

Chapter 31 Probabilistic Risk Assessment

Table of Contents

31.1	Introduction	1
31.2	Tiered Approach for Risk Assessment	2
31.3	Methods for Probabilistic Risk Assessment	4
31.4	Presenting Results for Probabilistic Risk Assessment	6
	References	10

31.1 Introduction

Probabilistic risk assessment (PRA) uses probability distributions to characterize variability or uncertainty in risk estimates. In a PRA, one or more variables in the risk equation is defined as a probability distribution rather than a single number. Similarly, the output of a PRA is a range or probability distribution of risks experienced by the receptors. Note that the ability to perform a PRA often is limited by the availability of distributional data that adequately describe one or more of the input parameters. For example, data often are insufficient to assess toxicity in a probabilistic manner (and therefore, dose-response values such as inhalation unit risks (IURs) and reference concentrations (RfCs) are included in a PRA analysis as point values). This general lack of data impacts both human health and ecological receptors.

The primary advantage of PRA is that it can provide a quantitative description of the degree of variability or uncertainty (or both) in risk estimates for both cancer and noncancer health effects and ecological hazards. The quantitative analysis of uncertainty and variability can provide a more comprehensive characterization of risk than is possible in the point estimate approach.

Another significant advantage of PRA is the additional information and potential flexibility it affords the risk manager. Risk management decisions are often based on an evaluation of high-end risk to an individual – for deterministic analyses, this is generally developed by the combination of a mix of central tendency and high-end point values for various exposure parameters (see Part II, Chapters 9 and 13). When using PRA, the risk manager can select a specific upper-bound level from the high-end range of percentiles of risk, generally between the 90th and 99.9th percentiles.

PRA may not be appropriate for every analysis. The primary disadvantages of PRA are that it generally requires more time, resources, and expertise on the part of the assessor, reviewer, and risk manager than a point estimate approach. The chief obstacle to using PRA in air toxics risk assessments is usually the lack of well-documented frequency distributions for many input variables.

A detailed discussion of PRA is beyond the scope of this document. Two documents provide more detailed introductory information and guidance and should be reviewed if a PRA is contemplated:

U.S. EPA. 2001. *Risk Assessment Guidance for Superfund (RAGS), Volume III - Part A, Process for Conducting Probabilistic Risk Assessment*. Office of Solid Waste and Emergency Response. December. EPA 540-R-02-002, OSWER 9285.7-45, PB2002 963302, available at: <http://www.epa.gov/superfund/programs/risk/rags3a/index.htm>.

National Council on Radiation Protection and Measurements (NCRP). 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*. NCRP Commentary No. 14, May 1996.

This chapter provides a general overview of PRA as it applies to air toxics risk assessment. It revisits the tiered approach to risk assessment, introduces calculation algorithms, and identifies advanced statistical methods currently available to support risk policy decisions.

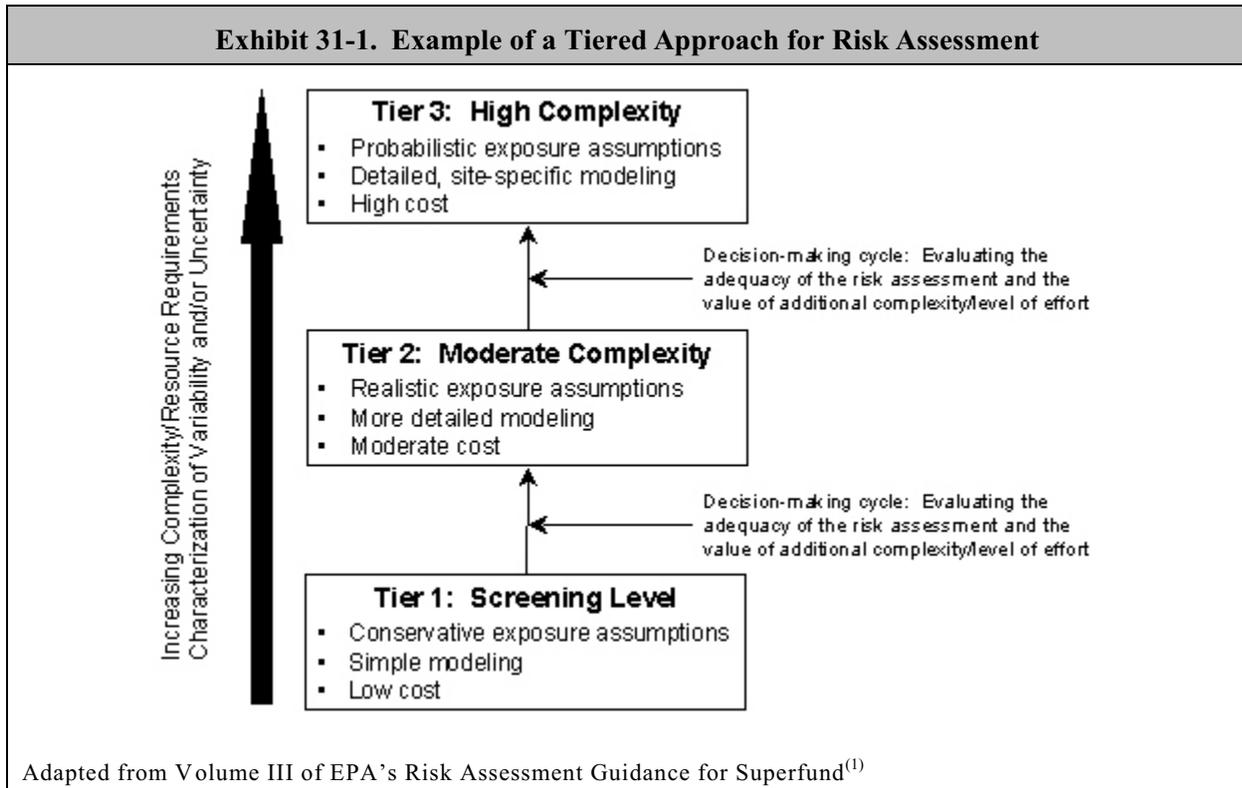
31.2 Tiered Approach for Risk Assessment

The tiered approach is a process for a systematic, informed progression to increasingly more complex risk assessment methods including PRA. Exhibit 31-1 presents a schematic representation of the tiered approach. Higher tiers reflect increasing complexity and, in many cases, will require more time and resources. Higher tiers also reflect increasing characterization of variability and/or uncertainty in the risk estimate, which may be important for making risk management decisions. Central to the concept of a systematic, informed progression is an iterative process of evaluation, deliberation, data collection, work planning, and communication. All of these steps should focus on deciding: (1) whether or not the risk assessment, in its current state, is sufficient to support risk management decisions (a clear path to exiting the tiered process is available at each tier), and (2) if the assessment is determined to be insufficient, whether or not progression to a higher tier of complexity (or refinement of the current tier) would provide a sufficient benefit to warrant the additional effort.

- The problem formulation step precedes Tier 1 and includes scoping and refinement of the conceptual site model, including exposure pathways/routes, and identifying chemicals of potential concern (COPCs).
- In Tier 1, deterministic (point estimate) risk assessment is then performed using the basic methodology described in Part II (inhalation) and/or Part III (multipathway) of this Reference Manual. In deciding whether the results of a deterministic risk assessment are sufficient for decision-making or whether more refined analyses should be implemented, two factors generally are considered: (1) the magnitude of the estimates of risk (i.e., the value of hazard indices [HIs] or cancer risks for COPCs), and (2) the level of confidence in these estimates. In a Tier I deterministic risk assessment, quantitative risk estimates can be easily calculated, but the level of confidence associated with these calculations can be difficult to assess. For example, variability in exposure levels among individual members of the population can generally only be assessed semi-quantitatively by considering central tendency and high-end exposure estimates. Uncertainty can often be evaluated only as confidence limits on certain point estimates (e.g., the concentration term).

In some cases, the results of a Tier 1 risk analysis may be sufficient for decision-making. For example, a deterministic analysis may indicate very low levels of risk for some air toxics. If the assessment is considered to be overly conservative (even in light of uncertainties), this may be sufficient for a “no action” decision for those chemicals. The same analysis may indicate a very high potential for risk for other air toxics. EPA generally recommends that the risk manager proceed to higher tiers only when site decision-making would benefit from additional analysis beyond the point-estimate risk assessment (i.e., when the risk manager needs more complete or certain information to complete the risk management process).

Thus, only the combinations of COPC-exposure pathway-receptors of highest potential concern are generally analyzed using higher level techniques such as PRA.



- Tier 2 is represented as an intermediate-level analysis using more realistic exposure assumptions (e.g., use of actual receptor locations) and more detailed modeling (e.g., a model that requires additional site-specific inputs). Although not depicted, Tier 2 could incorporate a sensitivity analysis to identify the most important parameters that are driving the risk estimate for specific receptors or population groups. Tier 2 also could incorporate limited (one-dimensional) Monte Carlo techniques.
- Tier 3 is represented as an advanced analysis using probabilistic techniques such as two-dimensional Monte Carlo analysis. Results of sensitivity analyses (Tier 2 or Tier 3) could be used to assess risk distributions for the high-end individuals within the population. The one-dimensional Monte-Carlo simulation does not separate variability and uncertainty associated with the risk estimates. If necessary, separate analyses of uncertainty and variability can be performed in Tier 3. Techniques such as two-dimensional Monte Carlo simulation can be used to estimate the relative impact of natural variability and lack of data on the overall uncertainty in the risk estimate, and can be used to direct additional data gathering or to support mitigation decisions.

The deliberation cycle provides an opportunity to evaluate the direction and goals of the assessment as new information becomes available. It may include evaluations of both scientific and policy information. (Also note that, while a three-tiered approach was provided in Exhibit

31-1, the tiered approach is really more of a continuum from a point where the analysis is done with little data and conservative assumptions to a point where there is an extensive data set and fewer assumptions. In between, there can be a wide variety of tiers of increasing complexity, or, as discussed in Chapter 3, there may only be a few reasonable choices between screening methods and highly refined analyses. The three tiered approach is only provided here as an illustration of the concept, not a prescriptive, fixed methodology.)

31.3 Methods for Probabilistic Risk Assessment

As discussed in previous chapters, there are a number of approaches available for analyzing uncertainty in risk assessments. For simple screening level analyses, or analyses where there are only a few major sources of uncertainty, sensitivity analyses may be used to estimate the impacts of likely variations in the key parameter values. Where scenario uncertainty is important (that is, there are multiple sequences of events that could contribute to risk), decision tree or Bayesian statistical analysis are commonly used. The most common numerical technique for PRA (analyses in which a large number of variables need to be evaluated simultaneously) in large-scale air risk assessments is Monte Carlo simulation. Monte Carlo simulation integrates varying assumptions, usually about exposure, to come up with possible distributions (or ranges) of risk instead of point estimates. A continuous probability distribution can be displayed in a graph in the form of either **probability density functions** (PDFs) or corresponding **cumulative distribution functions** (CDFs); however, for clarity, it is recommended that both representations be presented in adjacent (rather than overlaid) plots.

Exhibit 31-2 illustrates a PDF and CDF for a normal probability distribution for adult body weight. Both displays represent the same distribution, but are useful for conveying different information. PDFs are most useful for displaying (1) the relative probability of values; (2) the most likely values (e.g., modes); and (3) the shape of the distribution (e.g., skewness, kurtosis, multimodality). CDFs can be used to display (1) percentiles, including the median; (2) high-end risk range (e.g., 90th to 99th percentiles); (3) confidence intervals for selected percentiles; and (4) stochastic dominance (i.e., for any percentile, the value for one variable exceeds that of any other variable). Note that it is helpful to include a text box with summary statistics relevant to the distribution (e.g., mean, standard deviation).

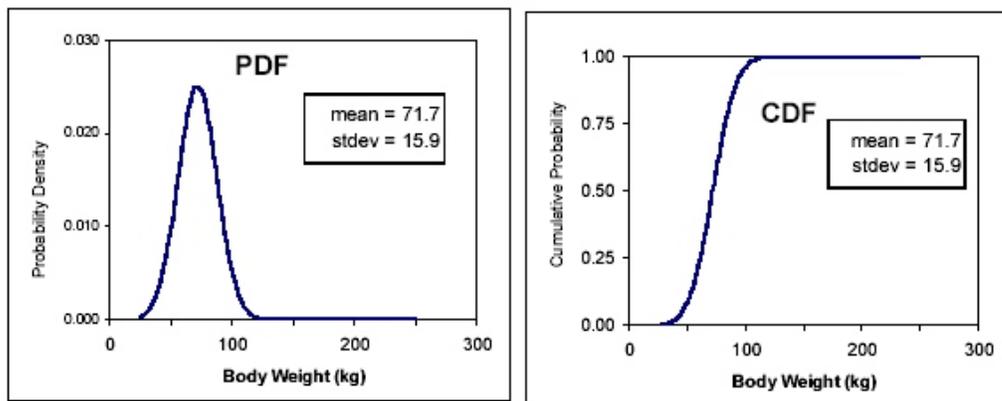
These results expressed as probability distributions help risk managers decide whether and what actions are necessary to reduce risk. Monte Carlo simulation has been widely used to explore problems in many disciplines of science as well as engineering, finance, and insurance.⁽¹⁾ The process for a Monte Carlo simulation is illustrated in Exhibit 31-3. In its general form, the risk equation can be expressed as a function of a toxicity term (as a point value) and multiple exposure variables (V_n) represented as distributions (not point values):

$$\text{Risk} = f(V_1, V_2, V_3, \dots V_n) \times \text{Toxicity} \quad \text{Equation 31-4}$$

The first decision(s) the risk assessor has to make is which of the “Vs” are going to be evaluated probabilistically. Ideally, every model input that is variable or uncertain should be evaluated to provide a comprehensive characterization of uncertainty in exposure estimates. In practice, the

number of variables that can be addressed systematically is severely limited by lack of data related to variability, uncertainty, or both. Sensitivity analyses can often be used to focus the analysis on the variables that contribute most to the overall uncertainty in risks.

Exhibit 31-2. Examples of Probability Density and Cumulative Distribution Functions



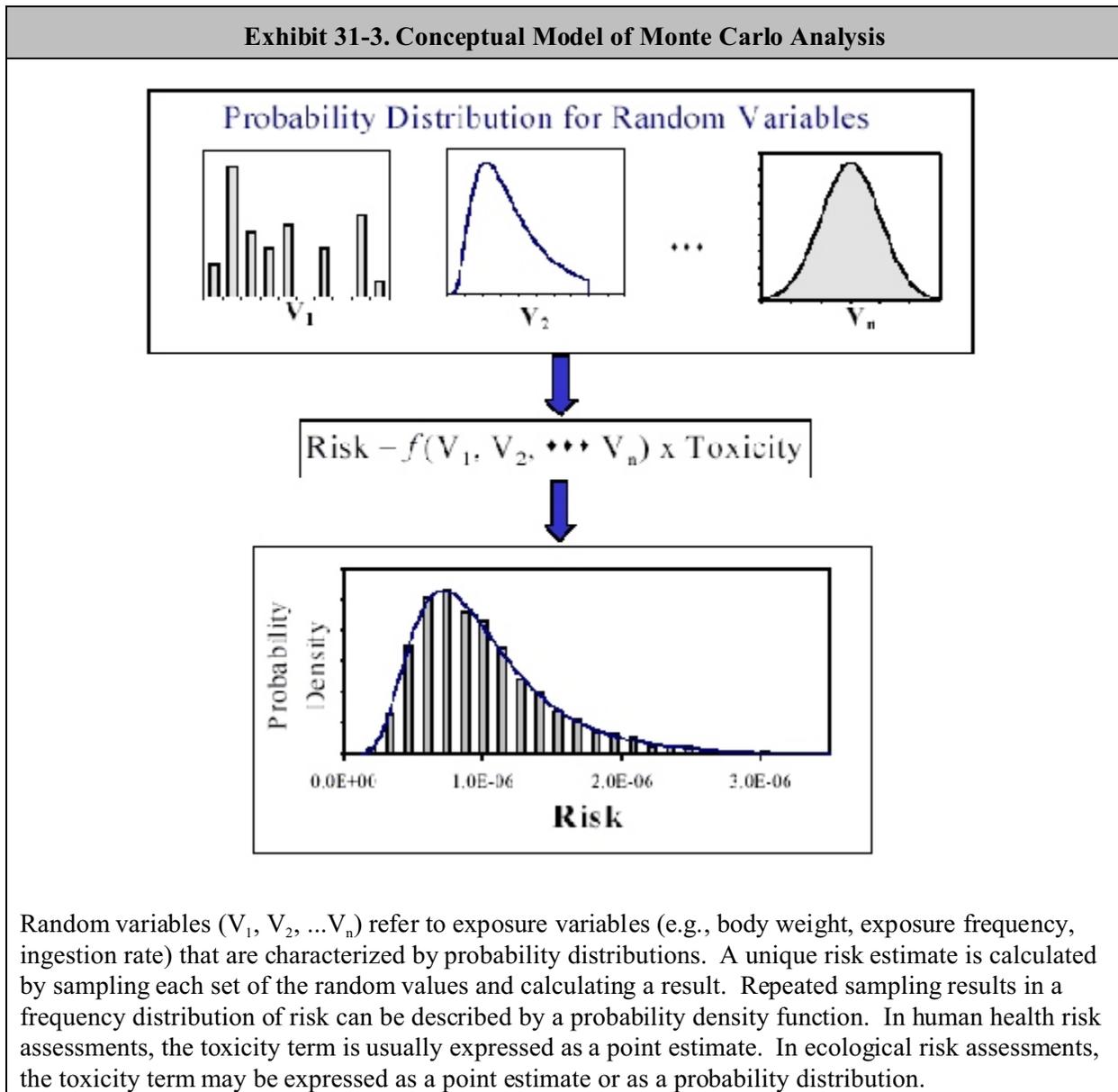
Example of a normal distribution that characterizes variability in adult body weight (males and females combined). The arithmetic mean = 71.7 kg, and standard deviation = 15.9 kg. Body weight may be considered a continuous random variable. The left panel shows a bell-shaped curve and represents the PDF, while the right panel shows an S-shaped curve and represents the CDF. Both displays represent the same distribution (including summary statistics), but are useful for conveying different information.

Source: Finley and Paustenbach⁽²⁾

Solutions for equations with PDFs are typically too complex for even an expert mathematician to calculate the risk distribution analytically. However, numerical techniques applied with the aid of computers can provide very close approximations of the solution. This is illustrated here for the simplified case in which the assessment variables are statistically independent, that is, the value of one variable has no relationship to the value of any other variable. In this case, the computer selects a value for each variable (V_n) at random from a specified PDF and calculates the corresponding risk. This process is repeated many times (e.g., 10,000), each time saving the set of input values and corresponding estimate of risk. For example, the first risk estimate might represent a hypothetical individual who drinks 2 L/day of water and weighs 65 kg, the second estimate might represent someone who drinks 1 L/day and weighs 72 kg, and so forth. Each calculation is referred to as an **iteration**, and a set of iterations is called a **simulation**.

Each iteration of a Monte Carlo simulation should represent a plausible combination of input values (i.e., exposure or ecotoxicity variables), which may require using bounded or truncated probability distributions. However, risk estimates are not intended to correspond to any one person. The “individuals” represented by Monte Carlo iterations are “virtual,” and the risk distributions derived from a PRA allow for inferences to be made about the likelihood or probability of risks occurring within a specified range for an exposed human or ecological

population. A simulation yields a set of risk estimates that can be summarized with selected statistics (e.g., arithmetic mean, percentiles) and displayed graphically using PDF and CDF for the estimated risk distribution.



31.4 Presenting Results for Probabilistic Risk Assessment

The complexity of risk evaluation, and particularly of probabilistic methods, may pose a significant barrier to understanding among the affected and interested parties (and thus to the utility of the analysis). In the past, regulatory decisions have been evaluated primarily in terms of point estimates of risk and simple dichotomous decision rules (e.g., "If the point estimate of risk

is above a certain level, take a certain action. If not, take another action.”). In contrast, it may not be intuitively obvious, even to relatively sophisticated audiences, how to relate the outputs of quantitative uncertainty evaluation to a particular decision. For example, important aspects of a regulatory decision may rest on relatively subtle statistical distinctions (e.g., the difference between a 95th percentile risk estimate and a 95th percent upper confidence limit on a risk estimate), and the challenges in presenting such information can be formidable. In its recent guidance, EPA has begun to define concrete approaches to presenting risks and uncertainty information to decision-makers and stakeholders.⁽⁵⁾

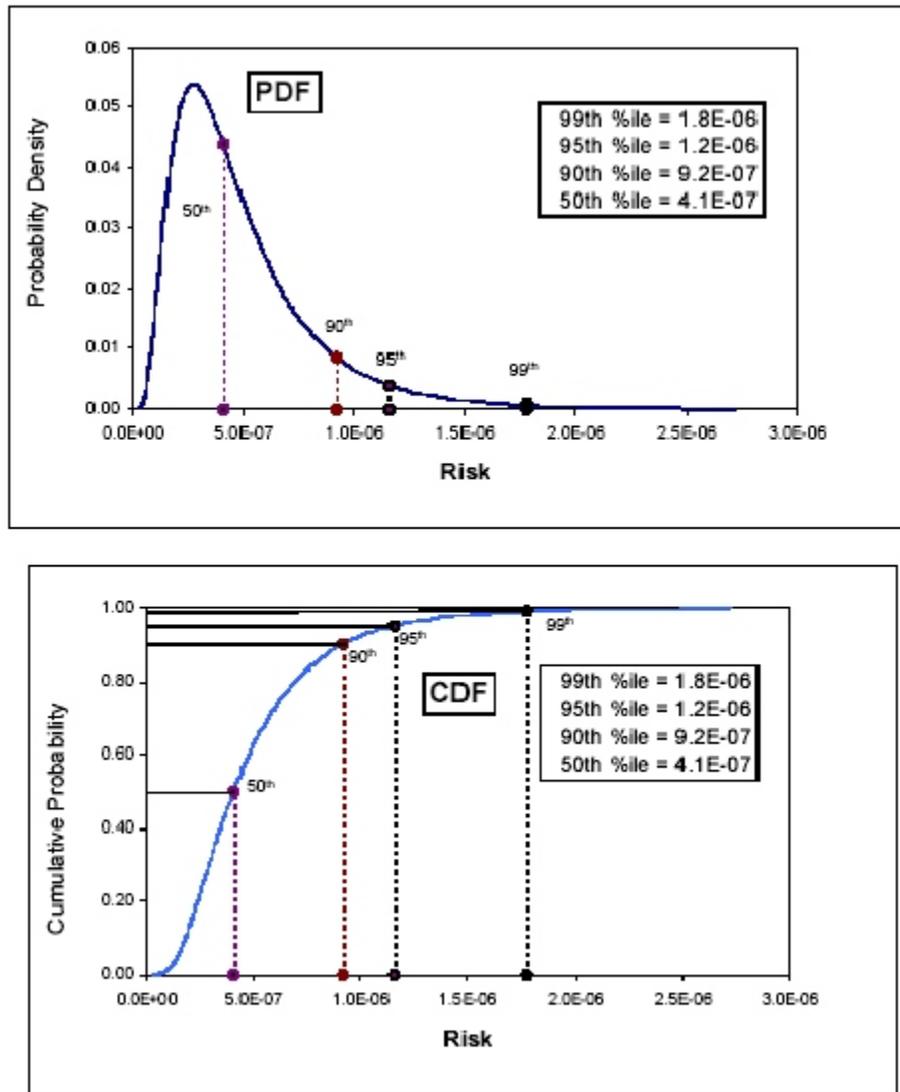
The key factors for successful communication of PRA include early and continuous involvement of affected and interested parties, a well-developed communication plan, good graphics, a working knowledge of the factors that may influence perceptions of risk and uncertainty, and a foundation of trust and credibility. A certain amount of training for interested stakeholders will likely be necessary to help them understand the complexities of not only risk assessment in general, but the intricacies of higher levels of analysis. Part III of this Reference Manual provides guidance on community involvement and risk communication.

When summarizing results of PRA, graphs and tables should generally also include the results of the point estimates of risk (e.g., central tendency and high-end).

Consistent with EPA’s guidance on risk characterization,⁽³⁾ the central tendency and high-end cancer risks and noncancer hazards, along with decision points, should be highlighted on graphics. The discussions accompanying the graph should emphasize that these values represent risks to the average and high-end individuals, respectively, and serve as a point of reference to EPA’s decision point. The distribution of risks should be characterized as representing variability among the population based on differences in exposure. Similarly, graphics that show uncertainty in risk estimates can be described using terms such as “confidence interval,” “credible interval,” or “plausible range,” as appropriate. The graphics need not highlight all percentiles. Instead, selected percentiles that may inform risk management decisions (such as the 5th, 50th, 90th, 95th, and 99th percentiles) should be the focus. Exhibit 31-4 presents an example of a PDF for variability in risk with an associated text box for identifying key risk descriptors.

By understanding the assumptions regarding the inputs and modeling approaches used to derive point estimates and probabilistic estimates of risk, a risk communicator will be better prepared to explain the significant differences in risk estimates that have been developed. Special emphasis should be given to the model and parameter assumptions that have the most influence on the risk estimates, as determined from the sensitivity analysis.

Exhibit 31-4. Example of Presenting the Results of a Probabilistic Risk Assessment



Hypothetical PRA results showing a PDF (top panel) for cancer risk with selected summary statistics for central tendency and high-end percentiles. This view of a distribution is useful for illustrating the shape of the distribution (e.g., slightly right-skewed) and explaining the concept of probability as the area under a curve (e.g., most of the area is below 1×10^{-6} , but there is a small chance of 2×10^{-6}). Although percentiles can also be overlaid on this graphic, a CDF (bottom panel) may be preferable for explaining the concept of a percentile.

Additional References on Uncertainty Analysis

- Burmaster, D.E. and Anderson, P.D. 1994. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Analysis* 14: 477-481.
- Cullen, A.C. and Frey, H.C. 1999. *Probabilistic Techniques in Exposure Assessment*. New York: Plenum Press.
- Fayerweather, W.E., Collins, J.J., Schnatter, A.R., Hearne, F.T., Menning, R.A., and Reyner, D.P. 1999. Quantifying uncertainty in a risk assessment using human data. *Risk Analysis* 19: 1077-1090
- Finkel, A.M. and Evans, J.S. 1987. Evaluating the benefits of uncertainty reduction in environmental health risk management. *Journal of the Air Pollution Control Association*. 37: 1164-1171.
- Frey, H.C. 1992. Quantitative analysis of uncertainty and variability in environmental policy making. Pittsburgh: Carnegie Mellon University.
- Hattis, D. and Burmaster, D.E. 1994. Assessment of variability and uncertainty distributions for practical risk assessments. *Risk Analysis* 14: 713-730.
- Hope, B. K. 1999. Assessment of risk to terrestrial receptors using uncertainty analysis - A case study. *Human and Ecological Risk Assessment* 5(1): 145-170.
- Moore, D.R.J., Sample, B.E., Suter, G.W., Parkhurst, B.R., and Teed, R.S. 1999. A probabilistic risk assessment of the effects of methylmercury and PCBs on mink and kingfishers along East Fork Poplar Creek, Oak Ridge, Tennessee, USA. *Environmental Toxicology and Chemistry* 18: 2941-2953.
- National Research Council (NRC). 1991. *Human Exposure Assessment for Airborne Pollutants*. Washington DC: National Academy Press.
- Roberts, S.M. 1999. Practical issues in the use of probabilistic risk assessment and its applications to hazardous waste sites. *Human and Ecological Risk Assessment*. 5(4): 729-868. Special Issue.
- Smith, R.L.. 1994. Use of Monte Carlo simulation for human exposure assessment at a Superfund site. *Risk Analysis* 14(4): 433-439.
- U.S. Environmental Protection Agency. 1985. *Methodology for Characterization of Uncertainty in Exposure Assessments*. Washington DC, EPA-600/8-85-009).

References

1. Rugen, P. and B. Callahan. 1996. An overview of Monte Carlo: A fifty year perspective. *Human Health and Ecological Risk Assessment* 2(4): 671-680.
2. Finley, B. and D. Paustenbach. 1994. The benefits of probabilistic exposure assessment: Three case studies involving contaminated air, water, and soil. *Risk Analysis* 14(1): 53-73.
3. U.S. Environmental Protection Agency. 1992. *Guidance on Risk Characterization for Risk Managers and Risk Assessors*. Risk Assessment Council, Washington, DC, February 26, 1992.