

This section presents a few questions and answers relating to probabilistic risk assessment (PRA). The purpose of the frequently asked questions (FAQs) is to facilitate the understanding of PRA using a comparison with the traditional point estimate approach to risk assessment.

The FAQs presented here provide an overview of PRA with pointers to more detailed, and often more technical, discussions in other parts of the guidance.

(1) What is a risk assessment?

Risk assessment is a tool for organizing available information to make inferences about the potential human health or ecological effects associated with exposure to hazardous materials. The National Contingency Plan (NCP) addresses the use of a baseline risk assessment at Superfund sites to determine whether risks to human health and the environment are unacceptable. The NCP implements the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980.

Risk assessments traditionally provide single point descriptors of risk (e.g., a central tendency exposure (CTE) risk descriptor or a reasonable maximum exposure (RME) risk descriptor). As such, these types of risk assessments have been referred to as point estimate risk assessments.

In 1983, the National Research Council (NRC) described the following four steps for conducting human health risk assessments:

- *Hazard identification*: the determination of whether a particular chemical is or is not causally linked to a particular health effect.
- *Dose-response assessment*: the determination of the relation between the magnitude of exposure and the probability of occurrence of the health effects in question.
- *Exposure assessment*: the determination of the extent of human exposure before or after application of regulatory controls.
- *Risk characterization*: a description of the nature and often the magnitude of human risk, including attendant uncertainty (NRC, 1983).

Readers are referred to risk assessment guidance documents such as *Risk Assessment Guidance for Superfund* (RAGS): *Volume I. Human Health Evaluation Manual* (HHEM) *(Part A, Baseline Risk Assessment)*(U.S. EPA, 1989a), *Risk Assessment Guidance for Superfund: Volume II. Environmental Evaluation Manual* (U.S. EPA, 1989b), and *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments* (U.S. EPA, 1997a) for more information about point estimate risk assessment methods and policies.

(2) What is a probabilistic risk assessment (PRA)?

Superfund risk assessments have traditionally provided single point estimates of risk. More recently, PRAs have been developed. A PRA is a risk assessment that provides a probability distribution, rather than a point estimate, of risk. A probability distribution conveys both a range of values and a likelihood of occurrence of each value. This may allow a risk assessor to make statements about the likelihood that risks will exceed a level of concern. The probability distribution for risk often represents variability in risk estimates for a potentially exposed population. This variability may be due to variability in exposure and/or toxicity. PRA may also be used to quantify uncertainty in risk estimates. This can be useful because it allows a risk assessor to make statements about the level of confidence in the likelihood that risks will exceed a level of concern.

Probabilistic methods often use computer simulations to combine multiple probabilistic distributions in a risk equation. Monte Carlo analysis (MCA) is perhaps the most widely used probabilistic method in PRA (see Question #7).

(3) How does PRA compare with the point estimate approach?

A single point estimate of risk does not explicitly characterize associated variability or uncertainty. However, multiple point estimates of risk (e.g., CTE or RME) can begin to characterize variability in risk as they use different points on each input distribution for exposure). A PRA can characterize variability in risk by using the full distribution of variability in exposure parameters in the risk equations. Advanced PRA techniques can also quantitatively characterize uncertainty. In appropriate circumstances, results of a PRA can lead to more informed risk management decisions.

A PRA can be more resource intensive than a point estimate risk assessment. Some PRAs can require greater effort than point estimate approaches to define model inputs (i.e., select and fit probability distributions), as well as additional steps in the planning, review, and communication of the risk assessment assumptions and results (see Chapter 6 and Appendix F). A PRA does not necessarily require more data than a point estimate approach, although it does provide a framework for incorporating more of the available information into the risk assessment. When information on important exposure variables is lacking, results from a point estimate approach and a probabilistic approach will be equally uncertain.

If a decision is made to conduct a PRA, this does not replace a point estimate risk assessment. Results of point estimate approaches should still be presented along with results of probabilistic approaches in Tier 2 or Tier 3.

(4) Why should I consider using PRA?

PRA can have several advantages over the traditional point estimate approach to risk assessment. PRA can often provide a more complete characterization of risk; a quantitative description of the uncertainties in the risk estimates; more informative sensitivity analysis; the ability to make probabilistic statements about risk; the ability to know where specific risk levels are on the potential distribution of risk; an increased understanding of risks; and opportunities for improved communication and risk management decision making.

(5) When should I consider using PRA?

A PRA may be considered as early as the planning stages of a point estimate risk assessment or as late as after the completion of a point estimate risk assessment. Ideally, PRA should be considered as early as possible in the planning of risk assessment activities at a site so that sampling plans and data collection efforts may be appropriately directed. A PRA may be used when the risk management decision is not apparent and when the results of a PRA may inform the risk management decision. Often a risk management decision is not apparent when the site-specific risk estimate is close to the regulatory level of concern. The NCP discusses a generally acceptable range for cumulative excess cancer risk of 1E-06 to 1E-04 for protecting human health (U.S. EPA, 1990). Noncancer risks to human health and ecological health are generally characterized by a ratio of exposure to toxicity, called a Hazard Quotient (HQ) or Hazard Index (HI) for multiple contaminants. The point of departure for evaluating noncancer risks may vary from site to site, but a HQ of 1 may be a good starting point for risk management decisions.

PRA may also be considered when the results of the point estimate risk assessment suggest that risks are clearly above a risk level of concern, and a preliminary remediation goal (PRG) is needed. Because PRA and point estimate risk assessments use different techniques for quantifying variability and uncertainty, they may support different PRGs. If the results are dramatically different, further steps may be warranted to reevaluate the choices for input variables - both the point estimates, and the probability distributions and parameters (including truncation limits) for the 1-D MCA.

PRA will not be needed in many cases. Point risk estimates often produce results which are sufficient for making remedial decisions (e.g., sites are usually either heavily contaminated or only marginally contaminated). A tiered approach to risk assessment has been developed by Environmental Protection Agency (EPA) and is recommended for use in deciding when to move from point estimate risk assessments to PRAs of varying complexities. A workplan should be developed and submitted for review before beginning a PRA at any stage in the tiered process. As a general rule, if the potential value added by a PRA outweighs the additional resource required to conduct it, PRA may be warranted (see Chapter 2).

(6) How is the risk distribution from PRA used for decision making?

The EPA's *RAGS Volume I* (U.S. EPA, 1989a) and the NCP Preamble (U.S. EPA, 1990) state that the RME will generally be the principal basis for evaluating potential human health risks at Superfund sites. Ecological assessments also often consider an RME endpoint. The point estimate Superfund risk assessments use a combination of average and high-end input values to arrive at the RME. In PRA, risks are described by a probability distribution instead of a point estimate. To use a risk distribution for decision making, one needs to identify a percentile value that corresponds to the RME. *EPA's Guidelines for Exposure Assessment* (U.S. EPA, 1992a) states that, "the high-end risk means risks above the 90th percentile of the population distribution", and "the high-end estimator should not exceed the 99.9th percentile" due to uncertainty in specifying the upper tail of the input distributions in a Monte Carlo analysis. Similarly, the 90th to 99.9th percentiles of the risk distribution are recommended in this guidance as the RME range for decision making in PRA. Selection of a single point within the RME range generally requires consideration of the level of uncertainty in the risk distribution. The EPA recommends that the 95th percentile of the risk distribution be used as a starting point for risk management decisions in the absence of site-specific information.

(7) What is Monte Carlo Analysis (MCA)?

MCA is a numerical technique for PRA. MCA was developed in the 1940's during the beginnings of the nuclear power industry. MCA combines statistical analysis with modern computational techniques to calculate risk estimates, by randomly choosing different sets of input values each time. Each calculation is an iteration and a set of iterations is called a simulation. The output of a simulation used for risk assessment is a continuous probability distribution, which can be displayed in a graph in the form of either a probability density function (PDF) or corresponding cumulative distribution function (CDF). Both displays represent the same distribution, but are useful for conveying different information. For example, the PDF for risk is a good way for displaying relative probability using an area under the bell-shaped curve. The CDF for risk is generally S-shaped and can be especially informative for illustrating the percentile corresponding to a particular risk level of concern (e.g., 95th percentile=1E-06). Other uses of PDFs and CDFs are presented in Chapter 1, Exhibit 1-3. In 1997, EPA published a policy accepting the use of MCA to perform human health and ecological risk assessments (U.S. EPA, 1997a). This guidance focuses on MCA as a method of quantifying variability and uncertainty.

(8) What is the policy on using PRA to characterize variability or uncertainty in toxicity or dose response?

In human health risk assessments, probability distributions for risk should reflect variability or uncertainty in exposure. In ecological risk assessments, risk distributions may reflect variability or uncertainty in exposure and/or toxicity (see Chapter 1, Sections 1.4 and 1.4.1, Item 3).

Approaches to characterizing variability and uncertainty in toxicological information should reflect both the latest developments in the science of hazard and dose-response evaluation and consistent application of EPA science policy. This statement is consistent with the *1997 EPA Policy Statement* presented in Section 1.4 above (U.S. EPA, 1997g). Probabilistic approaches to ecological dose-response assessment may be explored, as discussed and demonstrated in Chapter 4. This guidance does not develop or evaluate probabilistic approaches for dose-response in human health assessment and, further, *discourages undertaking such activities on a site-by-site basis*. Such activities require contaminant-specific national consensus development and national policy development. Parties wishing to undertake such activities should contact the OERR to explore ways in which they might contribute to a national process for the contaminant of interest to them.

(9) What is the policy on using PRA at EPA and in Superfund?

In the spring of 1997, EPA released the memorandum, *Use of Probabilistic Techniques (including Monte Carlo Analysis) in Risk Assessment* (U.S. EPA, 1997b). The policy states that probabilistic analysis techniques, "given adequate supporting data and credible assumptions, can be viable statistical tools for analyzing variability and uncertainty in risk assessments." As such, a PRA, "will be evaluated and utilized in a manner that is consistent with other risk assessments submitted to the Agency." Together with this Policy Statement, the Agency released a set of guiding principles for use and review of probabilistic analyses. Hence, both RAGS and Agency-wide guidance emphasize the importance of review of the scientific and technical merit of a probabilistic analysis to determine whether or not the assessment is of sufficient quality to support a remedial decision. This guidance, *RAGS Volume 3: Part A*, provides risk assessors with comprehensive guidance on when and how to conduct PRAs using MCA within the Superfund program (see Preface and Chapter 1).

(10) What are the challenges of using PRA?

Although PRA may have several advantages over the traditional point estimate approach to risk assessment, the use of PRA tends to be more resource intensive and may introduce some additional challenges to risk communication efforts. Risk communication helps build trust with the stakeholders and disseminate the risk information. In general, EPA staff and stakeholders are accustomed to a point estimate of risk and are unfamiliar with PRA and the quantitative estimates of uncertainty that PRA can support. Although, quantitative risk estimates may be more informative, they also may be more difficult to communicate and may not be well received due to stakeholder desires for certainty (Slovic, et al. 1979). Early and frequent communication with stakeholders is key in implementing PRA successfully. Often PRA requires additional data collection efforts as well as more time and resources to select and fit probability distributions.

REFERENCES FOR APPENDIX G

National Research Council (NRC). 1983. *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press. Washington, DC.

Slovic, P., B. Fischoff, and S. Lichtenstein. 1979. Rating the Risks. *Environ.* 21(3):14–20 and 36–39.

U.S. EPA. 1989a. *Risk Assessment Guidance for Superfund* (RAGS): *Volume I. Human Health Evaluation Manual* (HHEM) *(Part A, Baseline Risk Assessment).* Interim Final. Office of Emergency and Remedial Response, Washington, DC. EPA/540/1-89/002. NTIS PB90-155581.

U.S. EPA. 1989b. *Risk Assessment Guidance for Superfund.* (RAGS)*: Volume II. Environmental Evaluation Manual*. Interim Final. Office of Emergency and Remedial Response, Washington, DC. EPA/540/1-89/001.

U.S. EPA. 1990. National Oil and Hazardous Substances Pollution Contingency Plan. Final Rule. 40 CFR 300: *55 Federal Register*, 8666-8865, March 8.

U.S. EPA. 1992a. *Final Guidelines for Exposure Assessment*. EPA/600/Z-92/001. *57 Federal Register*, 22888-22938, May 29.

U.S. EPA. 1997a. *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments*. Interim Final. Environmental Response Team, Edison, NJ. EPA/540/R-97/006, OSWER Directive No. 9285.7-25, June.

U.S. EPA. 1997b. *Memorandum from Deputy Administrator Fred Hansen on the Use of ProbabilisticTechniques (including Monte Carlo Analysis) in Risk Assessment, and Guiding Principles for Monte Carlo Analysis*. Office of Research and Development, Washington, DC. EPA/630/R-97/001. May.

APPENDIX H

INDEX