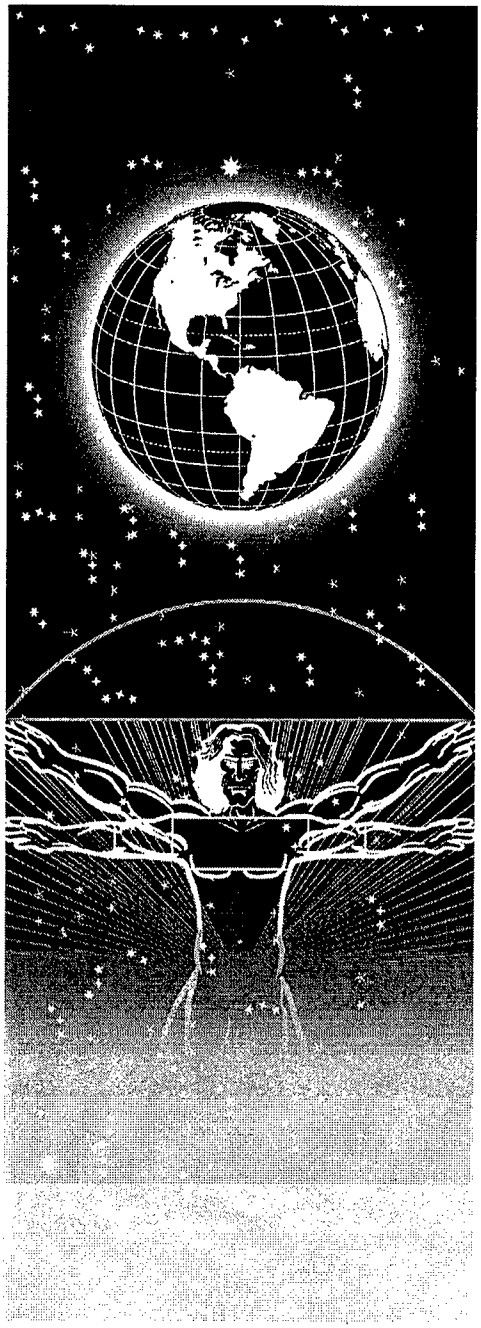
AL-OE-BR-TR-1998-0003



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

Methods to Quantify Uncertainty in Human Health Risk Assessment

Lea Aurelius

Parsons Engineering Science, Inc.
8000 Centre Park Drive, Suite 200
Austin, TX 78754

Brian L. Sassaman, Captain, USAF, BSC

February 1998

Approved for public release;

distribution is unlimited.

Occupational and Environmental Health Directorate
Occupational Medicine Division
2402 E Drive
Brooks Air Force Base TX 78235-5114

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The mention of trade names or commercial products in this publication is for illustration purposes and does not constitute endorsement or recommendation for use by the United State Air Force.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Government agencies and their contractors registered with Defense Technical Information Center (DTIC) should direct requests for copies to: Defense Technical Information Center, 8725 John J. Kingman Rd., STE 0944, Ft. Belvoir, VA 22060-6218.

Non-Government agencies may purchase copies of this report from: National Technical Information Services (NTIS), 5285 Port Royal Road, Springfield, VA 22161-2103.

ANDREW T. MACCABE, Maj, USAF, BSC
Chief, Health Risk Assessment Branch

KENT R. STRINGHAM, Lt Col, USAF, BSC
Director, Occupational Medicine Division

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1998	Final, 1996-1997

4. TITLE AND SUBTITLE
Methods to Quantify Uncertainty in Human Health Risk Assessment

5. FUNDING NUMBERS
C: F41624-95-D-9018

6. AUTHOR(S)
Lea Aurelius (Parsons Engineering Sciences, Inc.)
Brian L. Sassaman, Captain, USAF, BSC (Det 1, HSC/OEMH)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Parsons Engineering Science, Inc.
8000 Centre Park Drive, Suite 200
Austin, TX 78754

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Detachment 1, Human Systems Center
Occupational and Environmental Health Directorate
Occupational Medicine Division
2402 E Drive
Brooks Air Force Base TX 78235-5114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL-OE-BR-TR-1998-0003

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this document is to provide an introductory handbook for the Air Force remedial project manager (RPM) and other health professionals, such as the Bioenvironmental Engineer, to identify the appropriate use of probabilistic techniques for a site, and the methods by which probabilistic risk assessment can be used to quantify uncertainty. This document assumes that the RPM or other health professional is somewhat familiar with the basics of the risk assessment and risk management decision making process as implemented in hazardous waste site remediations. This document emphasizes the Monte Carlo probabilistic method and the exposure assessment step of the human health risk assessment process. This includes the techniques and methodology as provided in the United States Environmental Protection Agency's Risk Assessment Guidance for Superfund (RAGS). Probabilistic risk assessment should be viewed as one of a set of appropriate tools in a tiered approach to performing enhanced site specific risk assessment. Example calculations showing results of both deterministic and probabilistic risk assessments are provided for illustration as well as several appendices that contain supporting information.

14. SUBJECT TERMS
Uncertainty
Probabilistic Risk Assessment
Monte Carlo Simulation

15. NUMBER OF PAGES
220

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

PREFACE

Over the past two decades, scientists and public health policy professionals have developed a process to assess the extent of risk to human health and the environment at thousands of government and privately-owned contaminated sites throughout the country. Of particular interest, the U.S. Environmental Protection Agency's (USEPA's) Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA, or "Superfund") program requires an assessment of risk as one component in the remedial decision-making process.

As the knowledge base for human health risk assessment increases, the USEPA is placing continued effort on performing Agency risk assessments using the best science possible. USEPA is currently in the process of developing guidance for preparing and reviewing risk assessments based on the use of probabilistic techniques. A probabilistic risk assessment would be selected for a site as part of a tiered approach that progresses from a simpler (e.g., a screening level risk analysis), to a more quantitative (e.g., deterministic), and finally a more complex (e.g., probabilistic) risk assessment as the risk management situation requires.

As the risk assessor progresses through the risk assessment tiers, uncertainty associated with the risk assessment would be reduced, but the same level of health protection maintained. The deterministic risk assessment expresses human health risks as single numerical values, or "single-point" estimates of risk, and provides little information about the level of uncertainty and variability surrounding the risk estimate. The deterministic estimates also rely on reasonable maximum exposure (RME) assumptions to estimate a high-end risk descriptor. A high-end risk descriptor is defined as one which characterizes risk to an individual at the upper end of the risk distribution (i.e., the RME is the highest exposure that is reasonably expected to occur at a site).

Advancing to the probabilistic technique will allow for a quantitative analysis of uncertainty and variability and present the risk manager with ranges of risk instead of the high-end, single-point risk estimate. By showing the distribution of health risk, a more realistic picture of the actual risk posed to potential receptors will be provided. The key benefits of probabilistic risk assessments are that they are more informative and provide more relevant information upon which the risk manager can base decisions and identify more cost-effective solutions.

The purpose of this document is to provide an introductory handbook for the Air Force remedial project manager (RPM) for identifying the appropriate use of probabilistic techniques for a site, and the methods by which probabilistic risk assessments can be performed. Example calculations showing results of both deterministic and probabilistic risk

assessments are provided for illustration. This handbook emphasizes the Monte Carlo probabilistic method and the exposure assessment step of the risk assessment process.

This introductory handbook assumes that the Air Force RPM is somewhat familiar with the basics of the risk assessment and risk management decision processes as implemented in hazardous site remediations. This includes the techniques and methodology as provided in the USEPA's "Risk Assessment Guidance for Superfund" (RAGS).

Numerous background documents and technical publications, including supplemental guidance to RAGS, are available from the USEPA and other state or research organizations. In addition to references cited in the text of this handbook, references for further reading are listed in Appendix A. For questions or further information on the use of probabilistic risk assessments, the RPM can also contact the following Air Force resources:

AF Surgeon General: Health Risk Assessment Branch
Det 1, HSC/OEMH
2402 E Drive
Brooks AFB, Texas 78235-5114
DSN 240-2063; (210)536-2063
FAX (210)536-1130/2315

AF Civil Engineering: Risk Assessment Consultants
HQ AFCEE/ERC
Building 532
3207 North Road
Brooks AFB, Texas 78235
DSN 240-5244; (210)536-5244
FAX (210)536-5989

ACRONYMS AND ABBREVIATIONS

ABP	4-Aminobiphenyl
ARAR	Applicable or relevant and appropriate requirement
ASTM	American Society for Testing and Materials
C-term	Concentration term
CDF	Cumulative distribution function
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
COPC	Chemical of potential concern
CSM	Conceptual site model
CSF	Cancer slope factor
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
ECDF	Empirical cumulative distribution function
EDA	Exploratory data analysis
HEAST	Health Effects Assessment Summary Tables
HI	Hazard index
HQ	Hazard quotient
IQR	Interquartile range
IRIS	Integrated Risk Information System
IRP	Installation Restoration Program
L/day	liters per day
LHS	Latin hypercube sampling
mg/kg	milligrams per kilogram
mg/L	milligrams per liter
MAD	Median absolute deviation
MCL	Maximum contaminant level
NAS	National Academy of Sciences
NCP	National Oil and Hazardous Substances Pollution Contingency Plan
NCRP	National Council on Radiation Protection and Measures
OSWER	Office of Solid Waste and Emergency Response
PDF	Probability density function
PID	Photoionization detector

Q-Q plot	Quantile-quantile (normal probability) plot
RAGS	Risk Assessment Guidance for Superfund
RBCA	Risk-based corrective action
RCRA	Resource Conservation and Recovery Act
RfD	Reference dose
RFP	Request for proposal
RME	Reasonable maximum exposure
RMSE	Root mean square error
ROD	Record of Decision
RPM	Remedial project manager
SOW	Scope of work
SQL	Sample quantitation limit
SRS	Simple random sampling
UCL	Upper confidence limit
USEPA	U.S. Environmental Protection Agency

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 1
INTRODUCTION

1.1 BACKGROUND

Most environmental remedial programs, including U.S. Environmental Protection Agency's (USEPA's) Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA, or "Superfund") and Resource Conservation and Recovery Act (RCRA) programs, have a statutory mandate to protect human health and the environment. Initial considerations for determining the potential for adverse effects may include comparisons of site-detected concentrations of contaminants in different environmental media (e.g., groundwater, surface water, soil) to applicable or relevant and appropriate requirements (ARARs), such as maximum contaminant levels (MCLs) for chemicals detected in groundwater. Comparisons to background levels, represented by samples taken from media unaffected by waste management or industrial activities, may also be initially evaluated to determine possible environmental impact from site-related contaminants.

In recent years, the development and use of refined risk-based remedial action levels has gained substantial support among environmental regulators, responsible parties, and the public. These refined remediation levels are based on an assessment of risk to human and ecological receptors where, if it is determined that an unacceptable risk is present for a particular receptor, remedial cleanup levels or other remedial options (e.g., institutional controls such as capping or fencing) are developed to mitigate the risk. These levels may consider criteria and factors relating to the potential for exposure, uncertainty in the risk assessment, as well as technical feasibility of cleanup. At many sites, a combination of strategies (e.g., background levels, ARARs, risk-based levels, and institutional controls) may be used to develop an acceptable remedial strategy.

The use of risk assessment as a tool in remedial decision-making has been a key component of the Air Force's environmental Installation Restoration Program (IRP) since the program's inception in 1984. Air Force risk assessments have generally followed the lead of the Federal Superfund Program and have relied on that Program's guidance, protocols, and policies.

As the science of human health risk assessment advances, and methodologies, policies, and guidance evolve, the Air Force is committed to staying current with these developments, and to remaining a leader in the application of new tools that can enhance the cost-effective mitigation of risks associated with hazardous waste sites. This commitment includes an

assurance that remedial project managers (RPMs) have at their disposal the full range of state-of-the-art tools to support the decision-making process as it affects site restoration at Air Force facilities. A tiered approach, starting with a risk-based screening comparison, and advancing to a quantitative deterministic risk assessment, followed by a probabilistic risk assessment, as necessary, is essential to managing Air Force resources. For certain sites, the use of probabilistic techniques can provide relevant information to support more cost-effective risk management decisions.

1.2 PURPOSE AND SCOPE

This introductory handbook has been developed to illustrate the importance and usefulness of probabilistic approaches in estimating potential human health risks. The probabilistic techniques are discussed for use as part of a higher stage of a tiered approach in the evaluation of potential health risk associated with contaminants at a site. The probabilistic analysis would follow the more simplistic, yet health-protective, screening and/or quantitative deterministic risk assessment steps. This handbook has been developed for the Air Force RPM to help identify the appropriate need for such an analysis. It is also intended for use by the Bioenvironmental Engineer who provides risk assessment expertise in support of the Air Force RPM.

This document, which focuses on human health risk assessment rather than ecological risk assessment, reviews the most commonly used probabilistic methods. In particular, Monte Carlo simulation techniques are presented. The key feature of probabilistic methodologies – quantitatively assessing *uncertainty* and *variability* – is reviewed with emphasis on how such analyses can affect and enhance the results of the risk assessment. In addition, the advantages, disadvantages, and practical applications of using such statistical methods in risk-based decision-making is discussed.

More specifically, the primary objectives of this handbook are to:

- Describe a tiered approach in the evaluation of potential human health risk associated with contaminants at a site, and identify when and how probabilistic risk assessments can be performed using this approach;
- Summarize the current state of the science, including a description of deterministic and probabilistic risk assessment approaches, and how to decide which is the most appropriate tool to use;
- Summarize sources of uncertainty and variability in estimating human health risks and outline methods to quantify and distinguish between the two, with emphasis on the exposure assessment step of the risk assessment process;
- Illustrate the types of information that can be generated to support more rational, cost-effective risk management decisions, remedial designs, and remedial actions;

- Show how these more advanced tools complement the tiered risk-based corrective action (RBCA) methodologies; and
- Supplement forthcoming Air Force risk assessment guidance on state-of-the-art risk assessment techniques.

In addition, the difference between risk *assessment* and risk *management* is reviewed. Emphasis is placed on how risk managers (i.e., the Air Force RPMs) can use the results of quantitative uncertainty and variability analysis in risk-based decision-making.

1.3 TIERED APPROACH

The tiered approach to risk assessment commonly uses two main levels for evaluating the potential health risk associated with contaminants at a site, the first level (Tier 1) involving a screening comparison of chemical-specific site concentrations to risk-based concentrations, and the second level (Tier 2) involving a quantitative deterministic risk assessment. At the present time, deterministic risk assessments are most commonly used at hazardous waste sites to estimate potential health risks and establish cleanup standards. The deterministic method is generally based on use of high-end point values as input factors (i.e., variables) in the risk assessment calculations (e.g., exposure, intake, and toxicity factors used in the equations), thereby providing reasonable maximum estimates of potential risk. The third level to the tiered approach, a probabilistic analysis, takes into account the uncertainty and variability associated with these input factors.

Specifically, probabilistic risk assessment methodologies use distributions as inputs into the risk assessment equations, rather than just high-end point values as used in the deterministic risk assessment. An appropriate statistical technique is then applied (e.g., Monte Carlo simulation) and a distribution of risk (rather than a high-end point estimate) is calculated. Using this approach, a quantitative analysis of uncertainty and variability is provided. (For example, variations in the receptor population's actual exposure patterns are accounted for quantitatively.)

By showing the distribution of health risk, results of the probabilistic risk assessment provide a more realistic picture of the actual risk posed to potential receptor populations. The key benefits of probabilistic risk assessments are that they are more informative and provide more relevant information upon which the RPMs can base their risk management decisions.

1.3.1 Variability vs. Uncertainty

In addition to providing a single-point value risk estimate for a hypothetical, maximum-exposed individual (deterministic risk assessment), estimates may need to be prepared to better inform decision-makers about the realistic nature of the risk. As noted above, one way to do this is to perform a quantitative analysis of the uncertainty and variability of the risk for the population of exposed individuals. Throughout this handbook, such a quantitative analysis will simply be referred to as a "quantitative uncertainty analysis" since uncertainty

and variability are often both referred to as "uncertainty" (albeit different forms of uncertainty as discussed below).

The term *variability* represents true heterogeneity in characteristics within a population. Variability cannot be reduced by taking more samples. *Uncertainty*, on the other hand, means a lack of knowledge about characteristics within a population and may be reduced with additional study. For example, in residential exposure to contaminated drinking water, USEPA's deterministic method has historically assumed that an individual resident drinks 2 liters of water per day (2 L/day), based on total water intake. However, water intake rates vary from person to person, and taking more measurements is not going to change that fact. In other words, the individuals in the population exhibit *variability* in their water intake rates and this variability is not reducible through additional measurement.

It should be noted that when measuring a variable range, the measurements may not always be sufficient. Two common reasons are (1) use of an inaccurate measuring device, and (2) inconsistent use of a measuring device. Thus, the result could be a measured range of water intake rates that is *uncertain.* Unlike variability, the uncertainty about the water intake range can be reduced if a more accurate measuring tool is used or if consistent and correct techniques are practiced. Thus, unlike variability, it is possible to reduce uncertainty by taking more measurements.

It is important to understand that while probabilistic risk assessment allows for a quantitative analysis of uncertainty and variability, this issue is not entirely ignored in the current USEPA deterministic process. The deterministic paradigm has always provided for an evaluation of the potential variabilities and uncertainties inherent in any estimate of health risk.[1] These uncertainty evaluations are generally qualitative in nature.

Such qualitative evaluations of uncertainty and variability are important for all risk assessments, including probabilistic. These evaluations, usually based on inherent uncertainties associated with the risk assessment process and best professional judgment of the analyst, may not have a tangible, quantifiable basis. Generally, these evaluations include qualifying statements about potential sources of uncertainty and/or variability in the measured data or in the assumptions used to estimate potential exposure and risk. For example, the risk assessor might point out, without quantifying the statements, that potential human exposures could deviate from those used in the risk assessment equations through differences in exposure frequencies, contact rates, absorption efficiencies, exposure durations, body weight, and life span, and how each of these factors has a degree of uncertainty associated with it which could over- or underestimate risk.

1.3.2 Selecting a Risk Assessment Tier

The inclusion of a quantitative uncertainty analysis in the risk evaluation process typically will increase the complexity of the assessment. In many cases, the degree to which uncertainty and variability are quantitatively addressed will depend largely on the scope of the assessment and the resources available.[2] As such, the RPM should view probabilistic risk

assessment as one of a set of appropriate tools in which the relevance and usefulness will depend on a number of factors. Most hazardous site risk assessments do not begin with the complicated, resource-intensive probabilistic assessment. A number of more simple, yet conservative screening and/or quantitative steps are relied upon initially to evaluate whether an unacceptable risk potentially exists. The initial step (Tier 1) may include a screening risk analysis involving comparisons of site contaminant concentrations to risk-based screening levels of concern (maximum detected concentrations and/or statistically representative concentrations of the contaminants are compared to generic and/or site-specific screening levels). The second step (Tier 2) may include a quantitative deterministic risk assessment. Only at that point, if the situation warrants further analysis, would the RPM consider a more complex evaluation using advanced techniques. This third step (Tier 3) may include a probabilistic risk assessment.

Such a tiered approach is helpful in prioritizing and managing risk. Also to be considered are the costs associated with the effort. Figure 1.1 illustrates the hypothetical relationship between protectiveness, uncertainty, and cost for each of the risk assessment tiers. As the risk assessor progresses through the risk assessment tiers, uncertainty associated with the risk assessment is reduced, but the same level of health protection is maintained. As the uncertainty is decreased, it is likely that the remediation costs required to maintain the same level of health protection would be reduced.

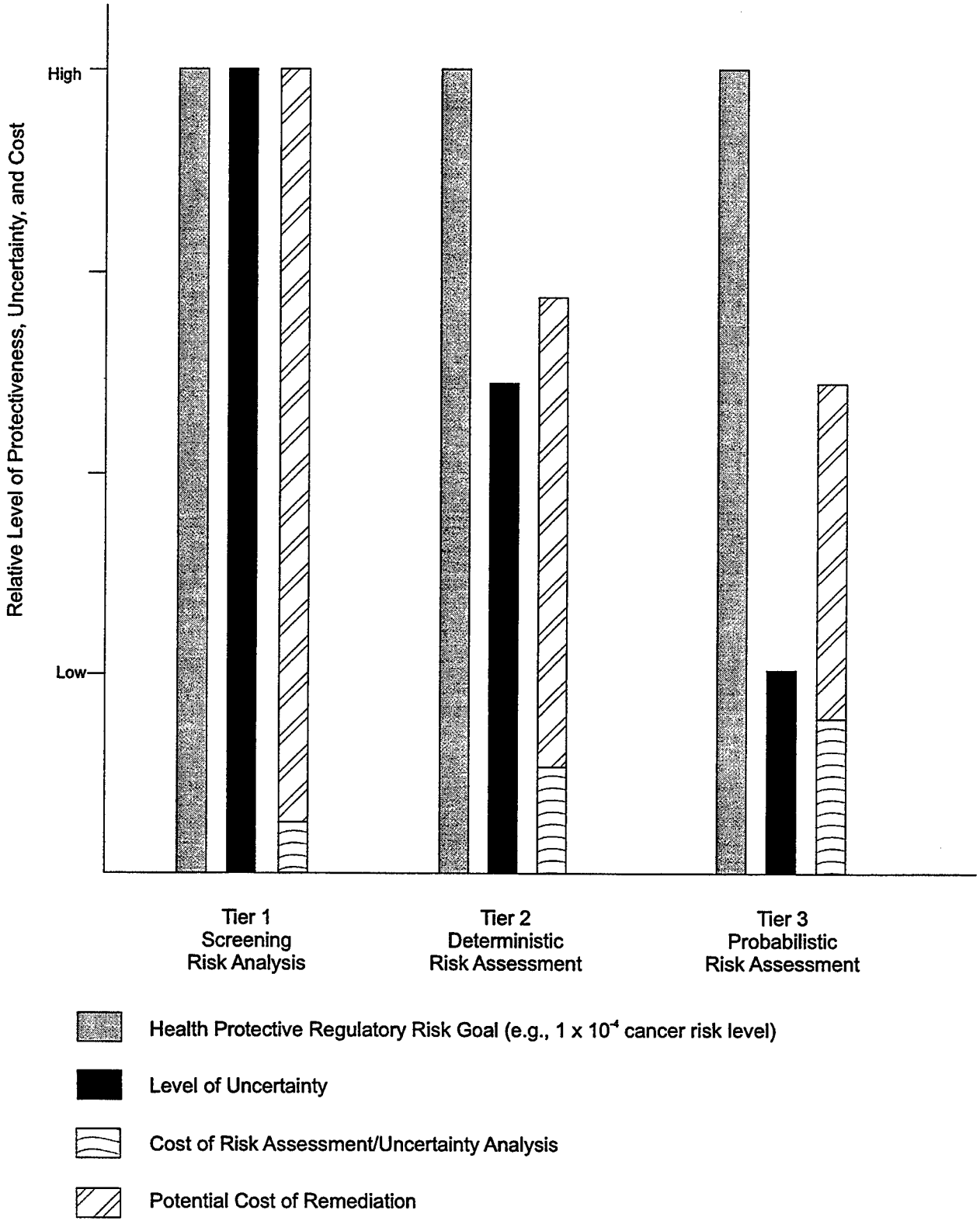
An example of how the tiered approach can reduce the area of remediation by reducing the uncertainty associated with the risk assessment is shown on Figure 1.2. A risk assessment attempts to calculate the unknown true risk associated with a site. However, due to the uncertainty inherent in the risk assessment, the estimated risk, to be protective of health, is chosen to be at the highest end of the risk range. Using the tiered approach, the uncertainty can be reduced and the unknown true risk can be more accurately represented by the estimated risk. Uncertainty analysis can change the number of contaminants which may require remediation, change the media to be remediated, refine the area to be remediated, help determine the method of remediation, or otherwise affect the cost of the remediation. (Reduction in the size of an area to be cleaned up to reach an acceptable risk level is provided as an example in Figure 1.2.)

The total cost of a project should be viewed as the sum of the cost of the risk assessment and the potential cost of the remediation. The RPM should choose a risk assessment tier which will result in the lowest total project cost. Therefore, a forward-looking consideration of the required remediation should be included in the determination of the risk assessment tier to be performed. In some cases, the progression to a higher risk assessment tier (e.g., from a Tier 2 deterministic risk assessment to a Tier 3 probabilistic risk assessment) may not be justified, because the cost of the remediation effort will not be reduced to a level that would justify the increased cost.

THIS PAGE INTENTIONALLY LEFT BLANK

Figure 1.1

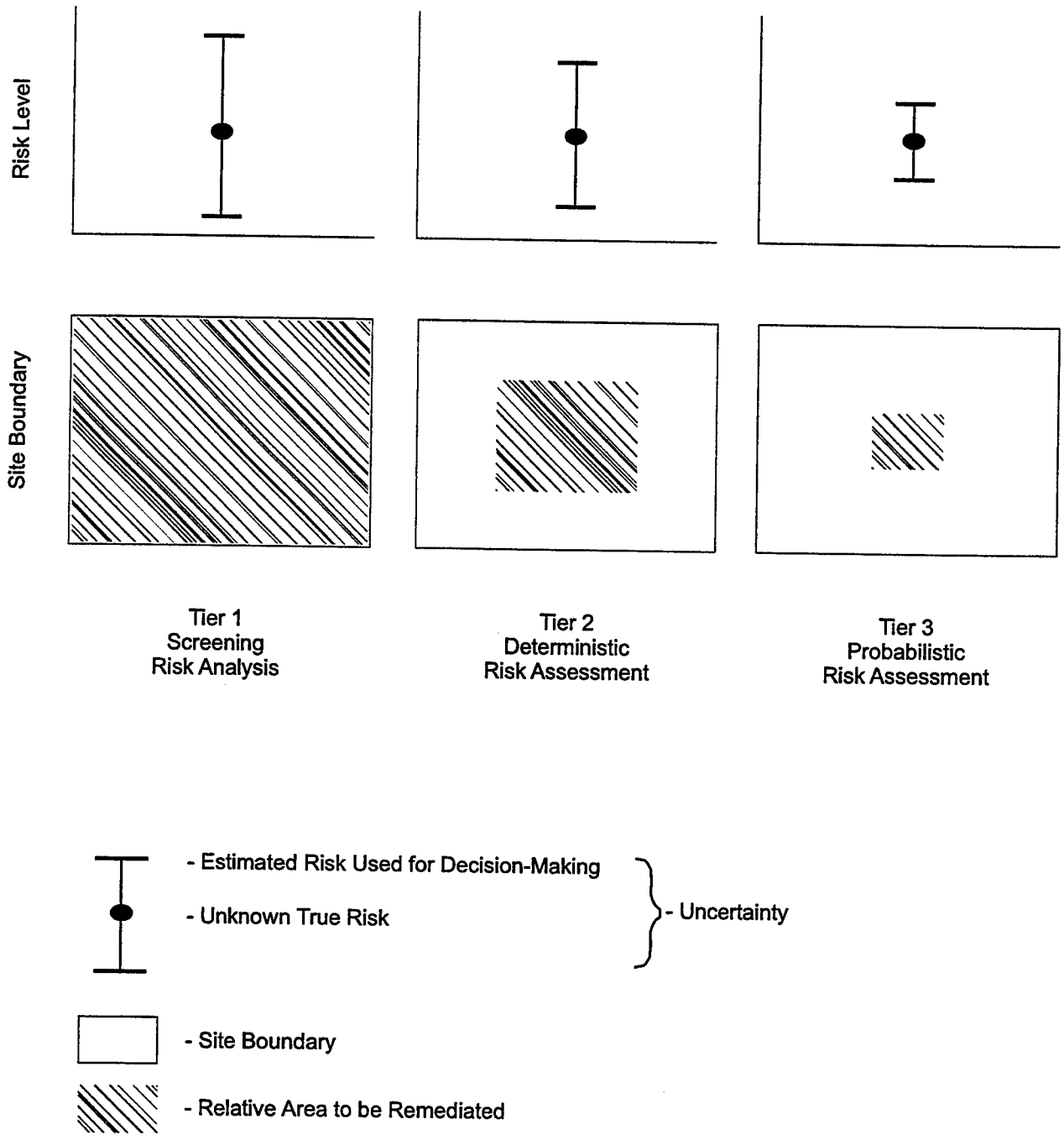
HYPOTHETICAL RELATIONSHIP BETWEEN PROTECTIVENESS, UNCERTAINTY, AND COST BY RISK ASSESSMENT TIER



THIS PAGE INTENTIONALLY LEFT BLANK

Figure 1.2

EXAMPLE OF TIERED APPROACH FOR REDUCING AREA OF CLEANUP *



* Other examples not shown may include changes in the number of contaminants to be remediated, the environmental media to be remediated, etc.

The costs associated with a probabilistic risk assessment may include additional early scoping and regulatory negotiation, collection of sufficient data, increased data manipulation and discussion requirements, enhanced graphics, possibly longer review times, and refined risk communication approaches. These costs and other planning requirements may be appropriate if it will allow the RPM to justify a more limited, cost-effective remedy/solution. For example, only in the probabilistic paradigm can sensitivity analysis be used to identify the input variables that are most responsible for the shape and form of the risk distribution output. Once the most sensitive variables are identified, the risk manager can decide whether the cost of collecting additional data to reduce uncertainty in those variables outweighs the potential extra remediation costs if such data are not collected. Under the deterministic paradigm, such a potential cost saving alternative is not available.

As a rule of thumb, quantitative uncertainty analysis is generally most appropriate when:

- Screening-level assessments and deterministic risk assessments indicate a potentially unacceptable problem;
- Remediation options under consideration may result in high costs; or
- It is necessary to establish the relative importance of site-related contaminants and exposure pathways in the risk estimates.[3]

Another "cost" of using the probabilistic tool is that it may be viewed by regulators and the public as a method to delay action or confuse stakeholders since the results of the assessment are generally more complicated to interpret and use as a decision-making tool. If the probabilistic route is taken, RPMs will need to be especially aware of these issues and take more time to educate stakeholders.[4] It is also worth restating that by deciding not to use the probabilistic approach, decisions based on more simplified techniques could result in the implementation of a remedial solution beyond the level required for adequate protection of human health.

1.3.3 Emerging Perspectives

In recent years, USEPA has acknowledged the need to incorporate quantitative uncertainty analysis into estimates of potential health risks.[5,6,7,8,9,10,11] Quantitative uncertainty analysis is now being recognized by USEPA as a useful approach to improving the decision-making process. The USEPA's Risk Assessment Forum recently approved a new agency policy on conducting human health risk assessments that incorporates a quantitative analysis of uncertainty and variability.[10]

The Risk Assessment Forum's endorsement of the use of Monte Carlo analysis should prompt all USEPA Regions to reevaluate the applicability of probabilistic risk assessment methods in remedial programs. Several USEPA Regional offices have also published guidance on the use of methods to quantify uncertainty and variability.[12,13] These regional policies allow for the preparation of a probabilistic assessment to be used in conjunction with

a deterministic assessment; generally, however, the results of the deterministic assessment are currently recommended for remedial decision-making.

1.4 DOCUMENT ORGANIZATION

This document is organized as follows. Section 1 provides an introduction to the handbook, giving an overview of the objectives and a brief discussion on the use of a tiered approach for estimating the potential risk or hazard associated with contaminants at a site.

Section 2 presents a brief review of basic risk assessment principles and current regulatory perspectives for performing a probabilistic risk assessment. Understanding regulatory perspectives is important when determining whether to pursue a quantitative uncertainty analysis. As regulatory expectations for risk assessment change, Air Force RPMs will need to decide when it may be appropriate to prepare risk information for planning purposes that reflects uncertainty and variability.

Section 3 focuses on uncertainty and variability in estimating human health risks. This section reviews the primary sources of uncertainty and variability that should be considered, with particular emphasis on the exposure assessment step in the risk assessment process. Section 3 also reviews some of the principal published sources of exposure information that may be relevant when incorporating uncertainty and variability into human health risk assessments.

Fundamental methods that RPMs, statisticians, and risk analysts may need in anticipation of a quantitative uncertainty analysis are introduced in Section 4. A familiarity with this information will be valuable when writing requests for proposals (RFPs) and scopes of work (SOWs) and when providing technical direction. Section 4 of this handbook also reviews how probabilistic analysis is conducted, with emphasis on completing a Monte Carlo evaluation as part of a human health risk assessment.

Section 5 presents a simple example of using probabilistic techniques to estimate risk. The output of this evaluation is compared to that obtained from a deterministic approach in which uncertainty and variability are not quantitatively considered. Recommendations on how best to assemble this information for review by all types of stakeholders (e.g., regulators and the public) are provided.

Section 6 concludes with a basic overview of how RPMs can incorporate probabilistic risk estimates into the decision-making process. Recommendations on early scoping and planning considerations, obtaining regulatory consensus, and facilitating effective communication of more detailed risk information are provided.

Several appendices are also provided in this handbook. Appendix A provides a list of references for further reading; Appendix B provides a glossary of common risk assessment terms; and Appendix C provides copies of four USEPA reference documents, including an

example of Monte Carlo simulation to supplement the example given in Section 5 of the handbook.

1 U.S. Environmental Protection Agency. 1989. *Risk Assessment Guidance for Superfund – Volume I: Human Health Evaluation Manual (Part A).* Office of Emergency and Remedial Response, Washington DC. EPA/540/1-89/002. December.

2 U.S. Environmental Protection Agency. 1996. *The USEPA's Perspective on Risk Probabilistic Assessments.* Presented by Dr. Timothy M. Barry. Office of Program Planning and Evaluation. March.

3 National Council on Radiation Programs. 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessment Related to Environmental Contamination.* NCRP Commentary No. 14, Bethesda, MD.

4 Hoffman, F. Owen. 1996. *"What is Probabilistic Risk Assessment?" Risk Assessment Issues in Toxics: The Probabilistic Approach.* USCB Extension Course. March 28-29.

5 U.S. Environmental Protection Agency. 1992. *Guidelines for Exposure Assessment; Notice, (Final).* Exposure Assessment Group, Office of Health and Environmental Assessment, Washington, DC. *Federal Register*, Vol. 57, No. 104, 22887-22938. May 29, 1992.

6 U.S. Environmental Protection Agency. 1992. *Guidance on Implementing the Deputy Administrator's Risk Characterization Memorandum.* Memo from H.L. Longest, Director, Office of Emergency and Remedial Response, Washington, DC. OSWER 9285.7-13. February 26.

7 U.S. Environmental Protection Agency. 1995. *Policy for Risk Characterization.* Science Policy Council, Washington, DC. February.

8 U.S. Environmental Protection Agency. 1995. *Proposed Guidelines for Ecological Risk Assessment.* Risk Assessment Forum, Washington, DC. EPA/630/R-95/002B. August.

9 National Academy of Sciences. 1994. *Science and Judgment in Risk Assessment.* Committee on Risk Assessment of Hazardous Air Pollutants, Board on Environmental Studies and Toxicology, Commission on Life Science, National Research Council, Washington, DC.

10 U.S. Environmental Protection Agency. 1997. *Guiding Principles for Monte Carlo Analysis.* Risk Assessment Forum, Washington, DC. EPA/630/R-97/001. March.

11 U.S. Environmental Protection Agency. 1995. *EPA Risk Characterization Program.* Office of the Administrator, Washington, DC. March.

12 U.S. Environmental Protection Agency, Region III. 1994. *Use of Monte Carlo Simulation in Risk Assessments.* Hazardous Waste Management Division, Office of Superfund Programs, Philadelphia, PA. EPA903-F-94-001. February.

13 U.S. Environmental Protection Agency, Region VIII. 1995. *Use of Monte Carlo Simulation in Performing Risk Assessments (Region 8 Superfund Technical Guidance).* Hazardous Waste Management Division, Superfund Management Branch, Technical Section, Denver, CO. RA-10. September.

SECTION 2

RISK ASSESSMENT APPROACHES

This section provides a basic overview of deterministic and probabilistic techniques in estimating human health risk. Included in this discussion is a review of the principles upon which most risk assessments are based, including the points at which different methodologies diverge. Also included is a historical perspective of the evolution of the risk assessment process.

2.1 OVERVIEW OF RISK ASSESSMENT

Based in part on recommendations made by the National Academy of Sciences (NAS) in 1983,[1] scientists and public health policy professionals have developed a process to assess the extent of risk to human health at hazardous sites including the Federal Superfund and other remedial oriented programs (e.g., RCRA Corrective Action Program).[2] The risk assessment process, as used by USEPA in these programs, is based on scientific information and public health policy considerations, and is intended to promote the development of risk assessments that are technically consistent and protective of human health. The USEPA has qualified the risk assessment process by stating that it (1) is not exact and continues to evolve as more information is gathered about the effects of various chemicals (man-made and/or naturally occurring),[2] and (2) was developed to "produce protective, rather than best, estimates of risk."[3,4]

The primary risk assessment methodology traditionally used and required by USEPA for decision-making purposes is called *deterministic* or single-point evaluation of risk. Such risk assessments are predicated on reasonable maximum exposure (RME), or "high-end" assumptions and criteria which are used as input factors in the risk calculations. USEPA has required development of RME input values to facilitate standardized and conservative (high-end) assessments for risk-based remedial decision-making. This methodology was clearly described in 1989 with the advent of the Superfund risk assessment process and the publication of USEPA's "Risk Assessment Guidance for Superfund" (RAGS).

Since then, however, two important USEPA risk assessment guidance documents have placed additional emphasis on principles that promote, in addition to the RME single-point estimates, other descriptors of risk (e.g., exposure and dose information, such as how many cases of a particular effect might be probabilistically estimated in a population during a specific time period, or what percentage of the population is above a certain exposure, dose, or risk level).[5,6] Another related principle adopted by the USEPA[4] (and based on NAS[1] recommendations) requires that risk assessment be free of any subjective input-variable

manipulation or bias aimed at ensuring one particular outcome or result versus another. These three key principles are summarized below:

- Key scientific data and methods and their uncertainties should be identified in the risk characterization, and a statement of confidence should be included that identifies all major uncertainties along with comment on their influence on the assessment;
- Information on the range of exposures derived from exposure scenarios and on the use of multiple risk descriptors should be presented; and
- Risk assessment information must be clearly presented separate from any non-scientific risk management considerations.

Because single-point (deterministic) risk assessments focus on the high-end estimate at the upper end of an exposure and dose distribution and not the range of possible outcomes, they do not quantitatively address or incorporate exposure and dose descriptors into the risk assessment process. Specifically, deterministic methods provide little information about uncertainty and variability surrounding the estimated high-end risks.[3] It has been common practice to use such "high-end" risk values without the benefit of a more realistic picture of the true nature of risk (i.e., risk as a range of possible outcomes). Even if the high-end exposure scenario is chosen as the decision-making endpoint, it is advantageous to compare the single-point value to other estimates of risk. The high-end of the distribution in the deterministic risk assessment conceptually means above the 90th percentile of the population distribution. The probabilistic analysis could be used to provide additional information on the percentage of the population above a particular exposure, dose, or risk level.

Prohibiting subjectivity in the risk assessment and separating risk assessment information from risk management information is often difficult. Risk assessors may use varying assumptions regarding analytical data, exposure scenarios, or toxicity-related information. Although these assumptions are based on scientific literature and use best professional judgement, differences in the selection of assumptions and input variables may result in very different characterizations of risk. Risk managers, on the other hand, should take the information provided by the risk assessment and subsequently ask "What is the best plan or course of action for dealing with the risks?" In making such decisions, the risk manager will have to factor in not only scientific, but economic, social, political, and other considerations.

It should be noted that, even though the goal is to separate risk assessment from risk management, in the real world the two often overlap. For example, it is often the case that risk estimates exceed USEPA[7] acceptable risk levels (e.g., an estimated cancer risk greater than $1x10^{-4}$). Evaluation of the same site by a different risk assessor using different assumptions and input factors, however, might result in an estimate of risk that is within acceptable bounds. This is especially true if the uncertainty and variability inherent in the estimates are relatively high. Thus, a risk assessor can introduce a certain amount of risk management into the process via the subjective selection of various risk assessment assumptions and input factors.[3,4] One solution to this problem is to provide risk managers with more information such as a range of possible risk outcomes, along with information

relating to the probability of each outcome occurring and the limitations of the estimates. This lessens the opportunities for risk assessors to act as risk managers, whether on purpose or not.

2.2 RISK ASSESSMENT BASICS

Given the three principles described above as a backdrop, this section provides a brief summary of current USEPA risk assessment components, and discusses USEPA-required deterministic and supplemental probabilistic risk assessment methodologies, highlighting some advantages and disadvantages with respect to the evaluation of variability and uncertainty. These comparisons are intended to show how risk assessment conclusions and recommendations from the two approaches can be used by Air Force RPMs to the fullest extent possible in the decision-making process. (It should be noted that risk assessment should be treated as a tool, not as an end in itself as has often been the case in the past. Limited resources should be focused on generating information that helps risk managers choose the best possible course of action among the available options.)[6]

The term "risk assessment" is defined as the objective process by which scientific data are analyzed to describe the form, dimension, and characteristics of risk (i.e., the likelihood of harm to humans or the environment).[1,8] It is primarily a scientific effort in which data from toxicology, epidemiology, and exposure studies are used to estimate the nature and probability of risk at a given site.

Risk assessment estimates the magnitude of the risk, but makes no judgment concerning the applicability of that risk. In other words, risk assessments cannot determine whether adverse health effects have actually occurred or will occur in the future. Risk assessments also cannot identify particular individuals likely to suffer health problems because of contamination at a site.[2] Risk assessment is most useful when those who rely on it to inform the risk management process understand its nature and limitations and can successfully explain those concepts to the risk manager.

Risk assessment and risk management are closely related but, as discussed above, are theoretically discrete processes – one supposedly objective (risk assessment), and the other more subjective (risk management). Specifically, risk management is the process by which decisions are made using all available information (including, but not limited to, the results and recommendations of the risk assessment).[5,8]

Put another way, risk management is primarily a policy-making process in which government officials and the public use the risk assessment as the foundation for making a value judgment about whether the risks are acceptable and, if not, how to manage and reduce such risks. Risk management takes the process from the realm of objective science into the realm of subjective policy, cost-benefit analysis, and value judgment. USEPA has described risk assessment as asking the question "How risky is this situation" while risk management subsequently asks "What shall we do about it?"[9]

As shown on Figure 2.1, human health risk assessment is defined as consisting of four subdisciplines or fields of analysis:[1,9]

1) Data collection and evaluation to identify chemicals of potential concern (COPCs);
2) Toxicity assessment;
3) Exposure assessment; and
4) Risk characterization.

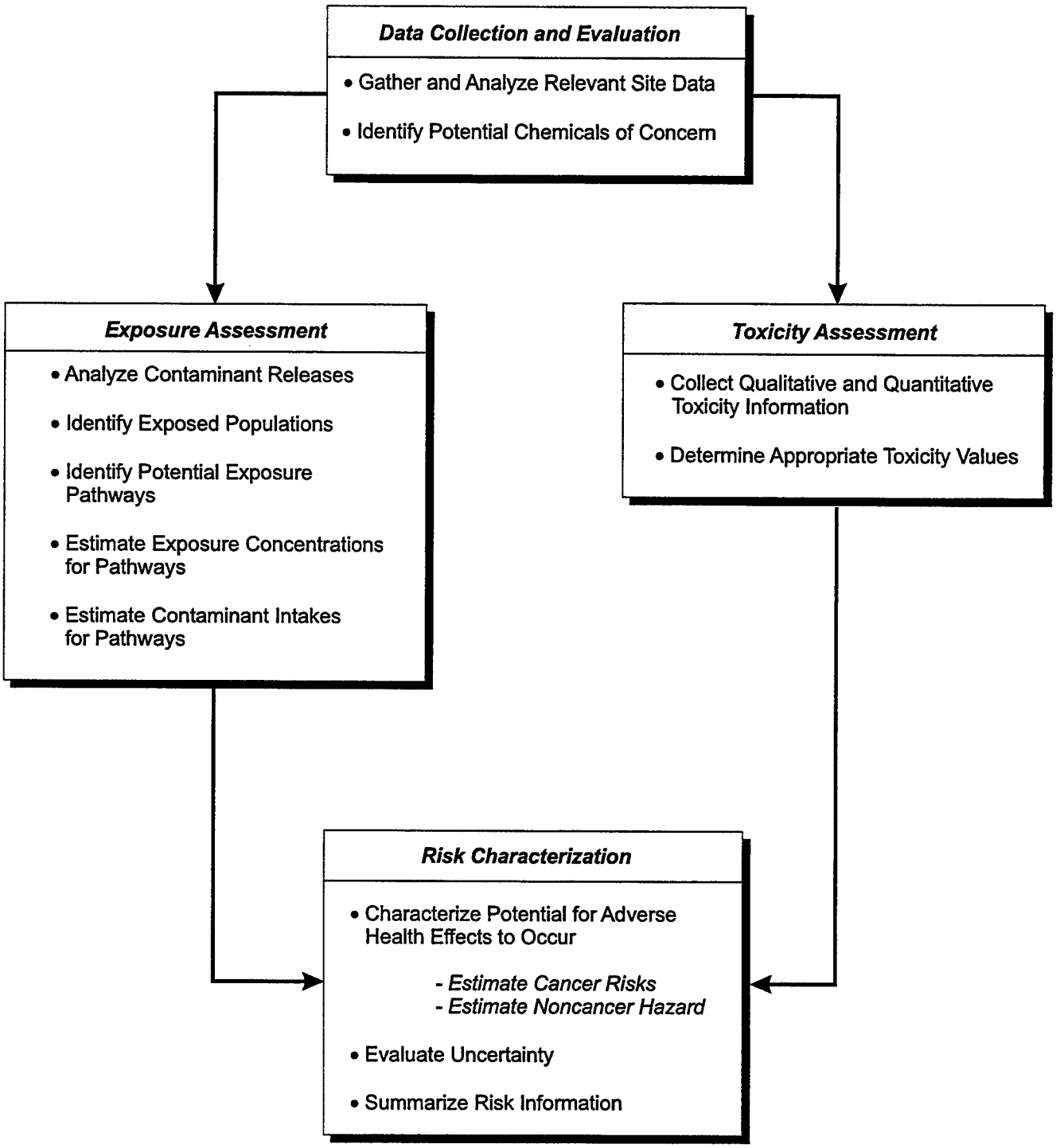
It should be noted that each of these four steps are performed in both deterministic and probabilistic methodologies. Differences arise due to the form the data take (e.g., single high-end point values versus distributions of values), the way the data are handled statistically, and the way the outputs are evaluated. The components of these four steps of the risk assessment process are described below.

Data collection and evaluation is the initial process of evaluating historic uses and potential releases of chemicals at a site and collecting and analyzing samples of environmental media (and in some instances, biological samples) to determine concentrations present in media of concern. The collected data are then evaluated to identify chemicals that may be of potential concern (i.e., the COPCs). A conservative risk-based screening method is often initially applied to the data to develop the list of COPCs (i.e., to reduce the number of chemicals to a subset that is likely to be of most concern). Of all chemicals detected, the COPCs are generally the only chemicals that are carried through the remaining steps of the risk assessment.[2]

Toxicity assessment requires an identification of the adverse effects associated with exposure to a specific chemical and the development of toxicity factors to describe the relationship between the dose of chemical an organism receives and the expected response. Such toxicity data and factors are based both on epidemiological studies of actual human exposures and on experimental animal studies.[2] The adverse health effects identified are classified as either a carcinogenic response or a noncarcinogenic response. Some chemicals cause both type of effects.

Generally, toxicity factors are obtained from USEPA's Integrated Risk Information System (IRIS) or Health Effects Assessment Summary Tables (HEAST). These references are useful resources for point estimates of toxicity factors. However, these references do not provide a complete list of chemicals, only a limited subset for which there are sufficient data to calculate a toxicity factor. Therefore, there are a large number of compounds for which risk is generally not evaluated, since there is not enough information to derive toxicity factors. Additionally, there is considerable uncertainty associated with the toxicity factors based on the study methodology, interspecies extrapolation (animal to human), and intraspecies extrapolation (sensitive subpopulations). While uncertainty in toxicity factors is present, point estimates are generally used because of their widespread acceptance among regulatory

Figure 2.1
PRIMARY COMPONENTS OF RISK ASSESSMENT



THIS PAGE INTENTIONALLY LEFT BLANK

agencies. Toxicity assessment is constantly evolving, with toxicity factors for additional compounds being derived, and existing toxicity factors being changed. To perform the risk assessment, the RPM should verify that the appropriate toxicity factors are being used.

Exposure assessments are performed to estimate the type and magnitude of receptor exposure to the chemicals detected at a site. The exposure assessment answers questions such as: "Under what scenarios are people exposed, how often, by what pathways, and to what concentrations?" Other relevant factors addressed in the exposure assessment include the physical characteristics of the receptor populations, such as body weights and age structure, and contaminant fate and transport. A conceptual site model (CSM), in the form of a table or figure, is generally developed to show the results of the exposure assessment. The CSM may include identification of the contamination sources, affected media, release mechanisms, migration pathways, exposure routes, potential receptors, land-use assumptions, potential for exposure, and whether or not an exposure pathway is potentially complete.

The USEPA considers exposure to sensitive subgroups of populations as applicable in the risk assessment Superfund program and has been estimating individual risk corresponding to the RME.[10] RAGS[9] defines the RME as the highest exposure that is reasonably expected to occur at a site and in practice is estimated by combining 90th and 95th percentile values for some but not all exposure variables. Additionally, to promote consistency in the evaluation of RME when site-specific data are missing, the USEPA has released supplemental guidance describing standard default exposure factors for use in quantitative risk assessments.[10,11] The standard default values presented in these documents provide a description of the high-end portion (the RME) of the exposure distribution. Using these standard exposure values provides an estimate of exposures in the upper range of the distribution. Conceptually, this would be above the 90th percentile of the population distribution, but not higher than an individual who may potentially have the highest exposure.

Generally, the risk assessment considers currently exposed populations and populations that may be present under reasonably anticipated future uses of the site. For example, an Air Force base may currently have a worker population, but if closed and converted to residential land use in the future, other exposures, such as childhood exposures, may become relevant.

The combination of data collection/evaluation and the exposure evaluation provides an understanding of who is exposed to which hazardous chemicals and what doses those receptors are estimated to receive (e.g., how often they are exposed, how the exposures occur, and the chemical concentrations to which they are exposed). At this point in the risk assessment process, no conclusions, either qualitative or quantitative, have yet been made regarding the potential risk to the receptors.

Risk characterization, the last step of the risk assessment, combines the dose estimate results from the exposure assessment with information developed in the toxicity assessment to make quantitative statements about risk.[8] Typically, the end result of the risk characterization (both deterministic and probabilistic) is a set of chemical-specific numerical risk estimates.

In the deterministic paradigm, these quantitative estimates are presented as cumulative cancer risk estimates for a receptor (e.g., cancer risk of $1x10^{-5}$) and cumulative hazard quotients (termed a "hazard index" or HI). The values derived are generally compared to "acceptable" levels of risk (a range of usually $1x10^{-4}$ to $1x10^{-6}$ for carcinogens, and an HI of 1.0 for noncarcinogens.) Because the deterministic values represent the RME individual, the deterministic estimates of risk are generally high-end risk descriptors.

The probabilistic paradigm, on the other hand, can evaluate a population in a variety of ways (rather than focusing on one highly exposed individual). For example, a population of workers at an Air Force base might be considered together. The exposure descriptors (e.g., inhalation rate, body weight, exposure duration) might be described by realistic ranges rather than high-end point values. The output of the probabilistic analysis is a distribution of risk for the entire worker population. This technique presents a clearer picture of the uncertainty and variability in the risk estimates and how that impacts the potential risk posed to facility workers.

In addition to making quantitative statements about the level of risk present, the risk characterization also identifies the limitations of the information collected in preceding steps and makes statements about uncertainty and variability in the risk estimates.[12] In the case of deterministic risk assessment, this takes the form of a qualitative discussion of uncertainty and variability and how these factors may cause the risk estimate to be over- or underestimated. In probabilistic methodologies, more quantitative evaluations of these two characteristics can be made. A basic understanding of uncertainty and variability inherent in risk estimations can help clarify assumptions and limitations. Consequently, Air Force RPMs may be able to make more informed remedial decisions utilizing this information.

2.3 DETERMINISTIC RISK ASSESSMENTS

As noted above, deterministic methods are generally used to produce RME single-point risk estimates. This method is currently recognized as the standard approach for quantifying risks to human health. This approach is consistent with the National Oil and Hazardous Substances Pollution Contingency Plan (NCP), which requires that RME scenarios be used "in the remedial decision in evaluating what is necessary to achieve protection against risk to human health."[5,13]

2.3.1 Advantages

There are several advantages to using a deterministic risk assessment approach. The deterministic approach:

- Uses relative straightforward calculations;
- Is the most widely-accepted and standard approach used by regulators and responsible parties;

- Provides an estimate of potential risk that may potentially occur to a small but definable high-end segment of the receptor population;
- Is useful for determining that a site or an exposure scenario at a site is not of concern;
- Provides a level of consistency for comparability between risk assessments (e.g., to prioritize response actions at Air Force facility sites based on potential relative risk); and
- Is easier to explain and understand (i.e., facilitates risk communication).

2.3.2 Disadvantages

There are also several disadvantages associated with the deterministic risk assessment approach. Some disadvantages are that the approach:

- Often results in estimates of risk that are biased high relative to the mean values of the uncertainties they represent;
- May portray a false degree of precision of the risk assessment process – "a single-point value is the actual risk;"
- May place the risk assessor in an inappropriate risk management role (i.e., different assumptions can result in different estimates of risk depending on interpretation and use of appropriate input values);[2]
- Provides little or no information to decision-makers regarding the distribution of possible risks, the magnitude of underlying uncertainties, or any quantitative indication of the key sources of uncertainty; and
- May result in unreasonable cleanup goals beyond what is necessary to protect human health, and which may not be technologically attainable or cost-effective, thereby reducing the amount of funding available for other risk reducing opportunities.[14,15]

2.4 PROBABILISTIC RISK ASSESSMENTS

In order to assess the full range of risk possibilities, Air Force RPMs can utilize probabilistic statistical analyses to develop a more realistic picture of risk posed to an exposed population. The most frequently used and perhaps best understood of the tools used to perform this statistical analysis is called Monte Carlo analysis and is generally run with the aid of software developed for this purpose. Probabilistic statistical techniques allow risk estimation to incorporate most of the potential exposure and dose scenarios rather than those associated with upper-end, conservative assumptions only.[3,4]

As used in probabilistic risk analysis, Monte Carlo simulation is a statistical technique by which a risk equation is solved numerous times (perhaps tens of thousands of iterations). The inputs to the risk equation, rather than conservative point values, are some combination of point values and distribution functions that more clearly define the variability and/or uncertainty associated with the variable. Each calculated risk estimate has an associated

likelihood of occurrence. The multiple results, when plotted graphically, represent a cumulative frequency that is useful in understanding the probability of hypothetical outcomes. This technique will be discussed more fully in Chapter 4.

In contrast to the deterministic analysis, probabilistic risk assessments use statistical simulation techniques to generate probability distributions of risks. This provides not only more information, but higher quality data to risk managers and the public than that provided by standard point estimates. Probabilistic risk assessments provide an explicit, quantifiable characterization of risk and uncertainty.[16]

A probabilistic risk assessment is a valuable tool for quantifying uncertainty because:

- The risk equation is solved numerous times to generate a range of possible answers (see Section 4);
- Each calculated risk value has an associated probability of occurrence; and
- The output of the analysis reflects the full distribution of the potential risk, not just the high-end, single-point estimate.

2.4.1 Overview of Probabilistic Risk Assessments

The basic goal of a probabilistic risk assessment is to conduct the assessment of exposure or risk in a "realistic manner" for a given assessment endpoint by accounting for all of the available information and the lack of knowledge. The realism is introduced through the language of probability – the probability of occurrence of an event in light of what is known and not known. Quantitative characterization of uncertainty and variability are tools to accomplish the goal.

Some general conditions where probabilistic methods are most appropriate for a site include the following:

- When it is necessary or desirable to characterize uncertainty and variability in the estimates of risk (i.e., whenever a more detailed, realistic risk estimate is needed);
- When results of more simplistic risk assessment methods (e.g., a tiered approach, such as a risk-based screening comparison and a deterministic single-point risk analysis) show that the potential risk from exposure at the site is above risk levels of concern.[17]
- When distributions of values are available or can be estimated for any type of variable used to develop the risk estimate;[15] and
- When the cost of regulatory or remedial action is high and the potential health risk associated with exposure is expected to be marginal.

There are two primary types of probabilistic risk assessments. The first, and most common, is based on the assumption that certain parameters used to evaluate exposure or risk

can vary in a population, but are constant over time. In this type of analysis, the distribution of risks in a population based on that variability is estimated. The second type of probabilistic risk assessment is based on the assumption that certain parameters used to evaluate exposure or risk vary in a population *and over time*. This type of analysis is more complex, resulting in an estimate of the distribution of risks for a population based on a lifetime exposure assessment for thousands of individuals. Both probabilistic methods are more detailed and require more training and expertise than deterministic methods. This handbook addresses the first, most common type of probabilistic risk assessment.

In addition to the discussions in this handbook, the following USEPA references should be used for further direction on risk assessment, and perspective on the use of probabilistic methods for assessing uncertainty and variability:

- The 1989 Superfund Risk Assessment Guidance,[9]
- The 1992 Guidance on Risk Characterization for Risk Managers and Risk Assessors,[4]
- The 1992 Exposure Assessment Guidelines,[5] and
- The 1995 Policy for Risk Characterization.[18]

2.4.2 Role of Probabilistic Risk Assessments

Probabilistic methods are very powerful and can be used for multiple applications, such as human health exposure assessments and risk characterizations (the focus of this handbook), ecological risk assessments, and pharmacokinetic models (the study and modeling of the disposition of chemicals in the body). Thus, a quantitative approach to uncertainty analysis can help decision-makers address many types of questions in a more robust way.

Probabilistic methods, such as Monte Carlo analysis, are viable statistical tools for probabilistic risk assessment.[16] There are also numerous other probabilistic techniques available for the Air Force RPM to choose from, including (1) analytic techniques such as variance propagation models, (2) response surface modeling, and (3) differential analysis. As previously stated, this handbook focuses on the Monte Carlo technique because it is one of the most commonly used and easily understood probabilistic methods.

2.4.3 Emerging Regulatory Notice of Probabilistic Approaches

The USEPA is placing continued effort in the process of developing guidance covering the use of probabilistic techniques in Agency risk assessments.[19] The USEPA Risk Assessment Forum was established to promote scientific consensus on risk assessment issues, including recommendations to advance the development of guiding principles on how to prepare and review an assessment based on use of probabilistic techniques, specifically Monte Carlo analysis.[20] USEPA Regions III and VIII have recently decided to accept Monte Carlo simulations, submitted as uncertainty and variability analyses, as part of baseline human health assessments.[3,21] For example, the Region VIII guidance states that:

> "When compared with alternative approaches for assessing parameter uncertainty or variability (e.g., analytical uncertainty propagation or classical statistical analysis), the Monte Carlo technique has the advantages of very general applicability, no inherent restrictions on input distributions or input-output relationship, and relatively straight-forward computations. In its application to risk assessment, Monte Carlo simulation not only generates results that can be expressed in a more easily understood graphical format, but it also permits the degree of conservativeness to be specified quantitatively (i.e., the Reasonable Maximum Exposure (RME) can be defined as the upper 90th or 95th percentile of output results). Furthermore, risk results are more easily justified statistically, and the uncertainty underlying them can be discussed quantitatively. In general, this approach can satisfactorily address the goals of uncertainty analysis outlined in recent EPA guidance."

The stipulation in both regional guidances is that all risk assessments must still include the single-point RME risk estimates prepared under current USEPA guidance (i.e., Monte Carlo-based risk analyses cannot act as a substitute for the currently accepted deterministic paradigm). In general, Monte Carlo simulations are only accepted as an optional addition to, not substitute for, current risk assessment methods. Furthermore, simulations are not accepted that are not approved beforehand or that do not adhere to guidelines.

The USEPA Risk Assessment Forum's recently published document on the guiding principles for Monte Carlo Analysis, "Summary Report for the Workshop on Monte Carlo Analysis," provides the conditions for an acceptable risk assessment that uses probabilistic analysis techniques.[20] The document presents USEPA's position that probabilistic techniques, such as Monte Carlo analysis, can be viable statistical tools for analyzing uncertainty and variability in risk assessments and establishes conditions that are to be satisfied by risk assessments that use such techniques. Specifically, the conditions relate to good scientific practices of clarity, consistency, transparency, reproducibility, and the use of sound methods.[19,20]

2.5 WHEN TO USE A PROBABILISTIC RISK ASSESSMENT

Because Monte Carlo analysis can be a resource-intensive activity, the level of sophistication should be appropriately tailored to the goals of the analysis. There are several potential advantages to using a probabilistic risk assessment approach, depending on the circumstances particular to a given site. Figure 2.2 illustrates the importance of weighing the pros and cons when considering coupling probabilistic risk assessments with a deterministic assessment. To systematically discern the most appropriate level of analysis, Air Force RPMs should follow a decision tree such as the one presented in Figure 2.3. This decision tree is intended to help RPMs implement a tiered approach. This figure also illustrates the additional steps involved in using a probabilistic approach once the RPM determines that the standard deterministic approach does not provide sufficient information.

Figure 2.2
ADVANTAGES AND DISADVANTAGES OF CONDUCTING A PROBABILISTIC ANALYSIS

ADVANTAGES

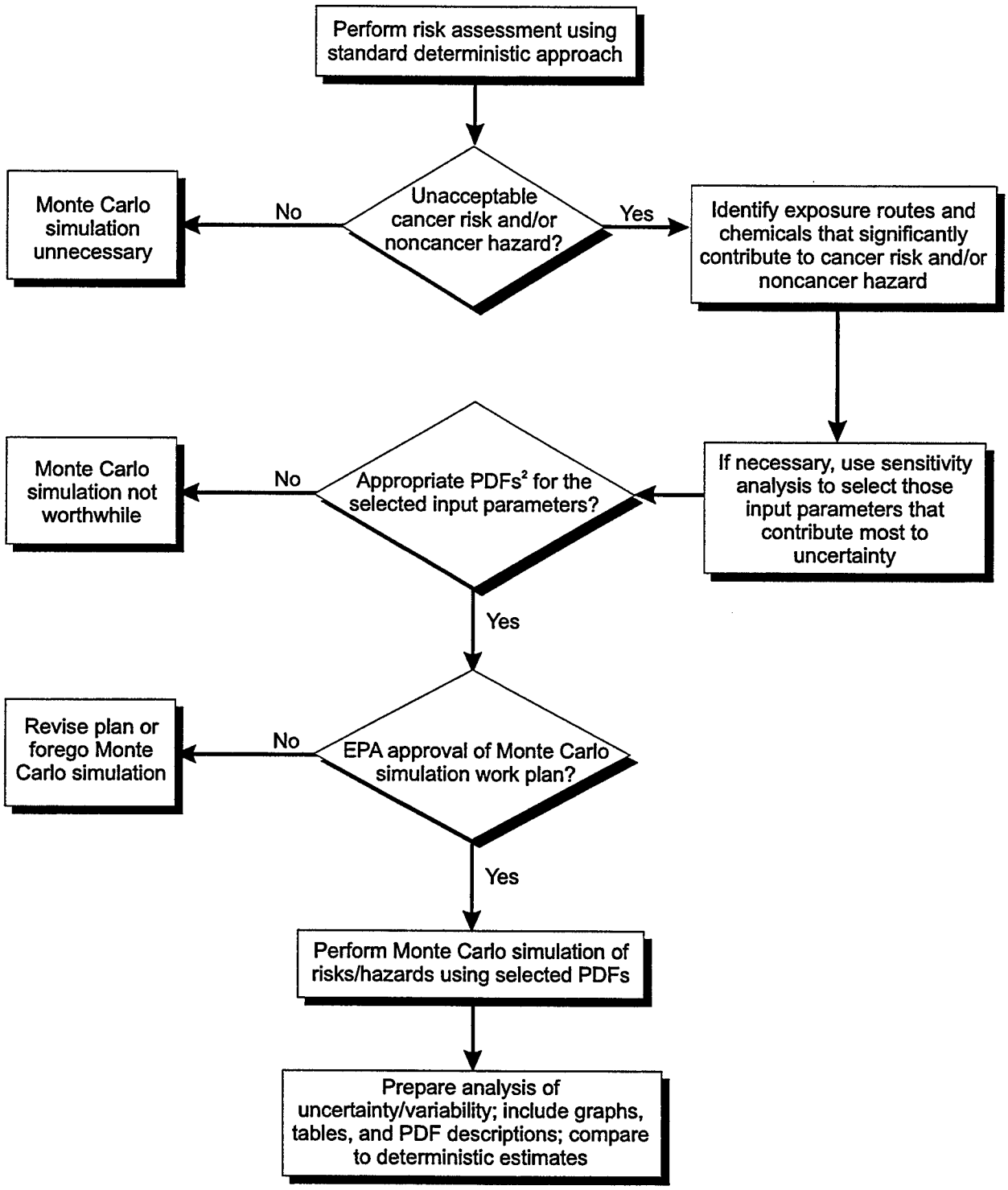
- Honors definition of risk
- Uses all information pertaining to uncertainty and variability
- Reveals the compounded conservatism inherent in deterministic methods
- Helps reestablish the line between risk assessment and risk management
- Can result in long-term cost savings (potential cost of remediation)
- Provides a "reality check" of USEPA default values
- Provides a detailed estimate of a range of risks via the use of distribution functions
- Illustrates the full range of exposures and risks
- Identifies variables and factors which have the greatest impact on producing uncertainty

DISADVANTAGES

- Involves more complex calculations
- Resource-intensive
- No long-standing tradition of regulatory use to date
- Inherent limitations in distinguishing between variability and uncertainty or potential correlations of variables
- Ignoring correlations among exposure variables can bias simulations
- More complex to communicate results versus single-point estimates

THIS PAGE INTENTIONALLY LEFT BLANK

Figure 2.3
RISK ASSESSMENT DECISION TREE [1]



[1] Figure modified from *Use of MONTE CARLO Simulation in Risk Assessments*, USEPA Region VIII Superfund Technical Guidance, September 1995.[21]

[2] PDF A probability density function is a statistical tool used to determine the distribution of values for a random variable, each value having a specific probability of occurrence. PDFs may be selected from available literature ("standard" data distributions that are not influenced by site-specific conditions) or may be developed if published distributions are not applicable to the site. (Refer to Section 4 of this handbook.)

THIS PAGE INTENTIONALLY LEFT BLANK

For sites that do not fall clearly under CERCLA or RCRA jurisdiction (such as petroleum release sites), Monte Carlo analysis may be incorporated into standard tiered approaches to risk assessment that progress from simpler (i.e., screening level analysis) to more complex analyses as the risk management situation requires.[16] Tiered approaches such as the American Society for Testing and Materials (ASTM's) risk-based corrective action models,[22] are gaining wide acceptance by States, Federal agencies, and industry as cost-effective strategies for environmental site management.

For example, a Tier 1 evaluation may simply require a comparison of maximum detected site concentrations of contaminants for a particular environmental media (e.g., soil) to generic (i.e., non-site specific) risk-based screening criteria. The outcome of this lower tier comparison typically helps the Air Force RPM determine whether additional study is required. Typically, the level of sophistication inherent in probabilistic risk assessments reflects an advanced tier level, such as ASTM's Tier 3.[22]

The higher Tier 3 level evaluation may be beneficial after early decision "tiers" indicate the necessity for further evaluation. The following documents should be referenced for more information on tiered approaches:

- ASTM's Risk-Based Corrective Action Program,[22] and
- USEPA's Soil Screening Methodology.[23,24]

Because the level of complexity increases as the risk evaluation progresses, the impacts on budgets and levels of understanding by the regulators and public must be carefully considered. Advantages derived from the application of good science does not necessarily correlate with public or regulatory acceptance.

The decision of whether and when to introduce probabilistic methods is critical. If complicated topics are introduced at the preliminary stages of the process, stakeholders have a chance to develop scientific-based opinions about the cost/benefits associated with more complex techniques. Introducing these techniques too late in the process may cause regulators and the public to view the use of probabilistic methods as an attempt to delay or misrepresent the results.

1 National Academy of Sciences. 1983. *Risk Assessment in the Federal Government: Managing the Process*, Washington DC.

2 U.S. Environmental Protection Agency, Region VIII. 1994. *Health Risk Assessments for Superfund Sites*: Fact Sheet. Region VIII Technical Guidance, Denver, CO. December.

3 U.S. Environmental Protection Agency, Region III. 1994. *Use of Monte Carlo Simulation in Risk Assessments*, Hazardous Waste Management Division, Office of Superfund Programs, Philadelphia, PA. EPA903-F-94-001. February

4 U.S. Environmental Protection Agency. 1992. Memorandum to Assistant Administrators from F.H. Habicht, Deputy Administrator. *Guidance on Risk Characterization for Risk Managers and Risk Assessors*. Washington, DC. February 26, 1992.

5 U.S. Environmental Protection Agency. 1992. *Guidelines for Exposure Assessment; Notice, (Final)*. Exposure Assessment Group, Office of Health and Environmental Assessment, Washington, DC. *Federal Register*, Vol. 57, No. 104, 22887-22938. May 29, 1992.

6 U.S. Environmental Protection Agency. 1991. *Role of the Baseline Risk Assessment in Superfund Remedy Selection Decisions*. Office of Solid Waste and Emergency Response, Washington, DC. OSWER Directive 9355.0-30. April.

7 National Academy of Sciences. 1994. *Science and Judgment in Risk Assessment*. Committee on Risk Assessment of Hazardous Air Pollutants, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council, Washington, DC.

8 Carnegie Commission. 1993. *Science, Risk, and Regulatory Decision-Making*. June.

9 U.S. Environmental Protection Agency. 1989. *Risk Assessment Guidance for Superfund – Volume I: Human Health Evaluation Manual (Part A)*. Office of Emergency and Remedial Response, Washington DC. EPA/540/1-89/002. December.

10 U.S. Environmental Protection Agency, Region VIII. 1993. *Superfund's Standard Default Exposure Factors for the Central Tendency and Reasonable Maximum Exposure*, Denver, CO.

11 U.S. Environmental Protection Agency. 1991. *Human Health Evaluation Manual, Supplemental Guidance: Standard Default Exposure Factors*. OSWER Directive 9285.6-03. March 25, 1991.

12 CalEPA. 1996. *The California EPA's Perspective on Probabilistic Risk Assessments*. Presented by Dr. Richard A. Becker. Office of Environmental Health Hazard Assessment. March.

13 U.S. Environmental Protection Agency. 1990. 40 CFR 300, National Oil and Hazardous Substances Pollution Contingency Plan, Final Rule, *Federal Register* 55(46):8665-8865. March 8, 1990.

14 Frey, H. Christopher. 1992. *Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making.* September.

15 American Society for Testing Materials. 1996. *Draft Standard Guide for Probabilistic Risk Assessment and Handling Uncertainties in Risk Assessments for Contaminated Sites.* ASTM Committee E-47 on Biological Effects and Environmental Fate. E47.13.05.A. February.

16 U.S. Environmental Protection Agency. 1996. *Policy Statement for the Use of Monte Carlo Analysis in Agency Risk Assessments.* November.

17 Crouch, Edmund et al. 1995. *Report to the Commission on Risk Assessment and Risk Management.* October.

18 U.S. Environmental Protection Agency. 1995. *Policy for Risk Characterization.* Science Policy Council, Washington, DC. February.

19 U.S. Environmental Protection Agency. 1997. *Guiding Principles for Monte Carlo Analysis.* Risk Assessment Forum, Washington, DC. March.

20 U.S. Environmental Protection Agency. 1996. *Summary Report for the Workshop on Monte Carlo Analysis.* Risk Assessment Forum, Office of Research and Development. EPA/630/R-96/010. September.

21 U.S. Environmental Protection Agency, Region VIII. 1995. *Use of Monte Carlo Simulation in Performing Risk Assessments (Region 8 Superfund Technical Guidance).* Hazardous Waste Management Division, Superfund Management Branch, Technical Section, Denver, CO. RA-10. September.

22 American Society for Testing and Materials. 1995. *Standard Guide for Risk-Based Corrective Action Applied at Petroleum Release Sites.* Annual book of ASTM Standards. E1739-95.

23 U.S. Environmental Protection Agency. 1996. *Soil Screening Guidance: User's Guide.* Office of Solid Waste and Emergency Response, Washington, DC. EPA/540/R-96/018.

24 U.S. Environmental Protection Agency. 1995. *Soil Screening Guidance: Technical Background Document.* Office of Solid Waste and Emergency Response, Washington, DC. EPA/540/R-95/126.

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 3

UNCERTAINTY AND VARIABILITY IN ESTIMATING RISK

This section discusses the differences between *uncertainty* and the different but related concept of *variability* and provides a discussion of the basic sources of uncertainty and variability in estimating risk. The section then briefly reviews how uncertainty and variability are introduced into risk estimates through data evaluation, exposure assessment, toxicity assessment, and risk characterization.

There are numerous exposure parameters which are combined mathematically to estimate risk to an exposed population using either the deterministic or probabilistic paradigms. As noted in the introduction, these parameters can vary from person to person due to environmental, lifestyle, and genetic differences. As such, it is important to identify available data distributions for specific exposure assessment parameters in order to perform a probabilistic assessment. This section will discuss these sources of information.

In addition, it is often the case in the probabilistic paradigm that adequate data distributions are not available. This section also addresses the elicitation of *expert judgment* in the absence of established distributions or sufficient information to develop distributions from scientific literature. Expert judgement can be used to fill in data gaps in order for the probabilistic risk assessment to proceed.

3.1 UNCERTAINTY

Uncertainty, as used in this document, refers to lack of knowledge. At least in theory, this lack of knowledge can be reduced through further data collection. In practice, the additional cost, time constraints, or minimal impact of the uncertainty may make further data collection impractical, unnecessary, or even impossible.

3.1.1 Sources of Uncertainty

During the risk assessment process, it is often desirable to assess the level of uncertainty incorporated into a determination of risk. The first step in this process is to identify the various sources of uncertainty.

Potential sources of uncertainty can be divided into two broad categories: uncertainties associated with model form and uncertainties associated with assigning values to the parameters of the model. Parameter uncertainty stems primarily from errors such as:

- Imprecision of analytical devices or historical information used to quantify a specific parameter;
- Sampling (random) error caused by making inferences from a limited database;
- Extrapolation and use of subjective information; and
- Systemic errors that can be interjected into the risk analysis process through flawed data gathering processes.[1]

In contrast, model uncertainty stems from using a model or mathematical formula to describe a physical process or to use measured data from one system as a surrogate for another system.[2] A prime example of the latter case is the use of animal cancer potency data to estimate human cancer potency potential. Parameter uncertainty and model uncertainty are discussed in more detail below.

3.1.2 Parameter Uncertainty

3.1.2.1 Measurement Error

Many data sets representing populations, particularly site characterization data, reflect both variability within the population and any additional uncertainty due to measurement errors. Distinguishing measurement-induced uncertainty from population variability may be important when making costly decisions based on conservative risk estimates.

Measurement error is due to the imprecision of the measurement device. The effects of measurement error can be reduced by taking repeated measurements. Measurement error is particularly common in the exposure assessment step of the risk assessment process, especially with regard to establishing average concentrations for site-related chemicals. The potential for repeated measurement (i.e., sampling and analysis) at this stage may be limited by practical and cost considerations. Standard sampling procedures such as taking duplicate or split samples can give some indication of the amount of measurement error likely in the data analysis.

3.1.2.2 Random Error

Random error is another potentially important source of uncertainty. Random error can play a significant role when attempting to establish a single-point value or a probability density function (PDF) from a limited number of samples (see Section 4 for a detailed discussion of PDFs).

Random error arises when a sample is used to represent the true, but unknown, distribution. For example, the sample average of a group of laboratory results can be used to estimate the true population mean. A new sample, however, might provide a different estimate. The difference between the estimates is known as random error and creates uncertainty about the knowledge of the true mean. In general, the larger the sample size, the lower the uncertainty due to random error (assuming there is no bias in the sampling method).

However, in most situations the number of samples is limited and uncertainties due to random error may play an important role in estimating risk.

Another area where random error can be significant is in the determination of cancer risk based upon a limited number of observations. For example, animal bioassays used to develop cancer potency factors are generally based on a relatively small number of test animals per study group.[1]

The effect of random error on the input PDF of a parameter will vary greatly. For some parameters, such as body weight, the data sets used are very large. The corresponding estimate of the true PDF is good because the random error is small. PDFs based on a limited data set, on the other hand, can be an unreliable estimate of the true PDF. Thus, in this type of situation, it would be important to quantify the amount of uncertainty.

3.1.2.3 Systematic Error

Systematic errors, also known as non-random errors or bias, are another source of uncertainty in risk assessment. Systematic errors arise through an inherent flaw in the data collection process.[1] Depending on the direction of the bias, the mean value of the sample data will be either consistently higher or lower than the true mean. Because the error is introduced due to some flaw in the data collection process, increasing the number of samples or the amount of information gathered generally will not reduce the uncertainty due to systematic errors. Systematic error can only be corrected by changes in the data collection or analytical methods.

Use of a surrogate measure to represent the parameter of interest can introduce systematic error if the surrogate and the parameter do not behave similarly under the same conditions, or if incorrect assumptions are made in order to infer the parameter of interest from the surrogate. An example of this problem is the "healthy worker effect".[1] In the "healthy worker effect," the extrapolation of risks determined for exposed workers to the general population are performed. This is an example of the potential for systematic error that can be generated when the population sampled is not representative of the population being modeled. A similar example is the use of data gathered on the general population to represent a subpopulation that differs in some significant way (e.g., a subpopulation more or less susceptible to a given chemical).

3.1.3 Model Uncertainty

The structure of mathematical models employed to represent scenarios and phenomena of interest is often an important source of uncertainty. Risk assessors generally use mathematical models to represent the interaction between parameters, and also to estimate the values of parameters that cannot be measured directly.

Models are simplified, idealized representations of physical processes that may be too complicated to express in any other way.[1] Simplifications are often an important part of the

assumptions upon which a model is based and insufficiently complex models may fail to capture important relationships among input parameters.

Generally, models are selected on the basis of physical principles, for their ability to describe or make sense of a limited set of observations, or for both of these reasons.[1] Often there may be more than one model to explain or characterize a given phenomenon. If the models are based on goodness-of-fit to some set of data, the differing models may fit the data equally well. However, these models may provide different results when applied to other situations. Thus, selection of a particular model can have a profound impact on the final risk distribution.

An example is USEPA's use of the "linearized multistage model" to assess cancer risk. This model extrapolates the risk from high doses used in animal bioassays to the risks from low doses typical of environmental exposures. Several other models have been proposed that fit the animal data equally well. Depending on the specific model used, the estimated risk at low doses can vary by several orders of magnitude.

3.1.3.1 Surrogate Variables

Problems can arise when a model uses information for a variable that is not exactly the same as the variable under study. For example, toxicity assessments of the dose-response relationship using rodents as surrogates to represent humans is very common. The uncertainty about the differences between humans and rodents (e.g., differences in how they metabolize the same chemical) could give rise to the false presumption that toxicity in rats equals toxicity in humans for a given chemical. As it turns out, this is the generally accepted presumption, even though any given chemical could induce very different responses in different species.[1]

In theory, one can minimize the problems introduced by surrogate variables by increasing the complexity of the equation that relates the modeled variable (risk) to the input variables. For example, if the dose-response model could be made to account for all the differences between rodents and humans, in terms of the toxic effects of chemical exposure, then the use of rodents as surrogates would no longer introduce potential error into the calculation. Unfortunately, the information often does not exist to allow this increased complexity of the model. In fact, this is usually the reason the surrogate must be used in the first place.

3.1.3.2 Excluded Variables

In general, models cannot include all of the factors that influence the output of interest (risk). There is an inherent trade-off between keeping a model manageably simple and yet including as many variables as possible. Analysts therefore take the chance that by making their models manageably simple, they will miss one or more important variables.

Assessment of the potential for error due to excluded variables can be difficult. If a simplified model is being used because the information is not available to develop more complex models, then quantitative estimates of model error may be impossible. If

information exists for a more complex model, then the output for the simpler model can be compared to the output from the more complex model.[1]

A variable also may be excluded merely because the variable is not thought to be important. This type of error is difficult to identify. If the model is developed using observed data, the importance of the excluded variable may not be apparent if none of the observed data indicate that the model is incomplete. In other words, if the observed data set is in some way limited, or not representative, it may give no indication that there is an additional variable that should be included.[1]

Consider a human health risk assessment in which there may be hundreds of potential input variables. Developing PDFs for each of these in a Monte Carlo-based risk assessment would be difficult and unnecessary. To limit the number of variables, the analyst must decide which factors to model probabilistically with input distribution. PDFs should be developed only for those input factors that are both *uncertain* and *influential* to the output risk value. A "sensitivity analysis" performed on the simulation results could be used to identify which input factors meet these criteria and the relative importance of the variables to each other. This presents a problem, however, since it requires defining the parameter in question as a PDF and performing the simulation (i.e., it does not help in the initial selection of a variable which should be modeled as a PDF at the start of the simulation). An approach to overcome this problem is to use "bounding" estimates of the PDFs for all input parameters that might be important and then perform the simulation and sensitivity analysis. PDFs for those input parameters shown to be important could then be refined; for non-influential parameters, point values could be used.

3.1.3.3 Extreme Events

Because models are simplified and combine representations of complicated processes, it is often difficult to determine how applicable they are to extreme events (i.e., unusual or aberrant conditions). In risk assessment, this is of particular concern if a distribution contains a "tail" that includes a small but significant portion of the population that may not be well characterized by the distribution. For practical purposes, ignoring possible extreme events is equivalent to building a model but excluding one or more variables.[1]

3.1.3.4 Incorrect Model Form

If the model involves multiple predictor (i.e., input) variables, additional uncertainty may be created through incorrectly modeling potential interactions among the variables. Failure to account for interdependence, or correlation, among variables is an example of this type of modeling problem. Correlation occurs when two or more variables vary in tandem; for example, extremely high levels of one variable are only seen with extremely low levels of another. In this situation, the variables are not statistically independent.

One potential means of addressing parameter correlation is through the use of correlated data distributions for receptor-related distributions.[3] An example of this type of correlation would prevent use of a large adult's body weight with a small adult's dermal surface area.

3.1.4 Understanding Uncertainty

Understanding the sources of uncertainty in estimating risk is critical to guiding additional, meaningful data collection or evaluation. Understanding the relative contribution of each source of uncertainty on the final risk estimate is also important in deciding when further data collection and analysis is not cost effective. If the source of uncertainty that dominates the final risk estimate for any given site is reducible and may result in a significant change in the decision, RPMs are encouraged to investigate means to reevaluate remedial decisions with new data and/or new tools. Inclusion of new data does not necessarily mean that the decision will radically change; however, such information can only foster more well-informed decisions and priority setting.

3.2 VARIABILITY

As noted previously, uncertainty and variability are usually mixed together in an environmental data set. Additional collection of data may reduce uncertainty. Variability, on the other hand, represents the inherent, natural heterogeneity of the population. Although further data collection may improve understanding of the variability in the population (and therefore may improve the ability to accurately incorporate variation into the risk estimate), it will not reduce the differences that exist in the population.

Any group of exposed or potentially exposed individuals will almost certainly display variation in most parameters. For instance, they will have different body weights, will be present at the exposure point for different periods of time, and will engage in different activities that promote or inhibit their exposure potential.

As an example, even if every potentially exposed individual in a population was weighed, giving complete information on the distribution of body weights, this would in no way change the fact that individuals in the population had different body weights (or, in fact, that each individual's body weight is subject to change over time). Similarly, if it was possible to perfectly characterize the levels of soil contamination throughout a site, this would not change the fact that "hot spots" have higher contaminant concentrations than less contaminated areas.

3.3 UNCERTAINTY AND VARIABILITY IN THE KEY STEPS OF RISK ASSESSMENT

As discussed in Section 2, deterministic risk assessments generally only involve qualitative estimates of uncertainty and variability. The inherent strength of probabilistic risk assessment techniques is that they provide a means to quantitatively incorporate and assess the impact of both uncertainty and variability on the final output (estimated risk to a human receptor).

3.3.1 Data Evaluation

As noted in Section 2, the data evaluation step in risk assessment involves examination and compilation of all data gathered for the site. The primary objective of this step in the risk assessment process is to develop a list of chemicals that will be quantitatively considered in the risk analysis, specifically by estimating the chemical-specific exposure-point concentrations. The primary source of variability in the data evaluation step arises from the temporal or spatial variation in the levels of the chemical contaminants in environmental media.

Because contaminant concentrations factor heavily into determining an individual's intake, it is imperative that the sampling data be representative of the environmental matrices (e.g., soil, groundwater, surface water) for the exposure pathways of concern. Often times, though, the analyst is forced to rely on data that have been generated for a completely different purpose, namely to identify the nature and extent of contamination.

Nature and extent sampling schemes are often biased and emphasize the sampling of expected areas of contamination (e.g., potential hot spots and potential release points). This biased sampling strategy is not designed to provide an accurate representation of the average contaminant levels at the site. This sampling methodology may tend to underestimate the variability within a site, and overestimate site-related concentrations.

Often, a sampling scheme designed to meet a specific objective (e.g., nature and extent determinations) will not necessarily meet the data needs of the risk assessor. As such, it is imperative that the risk assessor be initially included in development of work and sampling plans to characterize the site. Otherwise, the resulting data could have serious implications to data quality objectives of the risk assessment results.

Once sufficient chemical data are identified, current USEPA guidance stipulates that a single exposure-point concentration term (C-term) representative of the average site concentration be estimated.[4] This point estimate is then to be used to develop both the average and RME risk estimates. The USEPA Superfund Program has defined the C-term as the 95-percent upper confidence limit (UCL) of the arithmetic mean of the site characterization data. The USEPA further recommends that the maximum detected site concentration be used as the C-term in cases where sampling has been insufficient to calculate a 95-percent UCL or where the calculated 95-percent UCL exceeds the maximum detected site concentration (often the case when the sample size is small or when the proportion of nondetects is high and numerical values associated with the nondetects are greater than the maximum detected concentration). In either case, whether the C-term is the 95-percent UCL or the maximum detected concentration, this results in the use of a single-point estimate representing chemical contamination both spatially and temporally throughout the site.

Other authors who have completed a theoretical comparative analysis using a point estimate and a PDF to represent site contamination characteristics suggest that the single-point C-term may moderately overestimate the representative "average" site concentration of a given contaminant.[5] The impact of using a PDF to represent contaminant concentrations

rather than a single-point parameter estimator will vary considerably depending on the actual contaminant levels, the target environmental media, and the potentially complete exposure pathways used to estimate risk.

During the data evaluation process, natural or anthropogenic variability should not be confused with uncertainty. Specifically, contamination levels may vary either spatially or temporally due to anthropogenic or natural influences. For example, areas characterized by elevated contamination ("hot spots") may be attributable to past site activities.

In contrast to variability, several significant sources of uncertainty also exist that can complicate developing a point estimate or PDF which is representative of site contamination. For example, as previously discussed, sampling bias due to poor or inaccurate sampling plans, accidental contamination of samples during field sampling or laboratory analysis, and random error introduced during the sampling process are sources of uncertainty.[6]

Much of the potential uncertainty in the data evaluation process can be minimized through use of proper sampling techniques, well-designed sampling plans, and rigorous analytical procedures that minimize positive-detect biases.[6] Consequently, any proposed site sampling and analysis plans that will be used to collect data for use in risk and uncertainty analysis should clearly identify where uncertainty could be introduced into the end-use data set.

3.3.2 Exposure Assessment

The exposure assessment step involves estimating the types and magnitudes of chemical exposures to the potential receptors at a specific site. Potentially exposed populations and the pathways by which the exposure might occur are usually scoped out in a conceptual site model (CSM). The CSM is then used to select the mathematical models to describe the exposures for the site. There are a wide range of potential input parameters associated with this process which can loosely be grouped into two broad categories:

- Receptor parameters that describe the receptor and how the receptor interacts with the environment; and
- Environmental parameters that describe the source of exposure and the environmental media (site characterization data).

Receptor parameters are those which describe some physical characteristic of the individual. Body weight, inhalation rate, rate at which a chemical is excreted from the body, or other physical characteristics of how the receptor behaves, are examples of parameters that will vary among the individuals of a population. Environmental parameters, on the other hand, are those that describe some physical attribute of the environment that will introduce variability into the amount of contaminant that the individual potentially receives. Distribution of a contaminant throughout a site or throughout a soil-depth interval are examples of environmental parameters.

The "Exposure Factor Sourcebook"[7] and "Exposure Factor Handbook"[8] present numerous distributions for many of the common parameters used in the exposure assessment process. Both sources also discuss the research and data sets used to develop the distributions. RPMs are encouraged to use this information to the greatest extent possible since it increases the likelihood of regulatory acceptance of the approach. If more information is needed, there are other distributions presented in the technical literature which RPMs can use to supplement or question USEPA-recognized distributions. If a variable is identified for which a distribution is required, but for which one cannot be found in the technical literature, solicitation of expert judgment should be considered (see Section 3.4). USEPA's policy on Monte Carlo analysis pledges that more distributions will be developed in the future.[7,8]

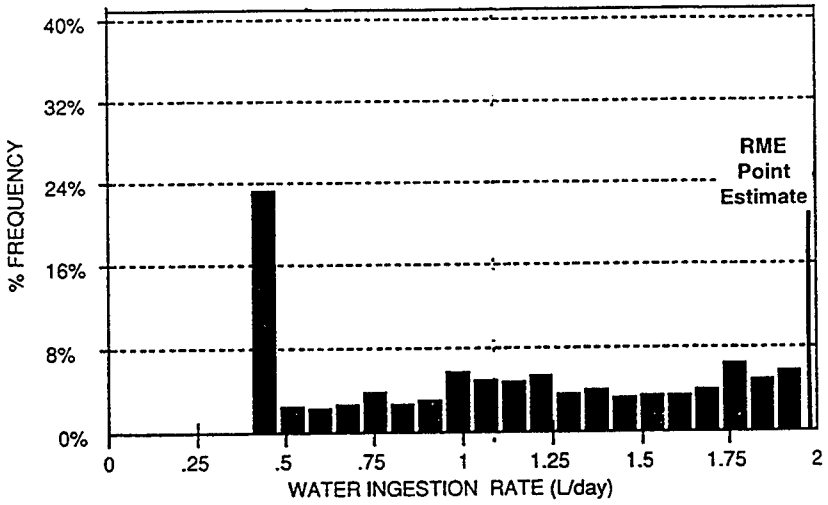
Common input parameters for which distributions have been developed are listed below. Examples of the distributions, recommended point estimates for calculation of the RME, and summary statistics for several of these parameters are presented in Figure 3.1. It should be noted that one would still need to determine the appropriate use of the parameter distributions and understand how they may be modified for site-specific conditions. The input parameters include:

- Adult Body Weight
- Child Body Weight
- Total Skin Surface Area
- Body-Part Specific Surface Area
- Inhalation Rates
- Exposure Duration – Adult Resident
- Exposure Duration – Child Resident
- Exposure Duration – Job Tenure
- Exposure Duration – Time/Activity Patterns
- Exposure Frequency – Showering
- Water Ingestion Rates
- Soil Ingestion Rates – Adult
- Soil Ingestion Rates – Children
- Fruit/Vegetable Ingestion Rates
- Fish Consumption Rates
- Soil Adherence Factors

Figure 3.1 Example Distributions for Exposure Parameters

Adult Water Ingestion Distribution

Cumulative Distribution
Median 1.13 L/day



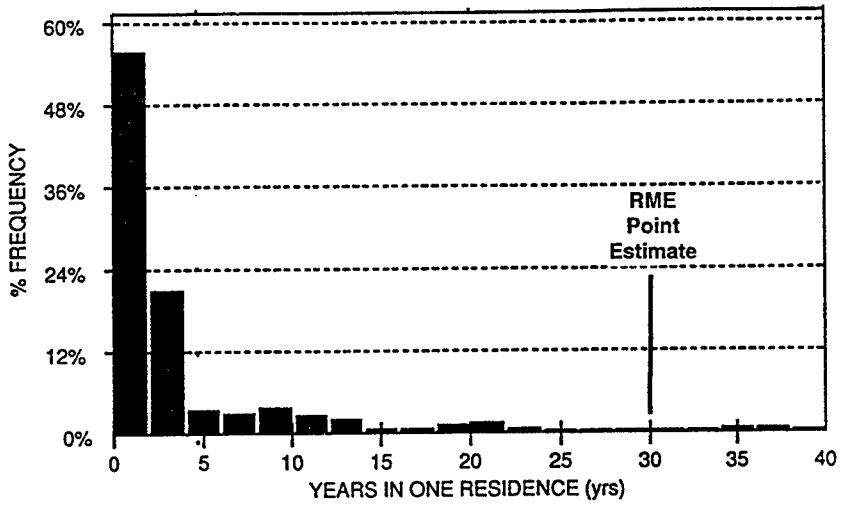
Selected Percentiles

Min	Max	19.2%	39.6%	60%	80%	100%
0.4	2	0.4	0.96	1.28	1.7	1.96

Source: Exposure Factors Handbook, 2-5
U.S. EPA, 1989a

Distribution of Years in One Residence - All Households

Cumulative Distribution

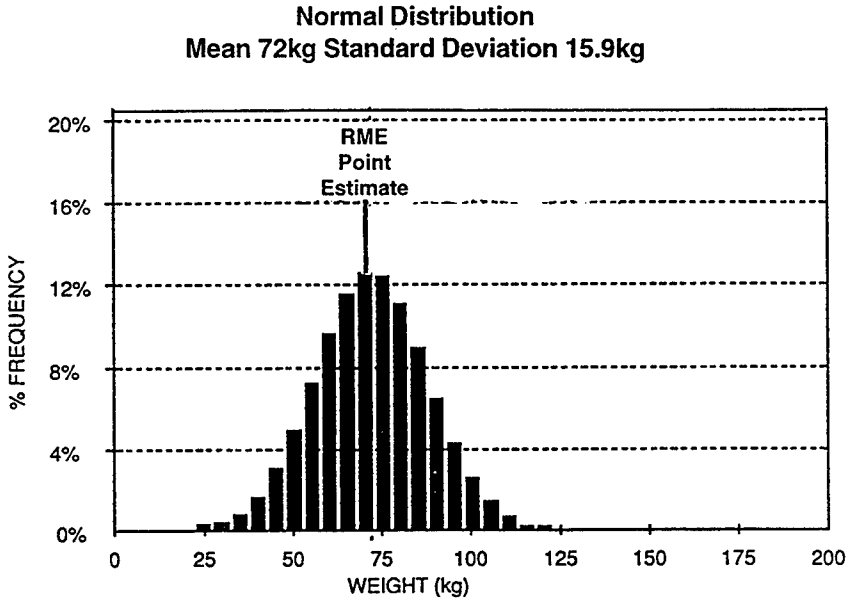


Selected Percentiles

Min	Max	25%	50%	75%	90%	95%
0	50	0.5	1.4	3.7	12.9	23.1

Source: Israeli and Nelson, 1992

Figure 3.1, cont.
Adult Body Weight Distribution - Both Sexes



Selected Percentiles

Mean	s.d.
72	15.9

Source: Brorby and Finley, 1993.

In addition to describing the inter-individual variability inherent in the population, it should also be noted that many of the distributions for the above-listed parameters may include uncertainty. For example, distributions may be based on models using small data sets Therefore, the distributions have a high potential for random error to be important in determining the distribution shape.

Several of the distributions are derived using mathematical models to relate the parameter of interest to the parameter actually measured, and thus incorporate potential problems with surrogacy. For example, information on current residency time was used to derive the distribution for total adult residency time.[9] Also, some parameters are inherently difficult to measure (e.g., soil adherence) and therefore measurement error may be important in determining the distribution shape.

A hypothetical example of how the exposure input parameters are incorporated into the risk assessment process, and comparison of point estimate-based risk estimates to probabilistically-based risk estimates is presented in Section 5.

3.3.3 Toxicity Assessment

The toxicity assessment in human health risk assessment generally requires an identification of whether the chemical causes an adverse effect and a dose-response evaluation. Potential adverse health effects include carcinogenic or noncarcinogenic effects. Generally, incorporating toxicity parameters described by a distribution into a probabilistic risk assessment is not pursued because it introduces a high level of complexity into the assessment. Rather, a point-value is selected for the toxicity value. It should be noted, however, that the level of complexity does not justify the use of a point value. Uncertainties in the toxicity values can be incorporated into the risk assessment without including the entire toxicity assessment. Experts in the field can perform uncertainty analysis for the dose-response and pharmacokinetic (also called toxicokinetic) models. The results can then be incorporated into a probabilistic risk assessment as distributions of toxicity values.

One of the goals of the uncertainty analysis is to improve the state of knowledge about key input parameters that contribute to the uncertainties in the risk estimates so that they can be reduced. It is possible that explicit inclusion of uncertainty in toxicity estimates may identify toxicity parameters as a key source of uncertainty (via sensitivity analysis) and justify additional research towards reducing the uncertainty.

3.3.3.1 Dose-Response Models

The dose-response evaluation step involves quantitatively evaluating the toxicity information for a specific compound and characterizing the relationship between the dose of the contaminant administered or absorbed, and the incidence of adverse health effects in the exposed population. From the dose-response information, toxicity values are derived and used in the risk characterization step to estimate the likelihood of adverse effects occurring in

potential receptors. Cancer slope factors (CSFs) are derived for carcinogenic effects, and reference doses (RfDs) are derived for noncarcinogenic effects.

Although the dose-response step may represent an important source of uncertainty and variability in the final risk estimate, CSFs and RfDs are conventionally calculated as point estimates. The USEPA defines the CSF as a "plausible upper-bound estimate of the probability of a response per unit intake of a chemical over a lifetime." The RfD is an estimate of the daily exposure that is unlikely to cause appreciable risk of adverse noncarcinogenic health effects during a lifetime.[10]

The use of a point estimate for the CSF or RfD does not reflect the expected variability in how individuals will respond to a given dose of a toxicant. Toxicodynamic variation is the different susceptibility of target tissues to toxic insult in different individuals. The variation in response by different individuals can be based on a number of toxicokinetic factors. Factors that would be expected to display inter-individual variation include: (1) rate at which a compound is absorbed from the gastrointestinal track into the blood stream, (2) rate at which compounds are detoxified, (3) rate at which non-toxic compounds are converted into toxic intermediates, and (4) rate at which compounds are excreted from the body.

The study and modeling of these four factors are encompassed by the field of pharmacokinetics. The specifics of a pharmacokinetic model will be dependent on the particular chemical or class of chemical. Information on a particular chemical, when available, can be found in the technical literature. An example is presented in the next section showing why use of a point value for a toxicity factor can cause significant problems in the risk assessment output.

3.3.3.2 Pharmacokinetic Models

The potential for variability among individuals has been concisely illustrated by a pharmacokinetic model developed for metabolism of the carcinogen 4-aminobiphenyl (ABP).[11] ABP was of interest because it has been recognized as causing cancer in both humans and a number of animal species. Also, the pharmacokinetics of ABP has been studied in animals.

The pharmacokinetic model developed for humans assumed lognormal distributions to describe several steps in the absorption, distribution, metabolism, and excretion of ABP. The model used one of the primary variability propagation techniques (i.e., Monte Carlo simulations) to predict the amount of ABP that would bind to deoxyribonucleic acid (DNA) in a particular theoretical individual. The results indicated a four-order-of-magnitude difference among the highest and lowest individuals, and a two order-of-magnitude difference between the 5th and 95th percentile of the population.[11]

The exact relationship between number of ABP molecules bound to DNA and the increased risk of developing cancer is not known. However, given the probable involvement of DNA lesions in the carcinogenicity of ABP, it is likely that large differences in

DNA-binding of ABP among individuals corresponds to large inter-individual differences in cancer susceptibility.

Adequate data are limited for most other chemicals to support analysis of this type. But clearly, accounting for the impact of this type of inter-individual variability could dramatically change the focus and conclusions of human health risk assessments.

3.3.4 Risk Characterization

The risk characterization process involves the combination of information from the data evaluation, the exposure assessment and the toxicity assessment to develop risk estimates. In general, for each exposure pathway for each receptor, compound-specific estimates of carcinogenic and noncarcinogenic risk are developed. These compound-specific risks are then summed to provide a pathway-specific total carcinogenic and noncarcinogenic risk for each receptor. When appropriate, risks from different pathways are also summed to give a total carcinogen and noncarcinogenic risk for each receptor.

In the probabilistic risk assessment paradigm, Monte Carlo techniques are used to combine the exposure parameter distributions along with any single-point values selected (e.g., toxicity factors). The result is a distribution of potential risks that reflects the information and assumptions concerning the exposed population. See Sections 4 and 5 for additional information.

The spread of this distribution will be determined, in part, by the amount of variation found in the various parameters used in the risk calculation and, in part, by uncertainties in the parameters. As noted previously, analysis can be performed to determine which input variables have the greatest impact on the resulting risk distribution. Techniques have also been developed to distinguish the amount of spread in the risk distribution that is due to variation, and the amount that is due to uncertainty.

The potential for correlation among input variables to influence the risk estimates and uncertainty analysis must also be considered. Two parameters that vary together are said to be correlated. There are specific situations in which correlation could be important, and would need to be accounted for in the risk assessment. The impact will be most pronounced when there is strong correlation among several sensitive parameters. Sensitive parameters are those that have the most influence on the risk distribution. Correlation also becomes more important when one is concerned with extreme values (i.e., with the tails of a distribution).[12] Section 4 discusses methods to explore correlation and how to incorporate correlation into the risk estimation process.

3.4 ELICITATION OF EXPERT JUDGMENT

The model used to describe the pharmacokinetics of ABP carcinogens (discussed above) incorporated parameter distributions for which there were limited data. This raises the issue of what action is to be taken if a parameter for which there is not an established distribution,

or sufficient information to develop a distribution, is identified as important to the risk calculation. A related situation arises when the site-specific population is thought to vary in some important aspect from the population upon which an established distribution is based. In these and similar situations, it may be necessary to estimate uncertainty and variability in one or more input parameters in the model.

An expert is someone who (1) has training and experience in the subject area resulting in extensive knowledge of the field, (2) has access to relevant information, (3) has an ability to process and effectively use the information, and (4) is recognized by his or her peers or those conducting the study as qualified to provide judgments about assumptions, models, and model parameters at the level of detail required.[13] In performing a probabilistic risk assessment, an expert might be called upon to develop key parameters about which insufficient data are available as subjective PDFs.

Generally, a distinction is made between informal and formal solicitation of expert judgment. Informal solicitation may include self-assessment, casual solicitation from an expert, brainstorming, and group discussions by staff or available experts.

If the above informal criteria do not apply (i.e., if the potential impact is large or the results are likely to undergo intense scrutiny) the formal solicitation process should be pursued. Formal solicitation methods have a predetermined structure for selecting and training experts and for eliciting, processing, and documenting expert judgments and their rationales. Some of the advantages of using a formal approach include:

- Improved quality of expert judgments,
- Reduced likelihood of critical mistakes,
- Improved accountability,
- Improved consistency of procedures,
- Enhanced communication, and
- Reduced chance of unexpected delay.[13]

Formal methods of eliciting expert judgments are more costly and time consuming than informal methods.

A process for formal solicitation of expert judgment has been developed based on methodology developed by the Nuclear Regulatory Committee to obtain expert opinion on uncertainty of off-site risk for nuclear power plants. The process consists of ten steps designed to elicit unbiased judgment from a panel of experts.

The process is accomplished through preparation of complete background information on the issue in question, and presentation of this information to the experts. The experts, either separately or in groups, will then analyze and discuss the issue, and render a judgment. If separate judgments are elicited from each expert, the judgments should be consolidated. All

judgments should be reviewed and the results should be communicated to risk management decision-makers and other stakeholders. A more detailed discussion of this subject is made in the NCRP Commentary No. 14 "A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination."[13]

1 Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers*. Center for Risk Management, Resources for the Future, Washington, DC.

2 Henrion, M. and Fischhoff, B. 1986. Assessing Uncertainty in Physical Constants. *American Journal of Physics*. 54(9):791-798.

3 Finley, B., Proctor, D., Scott, P.K., Harrington, N., Paustenbach, D.J., and Price, P. 1994. Recommended Distributions for Exposure Factors Frequently Used in Health Risk Assessment. *Risk Analysis*. 14(4):533-553.

4 U.S. Environmental Protection Agency, Region VIII. 1994. *Calculating the Concentration Term for Risk Assessment: Use of One "C" Term to Estimate Lower Average and Upper RME Risk Range, (Technical Section)*. Hazardous Waste Management Division, Superfund Management Branch Technical Guidance, Denver, CO. RA-02. January.

5 Thompson, K.M., Burmaster, D.E., and Crouch, E.A.C. 1992. Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments. *Risk Analysis*. 12(1):53-63.

6 U.S. Environmental Protection Agency. 1990. *A Rationale for the Assessment of Errors in the Sampling of Soils*. EPA600-04-90-013.

7 U.S. Environmental Protection Agency. 1997. *Guiding Principles for Monte Carlo Analysis*. Risk Assessment Forum, Washington, DC. March.

8 U.S. Environmental Protection Agency. 1996. *Summary Report for the Workshop on Monte Carlo Analysis*. Risk Assessment Forum, Office of Research and Development. EPA/630/R-96/010. September.

9 Israeli, M. and Nelson, C.B. 1992. Distribution and Expected Time of Residence for U.S. Households. *Risk Analysis*. 12(1):65-72.

10 U.S. Environmental Protection Agency. 1989. *Risk Assessment Guidance for Superfund – Volume I: Human Health Evaluation Manual (Part A)*. Office of Emergency and Remedial Response, Washington DC. EPA/540/1-89/002. December.

11 Bois, F.Y., Krowech, G., and Zeise, L. 1995. Modeling Human Interindividual Variability in Metabolism and Risk: The Example of 4-Aminobiphenyl. *Risk Analysis*. 15(2):205-213.

12 Frey, H.C., Burmaster, D.E., Hoffman, F.O., McKone, T.E., and Price, P.S. 1995. *Quantitative Techniques for Analysis of Variability and Uncertainty in Exposure and Risk Assessment*. Workshop Notebook. Society for Risk Analysis Annual Meeting, Honolulu, Hawaii. December 3, 1995.

13 National Council on Radiation Programs. 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessment Related to Environmental Contamination*. NCRP Commentary No. 14, Bethesda, MD.

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 4

PRINCIPAL METHODS OF PROBABILISTIC ANALYSIS

The purpose of this section is to provide Air Force RPMs with an introduction to the statistics necessary to understand the development, selection, and use of probability distributions in probabilistic human health risk assessments. The RPM should have a fairly strong background concerning relevant statistical methods in order to understand the analysis incorporated into a risk assessment and to be able to interpret the results.

This section presents several recommended statistical methods for characterizing information, selecting and/or developing input distributions, and interpreting the results of a probabilistic risk assessment. These include:

- Types of distributions commonly used to model parameters in human health risk assessments;
- Methods to establish distributions from data;
- Graphical methods to assess how well data fit a theoretical distribution;
- Traditional "goodness-of-fit" tests;
- Summary statistics used to describe distributions;
- Issues related to developing distributions for site characterization; and
- Methods for assessing uncertainty and variability in PDFs and how this propagates through the risk assessment.

4.1 PROBABILITY FUNCTIONS

A probability distribution is the set of outcomes of a random variable and their corresponding probabilities. Two commonly used functions to mathematically describe the probability distribution of a continuous random variable are the PDF and the cumulative distribution function (CDF). The PDF describes the probabilities of occurrence of particular outcomes. For example, a PDF could be used to describe the range of body weights in an adult population and their relative likelihood of occurrence. The CDF gives the cumulative probability of all outcomes at or below a specific value. For example, from a CDF, one could determine the probability of cancer risk due to exposure to chemicals at a hazardous waste site being less than a certain value, such as less than an acceptable cancer risk level of one in ten thousand ($1x10^{-4}$).

PDFs and CDFs are used to incorporate variability and uncertainty into the probabilistic risk assessment. Either function is a valid way of mathematically specifying the statistical distribution in probabilistic techniques. In Monte Carlo simulations, PDFs are used for specific input variables that are combined with appropriate point estimates to produce an output distribution for risk.

Monte Carlo simulations are very sensitive to the shape of the input distributions. Therefore, to ensure as accurate an estimate of risk as possible, it is important to have input PDFs that accurately represent an input variable. As discussed below, statistical methods are used to summarize and characterize the data to allow determination of a distribution which "best" represents that data. Statistical methods can also be used to determine how well data fit a theoretical distribution.

Sensitivity analysis is used to measure how the uncertainty and inherent variability of the PDFs used as input variables propagate through the risk equations to the risk output. Depending on the specifics of the risk calculation, different input parameters will have different levels of impact on the output parameter (risk). Determining which inputs "drive" the output is critical in determining which parameters should be modeled as distributions and how important it is to have an accurate distribution.

4.2 DISTRIBUTION CHARACTERIZATION

Distributions are a set of numbers collected from a well-defined universe of possible measurements arising from a property or relationship under study. Understanding what a distribution is and selecting appropriate input distributions is one of the most crucial aspects of Monte Carlo simulation since it is very sensitive to the shape of the input distributions and their interaction in the risk calculation. However, there is not universal agreement on one approach for developing these distributions or even the types of theoretical distributions that should be considered.

4.2.1 Distribution Types

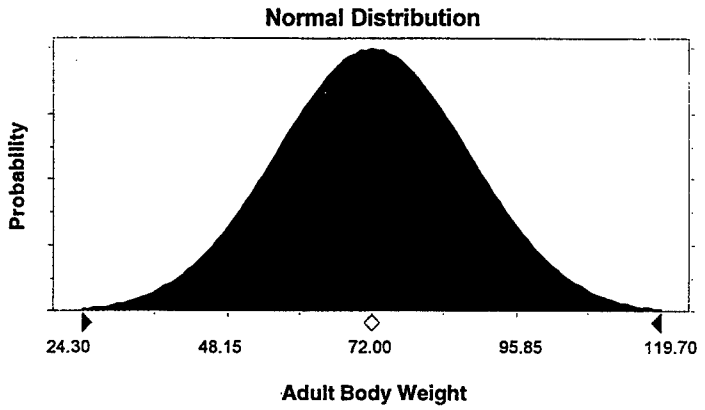
Probability distributions can be assigned to data via graphical interpretation (the preferred method) or formal statistical tests. Once sample data are categorized based on all available information, inferences regarding variability inherent in the populations (e.g., body weights, ingestion rates, etc.) can be approximated and input into the risk calculation.

There are a variety of theoretical distributions used to represent populations and data sets. Use of these distributions is an appropriate way to represent the uncertainty and/or variability in the population. The distributions most commonly seen in human health risk assessments are the normal, lognormal, triangular, beta, uniform, and empirical distributions. Examples of some of these distributions are described below.

4.2.1.1 Normal Distribution

The normal distribution is frequently used to describe natural populations and phenomena and is described by the well-known "bell-shaped" curve. An example is shown below. The normal curve is a convenient PDF because many physical measurements (such as depth to groundwater) as well as "additive processes" have distributions that are bell shaped. The normal distribution is also applied in inferential statistics. For example, the normal distribution can sometimes be used to describe the distribution of the mean of a population.[1]

The entire shape of a normal curve can be described by two summary statistics: the mean and the variance (see Section 4.4.1).

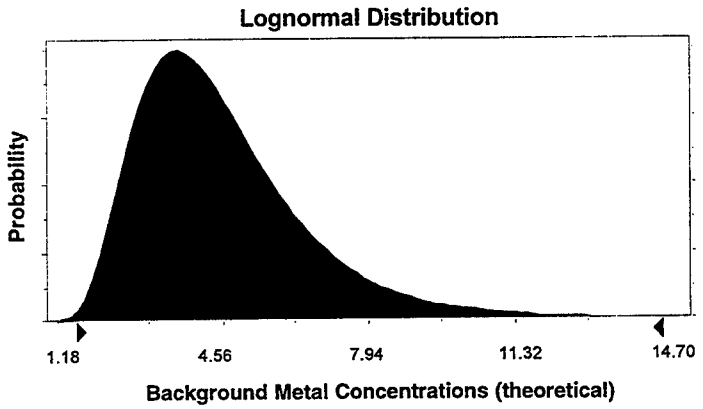


4.2.1.2 Lognormal Distribution

The lognormal distribution is similar to the normal distribution, except that the log-transformed values (i.e., logarithm of the values) are normally distributed. As shown in the example below, the shape of the distribution of the untransformed values is skewed (i.e., tapered) to the right. In general, "multiplicative" processes can follow a lognormal distribution, (e.g., laboratory analytical error). The lognormal distribution is often also used to represent natural phenomenon (e.g., concentrations of a contaminant in soil).

There are three ways to specify the statistical parameters of the lognormal distribution: (1) the mean and the variance, (2) the mean and variance in the log scale, and (3) the geometric mean and geometric variance. It is important to describe which set of these parameters are being used when the data are reported. There are also a variety of methods appropriate for estimating these parameters depending on different statistical conditions of the data set. For example, under some circumstances, the sample average is a good estimate of the true, but unknown population mean for a lognormal distribution. Evaluating the tail of the

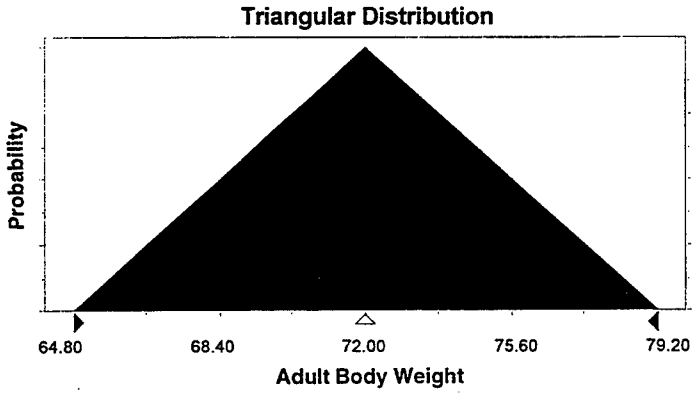
distribution is often important in environmental applications, such as when one is interested in the characteristics of the high end of concentrations of a contaminant.



4.2.1.3 Triangular Distribution

As its name suggests, the triangular distribution has a triangular shape. Often this distribution is used to represent natural populations or phenomenon which are not well characterized. The triangular distribution tends to overestimate the portion of the distribution found in the tails, for a population which is actually normally distributed. Thus, it is often used as a "conservative" estimator (although a beta distribution may be more appropriate in some cases).

The triangular distribution can be fully described by its minimum, maximum, and most likely values. The distribution is bounded by its maximum and minimum values.



4.2.1.4 Empirical (Cumulative) Distribution

Monte Carlo methods allow relatively easy use of empirically derived distributions. These distributions are often based on site-specific data which do not "fit" any of the standard theoretical distributions. The shape of each distribution is different. Usually these distributions will be described by listing percentiles of the population for which a given percent of the population is below a specified value.[2,3]

4.3 DERIVING DISTRIBUTIONS

There are a number of ways to include a set of data, or a population parameter, as a distribution in the risk assessment. Many parameters which describe natural populations, body weight, or inhalation rate for example, have already been modeled, and distributions are published in the technical literature or in a number of sources which have compiled this information. If however, the distribution will be applied to site-specific data, such as sampling results, or if site-specific conditions indicate that previously published distributions are not applicable, it will be necessary to determine if the site-specific data follow one of the theoretical distributions (e.g., a normal distribution), and if not, to develop a distribution which accurately represents the data.

4.3.1 Standard Data Distributions

Several authors in the literature have called for "standard" data distributions for input variables that are not significantly influenced by site-specific conditions. The goal is to use existing knowledge to support probabilistic descriptions of exposure variables. This will enable the development of standard distributions wherever possible and appropriate. Finley has proposed several distributions that can be considered standard for most settings.[4] Examples of standard distributions include residency time and body weight.

4.3.2 Deriving Distributions from Adequate Environmental Data

When standard data distributions are not applicable to an exposure scenario at a site, classical statistical methods (e.g., measures of central tendency, skewness, and precision) should be used to derive an appropriate input distribution from the environmental sampling results. (Descriptions of classical statistical methods are described in Section 4.5.) In particular, a graphical analysis of environmental data should be performed to determine if the data adequately fit a normal or lognormal distribution.

4.3.3 Deriving Distributions with Lack of Knowledge

Two approaches are recommended to deriving a statistical distribution for an input variable when little information is known. The first method is referred to as an informal approach to deriving defensible distributions based on a priori knowledge of the nature of the stochastic (i.e., random) variable. The second method is a formal approach to eliciting expert judgment. Both methods are based on the scientific method. A more detailed discussion of the two approaches is presented in Section 3.4.

Determining data distributions (such as uniform or triangular) on the basis of "not knowing very much" or that it is a "conservative" approach is not recommended.[5] These arguments are not statistically justifiable and often reflect inadequate research into the given input variable. The uniform and triangular distributions make assumptions that often go overlooked in their applications. However, they are sometimes acceptable for "data poor" situations. In general, selection of PDFs should follow a careful process of formulating a more "realistic" distribution from the current state of knowledge coupled with a sensitivity analysis (see Sensitivity Analysis Methods, Section 4.7). The beta distribution is recommended for developing "customized" distributions because its shape is very flexible and can assume a wide variety of forms by adjusting its statistical parameters.[6] The beta distribution is also bounded by its maximum and minimum values.

4.4 GRAPHICAL ANALYSIS

Statistical tools can be used to assess how well a set of data is represented by a particular distribution. It is recommended that graphical analyses be performed because the graphical techniques are often superior to more traditional statistical techniques in assessing a distribution for use in a probabilistic risk assessment. A "picture" of the data is simple to prepare yet can be more informative than a series of statistical computations.

Graphical displays provide a means for determining the distribution of the data, identifying outliers, and selecting appropriate statistical methods and tests. This process is often referred to as exploratory data analysis (EDA).

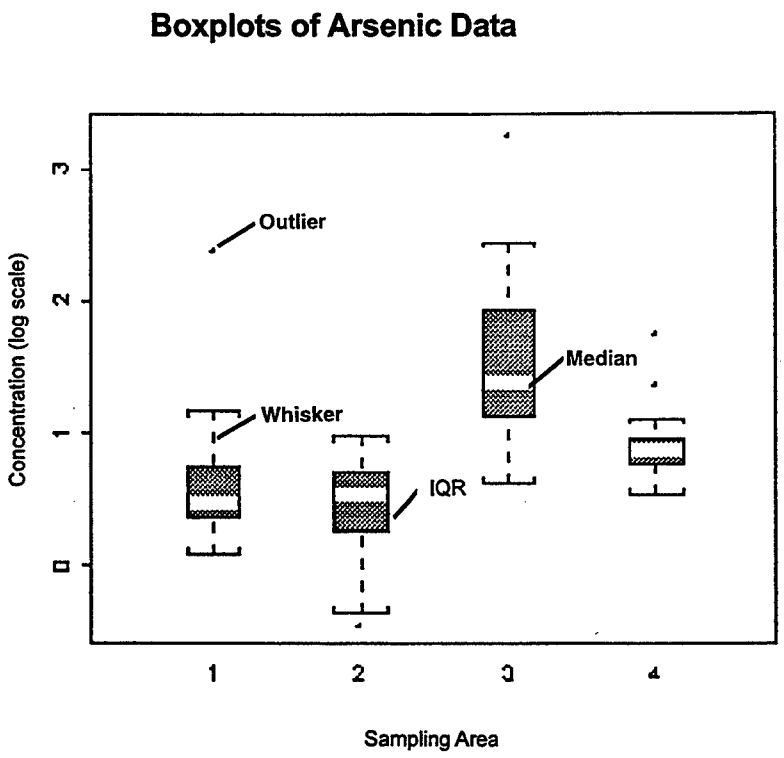
Graphs also provide a more complete picture of the data and convey information far beyond that of summary statistics (see Section 4.5.1 for a discussion of summary statistics). They are an invaluable tool for understanding the statistical characteristics of the data and presenting results.

Four types of graphs are presented in this discussion to describe the distribution of data: histograms, boxplots, quantile-quantile (Q-Q) plots (e.g., normal probability plots), and density estimation plots that describe the empirical PDFs. These graphs should be evaluated together to determine if a data set adequately follows a theoretical distribution (e.g., the normal distribution).

4.4.1 Boxplots

The boxplot is a very useful tool that gives a general overview of the data, regardless of its distribution. Boxplots show the location, spread, skewness, tail length, and outlying data points of the data in a compact form. It consists of a center line as representing the median (i.e., 50th percentile) of the data splitting a rectangle defined lengthwise by the 25th and 75th percentiles. The length of the box is the difference between the 25th and 75th percentiles and is called the interquartile range (IQR). "Whiskers" are drawn extending outside the box to show the tails of the distribution. Potential outliers are plotted as points beyond the whiskers.

Because the median and IQR are resistant measures of the data, the boxplot is particularly attractive for exploratory data analysis of environmental data.

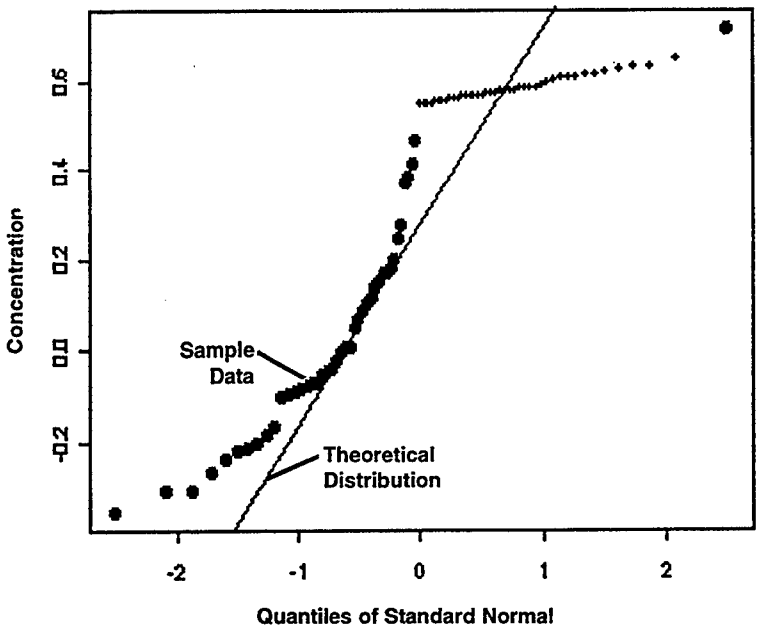


4.4.2 Q-Q Plots

The Q-Q plot is one of the best means for assessing if the data are normally or lognormally distributed. The Q-Q plot portrays the quantiles (percentiles divided by 100) of the sample data against the quantiles of another data set or theoretical distribution (e.g., normal distribution). By comparing the data to a theoretical distribution with a straight line, departures from the distribution are more easily perceived.[7] The diagonal line indicates a perfect fit.

The lognormal Q-Q plot shown below has been modified for environmental data. The points plotted as a "plus" indicate the result was a nondetect (in this example, it is clear how elevated detection limits affect the distribution of the sample). The extent, pattern, and locations of nondetects can be visually assessed. If the detected values appear to be approximately normally distributed, and the main departures from the line are nondetects, a normal distribution assumption may be reasonable and appropriate surrogate values that fit the distribution can be substituted for the nondetect results.[8,9]

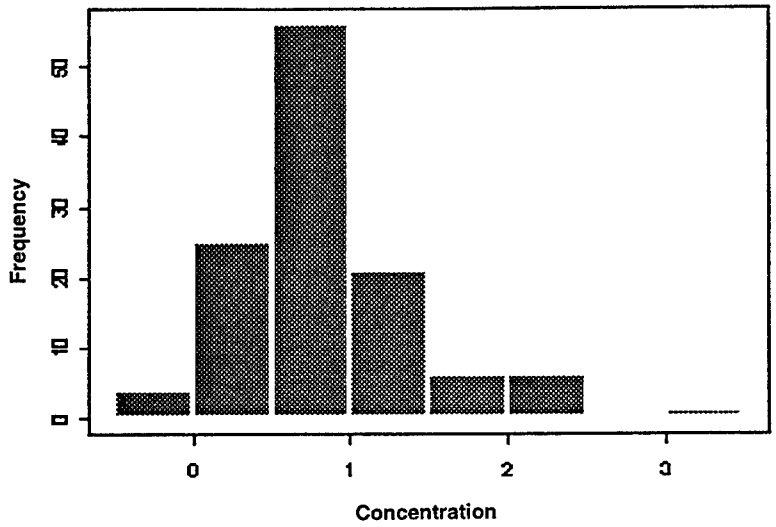
Lognormal Q-Q Plot of Thallium



4.4.3 Histograms

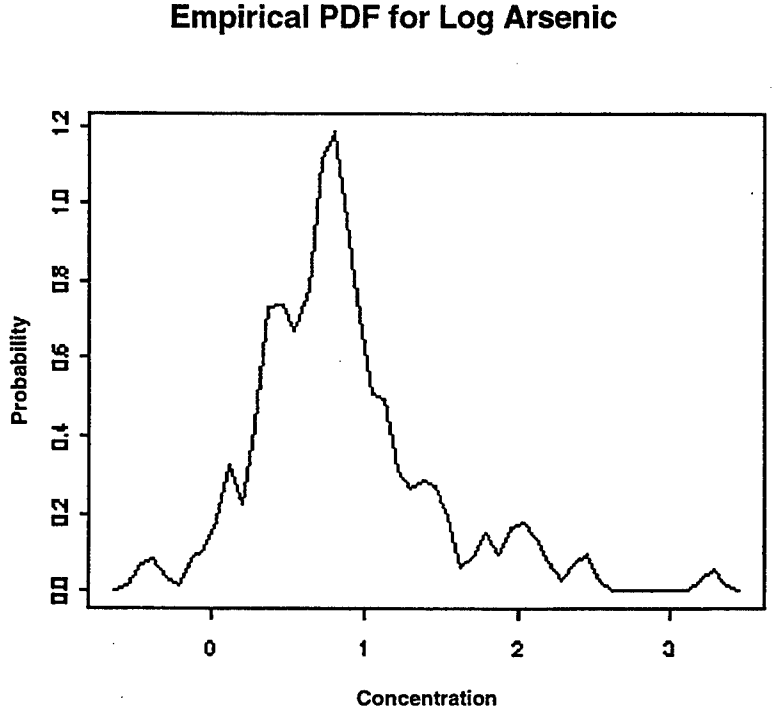
Histograms are used to display the general shape of the data distribution. A histogram breaks the range of values of a variable into intervals and displays the count (or percent) of the observations that fall into each interval. They approximate the shape of the PDF, but have one deficiency: their visual impression depends on the number of categories (i.e., vertical bars) selected for the plot. Varying the number of categories can have a remarkably large effect on the shape of the distribution.[10]

Histogram of Log Arsenic



4.4.4 Empirical Probability Density Functions

The density plot is a smoothed estimate of the distribution shape and does not have the category-width problems of a histogram. This "smoothing" technique can give a much clearer view of the distribution, and attempts to compromise between smoothing insignificant "bumps" in the data while not obscuring real peaks.[10]



4.4.5 Graphs vs. Formal Normality Tests

Formal statistical tests of normality alone are not recommended for assessing the distribution of the data. Unlike graphical analyses, these tests generally do not distinguish between the causes of departures from normality (e.g., substitutions of one-half the reporting limit for nondetects) and, more importantly, do not provide information on the pattern of the departures. If, however, a formal normality test is being used to supplement a graphical analysis, the Shapiro-Wilk test is recommended.[11] The Shapiro-Wilk test can be augmented with the Shapiro-Francia test for testing normality when more than fifty samples are being evaluated.[11] Alternative "goodness-of-fit" tests include the chi-square,[6] D'Agostino,[12] and Kologorov-Smirnov.[13]

4.5 CLASSICAL STATISTICAL METHODS

Classical statistical methods refers to traditional statistical theory and basic assumptions about a set of data (e.g., independent data points, random sampling, etc.). The assumptions which underlie classical statistical methods should, to the extent possible, be the goal of the risk assessor and the RPM. These assumptions include:

- Data are collected using random sampling techniques;
- Data represent the area or process being evaluated;
- Spatial and temporal dependencies have been accounted for or are not significant;
- Field sampling techniques and laboratory procedures are the same for data combined into the same group or for groups of data being compared; and
- Data are excluded only if they were shown to be in error and not solely on the basis of statistical outlier tests.

For environmental sampling data, it is often the case that many of the assumptions listed above do not apply. Commonly, the analyst cannot change or improve the sampling plan or results, but must work with the available data and acknowledge the limitations (e.g., small sample size, non-random methods of sample collection). Several classical statistical parameters, and their assumptions and limitations, are discussed below.

4.5.1 Summary Statistics

Summary statistics supply important information pertaining to a data set. They provide measures of central tendency (e.g., mean, median, and mode), skewness (e.g., coefficient of variation), and precision (e.g., variance, standard deviation, and percentiles) that are useful in describing the shape of a PDF. For many of the common distributions, the shape of the entire curve can be described with only a few parameters. For those data that are described by a parametric distribution, all of the information about the data can be conveyed simply by giving the distribution type and two or three key parameters. For example, if one wanted to model body weight with a distribution, and that distribution were normally distributed, the mean and standard deviation of the distribution would be used in the risk calculation. Summary statistics are required for use in the commercially available software packages which actually perform the Monte Carlo analysis. They are also used to describe PDFs in the technical literature, and in documents that summarize current information on exposure variables.[2,3] Therefore an understanding of these statistics is necessary to interpret the information presented. Commonly used summary statistics are discussed below.

It is first helpful to create a table of basic summary statistics (e.g., mean, median, standard deviation, etc.) for each data set to be used for the probabilistic analysis. Table 4.1 provides an example of the summary statistics for a hypothetical environment data set.

Table 4.1 Example Summary Statistics by Compound

Compound	n	% nd	range of SQLs	min	max	mean	median	s	MAD	CV
Aluminum	119	0	NA - NA	2920	29900	9790	8420	4980	3230	0.509
Antimony	55	0	NA - NA	1.54	6.46	2.74	2.11	1.26	0.756	0.46
Arsenic	119	0	NA - NA	0.642	26.3	2.94	2.18	2.9	0.964	0.986
Barium	119	0	NA - NA	13.9	252	65.3	58.5	34	17.5	0.521
Beryllium	119	0	NA - NA	0.171	1.88	0.482	0.43	0.249	0.141	0.517
Cadmium	119	16.8	0.408 - 0.443	0.0417	0.668	0.184	0.193	0.109	0.083	0.594
Calcium	119	0	NA - NA	1310	169000	12800	4810	25100	3450	1.96
Chromium	119	0	NA - NA	1.61	63.3	9.94	6.65	8.62	3.08	0.867
Cobalt	119	0	NA - NA	0.814	13.9	4.88	4.14	2.79	1.72	0.572
Copper	119	0	NA - NA	1.77	34.8	10.4	7.54	6.83	3.34	0.66
Chromium VI	59	96.6	0.218 - 1	0.109	0.62	NA	NA	NA	NA	NA
Iron	119	0	NA - NA	3710	42100	12400	11400	5600	3320	0.453
Lead	119	0	NA - NA	0.988	9.49	3.83	3.72	1.41	1.26	0.369
Magnesium	119	0	NA - NA	1140	13900	3850	3090	2380	1350	0.619
Manganese	118	0	NA - NA	80.3	402	203	200	66.2	65.2	0.326
Mercury	119	95.8	0.169 - 0.445	0.0587	0.39	NA	NA	NA	NA	NA
Molybdenum	15	0	NA - NA	1.15	3.5	1.87	1.66	0.777	0.593	0.416
Nickel	119	0	NA - NA	1.34	35.9	8.24	7.1	5.94	3.65	0.721
Nitrate/Nitrite	119	6.7	0.228 - 0.458	0.114	10.8	1.27	0.843	1.56	0.589	1.23
Phosphorus	119	0	NA - NA	110	1320	329	303	163	136	0.495
Potassium	119	0	NA - NA	716	11700	2870	2200	2000	1200	0.697
Sodium	119	0	NA - NA	46.9	4730	600	247	902	166	1.5
Thallium	83	49.4	3.47 - 3.83	0.702	2.05	NA	NA	NA	NA	NA
Vanadium	119	0	NA - NA	5.98	166	25.6	18.9	22	6.97	0.856
Zinc	119	0	NA - NA	15.6	79.2	29.2	26.4	11.3	7.41	0.385

Notes:

n = Number of samples.

%nd = Percentage of samples reported as nondetects.

range of SQLs (sample quantitation limits) = Range of reporting limits.

min = Minimum detected value.

max = Maximum detected value.

s = Standard deviation.

MAD = Median absolute deviation.

CV = Coefficient of variation.

* Per USEPA guidance,[20] one-half the SQL used to calculate summary statistics.

4.5.1.1 Measures of Location

The **arithmetic mean** (or average) is the classical measure of location of a data set. The mean is the sum of all the observations divided by the number of observations. It can be very sensitive to the magnitude of a small number of extreme points. More often the median is preferred for cases with outlying observations. The median is the 50th percentile of the ordered values; half the data points are above the median and half are below. The median is "resistant" to outlying observations because it is based only on the ranks of the data and not on their magnitude. The two statistics can be compared for each compound to help identify possible outliers and skewed data. When the data are symmetric, the mean and median will be equal.

4.5.1.2 Measures of Spread

Measures of spread indicate how much variability is in a data set or distribution. The **variance** is the classical measure of variability. However, the **standard deviation**, is more commonly reported because it has the same units as the original data or variable – it is the square root of the variance, which is in squared units. The variance is the average of the squared distance of each point from the mean. It is important to note that both of these measures of variability are very sensitive to extreme values or outliers.

Alternative measures of dispersion are available that are less sensitive to outlying data. Two examples are the IQR (Section 4.4.1) and the median absolute deviation (MAD). These are called "robust" estimates of variability because they are "resistant" to the influence of outliers. These statistics would be appropriate to use if outliers were significantly inflating the traditional estimates of variability, and this influential effect was not desired.

The **coefficient of variation** (CV) is another useful measure of variability. It equals the ratio of the standard deviation to the mean of a data set. It is thought of as a "relative" standard deviation, because it allows one to compare the variability between data sets. It is also sometimes used as a measure of skewness [14] of a data set that has a lower bound of zero (i.e., a data set with no negative values such as chemical concentration data) (Section 4.5.1.3). Generally, for these data, a CV value greater than 0.33 indicates that the data are sufficiently skewed to affect the estimates of the mean and standard deviation,[15] and a CV greater than 1.0 indicates a highly skewed data set.

4.5.1.3 Measures of Skewness

Skewness indicates asymmetry about the center of a data set or statistical distribution. The distribution of chemical concentration results is often positively skewed or skewed to the right (i.e., a longer right tail). The lognormal distribution is often chosen to model this characteristic of the data.

Similar to the variance, skewness is defined as the average of the cubed deviations from the mean. The ratio of this value divided by the standard deviation cubed is called the **coefficient of skewness** and is unitless value ranging from zero to one.

4.5.1.4 Percentiles

Percentiles, or quantiles, simply represent a percentage of the data set above or below some specified point. For example, if one was to calculate a 95th percentile, that value is equal to 95 percent of the sample points contained in the data set. As presented above, the median is a special case of a percentile or quartile, with half of the sample points above and the other below this value. Quartiles (a form of a percentile) represent the 25th, 50th, and 75th percentiles.

4.5.1.5 Outliers

Outliers are individual observations far removed from the pattern set by the majority of the data. Outliers are mainly due to gross errors like transcription errors, laboratory mistakes, etc., and legitimate extreme observations.[16] Outliers are usually influential observations, that when excluded from the data set, cause significant changes in summary statistics, confidence intervals, and even the outcome of a statistical test.

A careful evaluation of outlier influence is always warranted. For example, if outliers are related to the data analyses (e.g., aberrantly high reporting limits for nondetect results, also known as censored data) they can be excluded from the data set. Three main options are available to evaluate outliers: (1) allow the outliers to drive the statistical analysis, (2) exclude the outliers and allow the remaining data to drive the analysis, or (3) moderate the influence of the outliers with the use of specialized statistical methods designed for this purpose. Exclusion of outliers that cannot be explained as gross error is not appropriate.

Specialized methods to evaluate outliers include the "nonparametric" and "robust" statistical methods. Nonparametric methods are based on the ranks of the data, not their magnitude. Robust methods refer to a family of non-traditional statistical procedures specifically designed for censored data. When the proportion of nondetects is equal to or greater than 20 percent, robust methods should be considered. There are many robust statistical methods available which can be evaluated for applicability to a data set when the proportion of nondetects is high. For example, the "bootstrap" method is a robust method appropriate for handling nondetects when the proportion is greater than 50 percent. The following references provide detailed descriptions on the use of several robust methods: Gibbons;[12] Gilliom and Helsel;[17] Helsel and Gilliom;[18] Helsel and Cohn;[9] and Haas and Scheff.[19]

4.5.1.6 Sample Size

Sample size is an important attribute in assessing an environmental data set. As the sample size approaches the true population size, sample statistics performed on the data become more accurate (i.e., closer to the true population values). As an example, if one was to sample 10 percent of the possible surface soils at hazardous waste site for a given contaminant and calculate an average concentration, this value would be closer to the absolute

(but unknown) true concentration than would be the average calculated from samples collected from 5 percent of the possible surface soils. Although not feasible, if one were to sample all of the surface soil, the sample statistic would equal the true concentration.

Likewise, with all else being equal, larger sample sizes yield more "powerful" statistical tests. The power of a test refers to its ability to find a statistically significant result when one truly exists. The sample size necessary to yield a more powerful result will depend on the statistical method and the characteristics of the data set (e.g., variability, percent nondetects).

4.5.1.7 Nondetect Results

High percentages of **nondetects** in environmental sample analyses (also referred to as censored data) complicates estimates of summary statistics (e.g., arithmetic mean, median, standard deviation, coefficient of variation, etc.) and the characterization of distributions. There is considerable controversy regarding the appropriate methods for incorporating censored data. Different methods are appropriate depending on the percentage of nondetects, the pattern of nondetects in the data distribution, and whether or not multiple laboratory reporting limits exist. The issue of how to handle nondetects is complex and beyond the scope of this handbook. Risk assessors should work with the appropriate regulatory agency and reference documents for guidance.[20,21]

4.5.2 Estimators and Population Parameters

As previously discussed, the analyst does not generally have access to an entire population and must rely on a sample to draw *inferences* about the true but unknown population characteristic. This process uses "estimates" of the unknown population parameters (e.g., mean, standard deviation, etc.) derived from the sample data. Obviously, it is important to estimate the "true" value as accurately as possible based on the data available. Statistical techniques like confidence intervals (discussed in Section 4.5.3), give an idea of the uncertainty in the estimate of the population.

Statistical theory offers two principal approaches for developing estimates for population parameters. They are known as the method of moments and the method of maximum likelihood. Their descriptions and uses can be found in many statistics texts.[6] Statistical estimates for the mean and different percentiles of a distribution and intervals for quantifying their uncertainty are discussed below.

4.5.3 Confidence Intervals

Confidence intervals are one of the most common types of statistical inference. A statistical confidence interval expresses a level of "confidence" in how closely the estimator approximates the true unknown population parameter. As an example, suppose a value for the true, but unknown population mean of soil arsenic concentrations is desired for input into a risk assessment calculation. A set of samples is collected and the average of the data is 6.5 milligrams per kilogram (mg/kg). The question is asked how accurately 6.5 mg/kg estimates the population mean. A 90-percent confidence interval would provide this information.

Suppose this interval was calculated as 3.2 to 9.8 mg/kg. The upper bound of this interval or 95-percent UCL of 9.8 mg/kg could be selected as a conservative estimate of the true, but unknown population mean for arsenic at the site.

4.5.4 Random Sampling

Probabilistic sampling techniques are employed to ensure the collected sample has the "random" character required by the theory of statistics to draw inferences about the unknown population. Each member of the population has a specified (usually equal) probability of being sampled. Extreme caution is warranted if biased, non-random sampling results (e.g., screening with a photoionization detector (PID), hot-spot searches, etc.) are being used for developing input distributions representing the population of site concentrations or the average site concentration for input into a probabilistic risk assessment. Justifying this approach based on it being "conservative" is not defensible and risk characterizations can be invalid and misleading.

4.5.5 Correlation

One important aspect of correctly performing a Monte Carlo simulation is correlating the input variables when required. Failure to account for this aspect of the model can yield misleading results. Correlation is a measure of the strength of the linear association of two quantitative variables.[22] Correlation often is expressed as a dimensionless value between -1 and 1 inclusive, called the correlation coefficient. This concept is addressed again later in this section as to how to determine when correlation needs to be explicitly accounted for in the simulation.

4.6 UNCERTAINTY PROPAGATION METHODS

4.6.1 Monte Carlo and Other Uncertainty Propagation Methods

A variety of methods exist for quantifying the uncertainty associated with probabilistic risk assessments. These methods cover a broad range of complexity. A few of the more common approaches are: (1) analytic techniques such as variance propagation models, (2) response surface modeling, and (3) differential analysis.[23,24,25,26]

This handbook focuses on the appropriate use of Monte Carlo simulation for characterizing uncertainty, with emphasis on the exposure assessment step of the risk assessment process. Monte Carlo is a primary method of interest because it is considered by Agency sources to be more straightforward and easier to use than other probabilistic methods.[27] The Monte Carlo method also has the advantage of allowing the analyst to account for relationships between input variables and of providing the flexibility to investigate the effects of different modeling assumptions. In addition, the application of Monte Carlo and other probabilistic techniques for human health risk assessments has primarily been limited to exposure assessments. Current USEPA policy is not intended to apply to other evaluations (e.g., dose-response evaluations) until the application of probabilistic analysis has been further studied and refined.[28,29] Before dealing with dose-response uncertainties, USEPA questions

concerning interspecies and low-dose extrapolation issues will also need to be addressed.[27] However, as previously discussed in Section 3.3.3, this is a field that is currently being researched; uncertainty analysis for the dose-response and pharmacokinetic models are being performed by experts in the field.

The Monte Carlo simulation provides a numerical method for performing uncertainty analysis through computer simulation. This method is based on empirically estimating the uncertainty in the output variable or variables of the model by randomly sampling the uncertain input variables and performing deterministic model runs many times. This method is performed as follows. One value is randomly sampled from each uncertain input distribution and the corresponding risk result is calculated with the algebraic risk equation. This process is repeated many times to obtain risk results based on the different input values sampled from the input distributions. The end result is a probability distribution for the risk results. Statistics can then be calculated from the output risk distribution for making decisions that quantitatively account for uncertainty.

4.6.2 Simulation Techniques

Monte Carlo allows a direct simulation of the output variable(s) from the developed input distributions. The principal advantages of this technique for modeling the propagation of uncertainty are (1) its general applicability to models with even the most complex forms, and (2) the ease of which confidence intervals and other results can be calculated from the model output. Its disadvantages for application to risk assessment include a strong sensitivity to assumptions about the input distributions.

4.6.2.1 Sampling Methods

The probabilistic distributions developed for the input variables of a Monte Carlo risk assessment need to be sampled in some manner to perform each simulation run. Two common sampling approaches are called simple random sampling (SRS) and Latin hypercube sampling (LHS). A SRS procedure specifies that the probability of sampling a particular value is only determined by the shape of the PDF. LHS can be thought of as a stratified SRS procedure. It is usually employed when the cost of performing the many simulation runs required by SRS is high and a minimal number of simulation runs is desired. It attempts to retain the random nature of SRS, but forces equal coverage of each input PDF that might not otherwise be achieved with a small number of samplings. LHS divides each input distribution into non-overlapping intervals and randomly samples one value from each interval. The sampled values from the different input distributions are then randomly paired to form input data sets for each simulation run. The Monte Carlo analysis is then completed accordingly. One disadvantage of LHS compared to SRS is a greater possibility for inducing undesirable pairwise correlations among input variables.[23]

4.6.2.2 Computer Simulation

There are a number of commercially available programs which can perform Monte Carlo simulations. Some of the more commonly used include: Crystal Ball®, @RISK, and RiskQ.

(Note that this is not intended to provide an endorsement for any particular software package, nor is this meant to be a complete listing of all available software.) These software packages work in conjunction with standard commercially available spreadsheets such as Microsoft Excel or Lotus 1,2,3. The specifics of each software package differ; however, each input parameter is generally specified within the spreadsheet.

The formula relating the input parameters to the output parameter is also specified in the spreadsheet. The simulation software then allows distribution types to be specified, including those commonly used in risk assessment (e.g., normal, lognormal, triangular, empirical). It is possible to run a simulation with a combination of distributions and point estimates as input parameters. The simulation software then iteratively runs the specified number of simulations to develop the final risk distribution.

During each simulation run, the software randomly selects a value for each parameter based on the distribution specified for that parameter. The selected values are used to calculate the output parameter (risk). The level of risk calculated during each run is recorded, and all individual outputs are combined to develop the final risk distribution.

A Monte Carlo simulation will generally consist of several thousand iterations. The software also generally provides several additional functions such as sensitivity analysis. It should be noted that reliance on functions, such as sensitivity analysis, without an understanding of specifically how they are calculated and whether the calculation method is appropriate, is not recommended.

4.6.2.3 Inducing Correlation

One of the most important aspects of correctly setting up a Monte Carlo simulation is determining whether or not the input variables are correlated (i.e., interdependent), and if so, are the correlations important in the model. If significantly high correlations exist between input variables that play a large role in defining the variability of the output, then their correlation must be accounted for in the Monte Carlo simulation. Alternatively, if (1) the dependency is high, but the output is relatively insensitive to those variables, or (2) the variables are important, but the dependencies are small, their correlation can be ignored without much effect. The NCRP guide for uncertainty analysis gives some good examples of when correlations are important.[24] Correlations become more important if the stakeholders are interested in values occurring in the extreme tails of the distribution.

Correlated variables can be identified based on (1) a conceptual understanding of how the risk factors interact or (2) the physical scenario from which the uncertain distribution was derived. For example, one would expect body weight and body height to be correlated. A literature review can provide correlation information for many of the obvious cases. Alternatively, correlation can be estimated from empirical data or subjective information. In any case, if the correlation between variables can be quantified, a **correlation coefficient** can be used to account for it in the simulation. This will prevent sampling inappropriate

combinations of values from the correlated input distributions – for example, pairing a sampled body weight of 250 pounds with a body height of 5 feet in the same simulation run.

An alternative approach to minimizing errors due to interdependence of input distributions is to recognize that many correlated distributions have a strong age-related component. Using "age-specific" distribution functions based on different age groups can alleviate some of these concerns.[4] Specifically, the interdependent variables are linked by age group to reduce correlations in the model. This can be extremely important, for example, when calculating risk estimates for children.

4.7 SENSITIVITY ANALYSIS METHODS

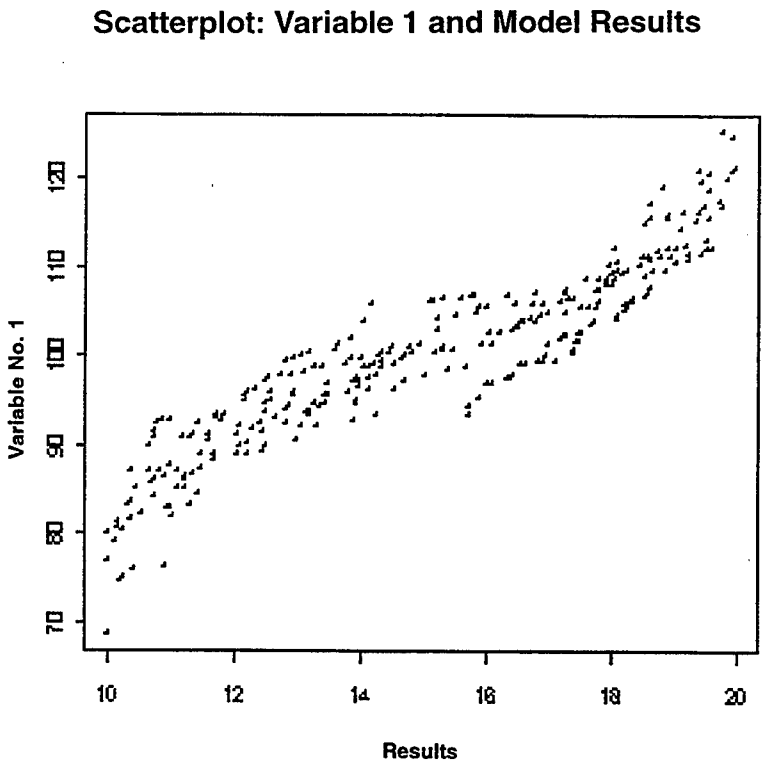
Decisions concerning allocating future resources to reduce lack-of-knowledge should take into consideration the most influential input factors in the model and the cost of gaining new information about these factors. Sensitivity analysis is used to identify which input factors are most important in the model. Input variables are considered important if they contribute significantly to the spread of the output risk distribution.

An input variable is considered to contribute significantly to the output risk distribution if it is *both* highly variable and the variability propagates through the algebraic risk equation to the risk result. Sensitivity analysis determines how this input variability propagates to the output risk distribution.

Once sensitivity analysis has identified an input variable as being important, the source of its spread or distribution should be determined. If an input variable has a significant *uncertainty* component, further research and/or data collection can be invested to reduce this uncertainty. This would, according to sensitivity analysis, reduce the spread of the output risk distribution. On the other hand, the output spread cannot be reduced if the sensitive input distribution represents inherent *variability* (e.g., inherent variability in the distribution of body weights of American adults cannot be reduced).

4.7.1 Graphical Techniques

Simple scatterplots of the simulated risk against the simulated uncertainty parameters from the Monte Carlo calculations can be used to qualitatively determine important lack-of-knowledge factors. A "tight" scatterplot suggests much influence in the total-risk uncertainty. An example scatterplot is shown on the following graph.



4.7.2 Regression Techniques

A statistical technique called multiple regression is a more formal approach to sensitivity analysis. Regression using either the simulated values[30] or their ranks[25] can be used to identify influential lack-of-knowledge factors. An index of sensitivity for the contribution of each input variable can be developed from multiple regression using the simulated values. This index quantifies the increase in the proportion of variability explained from the regression model by adding a selected factor to the model when all other factors are already being used. Spearman's Rank Correlation Coefficient can be used for the regression models based upon ranks. In either of these techniques, caution should be used in interpreting the individual contribution for factors that are explicitly correlated in the Monte Carlo simulations.

4.7.3 Analytic Techniques

There are several analytic approaches to sensitivity analysis.[25] The first approach, "differential sensitivity," requires deriving the partial derivatives of the risk function with respect to the uncertainty factors. Typically, these partial derivatives can become a computation problem in themselves, and hence are usually computationally expensive and/or computer code-development intensive. The second approach uses a functional analysis of variance decomposition of the total risk variability. That is, the total risk uncertainty is decomposed into the sum of individual contributions from each factor, plus the sum of pair-wise contributions. In practice, the computation of these individual parts can be expensive.

Another method of performing a sensitivity analysis is based on the first-order Taylor's series expansion of the risk model. Taylor Series are mathematical expressions used to

approximate certain algebraic functions and the associated error of the approximation. This "expansion" of the algebraic risk formula is used to decompose the total uncertainty of the model output into the contributions from each input variable. Each of these components can be approximately decomposed into additive components for each factor, plus pair-wise contributions. Rai, Krewski, and Bartlett present a technique that decomposes the model output uncertainty into components due to lack-of-knowledge and inherent variability.[31] Iman and Helton discuss a slightly less sophisticated use of Taylor's series approximations they call "differential analysis."[23]

4.8 NUMERICAL STABILITY

Results of the Monte Carlo simulation depend on performing calculations on computer-generated random numbers of the specified input distributions. As such, there can be some differences in the results strictly due to the computer modeling itself. This "artificial" source of uncertainty is referred to as "simulation noise." The effects of this error on the results is called numerical stability. Numerical stability is a function of multiple aspects of the simulation process including (1) the algebraic structure of the risk model, (2) the distributional forms of the input variables, and (3) the number of simulation runs or iterations performed. Also, the tails of the output distribution are more unstable than the middle. As such, the upper tail is usually evaluated because it is of greatest concern when evaluating risk.

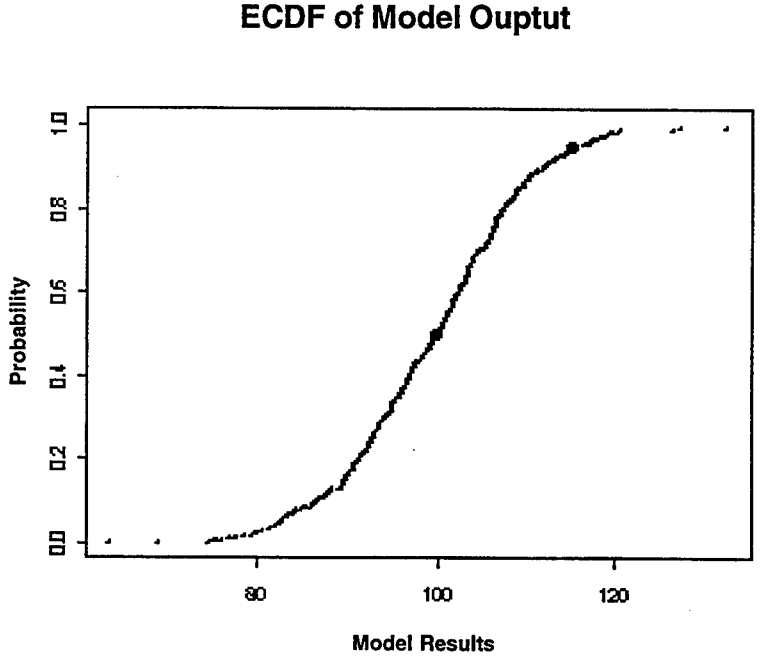
The nonparametric bootstrap method is appropriate for assessing the numerical stability of the output distribution from a simulation.[32] Specifically, the nonparametric bootstrap method for estimating the root mean square errors (RMSEs) is recommended.[33] The RMSE combines bias and variability, and is often used as the measure of the "goodness" of an estimator. The lower the RMSE, the better. An alternative approach is to calculate bootstrap confidence intervals for a particular statistic of interest of the output risk distribution. These methods provide a means of selecting or justifying a particular number of iterations to perform in the Monte Carlo simulation.

4.9 COMMUNICATING RESULTS

As discussed in more detail in Section 6, effective communication is an important step in the risk analysis. Graphical presentation of the results is recommended as the primary means of communicating simulation results. A review of the ability of graphical plots to communicate risk results was performed by Ibrekk and Morgan.[34] The study found that semi-technical and non-technical persons had significant difficulty discerning the mean risk estimate from a graph of all possible risk values output from the simulation. The study recommended clearly marking the mean or other important risk values on the graphs.

The two graphs commonly used to present simulation results are empirical PDFs and cumulative distribution functions (CDFs). "Empirical" is used to distinguish PDFs and CDFs using real data from those corresponding to theoretical statistical distributions (e.g., normal, uniform, etc.).

For an empirical CDF (ECDF), the results for each simulation run are ordered from high to low and plotted on the x-axis. The probability of each result is equally likely, so the cumulative probability of the ordered results is plotted on the y-axis. An example ECDF is presented below. The figure shows an ECDF of the simulation results with the median and 95th percentile marked in bold.



In addition, the curves should be accompanied by the results of the sensitivity analysis described earlier in this section to allow the end-users to make informed decisions. See Section 5 for examples of PDFs, CDFs, and sensitivity analysis graphics.

4.10 SUMMARY

Statistics provide the Air Force RPM or analyst with a powerful means for assessing uncertainty and variability. Although the statistical tools highlighted in this section are not without limitations, one can infer much about important risk-related issues (e.g., ingestion rates, representative soil concentrations for a given area, etc.) by statistically evaluating available sample data. Without statistics, uncertainty and variability in the risk estimation could not be quantitatively assessed and the RPM would be limited to making cleanup decisions using qualitative evaluations.

1 Weiss, Neil A. 1996. *Elementary Statistics*, 3rd Edition. Addison-Wesley Publishing Company, Inc., New York.

2 U.S. Environmental Protection Agency. 1996. *Exposure Factors Handbook (Update to Exposure Factors Handbook EPA600-08-89-043). Volume 1 – General Factors.* Office of Research and Development, National Center for Environmental Assessment, Washington, DC. August.

3 American Industrial Health Council. 1994. *Exposure Factors Sourcebook.* American Industrial Health Council, Washington, DC. August.

4 Finley, B., Proctor, D., Scott, P.K., Harrington, N., Paustenbach, D.J., and Price, P. 1994. Recommended Distributions for Exposure Factors Frequently Used in Health Risk Assessment. *Risk Analysis* 14(4):533-553.

5 Seiler, F.A., and Alvarez, J.L. 1996. On the Selection of Distributions for Stochastic Variables. *Risk Analysis.* 16(1):5-18.

6 Rice, John A. 1995. *Mathematical Statistics and Data Analysis.* Duxbury Press. Belmont, California.

7 Helsel, Dennis R., and Hirsch R.M. 1992. *Statistical Methods In Water Resources.* Elsevier, New York.

8 Helsel, Dennis R. 1990. Less than Obvious Statistical Treatment of Data Below the Detection Limit. *Environmental Science & Technology.* Vol. 24, No. 12.

9 Helsel, Dennis R., and Cohn, Timothy A. 1988. Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. *Water Resources Research.* 24(12):1997-2004.

10 Venables, W.N., and Ripley, B.B. 1994. *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

11 Shapiro, S.S., and Wilk, M.B. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika.* 52.

12 Gibbons, Robert D. 1994. *Statistical Methods for Groundwater Monitoring.* John Wiley & Sons, Inc., New York.

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

14 U.S. Environmental Protection Agency. 1992. *Methods for Evaluating the Attainment of Cleanup Standards Volume 1: Soils and Solid Media.* Statistical Policy Branch, Office of Policy, Planning, and Evaluation. EPA/230/02-89-042. February.

15 Anderson, Robert L. 1987. *Practical Statistics for Analytical Chemists.* Van Nostrand Reinhold, New York.

16 Marazzi, Alfio., Joss, Johann, and Randriamiharisoa, Alex. 1993. *Algorithms, Routines, and S Functions for Robust Statistics.* Chapman & Hall, New York.

17 Gilliom, Robert J., and Helsel, Dennis R. 1986. Estimation of Distributional Parameters for Censored Trace Level Water Quality Data 1: Estimation Techniques. *Water Resources Research.* Vol. 22, No. 2. February.

18 Helsel, Dennis R. and Gilliom, Robert J. 1986. Estimation of Distributional Parameters for Censored Trace Level Water Quality Data 2: Verification and Applications. *Water Resources Research.* Vol. 22, No. 2. February 1986.

19 Haas, C.N., and Scheff, P.A. 1990. Estimation of Averages in Truncated Samples. *Environmental Science & Technology.* 24.

20 U.S. Environmental Protection Agency. 1989. *Risk Assessment Guidance for Superfund – Volume I: Human Health Evaluation Manual (Part A).* Office of Emergency and Remedial Response, Washington DC. EPA/540/1-89/002. December.

21 U.S. Environmental Protection Agency. 1992. *Guidance for Data Useability in Risk Assessment (Part A).* Office of Emergency and Remedial Response, Washington, DC. 9285.7-9A. April.

22 Moore, David S., and McCabe, George P. 1993. *Introduction to the Practice of Statistics, 2nd Edition.* W.H. Freeman and Co., New York.

23 Iman, R.L., and Helton, J.C. 1988. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Analysis.* 8(1):71-90.

24 National Council on Radiation Programs. 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessment Related to Environmental Contamination.* NCRP Commentary No. 14, Bethesda, MD.

25 Cox, D.C., and Baybutt, P. 1981. Methods for Uncertainty Analysis: A Comparative Survey. *Risk Analysis.* 1(4):251-258.

26 Seiler, F.A. 1987. Error Propagation for Large Errors. *Risk Analysis.* 7(4)509-518.

27 U.S. Environmental Protection Agency. 1997. "Senior EPA Science Group Approves Probabilistic Analysis Policy. " Risk Policy Report. Vol. 3, No. 2. February 21, 1997.

28 U.S. Environmental Protection Agency. 1997. *Guiding Principles for Monte Carlo Analysis.* Risk Assessment Forum. EPA/630/R-97/001. March.

29 U.S. Environmental Protection Agency, Region VIII. 1995. *Use of Monte Carlo Simulation in Performing Risk Assessments (Region 8 Superfund Technical Guidance).* Hazardous Waste Management Division, Superfund Management Branch, Technical Section, Denver, CO. RA-10. September.

30 Cohen, J.T., Lampson, M.A., and Bowers, T.S. 1996. The Use of Two-Stage Monte Carlo Simulation Techniques to Characterize Variability and Uncertainty in Risk Analysis. *Human and Ecological Risk Assessment.* 2(4):939-971.

31 Rai, S.N., Krewski, D., and Bartlett, S. 1996. A General Framework for the Analysis of Uncertainty and Variability in Risk Assessment. *Human and Ecological Risk Assessment.* 2(4):972-989.

32 Efron, Bradley, and Tibshirani, Robert J. 1993. An Introduction to the Bootstrap. *Monographs on Statistics and Applied Probability 57*, Chapman & Hall, New York.

33 Ross, Sheldon M. 1997. Simulation, 2nd Ed. Academic Press, San Diego, CA.

34 Ibrekk, H., and Morgan, M.G. 1987. Graphical Communication of Uncertain Quantities to Nontechnical People. *Risk Analysis*. 7(4):519-529.

SECTION 5

EXAMPLE PROBABILISTIC AND DETERMINISTIC RISK CALCULATIONS

This section presents sample calculations for both a deterministic and probabilistic risk estimate using a hypothetical data set. The hypothetical data set includes three contaminants frequently detected at Air Force facilities – arsenic, benzene, and vinyl chloride. Each contaminant is modeled using both deterministic and probabilistic methods to illustrate the two approaches to estimating risk.

The purpose of this example is to:

- Demonstrate the use of the two risk estimation methods described by this handbook;
- Illustrate the resulting outcomes of each to highlight the differences between the methods (graphical presentations are included); and
- Further an understanding of how the use of the outcomes could impact the decision-making process.

Due to the cross-referencing of figures presented in this example, all tables and figures are presented at the end of Section 5. To enhance illustration of this approach, an additional published example of Monte Carlo simulation-based risk assessment is provided in Appendix C of this handbook. (The example was taken from USEPA Region VIII guidance.[1])

It should become apparent that the probabilistic technique using Monte Carlo simulation can be a powerful risk tool. It should also be noted, as previously discussed, that a risk-based screening comparison and a deterministic analysis are usually performed first (the first two levels of a tiered approach), prior to the probabilistic assessment (the third level of a tiered approach). The derived single-point value of the deterministic assessment will generally be explicitly included on the probabilistic range of risk for comparison.

A simplified risk assessment calculation is used in this example which includes two different exposure scenarios: (1) incidental ingestion of contaminated soil by an adult worker (Scenario I); and (2) ingestion of contaminated groundwater by an adult resident (Scenario II). Each of these scenarios are used to illustrate how a deterministic point estimate of risk compares to the probabilistic distribution of risks.

The intake equations used for these two scenarios are presented in Table 5.1 and are the standard ingestion algorithms for soil and groundwater recommended by USEPA.[2] Single-valued inputs into the equations represent reasonable maximum exposure (RME) assumptions. As previously described, RME assumptions describe the highest exposure and dose that is reasonably expected to occur for a site.[2,3] This high-end risk descriptor is estimated by combining upper-bound (90th to 95th percentile) values for some but not all of the exposure variables. These exposure variables for each scenario reflect those recommended by USEPA supplemental guidance to RAGS.[3,4,5] Variables used for Scenario I are representative of a soil ingestion pathway for a non-intrusive facility worker (e.g., a worker who does not perform soil invasive activities), and variables for Scenario II are representative of a groundwater ingestion pathway for a residential adult. The point estimate values and distribution parameters used in the calculations are shown in Table 5.2 for Scenario I and Table 5.3 for Scenario II.[1,6] (Graphical representations of the distribution parameters for both the soil and groundwater exposure variables are also presented at the end of this section.)

5.1 DETERMINISTIC RISK ESTIMATES

Recognized single-point estimates for specific variables were used as the input assumptions to develop a deterministic risk estimate. The exposure-point concentrations for each of the chemicals in soil and groundwater were assumed to be the 95-percent UCL on the sample data arithmetic mean. This is the method currently recommended by USEPA guidance for calculating a representative average value for site contamination.[7]

Results of the deterministic risk calculations are presented in Tables 5.4 and 5.5 for Scenario I, and in Tables 5.6 and 5.7 for Scenario II. Information is given for both cancer and noncancer effects. Benzene and vinyl chloride are classified as carcinogens and were evaluated for carcinogenic risk. Arsenic is capable of producing both carcinogenic and noncarcinogenic effects, each of which is evaluated separately. The carcinogenic risk estimate and noncarcinogenic hazard were estimated using the equations as illustrated in Table 5.1. The input variables used in the equations, including averaging times, cancer slope factors (CSF) for each chemical, and the noncancer reference dose (RfD) (for arsenic), were shown in Tables 5.2 and 5.3.

A carcinogenic risk estimate is derived by developing an estimate of the total dose averaged over an entire lifetime multiplied by the CSF (in this case, for soil and groundwater ingestion, an "oral" CSF). For exposures to multiple carcinogens for a given pathway, risk is calculated for each chemical and then summed to derive the total pathway cancer risk. Cancer risks from multiple exposure pathways are also assumed to be additive, provided the risks are summed for the same receptor and period of exposure. Under the USEPA National Oil and Hazardous Substances Pollution Contingency Plan (NCP), the target risk range for carcinogenic risk associated with a CERCLA Superfund site is one in ten thousand ($1x10^{-4}$) to one in one million ($1x10^{-6}$). Risks are considered acceptable within or below this range and unacceptable if above $1x10^{-4}$.[8]

The potential for an adverse noncarcinogenic effect, known as the hazard quotient (HQ), is calculated by dividing the chemical-specific average daily dose (intake) by the chemical-specific RfD (specifically, to calculate the ingestion pathway, an "oral" RfD). The hazard index (HI) is the sum of the results for all noncarcinogenic chemicals in the pathway (pathway HI), or the sum of the chemicals from various exposure pathways, as long as the pathways are for the same receptor and period of exposure (total exposure HI). An HI less than 1.0 indicates a very low threat of adverse health effects, whereas an HQ or HI in excess of 1.0 indicates that potential noncancer health effects may exist.

5.1.1 Scenario I Results

Based on the deterministic point-estimate results for carcinogenic effects (Table 5.4), the decision-maker(s) may conclude that some type of risk reduction is required to address the arsenic and vinyl chloride in soil. Although the predicted sum total carcinogenic risk estimate (2.33E-05) falls within the USEPA CERCLA acceptable risk range of $1x10^{-4}$ (1E-04) to $1x10^{-6}$ (1E-06), the fact that arsenic and vinyl chloride are categorized by USEPA as Class A (known human) carcinogens, may prompt a decision to require some level of remedial action. At this point, it is suggested that the RPM consider using a tiered approach to determine if risk reduction is necessary. The RPM can either accept the cumulative risk estimate or pursue the next "tier" by using probabilistic techniques (such as Monte Carlo) to recalculate the risk estimate. As illustrated in Table 5.4, Monte Carlo simulation results showed the mean, 90th, and 95th percentiles of the output risk distribution to be less than the deterministic point estimate.

As previously discussed, arsenic is the only compound contributing to the noncarcinogenic HI. The point estimate for the HI is 0.0636 (see Table 5.5). Further action is usually only considered if the HI is above 1.0. Since the deterministic risk estimate (high-end point estimate) indicates that there is no noncancer health threat due to the incidental ingestion of arsenic in soil, Monte Carlo simulation was not performed.

5.1.2 Scenario II Results

Given the results presented in Table 5.6, it is likely that some type of risk reduction would be required to address groundwater contamination. Action may be required because the predicted summed carcinogenic risk estimate (1.04E-03) exceeds the upper limit of the USEPA CERCLA acceptable range of $1x10^{-4}$ to $1x10^{-6}$. In addition, all three chemicals are categorized by USEPA as Class A (known human) carcinogens. When the risk estimate was recalculated using a Monte Carlo simulation, the cumulative risk estimate decreased by nearly one order of magnitude but still exceeded the upper limit of the acceptable range as defined by the USEPA.

The deterministic point estimate for noncarcinogenic risk (2.37) exceeded 1.0 and may be indicative that a noncancer health effect exists for this scenario (Table 5.7). As a result, further action may be considered to address noncarcinogenic arsenic in groundwater. At this point, before corrective action is taken, the RPM may decide to proceed with the next decision-making tier and recalculate the noncarcinogenic risk estimate using probabilistic

techniques. As illustrated in Table 5.7, the Monte Carlo simulation resulted in estimates ranging from 0.5 to 2.06.

5.2 MONTE CARLO SIMULATION

The Monte Carlo simulation was completed using the same equations as the deterministic approach. However, some of the input parameters were defined as PDFs instead of single-point values. The sources of these PDFs include site-specific data (hypothetical in this example), published information, and professional judgment.[9,10] As previously discussed in this handbook, PDFs incorporate both variability and uncertainty, which are not reflected in the single-point estimates.

Tables 5.2 and 5.3 describe the PDFs used in this example Monte Carlo simulation. Each PDF is listed as a distribution type with an associated mean and standard deviation (e.g., the distribution for body weight is given as "Normal (72, 15.9)," indicating that the actual distribution of adult body weight is modeled using a normal distribution with a mean of 72 kilograms (kg) and a standard deviation of 15.9 kg (see Section 4 for additional information about summary statistics and PDFs).

Output statistics from the Monte Carlo simulations are presented in Tables 5.8 and 5.9. Selected summary statistics and percentiles are shown for the risk distribution for each chemical. Selected percentiles are also presented in Tables 5.4 through 5.7 for comparison purposes. The figures corresponding to each distribution (Figures 5.1 and 5.2 for Scenario I and Figures 5.3 through 5.6 for Scenario II) give a better idea of the shape and location of the distribution. (These figures are presented at the end of this section.) The figures are labeled with the corresponding RME point estimate for comparison purposes.

5.2.1 Sensitivity Analysis

Figures 5.1(c) and 5.2(c) present the sensitivity analysis for Scenario I. Based on this analysis, the risk estimates for all chemicals are sensitive to two input variables: exposure duration and ingestion rate.

Figures 5.3(c) through 5.6(c) present the sensitivity analysis for Scenario II. Based on this analysis, the risk results for all chemicals are sensitive to both exposure duration and concentration of the chemical in the media. It should be noted that the sensitivity of a PDF is dependent on how the PDFs of input variables are defined and the algorithm used to estimate risk.

The risk estimates for each scenario were calculated using the Monte Carlo simulation software program, Crystal Ball® version 4.0. The program can be used to measure the sensitivity of the simulation output (i.e., risk) to the input variables by estimating the contribution of each input variable to the total variability of the output. This is important in determining which input variables have a large effect on the output variable. Sensitivity is calculated by computing the rank correlation coefficients between each PDF and the model

output while the simulation is running. Correlation coefficients provide a meaningful measure of the degree to which PDFs and model results change together.

If an input PDF and the model output have a high correlation coefficient, it means that the PDF has a significant impact on the model (both through its uncertainty and its model sensitivity). A positive coefficient indicates that an increase in the input PDF relates to an increase in the model output. A negative coefficient implies the reverse. The greater the absolute value of the correlation coefficient, the stronger the relationship.

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 minimal impact on the final distribution. This is important when trying to decide if gathering additional information for a variable is worthwhile. If the variable has a major impact on the final risk distribution (i.e., is sensitive and has a high uncertainty), then the additional expense of better defining the variable may be warranted.

While sensitivity analysis has been specifically studied in this example, a site-specific study would also consider other factors such as parameter correlation and numerical stability (see Section 4).

5.3 COMPARISON OF DETERMINISTIC AND PROBABILISTIC RISK ESTIMATES

Deterministic risk assessments use a combination of conservative point estimates and average values as inputs into the risk equation. The selection of these values is subjective and could vary for identically-scoped risk assessments performed by different professionals for the same site. As previously discussed, the deterministic approach generally yields a high-end risk estimate which may be unrealistic based on site-specific conditions. This problem is illustrated in Scenarios I and II where each point estimate is above 95 percent of its corresponding simulated risk distribution.

On the other hand, the point estimate derived from the deterministic risk assessment approach may underestimate risk. It is difficult for the RPM to know if the calculated point estimate potentially over- or underestimates risk without performing a quantitative uncertainty analysis. Hence, this is one of the main benefits of performing a Monte Carlo-based risk analysis.

5.3.1 Scenario I Comparison

In Scenario I, the RME point estimates correspond to locations in the risk distributions above the 95th percentile. In other words, more than 95 percent of the distribution of risk is below the RME point estimate. For both arsenic and benzene, the RME point estimates are approximately two times greater than the values corresponding to the 95th percentile of the risk distribution. This indicates that the point estimates may overestimate risk associated with the site, and the probabilistic risk estimates provide a more realistic description of the risk.

5.3.2 Scenario II Comparison

In Scenario II, the RME point estimates correspond to locations in the risk distributions above the 95th percentile. As with Scenario I, the results demonstrate that the vast majority of the risk distribution is below the RME point estimate.

It should be noted that pharmacokinetic factors were not modeled as distributions. The CSFs and the RfDs were modeled as point estimates because there is limited information concerning inter-individual variability concerning these factors. However, individuals can show considerable variation in pharmacokinetics. The incorporation of chemical-specific CSF and RfD distributions into the risk calculation, when they become available, could significantly affect the output.

5.4 CONCLUSIONS

The Monte Carlo simulations provided a range of risk estimates that represent all possible risk outcomes and their associated likelihood of occurring Therefore, the final risk distribution is likely to bound the true risk at the facility. The range of values produced by Monte Carlo analysis provides more information to use in the decision-making process. Monte Carlo analysis is a more dynamic tool for the RPM because it does not limit the RPM to a single-point risk estimate. Put another way, Monte Carlo analysis allows the RPM to make better informed decisions by seeing more of the "risk picture" than if only a single-point estimate were used.

By using the Monte Carlo simulation, it is possible to see that the uncertainty and variability which are incorporated into the deterministic risk estimate, but not explicitly accounted for, could be causing "unreasonable" and "unlikely" results when compared to the distribution of possible outcomes. Use of single-point estimates of risk, therefore, could possibly lead to the unnecessary expenditure of funds for remediation at the facility.

When the acceptable level of risk falls between the middle- and upper-end of the "acceptable" risk distribution, the concepts of variability and uncertainty can be important. As discussed earlier, variability is due to inherent differences in a population and cannot be reduced through collection of more data. Uncertainty, on the other hand, can presumably be reduced. In Scenarios I and II presented in this section, variability and uncertainty are mixed together in the simulation. For example, the concentration distribution represents the uncertainty in the true average concentration of the exposure area from collecting a limited number of samples. The body weight distribution, however, came from a large study and is, therefore, mostly composed of variability. The spread of the output risk distribution, therefore, is a convolution of both of these sources of input variation.

To better illustrate this, consider the following example. The 95th percentile of the risk distribution is compared to a "not-to-exceed" regulatory threshold and is slightly below the threshold value. If more data were collected, and a new simulation performed, would the new 95th percentile be below the regulatory limit? The answer depends on whether there is

"enough" *uncertainty* in the output risk distribution that would be reduced by including additional data. If so, the source of the uncertainty would need to be identified to decide if collecting more data is cost effective. In this example, the sources of uncertainty can only be identified qualitatively with professional judgment. Alternatively, more sophisticated simulations could be developed that explicitly model uncertainty and variability separately. More information on these types of models, known as "two-stage" simulations, can be found in "A Guide for Uncertainty Analysis in Dose and Risk Assessment Related to Environmental Contamination."[11,12]

1 U.S. Environmental Protection Agency, Region VIII. 1995. *Use of Monte Carlo Simulation in Performing Risk Assessments (Region 8 Superfund Technical Guidance).* Hazardous Waste Management Division, Superfund Management Branch, Technical Section, Denver, CO. RA-10. September.

2 U.S. Environmental Protection Agency. 1989. *Risk Assessment Guidance for Superfund – Volume I: Human Health Evaluation Manual (Part A).* Office of Emergency and Remedial Response, Washington DC. EPA/540/1-89/002. December.

3 U.S. Environmental Protection Agency, Region VIII. 1993. *Superfund's Standard Default Exposure Factors for the Central Tendency and Reasonable Maximum Exposure,* Draft, Denver, CO.

4 Micromedex, Inc. 1997. *Toxicology, Occupational Medicine, and Environmental Series (TOMES) Plus Database.* Denver, CO.

5 U.S. Environmental Protection Agency. 1996. *Health Effects Assessment Summary Tables.* Office of Emergency and Remedial Response. Office of Health Evaluation and Assessment, Washington, DC. ECAO-CIN-821.

6 American Industrial Health Council. 1994. *Exposure Factors Sourcebook.* American Industrial Health Council, Washington, DC. August.

7 U.S. Environmental Protection Agency, Region VIII. 1994. *Calculating the Concentration Term for Risk Assessment: Use of One "C" Term to Estimate Lower Average and Upper RME Risk Range. Region VIII Technical Guidance.* Denver, CO. January.

8 EPA, 1990. 40 CFR 300, National Oil and Hazardous Substances Pollution Contingency Plan, Final Rule, *Federal Register* 55(46):8665-8865. U.S. Environmental Protection Agency. March 8, 1990.

9 Smith, R. L. 1994. Use of Monte Carlo Simulation for Human Exposure Assessment at a Superfund Site. *Risk Analysis.* 14(4):433-440.

10 Israeli, M. and Nelson, C.B. 1992. Distribution and Expected Time of Residence for U.S. Households. *Risk Analysis.* 12(1):65-72.

11 National Council on Radiation Programs. 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessment Related to Environmental Contamination.* NCRP Commentary No. 14, Bethesda, MD.

12 Burmaster, D.E., McKone, T.E., Frey, C.H., Price, P., and Hoffman, F.O. 1995. *Workshop on Quantitative Techniques for Analysis of Variability and Uncertainty in Exposure and Risk Assessment.* Society for Risk Analysis Annual Meeting, Honolulu, HI. December 3, 1995.

13 Hoffman, F.O. and Hammonds, J.S. 1994. Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability. *Risk Analysis.* 14(5):707-712.

TABLE 5.1

EQUATIONS USED IN EXPOSURE MODELS

Intake equation for soil ingestion:

$$Intake = \frac{IR * FI * EF * ED * CF * Cs}{BW * AT}$$

Intake equation for groundwater ingestion:

$$Intake = \frac{IR * EF * ED * Cgw}{BW * AT}$$

Calculation of potential carcinogenic risk and noncarcinogenic health effects:

$$Risk = Intake * \text{CSF (carcinogenic)}$$

$$\text{Hazard Quotient} = \frac{Intake}{RfD} \text{ (noncarcinogenic)}$$

Where:

IR = Ingestion Rate

FI = Fraction Ingested

EF = Exposure Frequency

ED = Exposure Duration

CF = Conversion Factor

Cs = Contaminant Concentration in Soil

BW = Body Weight

AT = Averaging Time

AT (carcinogens) = 70 years (lifetime) * 365 days/year

AT (noncarcinogens) = ED x 365 days/year

Cgw = Contaminant Concentration in Groundwater

CSF = Carcinogenic Slope Factor

RfD = Reference Dose

TABLE 5.2
CONSTANTS AND VARIABLES USED IN RISK EXPOSURE MODEL FOR SCENARIO I (SOIL)

Name, Symbol	Units	Point Estimate	Distribution
Ingestion Rate, IR	mg/day	100	Lognormal (50.0, 50.0)[1,a]
Fraction Ingested, FI	unitless	1	
Exposure Frequency, EF	days/year	250	Triangular (125, 219, 300)[1,b]
Exposure Duration, ED	years	25	Lognormal (7.3, 12)[1,a]
Conversion Factor, CF	kg/mg	1.00E-06	
Body Weight, BW	kg	70	Normal (72, 15.9)[6,a]
Averaging Time, carcinogen, AT	days	25,550	
Averaging Time, noncarcinogen, AT	days	9125	
Concentration of Arsenic, Cs	mg/kg	19.49	Normal (16.2, 2)[a,c]
Concentration of Benzene, Cs	mg/kg	19.49	Normal (16.2, 2)[a,c]
Concentration of Vinyl Chloride, Cs	mg/kg	19.49	Normal (16.2, 2)[a,c]
Reference Dose Arsenic, RfD	mg/kg-day	3.00E-04	
Carcinogenic Slope Factor Arsenic, CSF	kg-day/mg	1.5	
CSF Benzene, CSF	kg-day/mg	2.90E-02	
CSF Vinyl Chloride, CSF	kg-day/mg	1.9	

[a/] Interpretation of lognormal distribution notation: Lognormal (7.3,12) indicates a mean of 7.3 and a standard deviation of 12 assuming a lognormal distribution.

[b/] Interpretation of triangular distribution notation: Triangular (125, 219, 300) indicates a minimum value of 125, a most likely value of 219, and a maximum value of 300 assuming a triangular distribution.

[c/] Site-specific data (hypothetical).

TABLE 5.3
CONSTANTS AND VARIABLES USED IN RISK EXPOSURE MODEL FOR SCENARIO II (GROUNDWATER)

Name, Symbol	Units	Point Estimate	Distribution
Ingestion Rate, IR	L/day	2	Lognormal (6.85, 0.53)[1,a]
Exposure Frequency, EF	days/year	350	Triangular (180, 345, 365)[1,b]
Exposure Duration, ED	years	30	Lognormal (4.55, 8.68)[1,a]
Body Weight, BW	kg	70	Normal (72, 15.9)[6,a]
Averaging Time, carcinogen, AT	days	25,550	
Averaging Time, noncarcinogen, AT	days	10,950	
Concentration of Arsenic, Cgw	mg/L	2.59E-02	Normal (1.25E-03, 1.5E-02)[a,c]
Concentration of Benzene, Cgw	mg/L	2.59E-02	Normal (1.25E-03, 1.5E-02)[a,c]
Concentration of Vinyl Chloride, Cgw	mg/L	2.59E-02	Normal (1.25E-03, 1.5E-02)[a,c]
Reference Dose Arsenic, RfD	mg/kg-day	3.00E-04	
Carcinogenic Slope Factor Arsenic, CSF	kg-day/mg	1.5	
CSF Benzene, CSF	kg-day/mg	2.90E-02	
CSF Vinyl Chloride, CSF	kg-day/mg	1.9	

[a/] Interpretation of lognormal distribution notation: Lognormal (6.85,0.53) indicates a mean of 6.85 and a standard deviation of 0.53 assuming a lognormal distribution.

[b/] Interpretation of triangular distribution notation: Triangular (180, 345, 365) indicates a minimum value of 180, a most likely value of 345, and a maximum value of 365 assuming a triangular distribution.

[c/] Site-specific data (hypothetical).

TABLE 5.4
Estimates of Carcinogenic Effects for Scenario I

Contaminant	Point Estimate	Mean	90th Percentile	95th Percentile
Arsenic	1.02E-05	1.14E-06	2.44E-06	4.50E-06
Benzene	1.97E-07	NA*	NA*	NA*
Vinyl Chloride	1.29E-05	1.44E-06	3.10E-06	5.70E-06
Cancer Risk (ICR):	2.33E-05	2.58E-06	5.54E-06	1.02E-05

* A Monte Carlo simulation was not performed given the point estimate was less than the USEPA criteria of 1E-06.

TABLE 5.5
Estimates of Noncarcinogenic Effects for Scenario I

Contaminant	Point Estimate	Mean	90th Percentile	95th Percentile
Arsenic	6.36E-02	NA*	NA*	NA*
Hazard Index (HI):	6.36E-02	NA*	NA*	NA*

* A Monte Carlo simulation was not performed given the point estimate was less than the USEPA criteria of 1.

TABLE 5.6
Estimates of Carcinogenic Effects for Scenario II

Contaminant	Point Estimate	Mean	90th Percentile	95th Percentile
Arsenic	4.56E-04	9.72E-05	2.23E-04	3.97E-04
Benzene	8.82E-06	1.88E-06	4.31E-06	7.68E-06
Vinyl Chloride	5.78E-04	1.23E-04	2.82E-04	5.03E-04
Cancer Risk (ICR):	1.04E-03	2.22E-04	5.09E-04	9.08E-04

TABLE 5.7
Estimates of Noncarcinogenic Effects for Scenario II

Contaminant	Point Estimate	Mean	90th Percentile	95th Percentile
Arsenic	2.37E+00	5.00E-01	1.16E+00	2.06E+00
Hazard Index (HI):	2.37E+00	5.00E-01	1.16E+00	2.06E+00

TABLE 5.8
MONTE CARLO SUMMARY STATISTICS FOR SCENARIO I

	Noncarcinogenic Arsenic	Arsenic	Benzene	Vinyl Chloride
Summary Statistics[a/]				
Trials	5000	5000	5000	5000
Mean	NA[b/]	1.14E-06	NA[b/]	1.44E-06
Median	NA[b/]	3.92E-07	NA[b/]	4.97E-07
Standard Deviation	NA[b/]	2.95E-06	NA[b/]	3.74E-06
Coeffecient of Variability	NA[b/]	2.60E+00	NA[b/]	2.60E+00
Range Minimum	NA[b/]	1.92E-09	NA[b/]	2.43E-09
Range Maximum	NA[b/]	8.89E-05	NA[b/]	1.13E-04
Percentiles				
0.0%	NA[b/]	1.92E-09	NA[b/]	2.43E-09
5.0%	NA[b/]	3.62E-08	NA[b/]	4.58E-08
50.0%	NA[b/]	3.92E-07	NA[b/]	4.97E-07
95.0%	NA[b/]	4.50E-06	NA[b/]	5.70E-06
100.0%	NA[b/]	8.89E-05	NA[b/]	1.13E-04

[a/] See discussion in Section 3.

[b/] A Monte Carlo simulation was not performed given the point estimate was less than the USEPA criteria of 1 for noncarcinogens and 1.0E-06 for for carcinogens.

TABLE 5.9
MONTE CARLO SUMMARY STATISTICS FOR SCENARIO II

	Noncarcinogenic Arsenic	Arsenic	Benzene	Vinyl Chloride
Summary Statistics*				
Trials	5000	5000	5000	5000
Mean	5.00E-01	9.72E-05	1.88E-06	1.23E-04
Median	1.70E-01	3.29E-05	6.35E-07	4.16E-05
Standard Deviation	1.19E+00	2.29E-04	4.43E-06	2.90E-04
Coeffecient of Variability	2.36E+00	2.36E+00	2.36E+00	2.36E+00
Range Minimum	0.00E+00	4.39E-09	8.48E-11	5.56E-09
Range Maximum	2.91E+01	5.61E-03	1.08E-04	7.11E-03
Percentiles				
0.0%	0.00E+00	4.39E-09	8.48E-11	5.56E-09
5.0%	1.00E-02	1.61E-06	3.11E-08	2.04E-06
50.0%	1.70E-01	3.29E-05	6.35E-07	4.16E-05
95.0%	2.06E+00	3.97E-04	7.68E-06	5.03E-04
100.0%	2.91E+01	5.61E-03	1.08E-04	7.11E-03

* See discussion in Section 3.

Figure 5.1 PDF, CDF, and Sensitivity Analysis for Carcinogenic Arsenic in Soil

Figure 5.1a
Probability Density Function (PDF)

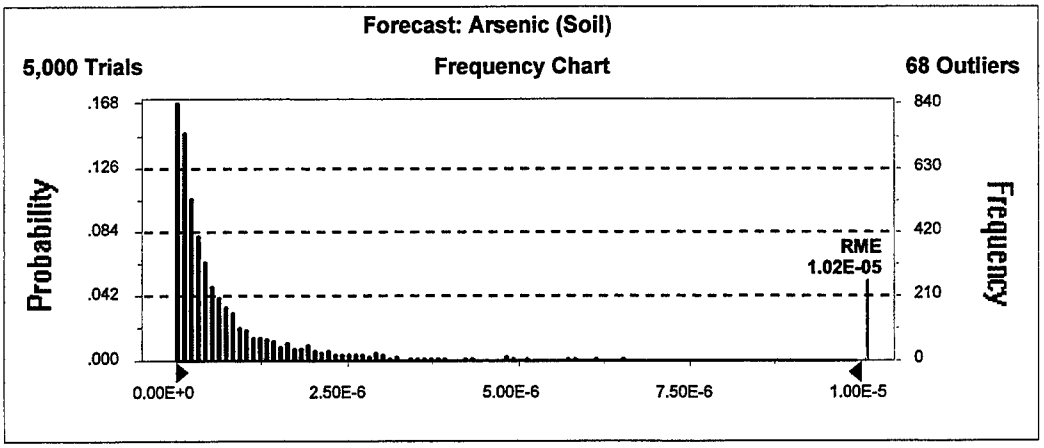


Figure 5.1b
Cumulative Distribution Function (CDF)

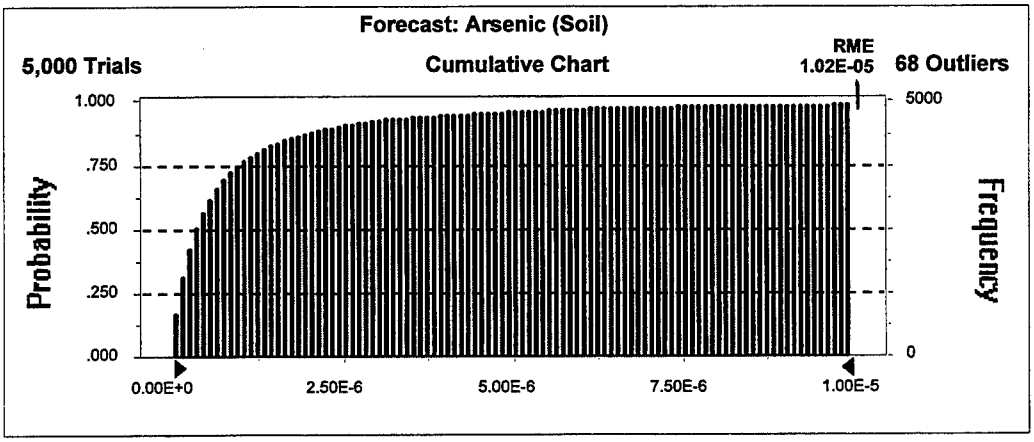


Figure 5.1c
Sensitivity Analysis
Sensitivity Chart

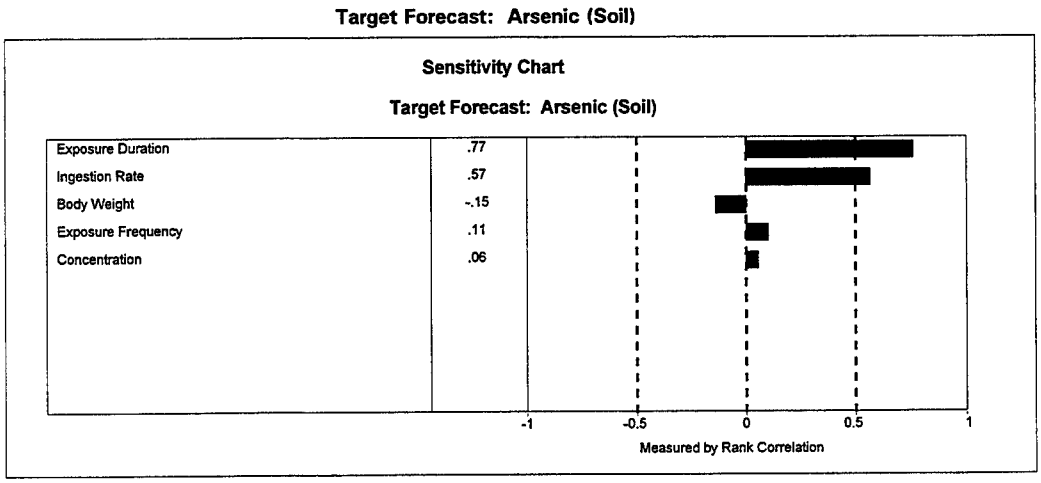


Figure 5.2 PDF, CDF, and Sensitivity Analysis for Vinyl Chloride in Soil

Figure 5.2a
Probability Density Function (PDF)

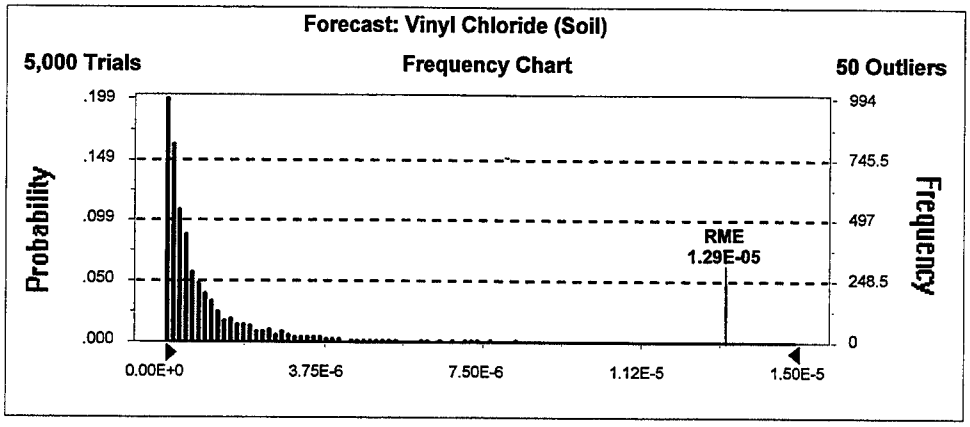


Figure 5.2b
Cumulative Distribution Function (CDF)

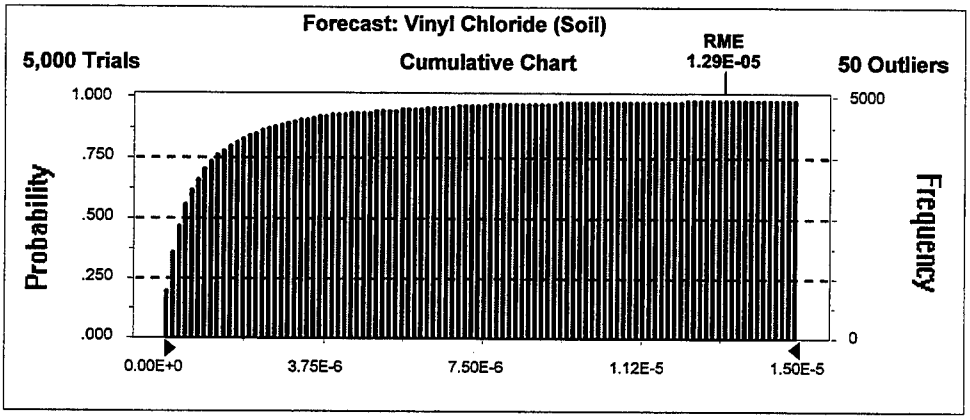


Figure 5.2c
Sensitivity Analysis
Sensitivity Chart

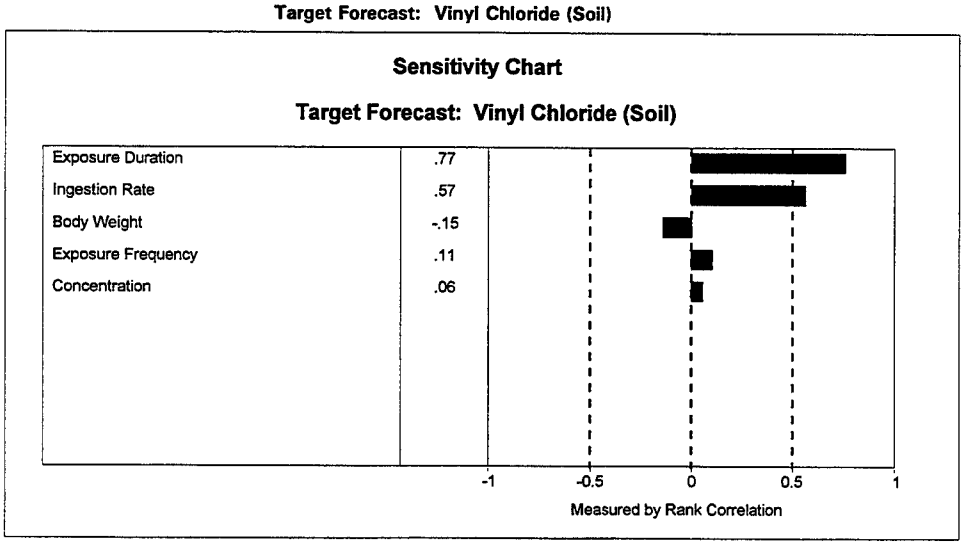


Figure 5.3 PDF, CDF, and Sensitivity Analysis for Carcinogenic Arsenic in Groundwater

Figure 5.3a
Probability Density Function (PDF)

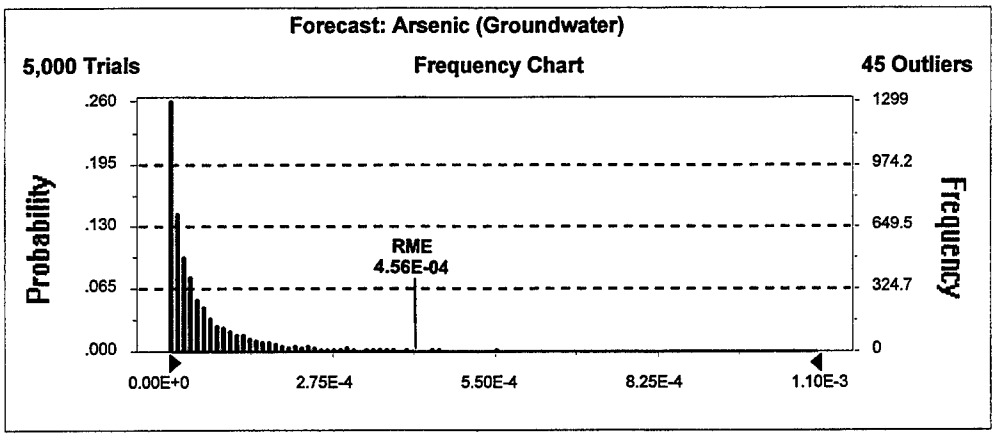


Figure 5.3b
Cumulative Distribution Function (CDF)

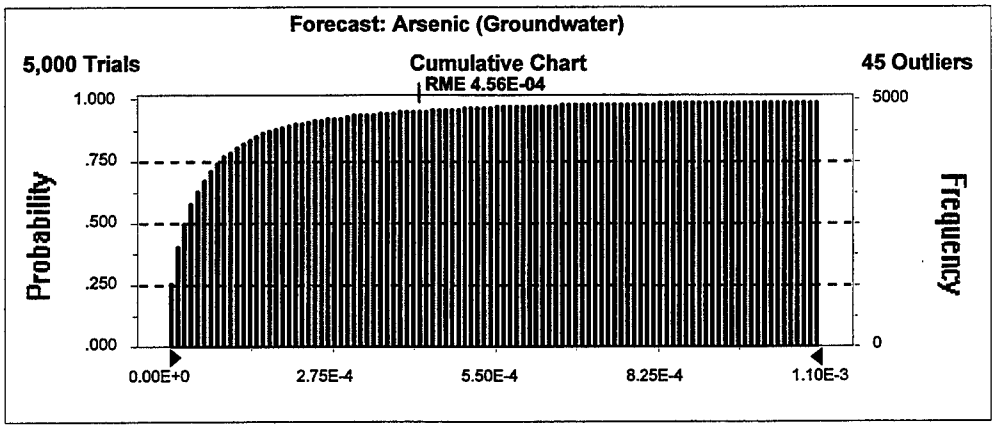


Figure 5.3c
Sensitivity Analysis
Sensitivity Chart

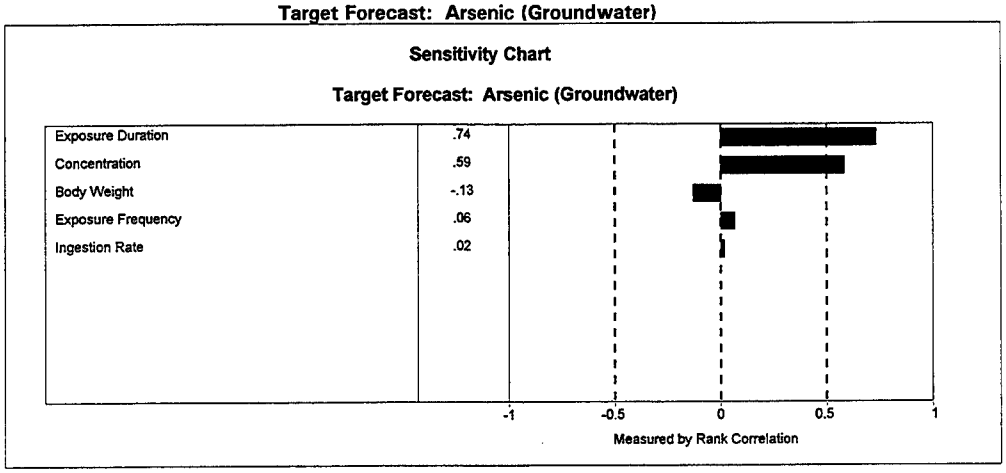


Figure 5.4 PDF, CDF, and Sensitivity Analysis for Benzene in Groundwater

Figure 5.4a
Probability Density Function (PDF)

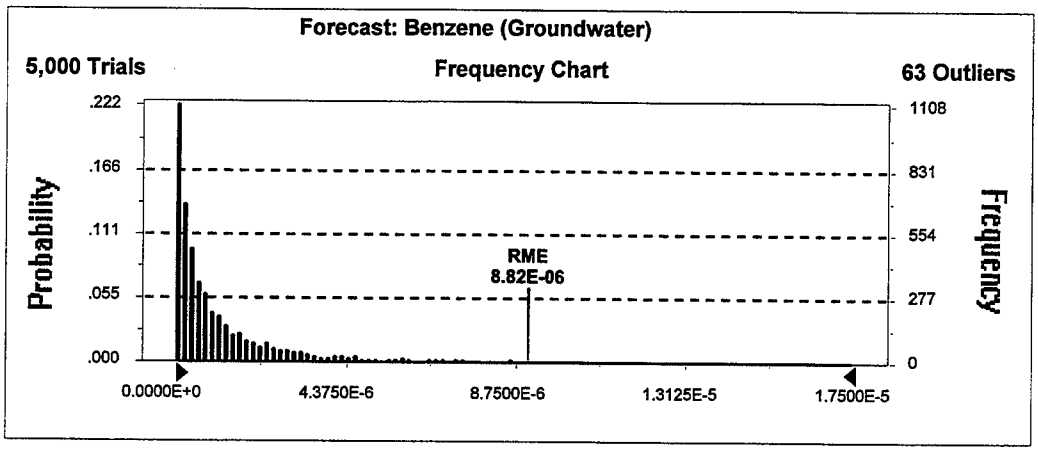


Figure 5.4b
Cumulative Distribution Function (CDF)

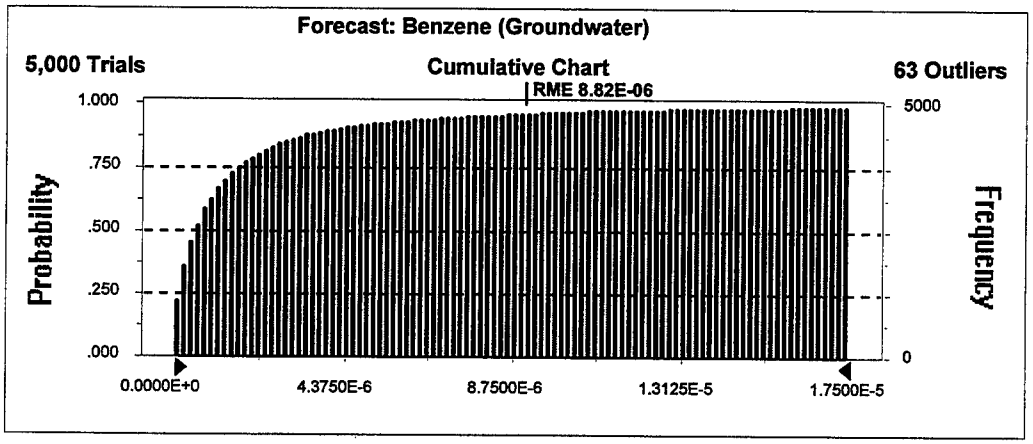


Figure 5.4c
Sensitivity Analysis
Sensitivity Chart

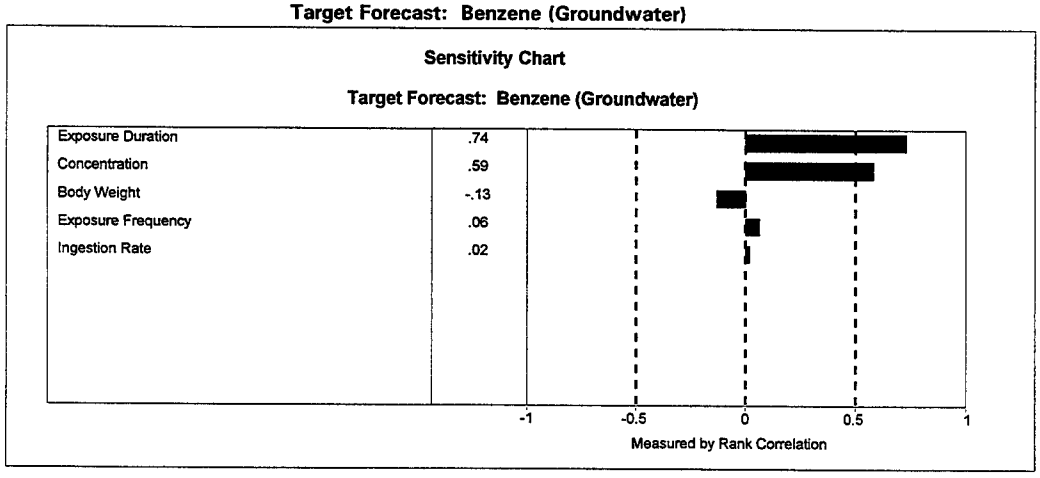


Figure 5.5 PDF, CDF, and Sensitivity Analysis for Vinyl Chloride in Groundwater

Figure 5.5a
Probability Density Function (PDF)

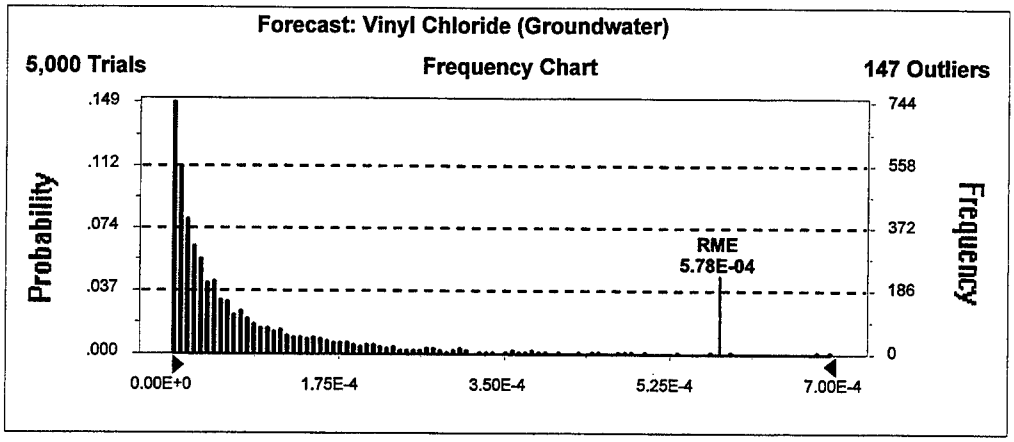


Figure 5.5b
Cumulative Distribution Function (CDF)

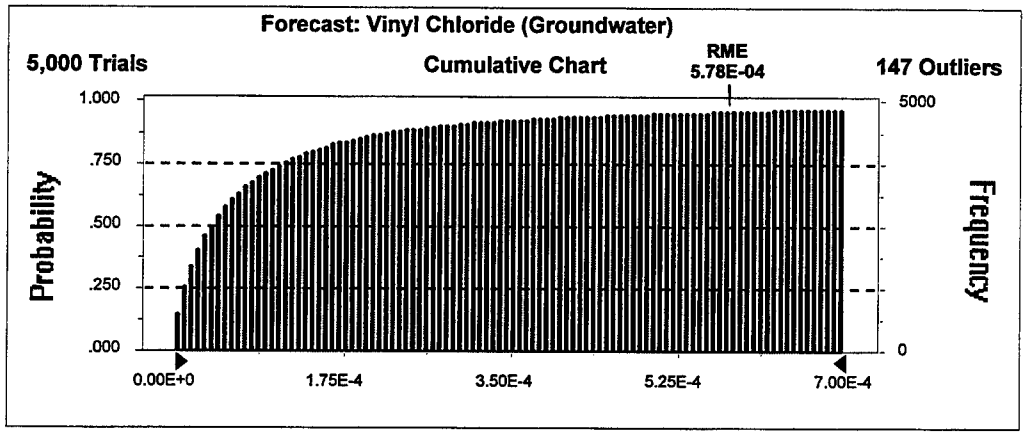


Figure 5.5c
Sensitivity Analysis
Sensitivity Chart

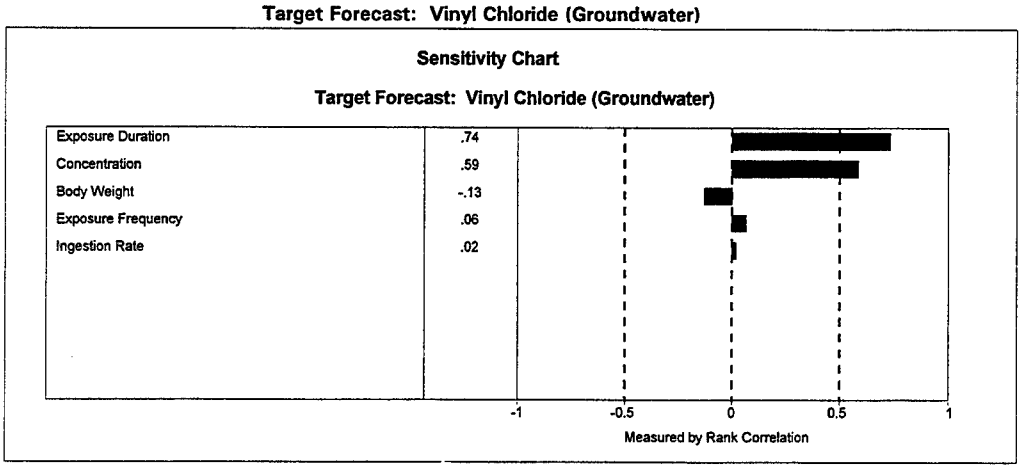


Figure 5.6 PDF, CDF, and Sensitivity Analysis for Noncarcinogenic Arsenic in Soil

Figure 5.6a
Probability Density Function (PDF)

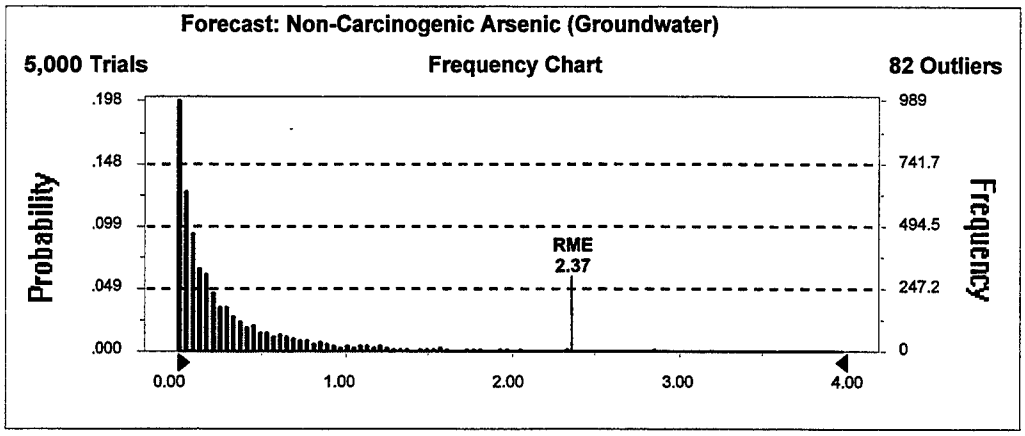


Figure 5.6b
Cumulative Distribution Function (CDF)

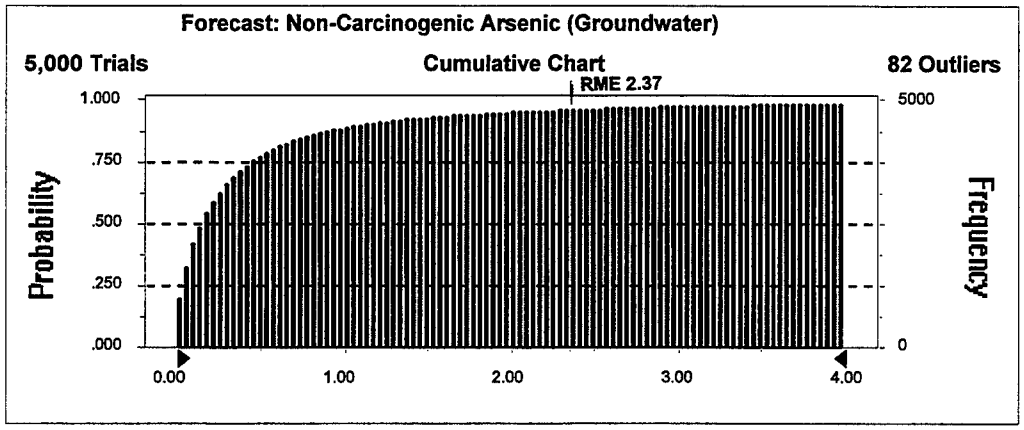


Figure 5.6c
Sensitivity Analysis
Sensitivity Chart

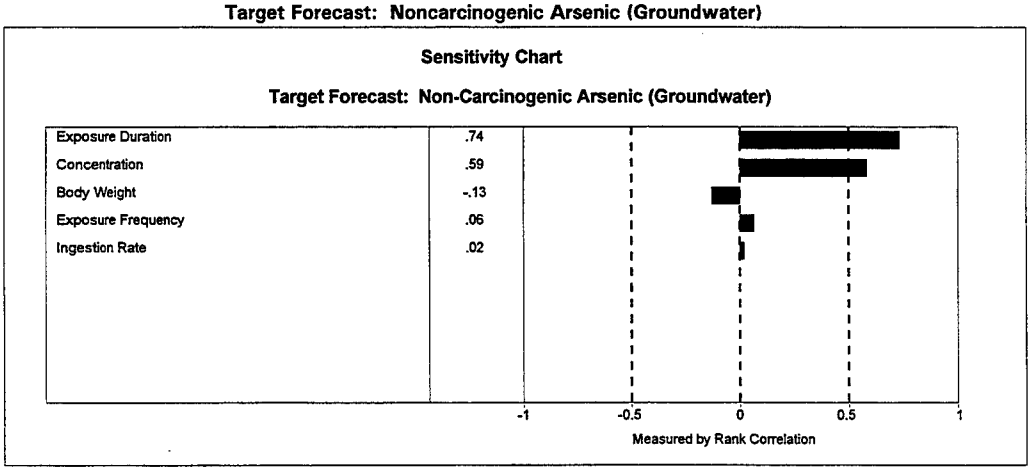


Figure 5.7 Graphical Representation of Distribution Parameters for Soil Exposure Variables

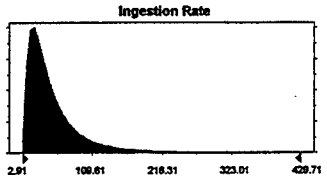
Soil Exposure Factor Assumptions

Assumption: Ingestion Rate

Lognormal distribution with parameters:

Mean	50.00
Standard Dev.	50.00

Selected range is from 0.00 to +Infinity
Mean value in simulation was 50.02

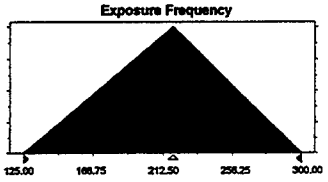


Assumption: Exposure Frequency

Triangular distribution with parameters:

Minimum	125.00
Likeliest	219.00
Maximum	300.00

Selected range is from 125.00 to 300.00
Mean value in simulation was 214.29

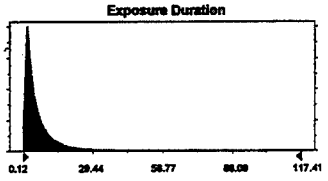


Assumption: Exposure Duration

Lognormal distribution with parameters:

Mean	7.30
Standard Dev.	12.00

Selected range is from 0.00 to +Infinity
Mean value in simulation was 7.37

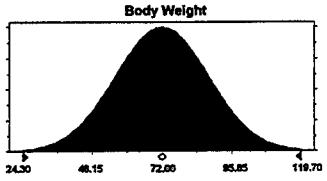


Assumption: Body Weight

Normal distribution with parameters:

Mean	72.00
Standard Dev.	15.90

Selected range is from -Infinity to +Infinity
Mean value in simulation was 72.15

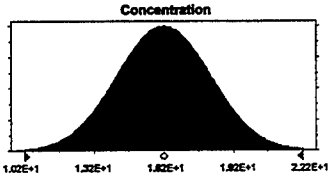


Assumption: Concentration

Normal distribution with parameters:

Mean	1.62E+01
Standard Dev.	2.00E+00

Selected range is from -Infinity to +Infinity
Mean value in simulation was 1.63E+1



End of Assumptions

Figure 5.8 Graphical Representation of Distribution Parameters for Groundwater Exposure Variables

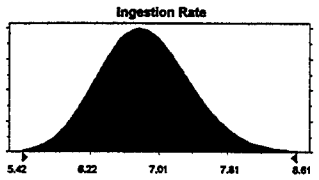
Groundwater Exposure Factor Assumptions

Assumption: Ingestion Rate

Lognormal distribution with parameters:

Mean	6.85
Standard Dev.	0.53

Selected range is from 0.00 to +Infinity
Mean value in simulation was 6.85

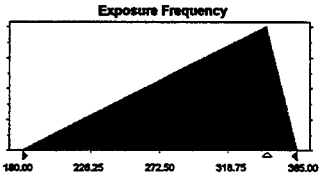


Assumption: Exposure Frequency

Triangular distribution with parameters:

Minimum	180.00
Likeliest	345.00
Maximum	365.00

Selected range is from 180.00 to 365.00
Mean value in simulation was 295.60

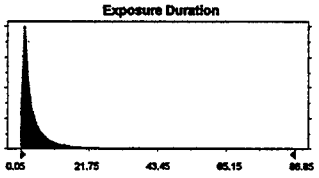


Assumption: Exposure Duration

Lognormal distribution with parameters:

Mean	4.55
Standard Dev.	8.68

Selected range is from 0.00 to +Infinity
Mean value in simulation was 4.51



Assumption: Body Weight

Normal distribution with parameters:

Mean	72.00
Standard Dev.	15.90

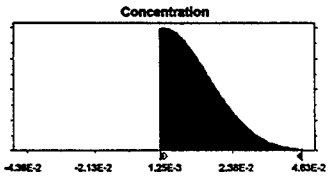
Selected range is from -Infinity to +Infinity
Mean value in simulation was 72.07

Assumption: Concentration

Normal distribution with parameters:

Mean	1.25E-03
Standard Dev.	1.50E-02

Selected range is from 0.00E+0 to +Infinity
Mean value in simulation was 1.23E-2



End of Assumptions

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 6

RISK MANAGEMENT

6.1 INCORPORATING PROBABILISTIC RISK ESTIMATES INTO RISK MANAGEMENT DECISIONS

The previous sections of this handbook have summarized the need for and basic approaches to developing risk estimates that quantitatively incorporate both uncertainty and variability. This information is critical to the decision-making process for Air Force RPMs because it provides:

- A better understanding of the possible distribution of risks;
- A means to identify what additional information or action could be taken at a site to reduce uncertainty and develop a more realistic picture of the actual risk posed to potential receptors;
- A clearer understanding of when a tiered approach would be selected for a site as the risk management situation requires;
- More relevant information upon which decisions can be based and more cost-effective solutions identified; and
- A means to effectively communicate and justify the decision made to any stakeholder.

Because an Air Force RPM is responsible for ensuring that work at a site progresses according to the priorities and objectives established during scoping and project planning, the RPM facilitates interactions between numerous involved parties, including Air Force personnel and their subcontractors, appropriate state and federal regulators, and the public. Air Force RPMs routinely act as a filter between the various stakeholders.

Air Force RPMs or other health professionals, such as the Bioenvironmental Engineer, may be required to explain what is known and unknown about human health risks at a particular site to different types of audiences. The purpose of doing probabilistic risk assessments is to provide more information about the nature of the risk estimate. Armed with the results of a probabilistic risk assessment, an Air Force RPM may be better equipped to demonstrate that the decision of choice was appropriate, and based on good science. This information can be used to more effectively communicate the scientific and policy basis for remedial decisions to different audiences.

In particular, probabilistic risk assessments can provide valuable input into the remedial decision-making process in a number of areas. For example, this methodology is useful in baseline risk assessments, ASTM Tier 3 evaluations,[1] comparative remedial alternatives evaluations, risk communication, and revision of Record of Decisions (RODs).

This final section of the handbook is not intended to exhaustively explain how risk information can be incorporated into the decision process. Rather, this section has been included to highlight how best to incorporate probabilistic risk assessments into different stages of the remedial planning process.

6.2 EARLY PLANNING AND SCOPING

Project scoping is the initial phase of the remedial planning process. Scoping activities typically begin with the collection of existing site characterization data, including any data from previous investigations. On the basis of this information, site management planning is undertaken to identify:

- Boundaries of the study area;
- Possible regulatory requirements and remedial action objectives;
- Need for any immediate interim action; and
- Other site management strategy concerns (e.g., address site as a single area or as several subsites).

Once an overall management strategy has been approved, specific plans for appropriate remedial planning documents are developed. This level of scoping activity can involve:

- Determining the types of decisions to be made and identifying the data and other information needed or desired to support those decisions;
- Assembling a technical advisory committee or closure team to assist in the review of initial project strategy through final remedial recommendations and implementation; and
- Preparation of technical scopes of work (SOWs) and any applicable workplans to guide data collection and analysis activities.

Understanding the basic data and analysis requirements of a probabilistic risk assessment is key to developing clear SOWs and early workplans for regulatory review and approval. As part of the early planning process, an Air Force RPM needs to define how uncertainty will be factored into the risk estimation step. As noted in previous sections of this handbook, a quantitative uncertainty analysis is not always warranted or recommended. However, if an Air Force RPM plans to potentially incorporate uncertainty into the risk management process at a specific site, a quantitative uncertainty analysis should be scoped as part of the risk estimation step of the project.

This scoping effort will have to take into account the existing regulatory perspectives on uncertainty analysis in risk assessments. Air Force RPMs are encouraged to openly discuss the need for and technical requirements of a quantitative uncertainty analysis or probabilistic risk assessment with all involved parties prior to even establishing the types of decisions to be made at a particular site. As described in Section 2, the results of a probabilistic risk assessment are not likely to be relied upon as stand-alone documents by decision-makers. Such information is currently primarily used to supplement deterministic risk estimates. Consequently, probabilistic risk estimates may be of the greatest use to Air Force RPMs in those cases where deterministic risk estimates have or may prompt costly "across-the-board" solutions. This supplemental information could provide the scientific and technical cornerstone for dialogues on how to cost-effectively manage risk.

When planning to include probabilistic risk estimates in the remedial decision process, an Air Force RPM needs to:

- Identify the measure of acceptable risk for a site (e.g., when the cost of regulatory or remedial action is expected to be high and the potential health risk is expected to be marginal);
- Specify the component variables of the risk measure that will be included in the analysis (e.g., exposure assessment step only);
- Identify acceptable sources of information to account for variability and uncertainty (e.g., USEPA references, expert judgment);
- Determine how uncertainty will be assessed (e.g., Monte Carlo simulation); and
- Define how the results will be included in the technical report, briefing materials, and the decision-making process.

These elements should be clearly discussed in early site strategy development meetings and referenced in any SOW written to solicit technical support to develop probabilistic risk estimates.

It should also be noted that an Air Force RPM will need to carefully balance the costs associated with obtaining credible information to complete a quantitative uncertainty analysis with the costs of not having such information (i.e., the cost of not being able to account for uncertainty and variability in the risk estimate). As this handbook illustrates, uncertainty in risk estimates is rarely caused wholly or even in large part by a single factor. Consequently, efficient allocation of resources among different uncertainty reductions becomes crucial. In a real sense, the RPM acts as a risk manager even in the early stages by determining the value of information.

6.3 OBTAINING EARLY CONSENSUS

Securing early consensus from all involved parties may also play a significant role in how well probabilistic risk estimates can be incorporated into the decision process at a specific site.

The RPM should point out that the goal of this type of supplemental information is to provide better information on the state of knowledge about potential risks than that which is traditionally provided by deterministic risk estimates.

In order for such information to be meaningfully incorporated into the decision process, however, an Air Force RPM should strive to facilitate a common understanding among involved parties about the objectives and goals of such information. All involved parties should be encouraged to bring their interests and concerns to the group for consideration throughout the decision process. Reaching early consensus facilitates group acceptance and approval of how best to use this information to make better informed decisions.

The following general issues should be considered as starting points for establishing group consensus:

- Identify goals and objectives of probabilistic risk estimates, with special emphasis on the potential technical, political, social, and economic ramifications of basing a decision on such information;
- Define the level of complexity of the effort, specifically in terms of increasing data analysis and review requirements and public communication requirements;
- Develop the data analysis approach, including the statistical methodology and the intended use of the results; and
- Determine how the data collection program may need to be amended to meet established project objectives and the requirements of the data analysis approach(s).

6.4 PRESENTING RESULTS TO VARIOUS AUDIENCES

Throughout the remedial decision process, an Air Force RPM acts as a liaison between technical analysts, regulators, and the public. Once probabilistic risk estimates have been developed, an Air Force RPM must decide how best to communicate a range of possible risk estimates from the already confusing activity of communicating summary information about deterministic risk estimates. There is more to risk communication than simply conveying information about potential harm. RPMs are typically required to direct dialogue about the:

- Type and associated probability of potential risks;
- Degree of voluntariness of the risk;
- Costs and feasibility of mitigating potential risks; and
- State of knowledge and confidence in predictions about risk.

With a probabilistic risk assessment, an Air Force RPM is better able to explain the state of knowledge and, therefore, to include information about the trade-offs that may be acceptable to different parties. For example, probabilistic risk estimates summarize the likelihood of certain risks occurring at a site. With this information, the group can jointly

debate and decide if the likelihood of unacceptable risk justifies implementation of certain costly remedies. Additionally, these data can add to the credibility of the presentation of risk information. The uncertainty analysis may not change the final decision, but it can increase the confidence that all involved parties have in making any particular decision.

There are various verbal, mathematical, and pictorial methods available to communicate the results of a quantitative uncertainty analysis. The scope of the uncertainty analysis will drive the verbal and mathematical methods used in summarizing risk information. However, the best pictorial method for depicting uncertainty analysis results is the same regardless of the scope of the analysis.[2] To facilitate discussions about both variability and probability, the PDF should be plotted directly above the CDF on the same horizontal scale to facilitate comparisons (see Figures 5.1 through 5.6 in Section 5). The mean of the distribution should be clearly indicated on both plots. Ideally, the following information could be included on or associated with these graphs:

- Description of the shape of the distribution;
- Median;
- Mean;
- Mode of the distribution;
- Value of the 5th and 95th percentile;
- Percentile location of the mean; and
- Estimate of the standard deviation or the CV.

Once these pictorial descriptions have been prepared to support discussions, an Air Force RPM should follow the basic four-step iterative process for communicating risk information:[3]

- Facilitate open-ended elicitation of people's beliefs about risk, allowing expression of both accurate and inaccurate concepts (i.e., first allow them to discuss their concerns about the proposed action or inaction);
- Develop structured questionnaires designed to determine the prevalence of these beliefs and current level of knowledge (e.g., ask specifically what they would do to address those concerns, with emphasis on state of knowledge about risk and cost);
- Refine approach to describing "pictorial" results based on what people need to know to make an informed decision and an assessment of their current beliefs (e.g., revisit the graphs to emphasize the results in terms of their concerns); and
- Summarize important information using both open-ended and closed form deliveries (e.g., ask for feedback on how they would use such information to make an informed decision).

6.5 CONCLUSION

The basic requirements for incorporating quantitative uncertainty analysis into human health risk assessments have been provided in this handbook. This type of analysis would be selected for a site as part of a tiered approach that progresses, as the risk management situation requires, from a simpler to a more quantitative evaluation. The probabilistic analysis is presented as the highest tier of such an approach.

This handbook provides an overview of the use of probabilistic methods to estimate potential risks and a base for directing Air Force RPMs to additional resources for further information. Additional references for more detailed guidance or technical direction are provided as an attachment in Appendix A.

1 American Society for Testing and Materials. 1995. *Standard Guide for Risk-Based Corrective Action Applied at Petroleum Release Sites.* Annual book of ASTM Standards. E1739-95.

2 Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers.* Center for Risk Management, Resources for the Future, Washington, DC.

3 Morgan, M. Granger, Fishhoff, Baruch, Bostrom, Ann, Lave, Lester, and Atman, Cynthia J. 1992. Communicating Risk to the Public. *Risk Analysis.* 26(11):2048-2056.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

References for Further Reading

REFERENCES FOR FURTHER READING

Allen, B.C., Crump, K.S., and Shipp, A.M. 1988. Correlation Between Carcinogenic Potency of Chemicals in Animals and Humans. *Risk Analysis*. 8(4):531-544.

Amaral, D. 1988. Including Uncertainty in Assessments of Sulfur Oxide Health Risks. *Journal of the Air Pollution Control Association*. .38:399-405.

American Industrial Health Council. *Risk Assessment for Contaminated Site Remediation: What It Is, What It Can Do.* The Environmental Health and Risk Assessment Subcommittee, Washington, DC.

American Industrial Health Council. 1994. *Exposure Factors Sourcebook.* Washington, DC. American Industrial Health Council. August.

American Society for Testing and Materials. 1995. *Standard Guide for Risk-Based Corrective Action Applied at Petroleum Release Sites*. Annual book of ASTM Standards. E1739-95.

American Society for Testing Materials. 1996. *Draft Standard Guide for Probabilistic Risk Assessment and Handling Uncertainties in Risk Assessments for Contaminated Sites*. ASTM Committee E-47 on Biological Effects and Environmental Fate. E47.13.05.A. February.

Anderson, Robert L. 1987. *Practical Statistics for Analytical Chemists*. Van Nostrand Reinhold, New York.

Baird, Sandra J.S., Cohen, Joshua, T., Graham, J.D., Shlyakhter, A.I., and Evans, J.S. 1996. Noncancer Risk Assessment: A Probabilistic Alternative to Current Practice. *Human and Ecological Risk Assessment*. 2(1):79-102.

Balch, G.I. and Sutton, S.M. 1995. Putting the First Audience First: Conducting Useful Evaluation for a Risk-Related Governmental Agency. *Risk Analysis*. 15(2):163-168.

Bogen, K.T. and Spear, R.C. 1987. Integrating Uncertainty and Interindividual Variability in Environmental Risk Assessment. *Risk Analysis*. 7(4):427-436.

Bogen, K.T. 1990. *Uncertainty in Environmental Health Risk Assessment*. Garland Publishing, New York.

Bogen, K.T. 1995. Methods to Approximate Joint Uncertainty and Variability in Risk. *Risk Analysis*. 15(3):411-419.

Bois, F.Y., Krowech, G., and Zeise, L. 1995. Modeling Human Interindividual Variability in Metabolism and Risk: The Example of 4-Aminobiphenyl. *Risk Analysis*. 15(2):205-213.

Bowers, T.S., Shifrin, N.S., and Murphy, B.L. 1996. Statistical Approach to Meeting Soil Cleanup Goals. *Environmental Science & Technology*. 30(5):1437-1444.

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

Brand, K.P. and Small, M.J. 1995. Updating Uncertainty in an Integrated Risk Assessment: Conceptual Framework and Methods. *Risk Analysis.* 15(6).

Brattin, W.J., Barry, T.M., and Chiu, N. 1996. Monte Carlo Modeling with Uncertain Probability Density Functions. *Human and Ecological Risk Assessment.* 2(4):820-840.

Bukowski, J., Korn, L., and Wartenberg, D. 1995. Correlated Inputs in Quantitative Risk Assessment: The Effects of Distributional Shape. *Risk Analysis.* 15(2):215-219.

Burmaster, D.E. and von Stackelberg, K. 1988. *A New Method for Uncertainty and Sensitivity Analysis in Public Health Risk Assessments at Hazardous Waste Sites Using Monte Carlo Techniques in a Spreadsheet.* Superfund '88, Proceedings of the 9th National Conference: Health and Endangerment, Washington, DC.

Burmaster, D.E. and von Stackelberg, K. 1991. Using Monte Carlo Simulations in Public Health Risk Assessments: Estimating and Presenting Full Distributions of Risk. *J. of Expos. Anal. Environ. Epidemiol.* 1(4):491-512.

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(4):477-482.

Burmaster, D.E., McKone, T.E., Frey, C.H., Price, P., and Hoffman, F.O. 1995. *Workshop on Quantitative Techniques for Analysis of Variability and Uncertainty in Exposure and Risk Assessment.* Society for Risk Analysis Annual Meeting, Honolulu, HI. December 3, 1995.

Burmaster, D.E. and Wilson, A.M. 1996. An Introduction to Second-Order Random Variables in Human Health Risk Assessments. *Human and Ecological Risk Assessment.* 2(4):892-919.

CalEPA. 1996. *The California EPA's Perspective on Probabilistic Risk Assessments*, Presented by Dr. Richard A. Becker, Office of Environmental Health Hazard Assessment. March.

Carnegie Commission. 1993. *Science, risk, and Regulatory Decision-Making.* June.

Chess, C. and Salmone, K.L. 1992. Rhetoric and Reality: Risk Communication in Government Agencies. *Journal of Environmental Education.* 23(3):28-33.

Chess, C., Salmone, K.L., Hance, B.J., and Saville, A. 1995. Results of a National Symposium on Risk Communication: Next Steps for Government Agencies. *Risk Analysis.* 15(2):115-125.

Cohen, J.T., Lampson, M.A., and Bowers, T.S. 1996. The Use of Two-Stage Monte Carlo Simulation Techniques to Characterize Variability and Uncertainty in Risk Analysis. *Human and Ecological Risk Assessment.* 2(4):939-971.

Commission on Risk Assessment and Risk Management. 1996. *Risk Assessment and Risk Management in Regulatory Decision-Making, Draft Report.* Commission on Risk Assessment and Risk Management, Washington, DC. June.

Copeland, T.L., Paustenbach, D.J., Harris, M.A., and Lau, V. 1993. Comparing the Results of a Monte Carlo Analysis with EPA's Reasonable Maximal Exposed Individual (RMEI): A Case Study of a Former Wood Treatment Site. *Regul. Toxicol. Pharm.* 18:275-312.

Cox, D.C. and Baybutt, P. 1981. Methods for Uncertainty Analysis: A Comparative Survey. *Risk Analysis*. 1(4):251-258.

Cronin IV, W.J., Oswald, E.J., Shelley, M.L., Fisher, J.W., and Flemming, C.D. 1995. A Trichloroethylene Risk Assessment Using a Monte Carlo Analysis of Parameter Uncertainty in Conjunction with Physiologically-Based Pharmacokinetic Modeling. *Risk Analysis*. 15(5):555-565.

Crouch, E.A.C. 1983. Uncertainties in Interspecies Extrapolations of Carcinogenicity. *Environmental Health Perspectives*. 50:321-327.

Crouch, Edmund et. al. 1995. *Report to the Commission on Risk Assessment and Risk Management.* October.

Cullen, A.C. 1994. Measures of Compounding Conservatism in Probabilistic Risk Assessment. *Risk Analysis*. 14(4):389-393.

Dakins, M.E., Toll, J.E., and Small, M. J. 1994. Risk-Based Environmental Remediation: Decision Framework and Role of Uncertainty. *Environ. Toxicol. Chem.* 13(12).

Efron, Bradley and Tibshirani, Robert J. 1993. *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability 57*, Chapman & Hall, New York.

Eschenroeder, A.Q. and Faeder, E.J. 1988. A Monte Carlo Analysis of Health Risks from PCB-Contaminated Mineral Oil Transformer Fires. *Risk Analysis*. 8(2):291-299.

Ferson, S. 1996. What Monte Carlo Methods Cannot Do. *Human and Ecological Risk Assessment.* 2(4):990-1007.

Fiering, M.B., Wilson, R., Kleiman, E., and Zeise, L. 1984. Statistical Distributions of Health Risks. *Journal of Civil Eng. Syst.* 1:129-138.

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.

Finkel, A.M. 1988. *Computing Uncertainty in Carcinogenic Potency: A Bootstrap Approach Incorporating Bayesian Prior Information.* Report to the Office of Policy, Planning, and Evaluation, US Environmental Protection Agency. August 8, 1988.

Finkel, A.M. 1989. Is Risk Assessment Really Too "Conservative"?: Revising the Revisionists. *Columbia Journal of Environmental Law.* 14(2):427-467.

Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers.* Center for Risk Management, Resources for the Future, Washington, DC.

Finkel, A.M. 1990. A Simple Formula for Calculating the "Mass Density" of a Lognormally Distributed Characteristic: Applications to Risk Analysis. *Risk Analysis*. 10(2):291-301.

Finkel, A.M. 1994. Stepping Out of Your Own Shadow: A Didactic Example of How Facing Uncertainty Can Improve Decision-Making. *Risk Analysis*. 14(5):751-761.

Finley, B.L., Scott, P.K., and Paustenbach, D.J. 1993. Evaluating the Adequacy of Maximum Contaminant Levels as Health-Protective Cleanup Goals: An Analysis Based on Monte Carlo Techniques. *Regulatory Toxicology and Pharmacology*. 18:438-455.

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

Finley, B.L., Proctor, D., Scott, P.K., Harrington, N., Paustenbach, D.J., and Price, P. 1994. Recommended Distributions for Exposure Factors Frequently Used in Health Risk Assessment. *Risk Analysis*. 14(4):533-553.

Finley, B.L., Scott, P.K., and Mayhall, D.A. 1994. Development of a Standard Soil-to-Skin Adherence Probability Density Function for Use in Monte Carlo Analyses of Dermal Exposure. *Risk Analysis*. 14(4):555-569.

Fischhoff, B. 1995. Risk Perception and Communication Unplugged: Twenty Years of Process. *Risk Analysis*. 15(2):137-145.

Freudenburg, W.R. 1988. Perceived Risk, Real Risk: Social Science and the Art of Probabilistic Risk Assessment. *Science*. 242:44-49.

Frey, H.C. 1992. *Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making*. Directorate for Science and Policy Programs, American Association for the Advancement of Science, Washington, DC. September.

Frey, H.C., Burmaster, D.E., Hoffman, F.O., McKone, T.E., and Price, P.S. 1995. *Quantitative Techniques for Analysis of Variability and Uncertainty in Exposure and Risk Assessment*. Workshop Notebook. Society for Risk Analysis Annual Meeting, Honolulu, Hawaii. December 3, 1995.

Frey, H.C. and Rhodes, D.S. 1996. Characterizing, Simulating, and Analyzing Variability and Uncertainty: An Illustration of Methods Using an Air Toxics Emissions Example. *Human and Ecological Risk Assessment*. March.

Gibbons, Robert D. 1994. *Statistical Methods for Groundwater Monitoring*. John Wiley & Sons, Inc. New York.

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

Goodrum, P.E., Diamond, G.L., Hassett, J.M., and Johnson, D.L. 1996. Monte Carlo Modeling of Childhood Lead Exposure: Development of a Probabilistic Methodology for Use with the US EPA IEUBK Model for Lead in Children. *Human and Ecological Risk Assessment*. 2(4):681-708.

Gori, G.B. 1995. Risk Assessment Reform: Is It for Real? *Risk Analysis*. 15(2):105-106.

Graham, J.D. and Vaupel, J.W. 1981. Value of a Life: What Difference Does It Make? *Risk Analysis*. 1(1):89-95.

Graham, J.D. 1995. The Future of Risk Regulation. *Environmental Engineer*. 31(2):22-23, 26, 28.

Gregory, R. and Lichtenstein, S. 1994. A Hint of Risk: Tradeoffs Between Quantitative and Qualitative Risk Factors. *Risk Analysis.* 14(2):199-206.

Haimes, Y.Y., Barry, T., and Lambert, J.H. 1994. When and How You Can Specify a Probability Distribution When You Don't Know Much. Workshop Proceedings. *Risk Analysis.* 14(5):661-706.

Hamby, D.M. 1994. A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environmental Monitoring Assessment.* 32:135-154.

Hattis, D., Erdreich, L., and Ballew, M. 1987. Human Variability in Susceptibility to Toxic Chemicals A Preliminary Analysis of Pharmacokinetic Data from Normal Volunteers. *Risk Analysis.* 7(4):415-426.

Hattis, D. 1990. Three Candidate 'Laws' of Uncertainty Analysis. *Risk Analysis.* 10(11).

Hattis, D. and Silver, K. 1994. Human Interindividual Variability - A Major Source of Uncertainty in Assessing Risks for Noncancer Health Effects. *Risk Analysis.* 14(4):421-431.

Hattis, D.B. and Burmaster, D.E. 1994. Assessment of Variability and Uncertainty Distributions for Practical Risk Analyses. *Risk Analysis.* 14(5):713-730.

Hawkins, N.C. and Graham, J.D. 1988. Expert Scientific Judgment and Cancer Risk Assessment: A Pilot Study of Pharmacokinetic Data. *Risk Analysis.* 8(4):615-625.

Helsel, Dennis R. and Cohn, Timothy A., 1988. Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. *Water Resources Research.* 24(12):1997-2004.

Helsel, Dennis R. 1990. Less than Obvious Statistical Treatment of Data Below the Detection Limit. *Environmental Science & Technology.* Vol. 24, No. 12.

Helsel, Dennis R. and Hirsch R.M. 1992. *Statistical Methods In Water Resources.* Elsevier, New York.

Helton, J.C. 1993. Risk, Uncertainty in Risk, and the EPA Release Limits for Radioactive Waste Disposal. *Nuclear Technology.* 101:18-39.

Helton, J.C. 1994. Treatment of Uncertainty in Performance Assessments for Complex Systems. *Risk Analysis.* 14(4):483-511.

Henrion, M. and Fischhoff, B. 1986. Assessing Uncertainty in Physical Constants. *American Journal of Physics.* 54(9):791-798.

Hofer, E. 1990. *On Some Distinctions in Uncertainty Analysis. Methods for Treatment of Different Types of Uncertainty.* Organization for Economic Cooperation and Development, Nuclear Energy Agency, Paris. PSAC/DOC (90)11.

Hoffman, F.O. and Hammonds, J.S. 1994. Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability. *Risk Analysis.* 14(5):707-712.

Hudak, D.G. 1994. Adjusting Triangular Distributions for Judgmental Bias. *Risk Analysis.* 14(6):1025-1030.

Ibrekk, H. and Morgan, M.G. 1987. Graphical Communication of Uncertain Quantities to Non-technical People. *Risk Analysis.* 7(4):519-529.

Iman, R.L. and Helton, J.C. 1988. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Analysis.* 8(1):71-90.

Iman, R.L., Helton, J.C., and Johnson, J.D. 1990. A Methodology for Grouping Source Terms for Consequence Calculations in Probabilistic Risk Assessments. *Risk Analysis.* 10(4):507-520.

Iman, R.L. and Helton, J.C. 1991. The Repeatability of Uncertainty and Sensitivity Analysis for Complex Probabilistic Risk Assessments. *Risk Analysis.* 11(4):591-606.

Israeli, M. and Nelson, C.B. 1992. Distribution and Expected Time of Residence for U.S. Households. *Risk Analysis.* 12(1):65-72.

Johnson, B.B. and Slovic, P. 1995. Presenting Uncertainty in Health Risk Assessment: Initial Studies of Its Effects on Risk Perception and Trust. *Risk Analysis.* 15(4).

Journal of Human Health and Ecological Assessment: An International Journal. 1996. Volume 2, Number 4. December.

Karlson, P. and Haimes, Y.Y. 1988. Risk-Based Analysis of Extreme Events. *Water Resources Research.* 24(1):9-20.

Koehler, James. 1997. *Personal Communication.* Associate Professor of Statistics, Department of Mathematics University of Colorado at Denver, Denver, CO.

Lambert, J.H., Matalas, N.C., Ling, C.W., Haimes, Y.Y., and Li, D. 1994. Selection of Probability Distributions in Characterizing Risk of Extreme Events. *Risk Analysis.* 14(5):731-742.

Latin, H. 1988. Good Science, Bad Regulation, and Toxic Risk Assessment. *Yale Journal on Regulation.* 5(1):89-148.

Linnerooth-Bayer, J. and Wahlstrom, B. 1991. Applications of Probabilistic Risk Assessments: The Selection of Appropriate Tools. *Risk Analysis.* 11(2):239-248.

Lynn, F.M. and Busenberg, G.J. 1995. Citizen Advisory Committees and Environmental Policy: What We Know, What's Left to Discover. *Risk Analysis.* 15(2):147-162.

MacIntosh, D.M., Suter II, G.W., and Hoffman, F.O. 1994. Uses of Probabilistic Exposure Models in Ecological Risk Assessments of Contaminated Sites. *Risk Analysis.* 14:405-419.

Magee, B., Anderson, P., and Burmaster, D. 1996. Absorption Adjustment Factor (AAF) Distributions for Polycyclic Aromatic Hydrocarbons (PAHs). *Human and Ecological Risk Assessment.* 2(4):841-873.

Manton, K.G., Stallard, E., and Vaupel, J. 1986. Alternative Models for the Heterogeneity of Mortality Risks Among the Aged. *Journal of the American Statistical Association.* 81:635-644.

Marazzi, Alfio., Joss, Johann., and Randriamiharisoa, Alex. 1993. *Algorithms, Routines, and S Functions for Robust Statistics.* Chapman & Hall, New York.

McKone, T.E. and Ryan, P.B. 1989. Human Exposure to Chemicals Through Food Chains: An Uncertainty Analysis. *Environmental. Science & Technology.* 23(9):1154-1163.

McKone, T.E. and Bogen, K.T. 1991. Predicting the Uncertainties in Risk Assessment. *Environmental. Science & Technology.* 25(10):1674-1681.

McKone, T.E. and Bogen, K.T. 1992. Uncertainties in Health-Risk Assessment: An Integrated Case Study Based on Tetrachloroethylene in California Groundwater. *Reg. Tox. Pharm.* 15:86-103.

Mitsiopoulos, J. and Haimes, Y.Y. 1989. Generalized Quantification of Risk Associated with Extreme Events. *Risk Analysis.* 9(2):243-254.

Moore, David S. and McCabe, George P. 1993. *Introduction to the Practice of Statistics*, 2nd Edition. W. H. Freeman and Co., New York.

Morgan, M.G. and Henrion, M. 1988. *Uncertainty: A Guide for Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Department of Engineering and Public Policy, Carnegie-Mellon University. Pittsburgh, PA.

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

Morgan, M.G. and Lave, L. 1990. Ethical Considerations in Risk Communication Practice and Research. *Risk Analysis.* 10(3):355-358.

Morgan, M. Granger, Fishhoff, Baruch, Bostrom, Ann, Lave, Lester, and Atman, Cynthia J. 1992. Communicating Risk to the Public. *Risk Analysis.* 26(11):2048-2056.

Murray, D.M. and Burmaster, D.E. 1992. Estimated Distributions for Total Body Surface Area of Men and Women in the United States. *J. Expos. Anal. and Environ. Epidemiol.* 2(4):451-461.

National Academy of Sciences. 1983. *Risk Assessment in the Federal Government: Managing the Process*, Committee on the Institutional Means for Assessment of Risks to Public Health. National Academy Press, Washington, DC.

National Academy of Sciences. 1994. *Science and Judgment in Risk Assessment.* Committee on Risk Assessment of Hazardous Air Pollutants, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council, Washington, DC.

National Academy of Sciences. 1996. *Understanding Risk: Informing Decisions in a Democratic Society.* National Academy Press, Washington, DC.

National Council on Radiation Programs. 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessment Related to Environmental Contamination.* NCRP Commentary No. 14, Bethesda, MD.

Paustenbach, D.J. 1989. Health Risk Assessments: Opportunities and Pitfalls. *Columbia Journal of Environmental Law.* 14(2):379-410.

Portier, C.J. and Kaplan, N.L. 1989. Variability of Safe Dose Estimates When Using Complicated Models of the Carcinogenic Process. A Case Study: Methylene Chloride. *Fundamental and Applied Toxicology.* 13:533-544.

Price, P.S., Sample, J., and Strieter, R. 1992. Determination of Less-Than-Lifetime Exposures to Point Source Emissions. *Risk Analysis.* 12(3):367-382.

Price, P.S., Curry, C.L., Goodrum, P.E., Gray, M.N., McCrodden, J.I., Harrington, N.W., Carlson-Lynch, H., and Keenan, R.E. 1996. Monte Carlo Modeling of Time-Dependent Exposures Using a Microexposure Event Approach. *Risk Analysis.* 16(3):339-348.

Rai, S.N., Krewski, D., and Bartlett, S. 1996. A General Framework for the Analysis of Uncertainty and Variability in Risk Assessment. *Human and Ecological Risk Assessment.* 2(4):972-989.

Raiffa, H. 1968. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Addison-Wesley Publishing Co., Reading, MA.

Rice, John A. 1995. *Mathematical Statistics and Data Analysis.* Duxbury Press, Belmont, California.

Richardson, G.M. and Allan, M. 1996. A Monte Carlo Assessment of Mercury Exposure and Risks from Dental Amalgam. *Human and Ecological Risk Assessment.* 2(4):709-761.

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

Rugen, P. and Callahan, B. 1996. An Overview of Monte Carlo, A 50 Year Perspective. *Human and Ecological Risk Assessment.* 2(4):671-680.

Ryan, E.A., Hawkins, E.T., Magee, B., and Santos, S.L. 1987. Assessing Risk from Dermal Exposure at Hazardous Waste Sites. *US Environmental Protection Agency, Superfund Public Health Assessment Manual.* Washington, DC. November, pg. 166-168.

Salmento, J.S., Rubin, E.S., and Finkel, A.M. 1989. A Review of @RISK. *Risk Analysis.* 9(2):255-257.

Seiler, F.A. 1987. Error Propagation for Large Errors. *Risk Analysis.* 7(4)509-518.

Seiler, Fritz A., and Alvarez, Joseph L. 1996. On the Selection of Distributions for Stochastic Variables. *Risk Analysis.* 16(1):5-18.

Shapiro, S.S. and Wilk, M.B. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika.* 52.

Shaw, C.D. and Burmaster, D.E. 1996. Distributions of Job Tenure for U.S. Workers in Selected Industries and Occupations. *Human and Ecological Risk Assessment.* 2(4):798-819.

Shlyakhter, Alexander I. 1994. An Improved Framework for Uncertainty Analysis: Accounting for Unsuspected Errors. *Risk Analysis.* 14(4):441-447.

Shlyakhter, Alexander I., Mirny, L., Vlasov, A., and Wilson, R. 1996. Monte Carlo Modeling of Epidemiological Studies. *Human and Ecological Risk Assessment.* 2(4):920-938.

Slob, W. 1994. Uncertainty Analysis in Multiplicative Models. *Risk Analysis.* 14(4):571-576.

Slovic, P. 1993. Perceived Risk, Trust, and Democracy. *Risk Analysis.* 13(6):675-682.

Smith, A.E., Ryan, P.B., and Evans, J.S. 1992. The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating the Population Distribution of Risk. *Risk Analysis.* 12(4):467-474.

Smith, E.P. and Shugart, H.H. 1994. Issue Paper on Uncertainty in Ecological Risk. Assessment. *Ecological Risk Assessment Issue Papers.* U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC. EPA/630-R-94-009.

Smith, Roy L. 1994. Use of Monte Carlo Simulation for Human Exposure Assessment at a Superfund Site. *Risk Analysis.* 14(4):433-439.

Stanek III, E.J. 1996. Estimating Exposure Distributions: A Caution for Monte Carlo Risk Assessment. *Human and Ecological Risk Assessment.* 2(4):874-891.

Taylor, A.C. 1993. Using Objective and Subjective Information to Develop Distributions for Probabilistic Exposure Assessment. *J. Expos. Anal. Environ. Epidemiol.* 3(3):285-298.

Thompson, K.M., Burmaster, D.E., and Crouch, E.A.C. 1992. Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments. *Risk Analysis.* 12(1):53-63.

Thompson, K.M. and Burmaster, D.E. 1991. Parametric Distributions for Soil Ingestion by Children. *Risk Analysis.* 11(12):339-342.

Thompson, K.M. and Graham, J.D. 1996. Going Beyond the Single Number: Using Probabilistic Risk Assessment to Improve Risk Management. *Human and Ecological Risk Assessment.* 2(4):1008-1034.

U.S. Environmental Protection Agency. 1985. *Methodology for Characterization of Uncertainty in Exposure Assessments.* Research Triangle Institute. EPA600-08-85-009.

U.S. Environmental Protection Agency. 1985. *Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments.* Office of Health and Environmental Assessment, Washington, DC. EPA600-08-85-010.

U.S. Environmental Protection Agency. 1989. *Risk Assessment Guidance for Superfund – Volume I: Human Health Evaluation Manual (Part A).* Office of Emergency and Remedial Response, Washington DC. EPA/540/1-89/002. December.

U.S. Environmental Protection Agency. 1990. *A Rationale for the Assessment of Errors in the Sampling of Soils.* EPA600-04-90-013.

U.S. Environmental Protection Agency. 1991. Memorandum from Don R. Clay, re: *Role of the Baseline Risk Assessment in Superfund Remedy Selection Decisions.* Office of Solid Waste and Emergency Response, Washington, DC. OSWER 9355.0-30. April.

U.S. Environmental Protection Agency. 1992. *Methods for Evaluating the Attainment of Cleanup Standards Volume 1: Soils and Solid Media.* Statistical Policy Branch, Office of Policy, Planning, and Evaluation. EPA/230/02-89-042. February.

U.S. Environmental Protection Agency. 1992. *Guidance on Implementing the Deputy Administrator's Risk Characterization Memorandum* (OSWER 9285.7-13). Memo from H.L. Longest, Director, Office of Emergency and Remedial Response, Washington, DC. OSWER 9285.7-13. February 26, 1992.

U.S. Environmental Protection Agency. 1992. Memorandum to Assistant Administrators from F.H. Habicht, Deputy Administrator. *Guidance on Risk Characterization for Risk Managers and Risk Assessors*. Washington, DC. February 26, 1992.

U.S. Environmental Protection Agency. 1992. *Guidance for Data Useability in Risk Assessment (Part A).* Office of Emergency and Remedial Response, Washington, DC. 9285.7-09A. April.

U.S. Environmental Protection Agency. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term. Intermittent Bulletin. 1(1).* Office of Solid Waste and Emergency Response, Washington, DC. May.

U.S. Environmental Protection Agency. 1992. *Guidelines for Exposure Assessment; Notice, (Final).* Exposure Assessment Group, Office of Health and Environmental Assessment, Washington, DC. *Federal Register*, Vol. 57, No. 104, 22887-22938. May 29, 1992.

U.S. Environmental Protection Agency, Region VIII. 1993. *Superfund's Standard Default Exposure Factors For the Central Tendency and Reasonable Maximum Exposure,* Draft. Denver, CO.

U.S. Environmental Protection Agency. 1993. The ABCs of Risk Assessment: Some Basic Principles Can Help People Understand Why Controversies Occur. *EPA Journal.* January/February/March.

U.S. Environmental Protection Agency, Region III. 1994. *Use of Monte Carlo Simulation in Risk Assessments.* Hazardous Waste Management Division, Office of Superfund Programs, Philadelphia, PA. February. EPA903-F-94-001.

U.S. Environmental Protection Agency, Region VIII. 1994. *Calculating the Concentration Term for Risk Assessment: Use of One "C" Term to Estimate Lower Average and Upper RME Risk Range.* Region VIII Technical Guidance, Denver, CO. January.

U.S. Environmental Protection Agency, Region VIII. 1994. *Health Risk Assessments for Superfund Sites*: Fact Sheet. Region VIII Technical Guidance, Denver, CO. December.

U.S. Environmental Protection Agency. 1995. *Soil Screening Guidance: Technical Background Document.* Office of Solid Waste and Emergency Response, Washington, DC. EPA/540/R-95/126.

U.S. Environmental Protection Agency. 1995. *Policy for Risk Characterization.* Science Policy Council, Washington, DC. February.

U.S. Environmental Protection Agency, Region VIII. 1995. *Use of Monte Carlo Simulation in Performing Risk Assessments (Region 8 Superfund Technical Guidance).* Hazardous Waste Management Division, Superfund Management Branch, Technical Section, Denver, CO. RA-10. September.

U.S. Environmental Protection Agency. 1996. *Soil Screening Guidance: User's Guide.* Office of Solid Waste and Emergency Response, Washington, DC. EPA/540/R-96/018.

U.S. Environmental Protection Agency. 1996. *The US EPA's Perspective on Risk Probabilistic Assessments*, Presented by Dr. Timothy M. Barry, Office of Program Planning and Evaluation. March.

U.S. Environmental Protection Agency. 1996. *Proposed Guidelines for Carcinogen Risk Assessment.* Office of Research and Development, Washington, DC. EPA600-P-92-003C. April.

U.S. Environmental Protection Agency. 1996. *Discussion Issues - Workshop on Monte Carlo Analysis.* Risk Assessment Forum, New York, NY. EPA630-R-96-010. May.

U.S. Environmental Protection Agency. 1996. *Exposure Factors Handbook (Update to Exposure Factors Handbook EPA600-08-89-043). Volume 1 - General Factors.* Office of Research and Development, National Center for Environmental Assessment, Washington, DC. August.

U.S. Environmental Protection Agency. 1996. *Summary Report for the Workshop on Monte Carlo Analysis.* Risk Assessment Forum, Office of Research and Development. EPA/630/R-96/010. September.

U.S. Environmental Protection Agency. 1996. *Policy Statement for the Use of Monte Carlo Analysis in Agency Risk Assessments.* Risk Assessment Forum, Science Policy Council Steering Committee. November.

U.S. Environmental Protection Agency. 1997. *Guiding Principles for Monte Carlo Analysis.* Risk Assessment Forum. Washington, DC. EPA/630/R-97/001. March.

Venables, W.N. and Ripley, B.B. 1994. *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

Wallace, L.A., Duan, N., and Ziegenfus, R. 1994. Can Long-Term Exposure Distributions Be Predicted from Short-Term Measurements? *Risk Analysis.* 14(1):75-85.

Weiss, Neil A. 1996. *Elementary Statistics*, 3rd Edition. Addison-Wesley Publishing Company, Inc., New York.

Appendix B

Glossary

GLOSSARY

ABSORBED DOSE — The amount of a substance penetrating the exchange boundaries of an organism after contact. Absorbed dose is calculated from the intake and the absorption efficiency. It usually is expressed as mass of a substance absorbed into the body per unit body weight per time (e.g., mg/kg-day).

ACCEPTABLE DAILY INTAKE (ADI) — The daily intake of chemical which, during a lifetime, appears to be without appreciable risk. An estimate similar in concept to the RfD, but derived using a less strictly defined methodology. RfDs have replaced ADIs as the Agency's preferred values for use in evaluating potential noncarcinogenic health effects resulting from exposure to a chemical.

ACCEPTABLE INTAKE FOR CHRONIC EXPOSURE (AIC) — An estimate similar in concept to the RfD, but derived using a less strictly defined methodology. Chronic RfDs have replaced AICs as the Agency's preferred values for use in evaluating potential noncarcinogenic health effects resulting from chronic exposure to a chemical.

ACCEPTABLE INTAKE FOR SUBCHRONIC EXPOSURE (AIS) — An estimate similar in concept to the subchronic RfD, but derived using a less strictly defined methodology. Subchronic RfDs have replaced AISs as the Agency's preferred values for use in evaluating potential noncarcinogenic health effects resulting from subchronic exposure to a chemical.

ADMINISTERED DOSE — The mass of substance given to an organism and in contact with an exchange boundary (e.g., gastrointestinal tract) per unit body weight per unit time (e.g., mg/kg-day).

BAYESIAN APPROACH — The Bayesian or subjective view is that the probability of an event is the degree of belief that a person has, given some state of knowledge, that the event will occur.

BETA DISTRIBUTION — A flexible, bounded PDF described by two shape parameters. It is commonly used when a range of the random variable is known.

BOXPLOT — Graphical representation showing the center and spread of a distribution, along with a display of outliers.

C-TERM	Concentration term used in risk calculations. Per USEPA guidance, this is the 95-percent UCL of the mean sample concentrations.
CANCER SLOPE FACTOR	A plausible upper-bound estimate of the probability of a response per unit intake of a chemical over a lifetime. The slope factor 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.
CARCINOGENESIS	The production of cancer.
CENTRAL LIMIT THEOREM	For a relatively large sample size, the random variable $\bar{x}$ (the mean of the samples) is normally distributed, regardless of the population's distribution.
CHRONIC DAILY INTAKE (CDI)	Exposure expressed as mass of a substance contacted per unit body weight per unit time, averaged over a long period of time (as a Superfund program guideline, 7 years to a lifetime).
CHRONIC REFERENCE DOSE (RfD)	An estimate (with uncertainty spanning perhaps an order of magnitude or greater) 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 long-term exposure to a compound (as a Superfund program guideline, 7 years to lifetime).
COEFFICIENT OF VARIATION	Estimate of relative standard deviation. Equals the standard deviation divided by the mean.
CONFIDENCE INTERVAL	The range within which one has a given level of confidence that the range includes the true value of the unknown parameter (e.g. a 95-percent confidence interval for a parameter means that 95 percent of the time the true value of that parameter will be within the interval).
CONTACT RATE	The amount of a substance given to an organism, especially through dermal contact.
CORRELATION	Simultaneous increase or decrease in value of two numerically valued random variables.

CUMULATIVE DISTRIBUTION FUNCTION (CDF)	A cumulative distribution function (CDF) is a commonly used function to mathematically describe a random variable. The CDF gives the cumulative probability of all outcomes of the random variable at or below a specific value.
DATA COLLECTION (FOR RISK ASSESSMENT)	Component of the risk assessment process entailing identification of chemicals of potential concern, through evaluation of historic uses and potential release/disposal of chemicals, and data collection and evaluation.
DATA QUALITY OBJECTIVES (DQOs)	Qualitative and quantitative statements to ensure that data of known and documented quality are obtained during an RI/FS to support an Agency decision.
DENSITY PLOT	An empirical estimate of the PDF from sample data.
DETECTION LIMIT (DL)	The lowest amount that can be distinguished from the normal "noise" of an analytical instrument or method. Often used synonymously for "quantitation limit" or the lowest concentration that can be routinely quantified under specified limits of precision and accuracy.
DETERMINISTIC APPROACH	An approach which uses single values for each variable to estimate the risk.
DISTRIBUTION	The pattern of variation of a random variable.
DOSE-RESPONSE EVALUATION	The process of quantitatively evaluating toxicity information and characterizing the relationship between the dose of a contaminant administered or received and the incidence of adverse health effects in the exposed population. From the quantitative dose-response relationship, toxicity values are derived that are used in the risk characterization step to estimate the likelihood of adverse effects occurring in humans at different exposure levels.

DOUBLE LOOP APPROACH An application of stochastic analyses where both variability and uncertainty are being represented with probability distributions. In such analyses, uncertainty and variability distributions for input variables are applied to characterize the distribution of variability in the population and to estimate uncertainty bounds for this distribution. It is "double loop" in the sense that in computer programs for stochastic analysis, the values of the uncertain parameters are selected first in the "outer loop" of the code and then the "inner loop" of code treats these values as fixed quantities and generates an estimate of the population distribution under these. The population statistics corresponding to multiple "outer loop" parameter values are then compiled to represent uncertainty in the estimates of the population distribution conditions (i.e., variability calculations are "nested" within uncertainty calculations).

EXPERT Someone who (1) has training and experience in the subject area resulting in extensive knowledge of the field, (2) has access to relevant information, (3) has an ability to process and effectively use the information, and (4) is recognized by his or her peers or those conducting the study as qualified to provide judgments about assumptions, models, and model parameters at the level of detail required

EXPERT JUDGMENT A qualitative or quantitative inference or evaluation based on an assessment of data, assumptions, criteria, models and parameters in response to questions posed in the expert's area of expertise.

EXPOSURE Contact of an organism with a chemical or physical agent. Exposure is quantified as the amount of the agent available at the exchange boundaries of the organism (e.g., skin, lungs, gut) and available for absorption.

EXPOSURE ASSESSMENT The determination or estimation (qualitative or quantitative) of the magnitude, frequency, duration, and route of exposure. A process that integrates information on chemical releases, environmental measurements, human behavior, and human physiology to estimate the exposure levels of doses of chemicals received by humans.

EXPOSURE EVENT An incident of contact with a chemical or physical agent. An exposure event can be defined by time (e.g., day, hour) or by the incident (e.g., eating a single meal of contaminated fish).

EXPOSURE PATHWAY The course a chemical or physical agent takes from a source to an exposed organism. An exposure pathway describes a unique mechanism by which an individual or population is exposed to chemicals or physical agents at or originating from a site. Each exposure pathway includes a source or release for a source, an exposure point, and an exposure route. If the exposure point differs from the source, a transport/exposure medium (e.g., air) or media (in cases of intermedia transfer) also is included.

EXPOSURE POINT A location of potential contact between an organism and a chemical or physical agent.

EXPOSURE ROUTE The way a chemical or physical agent comes in contact with an organism (i.e., by ingestion, inhalation, dermal contact).

EXTRAPOLATE To make use of a regression line or other model outside the range of the data to which the model is fitted.

GOODNESS-OF-FIT TEST A formal way to verify that the chosen distribution is consistent with the sample data.

HAZARD IDENTIFICATION The process of determining whether exposure to an agent can cause an increase in the incidence of a particular adverse health effect (e.g., cancer, birth defect) and whether the adverse health effect is likely to occur in humans.

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 The ratio of a single substance exposure level over a specified time period (e.g., subchronic) to a reference dose for that substance derived from a similar exposure period.

HEALTHY WORKER EFFECT	Potential bias introduced when applying information gathered in industrial settings to the general population. The bias is based in the fact that the working population is generally healthier than the general population. Most often this is encountered in epidemiological studies.
HISTOGRAM	A plot of the range of values of a variable into intervals and displays only the count of the observations that fall into each interval.
INTAKE	A measure of exposure expressed as the mass of a substance in contact with the exchange boundary per unit body weight per unit time (e.g., mg chemical/kg-day). Also, termed the normalized exposure rate; equivalent to administered dose.
INTERQUARTILE RANGE (IQR)	Difference between the third quartile (75th percentile) and the first quartile (25th percentile).
KRIGING	A statistical interpolation method that chooses the best linear unbiased estimate for the variable in question.
LATIN HYPERCUBE SAMPLING (LHS)	A technique that uses random sampling within equal intervals of the distribution.
LIFETIME AVERAGE DAILY INTAKE	Exposure expressed as mass of a substance contacted per unit body weight per unit time, averaged over a lifetime.
LOGNORMAL DISTRIBUTION	A frequency distribution for a set of variable data where the log values are normally distributed (data are transformed using $Y = \ln X$).
MAXIMUM	Highest value that a parameter can have.
MEAN	The sum of all the observations divided by the number of observations.
MEASUREMENT ERROR	Error introduced through imperfections in measurement techniques or equipment.
MEDIAN	That value above which and below which half the population lies.
MINIMUM	The minimum is the smallest value of a data set or process.

MODEL UNCERTAINTY	Uncertainty created through the use of a mathematical model to represent physical processes or phenomenon.
MONTE CARLO SIMULATION (MONTE CARLO ANALYSIS)	A method of calculating the probability of an event using values randomly selected from sets of data, repeating the process many times and deriving the probability from the distributions of the aggregated data.
NONDETECTS	Chemicals that are not detected in a particular sample above a certain limit, usually the quantitation limit for the chemical in that sample.
NONPARAMETRIC APPROACH	One that does not depend for its validity upon the data being drawn from a specific distribution, such as the normal or lognormal. A distribution-free technique.
NORMAL DISTRIBUTION	A frequency distribution for a set of variable data represented by a bell-shaped curve symmetrical about the mean.
PARAMETER	The constants and independent variables which define a mathematical equation or model.
PARAMETRIC APPROACH	A method of probabilistic analysis in which defined analytic probability distributions are used to represent the random variables, and mathematical techniques (e.g., calculus) are used to get the resultant distribution for a function of these random variables.
PERCENTILES	The value that exceeds X percent of the observations.
PHARMACOKINETICS/ TOXICOKINETICS	The study and modeling of the disposition of chemicals in the body. The principle purpose of pharmacokinetic modeling is to predict the concentration of chemical at the target site and describe the relationship between exposed dose and target dose.
POPULATION	The total collection of observations that is of interest.
POSITIVE DATA	Analytical results for which measurable concentrations (i.e., above a quantitation limit) are reported.
PROBABILISTIC APPROACH	An approach which uses a group of possible values for each variable to estimate risk.

PROBABILITY DENSITY FUNCTION (PDF)	Distribution of values for a random variable, each value having a specific probability of occurrence.
QUANTILE-QUANTILE (Q-Q) PLOT	The Q-Q plot portrays the quantiles (percentiles divided by 100) of the sample data against the quantiles of another data set or theoretical distribution (e.g., normal distribution). By comparing the data to a theoretical distribution with a straight line, departures from the distribution are more easily perceived.
QUANTITATION LIMIT (QL)	The lowest level at which a chemical can be accurately and reproducibly quantitated. Usually equal to the instrument detection limit multiplied by a factor of three to five, but varies for different chemicals and different samples.
RANDOM ERROR	Error caused by making inferences from a limited database.
REASONABLE MAXIMUM EXPOSURE (RME)	The highest exposure that is reasonably expected to occur at a particular site.
RECEPTOR	The person, whether real or theoretical, that is exposed to a chemical.
REGRESSION ANALYSIS	Derivation of an equation which can be used to estimate the unknown value of one variable on the basis of the known value of the other variable.
RISK	The likelihood of injury, disease, or death.
RISK ASSESSMENT	The objective process by which scientific data are analyzed to describe the form, dimension, and characteristics of risk.
RISK CHARACTERIZATION	Component of the risk assessment process entailing summarization and interpretation of information gathered in previous steps (data evaluation, exposure assessment, toxicity assessment).
RISK MANAGEMENT	The process by which regulatory decisions are made using all available information (including, but not limited to, the results and recommendations of the risk assessment).
SCATTERPLOT	Diagram that plots each data point as a distinct point on the plot.

SENSITIVITY ANALYSIS	An analysis that attempts to provide a ranking of the model's input parameters with respect to their contribution to model output variability or uncertainty.
SIMPLE RANDOM SAMPLING (SRS)	Sampling procedure by which each possible sample is equally likely to be the one selected.
SKEWNESS	The measure of asymmetry of a frequency distribution of data.
STANDARD DEVIATION (s)	A measure of dispersion which is expressed in the same units as the measurements. The positive square root of the variance.
SLOPE FACTOR (SF)	A plausible upper-bound estimate of the probability of a response per unit intake of a chemical over a lifetime. The slope factor 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.
STOCHASTIC	Term referring to a process involving a random variable.
SUBCHRONIC DAILY INTAKE (SDI)	Exposure expressed as mass of a substance contacted per unit body weight per unit time, averaged over a portion of a lifetime (as a Superfund program guideline, 2 weeks to 7 years).
SURROGATE	A variable or parameter that is used to model the variable one actually needs to measure. Usually the surrogate variable is one which can be more easily measured than the variable it is replacing.
SYSTEMIC ERROR (BIAS)	Non-random error introduced through flaws in the data gathering process.
TOXICITY ASSESSMENT	Component of the Risk Assessment process entailing an extensive review of the adverse (known or suspected) effects associated with exposure to a specific chemical, and the dose required to induce the effect.
TOXICITY VALUE	A numerical expression of a substance's dose-response relationship that is used in risk assessments. The most common toxicity values used in Superfund program risk assessments are reference doses (for noncarcinogenic effects) and slope factors (for carcinogenic effects).

TRIANGULAR DISTRIBUTION	Distribution with a triangular shape. It is characterized by its minimum, maximum and mode (most likely) values. It is often used to represent a truncated log-normal or normal distribution if the there is little information available on the parameter being modeled.
UNCERTAINTY	Arising from lack of complete information about contaminant levels and exposure factors that may be reducible through further measurements.
UNCERTAINTY ANALYSIS	See Monte Carlo analysis.
UPPER CONFIDENCE LIMIT (UCL)	An upper confidence limit (UCL) is an upper bound on an unknown statistical parameter with a specified probability. This expression describes the outcome of any one sample. For example, the 95 percent UCL on the mean, computed from a sample of data, is the value at which the true (but unknown) mean is less than with 95 percent confidence.
VARIABLE	A quantity capable of assuming any of a set of values.
VARIABILITY	Represents true heterogeneity in contaminant levels and human exposure factors and cannot be reduced by additional data collection.
WEIGHT-OF-EVIDENCE CLASSIFICATION	An USEPA classification system for characterizing the extent to which the available data indicate that an agent is a human carcinogen. Recently, USEPA has developed weight-of-evidence classification systems for some other kinds of toxic effects, such as developmental effects.

Appendix C

USEPA Reference Documents

POLICY FOR USE OF PROBABILISTIC ANALYSIS IN RISK ASSESSMENT
at the U.S. Environmental Protection Agency

May 15, 1997

Guiding Principles for Monte Carlo Analysis (EPA/630/R-97/001)

INTRODUCTION

The importance of adequately characterizing variability and uncertainty in risk assessments has been emphasized in several science and policy documents. These include the 1992 U.S. Environmental Protection Agency (EPA) Exposure Assessment Guidelines, the 1992 EPA Risk Assessment Council (RAC) Guidance, the 1995 EPA Policy for Risk Characterization, the EPA Proposed Guidelines for Ecological Risk Assessment, the EPA Region 3 Technical Guidance Manual on Risk Assessment, the EPA Region 8 Superfund Technical Guidance, the 1994 National Academy of Sciences "Science and Judgment in Risk Assessment," and the report by the Commission on Risk Assessment and Risk Management. As part of the implementation of the recommendations contained in these reports, the Agency is issuing guidance on the appropriate use of an application for analyzing variability and uncertainty in Agency risk assessments.

This policy and the guiding principles attached are designed to support the use of various techniques for characterizing variability and uncertainty. Further, the policy defines a set of Conditions for Acceptance. These conditions are important for ensuring good scientific practice in quantifying uncertainty and variability. In accordance with EPA's 1995 Policy for Risk Characterization, this policy also emphasizes the importance of clarity, transparency, reasonableness, and consistency in risk assessments.

There are a variety of different methods for characterizing uncertainty and variability. These methods cover a broad range of complexity from the simple comparison of discrete points to probabilistic techniques like Monte Carlo analysis. Recently, interest in using Monte Carlo analysis for risk assessment has increased. This method has the advantage of allowing the analyst to account for relationships between input variables and of providing the flexibility to investigate the effects of different modeling assumptions. Experience has shown that to benefit fully from the advantages of such probabilistic techniques as Monte Carlo analysis, certain standards of practice are to be observed. The Agency is issuing, therefore, this policy statement and associated guiding principles. While Monte Carlo analysis is the most frequently encountered probabilistic tool for analyzing variability and uncertainty in risk assessments, the intent of this policy is not to indicate that Monte Carlo analysis is the only acceptable approach for Agency risk assessments. The spirit of this policy and the Conditions for Acceptance described herein are equally applicable to other methods for analyzing variability and uncertainty.

POLICY STATEMENT

It is the policy of the U.S. Environmental Protection Agency that such probabilistic analysis techniques as Monte Carlo analysis, given adequate supporting data and credible assumptions, can be viable statistical tools for analyzing variability and uncertainty in risk assessments. As such, and provided that

statistical tools for analyzing variability and uncertainty in risk assessments. As such, and provided that the conditions described below are met, risk assessments using Monte Carlo analysis or other probabilistic techniques will be evaluated and utilized in a manner that is consistent with other risk assessments submitted to the Agency for review or consideration. It is not the intent of this policy to recommend that probabilistic analysis be conducted for all risk assessments supporting risk management decisions. Such analysis should be a part of a tiered approach to risk assessment that progresses from simpler (e.g., deterministic) to more complex (e.g., probabilistic) analyses as the risk management situation requires. Use of Monte Carlo or other such techniques in risk assessments shall not be cause, *per se*, for rejection of the risk assessment by the Agency. For human health risk assessments, the application of Monte Carlo and other probabilistic techniques has been limited to exposure assessments in the majority of cases. The current policy, Conditions for Acceptance and associated guiding principles are not intended to apply to dose response evaluations for human health risk assessment until this application of probabilistic analysis has been studied further. In the case of ecological risk assessment, however, this policy applies to all aspects including stressor and dose-response assessment.

CONDITIONS FOR ACCEPTANCE

When risk assessments using probabilistic analysis techniques (including Monte Carlo analysis) are submitted to the Agency for review and evaluation, the following conditions are to be satisfied to ensure high quality science. These conditions, related to the good scientific practices of transparency, reproducibility, and the use of sound methods, are summarized here and explained more fully in the Attachment, "Guiding Principles for Monte Carlo Analysis."

1. The purpose and scope of the assessment should be clearly articulated in a "problem formulation" section that includes a full discussion of any highly exposed or highly susceptible subpopulations evaluated (e.g., children, the elderly). The questions the assessment attempts to answer are to be discussed and the assessment endpoints are to be well defined.
2. The methods used for the analysis (including all models used, all data upon which the assessment is based, and all assumptions that have a significant impact upon the results) are to be documented and easily located in the report. This documentation is to include a discussion of the degree to which the data used are representative of the population under study. Also, this documentation is to include the names of the models and software used to generate the analysis. Sufficient information is to be provided to allow the results of the analysis to be independently reproduced.
3. The results of sensitivity analyses are to be presented and discussed in the report. Probabilistic techniques should be applied to the compounds, pathways, and factors of importance to the assessment, as determined by sensitivity analyses or other basic requirements of the assessment.
4. The presence or absence of moderate to strong correlations or dependencies between the input variables is to be discussed and accounted for in the analysis, along with the effects these have on the output distribution.
5. Information for each input and output distribution is to be provided in the report. This includes tabular and graphical representations of the distributions (e.g., probability density function and cumulative distribution function plots) that indicate the location of any point estimates of interest (e.g., mean, median, 95th percentile). The selection of distributions is to be explained and justified. For both the input and output distributions, variability and uncertainty are to be differentiated where possible.
6. The numerical stability of the central tendency and the higher end (i.e., tail) of the output distributions are to be presented and discussed.
7. Calculations of exposures and risks using deterministic (e.g., point estimate) methods are to be reported if possible. Providing these values will allow comparisons between the probabilistic analysis and past or screening level risk assessments. Further, deterministic estimates may be used to answer scenario specific questions and to facilitate risk communication. When comparisons are made, it is important to explain the similarities and differences in the underlying data, assumptions, and models.
8. Since fixed exposure assumptions (e.g., exposure duration, body weight) are sometimes embedded in the toxicity metrics (e.g., Reference Doses, Reference Concentrations, unit cancer risk factors), the exposure estimates from the probabilistic output distribution are to be aligned with the toxicity metric.

LEGAL EFFECT

This policy and associated guidance on probabilistic analysis techniques do not establish or affect legal rights or obligations. Rather, they confirm the Agency position that probabilistic techniques can be viable statistical tools for analyzing variability and uncertainty in some risk assessments. Further, they outline relevant Conditions for Acceptance and identify factors Agency staff should consider in implementing the policy.

The policy and associated guidance do not stand alone; nor do they establish a binding norm that is finally determinative of the issues addressed. Except where otherwise provided by law, the Agency's decision on conducting a risk assessment in any particular case is within the Agency's discretion. Variations in the application of the policy and associated guidance, therefore, are not a legitimate basis for delaying action on Agency decisions.

IMPLEMENTATION

Assistant Administrators and Regional Administrators are responsible for implementation of this policy within their organizational units. The implementation strategy is divided into immediate and follow-up activities.

Immediate Activities

To assist EPA program and regional offices with this implementation, initial guidance on the use of one probabilistic analysis tool, Monte Carlo analysis, is provided in the Attachment, "Guiding Principles for Monte Carlo Analysis" (EPA/630/R-97/001). The focus of this guidance is on Monte Carlo analysis because it is the most frequently encountered technique in human health risk assessments. Additional information may be found in the "Summary Report for the Workshop on Monte Carlo Analysis" (EPA/630/R-96/010). This report summarizes discussions held during the May 1996 Risk Assessment Forum sponsored workshop that involved leading experts in Monte Carlo analysis.

Follow-Up Activities

To prepare for the use and evaluation of probabilistic analysis methods, including Monte Carlo analysis, within the next year, EPA's Risk Assessment Forum (RAF) will develop illustrative case studies for use as guidance and training tools. Further, the RAF will organize workshops or colloquia to facilitate the development of distributions for selected exposure factors. EPA's National Center for Environmental Assessment (NCEA) will develop an Agency training course on probabilistic analysis methods, including Monte Carlo analysis for both risk assessors and risk managers which will become available during Fiscal Year (FY) 1997 or FY 1998. Also, NCEA will develop detailed technical guidance for the quantitative analysis of variability and uncertainty.

In the longer term, various Regions, Programs and the Office of Research and Development (ORD) may need to modify existing or develop new guidelines or models to facilitate use of such techniques as Monte Carlo analysis. Also, the NCEA will revise or update the Exposure Factors Handbook to include distributional information. ORD's National Exposure Research Laboratory

(NERL) has formed a modeling group that may provide assessment and analysis advice to Program and Regional Offices. The issue of using probabilistic techniques, including Monte Carlo analysis in the dose response portion of human health risk assessments requires further study. NCEA will conduct research in this area and additional guidance will be provided if necessary.

Fred Hansen

Deputy Administrator

EPA/630/R-97/001
March 1997

Guiding Principles for Monte Carlo Analysis

Technical Panel

Office of Prevention, Pesticides, and Toxic Substances

Michael Firestone (Chair) Penelope Fenner-Crisp

Office of Policy, Planning, and Evaluation

Timothy Barry

Office of Solid Waste and Emergency Response

David Bennett Steven Chang

Office of Research and Development

Michael Callahan

Regional Offices

AnneMarie Burke (Region I) Jayne Michaud (Region I)
Marian Olsen (Region II) Patricia Cirone (Region X)

Science Advisory Board Staff

Donald Barnes

Risk Assessment Forum Staff

William P. Wood Steven M. Knott

Risk Assessment Forum
U.S. Environmental Protection Agency
Washington, DC 20460

DISCLAIMER

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

PREFACE

The U.S. Environmental Protection Agency (EPA) Risk Assessment Forum was established to promote scientific consensus on risk assessment issues and to ensure that this consensus is incorporated into appropriate risk assessment guidance. To accomplish this, the Risk Assessment Forum assembles experts throughout EPA in a formal process to study and report on these issues from an Agency-wide perspective. For major risk assessment activities, the Risk Assessment Forum has established Technical Panels to conduct scientific reviews and analyses. Members are chosen to assure that necessary technical expertise is available.

This report is part of a continuing effort to develop guidance covering the use of probabilistic techniques in Agency risk assessments. This report draws heavily on the recommendations from a May 1996 workshop organized by the Risk Assessment Forum that convened experts and practitioners in the use of Monte Carlo analysis, internal as well as external to EPA, to discuss the issues and advance the development of guiding principles concerning how to prepare or review an assessment based on use of Monte Carlo analysis. The conclusions and recommendations that emerged from these discussions are summarized in the report "Summary Report for the Workshop on Monte Carlo Analysis" (EPA/630/R-96/010). Subsequent to the workshop, the Risk Assessment Forum organized a Technical Panel to consider the workshop recommendations and to develop an initial set of principles to guide Agency risk assessors in the use of probabilistic analysis tools including Monte Carlo analysis. It is anticipated that there will be need for further expansion and revision of these guiding principles as Agency risk assessors gain experience in their application.

Introduction

The importance of adequately characterizing variability and uncertainty in fate, transport, exposure, and dose-response assessments for human health and ecological risk assessments has been emphasized in several U.S. Environmental Protection Agency (EPA) documents and activities. These include:

- the 1986 Risk Assessment Guidelines;
- the 1992 Risk Assessment Council (RAC) Guidance (the *Habicht memorandum*);
- the 1992 Exposure Assessment Guidelines; and
- the 1995 Policy for Risk Characterization (the *Browner memorandum*).

As a follow up to these activities EPA is issuing this policy and preliminary guidance on using probabilistic analysis. The policy documents the EPA's position "that such probabilistic analysis techniques as Monte Carlo analysis, given adequate supporting data and credible assumptions, can be viable statistical tools for analyzing variability and uncertainty in risk assessments." The policy establishes conditions that are to be satisfied by risk assessments that use probabilistic techniques. These conditions relate to the good scientific practices of clarity, consistency, transparency, reproducibility, and the use of sound methods.

The EPA policy lists the following conditions for an acceptable risk assessment that uses probabilistic analysis techniques. These conditions were derived from principles that are presented later in this document and its Appendix. Therefore, after each condition, the relevant principles are noted.

1. The purpose and scope of the assessment should be clearly articulated in a "problem formulation" section that includes a full discussion of any highly exposed or highly susceptible subpopulations evaluated (e.g., children, the elderly, etc.). The questions the assessment attempts to answer are to be discussed and the assessment endpoints are to be well defined.

2. The methods used for the analysis (including all models used, all data upon which the assessment is based, and all assumptions that have a significant impact upon the results) are to be documented and easily located in the report. This documentation is

to include a discussion of the degree to which the data used are representative of the population under study. Also, this documentation is to include the names of the models and software used to generate the analysis. Sufficient information is to be provided to allow the results of the analysis to be independently reproduced. (Principles 4, 5, 6, and 11)

3. The results of sensitivity analyses are to be presented and discussed in the report. Probabilistic techniques should be applied to the compounds, pathways, and factors of importance to the assessment, as determined by sensitivity analyses or other basic requirements of the assessment. (Principles 1 and 2)

4. The presence or absence of moderate to strong correlations or dependencies between the input variables is to be discussed and accounted for in the analysis, along with the effects these have on the output distribution. (Principles 1 and 14)

5. Information for each input and output distribution is to be provided in the report. This includes tabular and graphical representations of the distributions (e.g., probability density function and cumulative distribution function plots) that indicate the location of any point estimates of interest (e.g., mean, median, 95th percentile). The selection of distributions is to be explained and justified. For both the input and output distributions, variability and uncertainty are to be differentiated where possible. (Principles 3, 7, 8, 10, 12, and 13)

6. The numerical stability of the central tendency and the higher end (i.e., tail) of the output distributions are to be presented and discussed. (Principle 9)

7. Calculations of exposures and risks using deterministic (e.g., point estimate) methods are to be reported if possible. Providing these values will allow comparisons between the probabilistic analysis and past or screening level risk assessments. Further, deterministic estimates may be used to answer scenario specific questions and to facilitate risk communication. When comparisons are made, it is important to explain the similarities and differences in the underlying data, assumptions, and models. (Principle 15).

8. Since fixed exposure assumptions (e.g., exposure duration, body weight) are sometimes embedded in the toxicity metrics (e.g., Reference Doses, Reference Concentrations, unit cancer risk factors), the exposure estimates from the probabilistic output distribution are to be aligned with the toxicity metric.

The following sections present a general framework and broad set of principles important for ensuring good scientific practices in the use of Monte Carlo analysis (a frequently encountered tool for evaluating uncertainty and variability). Many of the principles apply generally to the various techniques for conducting quantitative analyses of variability and uncertainty; however, the focus of the following principles is on Monte Carlo analysis. EPA recognizes that quantitative risk assessment methods and quantitative variability and uncertainty analysis are undergoing rapid development. These guiding principles are intended to serve as a minimum set of principles and are not intended to constrain or prevent the use of new or innovative improvements where scientifically defensible.

Fundamental Goals and Challenges

In the context of this policy, the basic goal of a Monte Carlo analysis is to chatacterize, quantitatively, the uncertainty and variability in estimates of exposure or risk. A secondary goal is to identify key sources of variability and uncertainty and to quantify the relative contribution of these sources to the overall variance and range of model results.

Consistent with EPA principles and policies, an analysis of variability and uncertainty should provide its audience with clear and concise information on the variability in individual exposures and risks; it should provide information on population risk (extent of harm in the exposed population); it should provide information on the distribution of exposures and risks to highly exposed or highly susceptible populations; it should describe qualitatively and quantitatively the scientific uncertainty in the models applied, the data utilized, and the specific risk estimates that are used.

Ultimately, the most important aspect of a quantitative variability and uncertainty analysis may well be the process of interaction between the risk assessor, risk manager and other interested parties that makes risk assessment into a dynamic rather than a static process. Questions for the risk assessor and risk manager to consider at the initiation of a quantitative variability and uncertainty analysis include:

- *Will the quantitative analysis of uncertainty and variability improve the risk assessment?*

- *What are the major sources of variability and uncertainty? How will variability and uncertainty be kept separate in the analysis?*

- *Are there time and resources to complete a complex analysis?*

- *Does the project warrant this level of effort?*

- *Will a quantitative estimate of uncertainty improve the decision? How will the regulatory decision be affected by this variability and uncertainty analysis?*

- *What types of skills and experience are needed to perform the analysis?*

- *Have the weaknesses and strengths of the methods been evaluated?*

- *How will the variability and uncertainty analysis be communicated to the public and decision makers?*

One of the most important challenges facing the risk assessor is to communicate, effectively, the insights an analysis of variability and uncertainty provides. It is important for the risk assessor to remember that insights will generally be qualitative in nature even though the models they derive from are quantitative. Insights can include:

- *An appreciation of the overall degree of variability and uncertainty and the confidence that can be placed in the analysis and its findings.*

- *An understanding of the key sources of variability and key sources of uncertainty and their impacts on the analysis.*

- *An understanding of the critical assumptions and their importance to the analysis and findings.*

- *An understanding of the unimportant assumptions and why they are unimportant.*

- *An understanding of the extent to which plausible alternative assumptions or models could affect any conclusions.*

- *An understanding of key scientific controversies related to the assessment and a sense of what difference they might make regarding the conclusions.*

The risk assessor should strive to present quantitative results in a manner that will clearly communicate the information they contain.

When a Monte Carlo Analysis Might Add Value to a Quantitative Risk Assessment

Not every assessment requires or warrants a quantitative characterization of variability and uncertainty. For example, it may be unnecessary to perform a Monte Carlo analysis when screening calculations show exposures or risks to be clearly below levels of concern (and the screening technique is known to significantly over-estimate exposure). As another example, it may be unnecessary to perform a Monte Carlo analysis when the costs of remediation are low.

On the other hand, there may be a number of situations in which a Monte Carlo analysis may be useful. For example, a Monte Carlo analysis may be useful when screening calculations using conservative point estimates fall above the levels of concern. Other situations could include when it is necessary to disclose the degree of bias associated with point estimates of exposure; when it is necessary to rank exposures, exposure pathways, sites or contaminants; when the cost of regulatory or remedial action is high and the exposures are marginal; or when the consequences of simplistic exposure estimates are unacceptable.

Often, a "tiered approach" may be helpful in deciding whether or not a Monte Carlo analysis can add value to the assessment and decision. In a tiered approach, one begins with a fairly simple screening level model and progresses to more sophisticated and realistic (and usually more complex) models only as warranted by the findings and value added to the decision. Throughout each of the steps in a tiered approach, soliciting input from each of the interested parties is recommended. Ultimately, whether or not a Monte Carlo analysis should be conducted is a matter of judgment, based on consideration of the intended use, the importance of the exposure assessment and the value and insights it provides to the risk assessor, risk manager, and other affected individuals or groups.

Key Terms and Their Definitions

The following section presents definitions for a number of key terms which are used throughout this document.

Bayesian

The Bayesian or subjective view is that the probability of an event is the degree of belief that a person has, given some state of knowledge, that the event will occur. In the classical or frequentist view, the probability of an event is the frequency with which an event occurs given a long sequence of identical and independent trials. In exposure assessment situations, directly representative and complete data sets are rarely available; inferences in these situations are inherently subjective. The decision as to the appropriateness of either approach (Bayesian or Classical) is based on the available data and the extent of subjectivity deemed appropriate.

Correlation, Correlation Analysis

Correlation analysis is an investigation of the measure of statistical association among random variables based on samples. Widely used measures include the *linear correlation coefficient* (also called the *product-moment correlation coefficient* or *Pearson's correlation coefficient*), and such non-parametric measures as *Spearman rank-order correlation coefficient*, and *Kendall's tau*. When the data are nonlinear, non-parametric correlation is generally considered to be more robust than linear correlation.

Cumulative Distribution Function (CDF)

The CDF is alternatively referred to in the literature as the *distribution function*, *cumulative frequency function*, or the *cumulative probability function*. The cumulative distribution function, F(x), expresses the probability the random variable X assumes a value less than or equal to some value x, $F(x) = \text{Prob}(X \leq x)$. For continuous random variables, the cumulative distribution function is obtained from the probability density function by integration, or by summation in the case of discrete random variables.

Latin Hypercube Sampling

In Monte Carlo analysis, one of two sampling schemes are generally employed: simple random sampling or Latin Hypercube sampling. Latin hypercube sampling may be viewed as a stratified sampling scheme designed to ensure that the upper or lower ends of the distributions

used in the analysis are well represented. Latin hypercube sampling is considered to be more efficient than simple random sampling, that is, it requires fewer simulations to produce the same level of precision. Latin hypercube sampling is generally recommended over simple random sampling when the model is complex or when time and resource constraints are an issue.

Monte Carlo Analysis, Monte Carlo Simulation

Monte Carlo Analysis is a computer-based method of analysis developed in the 1940's that uses statistical sampling techniques in obtaining a probabilistic approximation to the solution of a mathematical equation or model.

Parameter

Two distinct, but often confusing, definitions for parameter are used. In the first usage (preferred), parameter refers to the constants characterizing the probability density function or cumulative distribution function of a random variable. For example, if the random variable W is known to be normally distributed with mean μ and standard deviation σ, the characterizing constants μ and σ are called parameters. In the second usage, parameter is defined as the constants and independent variables which define a mathematical equation or model. For example, in the equation $Z = \alpha X + \beta Y$, the independent variables (X,Y) and the constants (α,β) are all parameters.

Probability Density Function (PDF)

The PDF is alternatively referred to in the literature as the *probability function* or the *frequency function*. For continuous random variables, that is, the random variables which can assume any value within some defined range (either finite or infinite), the probability density function expresses the probability that the random variable falls within some very small interval. For discrete random variables, that is, random variables which can only assume certain isolated or fixed values, the term *probability mass function* (PMF) is preferred over the term probability density function. PMF expresses the probability that the random variable takes on a specific value.

Random Variable

A random variable is a quantity which can take on any number of values but whose exact value cannot be known before a direct observation is made. For example, the outcome of the toss

of a pair of dice is a random variable, as is the height or weight of a person selected at random from the New York City phone book.

Representativeness

Representativeness is the degree to which a sample is characteristic of the population for which the samples are being used to make inferences.

Sensitivity, Sensitivity Analysis

Sensitivity generally refers to the variation in output of a mathematical model with respect to changes in the values of the model's input. A sensitivity analysis attempts to provide a ranking of the model's input assumptions with respect to their contribution to model output variability or uncertainty. The difficulty of a sensitivity analysis increases when the underlying model is nonlinear, nonmonotonic or when the input parameters range over several orders of magnitude. Many measures of sensitivity have been proposed. For example, the partial rank correlation coefficient and standardized rank regression coefficient have been found to be useful. Scatter plots of the output against each of the model inputs can be a very effective tool for identifying sensitivities, especially when the relationships are nonlinear. For simple models or for screening purposes, the sensitivity index can be helpful.

In a broader sense, sensitivity can refer to how conclusions may change if models, data, or assessment assumptions are changed.

Simulation

In the context of Monte Carlo analysis, simulation is the process of approximating the output of a model through repetitive random application of a model's algorithm.

Uncertainty

Uncertainty refers to *lack of knowledge* about specific factors, parameters, or models. For example, we may be uncertain about the mean concentration of a specific pollutant at a contaminated site or we may be uncertain about a specific measure of uptake (e.g., 95th percentile fish consumption rate among all adult males in the United States). Uncertainty includes *parameter uncertainty* (measurement errors, sampling errors, systematic errors), *model uncertainty* (uncertainty due to necessary simplification of real-world processes, mis-specification of the model structure, model misuse, use of inappropriate surrogate variables), and *scenario uncertainty* (descriptive errors, aggregation errors, errors in professional judgment, incomplete analysis).

Variability

Variability refers to observed differences attributable to *true heterogeneity* or diversity in a population or exposure parameter. Sources of variability are the result of natural random processes and stem from environmental, lifestyle, and genetic differences among humans. Examples include human physiological variation (e.g., natural variation in bodyweight, height, breathing rates, drinking water intake rates), weather variability, variation in soil types and differences in contaminant concentrations in the environment. Variability is usually not reducible by further measurement or study (but can be better characterized).

Preliminary Issues and Considerations

Defining the Assessment Questions

The critical first step in any exposure assessment is to develop a clear and unambiguous statement of the purpose and scope of the assessment. A clear understanding of the purpose will help to define and bound the analysis. Generally, the exposure assessment should be made as simple as possible while still including all important sources of risk. Finding the optimum match between the sophistication of the analysis and the assessment problem may be best achieved using a "tiered approach" to the analysis, that is, starting as simply as possible and sequentially employing increasingly sophisticated analyses, but only as warranted by the value added to the analysis and decision process.

Selection and Development of the Conceptual and Mathematical Models

Some Considerations in the Selection of Models

- appropriateness of the model's assumptions *vis-à-vis* the analysis objectives
- compatibility of the model input/output and linkages to other models used in the analysis
- the theoretical basis for the model
- level of aggregation, spatial and temporal scales
- resolution limits
- sensitivity to input variability and input uncertainty
- reliability of the model and code, including peer review of the theory and computer code
- verification studies, relevant field tests
- degree of acceptance by the user community
- friendliness, speed and accuracy
- staff and computer resources required

To help identify and select plausible models, the risk assessor should develop selection criteria tailored to each assessment question. The application of these criteria may dictate that different models be used for different subpopulations under study (e.g., highly exposed individuals vs. the general population). In developing these criteria, the risk assessor should consider all significant assumptions, be explicit about the uncertainties, including technical and scientific uncertainties about specific quantities, modeling uncertainties, uncertainties about functional forms, and should identify significant scientific issues about which there is uncertainty.

At any step in the analysis, the risk assessor should be aware of the manner in which alternative selections might influence the conclusions reached.

Selection and Evaluation of Available Data

After the assessment questions have been defined and conceptual models have been developed, it is necessary to compile and evaluate existing data (e.g., site specific or surrogate data) on variables important to the assessment. It is important to evaluate data quality and the extent to which the data are representative of the population under study.

Guiding Principles for Monte Carlo Analysis

This section presents a discussion of principles of good practice for Monte Carlo simulation as it may be applied to environmental assessments. It is not intended to serve as detailed technical guidance on how to conduct or evaluate an analysis of variability and uncertainty.

Selecting Input Data and Distributions for Use in Monte Carlo Analysis

1. **Conduct preliminary sensitivity analyses or numerical experiments to identify model structures, exposure pathways, and model input assumptions and parameters that make important contributions to the assessment endpoint and its overall variability and/or uncertainty.**

The capabilities of current desktop computers allow for a number of "what if" scenarios to be examined to provide insight into the effects on the analysis of selecting a particular model, including or excluding specific exposure pathways, and making certain assumptions with respect to model input parameters. The output of an analysis may be sensitive to the structure of the exposure model. Alternative plausible models should be examined to determine if structural differences have important effects on the output distribution (in both the region of central tendency and in the tails).

Numerical experiments or sensitivity analysis also should be used to identify exposure pathways that contribute significantly to or even dominate total exposure. Resources might be saved by excluding unimportant exposure pathways (e.g., those that do not contribute appreciably to the total exposure) from full probabilistic analyses or from further analyses altogether. For important pathways, the model input parameters that contribute the most to overall variability and uncertainty should be identified. Again, unimportant parameters may be excluded from full probabilistic treatment. For important parameters, empirical distributions or parametric distributions may be used. Once again, numerical experiments should be conducted to determine the sensitivity of the output to different assumptions with respect to the distributional forms of the input parameters. Identifying important pathways and parameters where assumptions about distributional form contribute significantly to overall uncertainty may aid in focusing data gathering efforts.

Dependencies or correlations between model parameters also may have a significant influence on the outcome of the analysis. The sensitivity of the analysis to various assumptions about known or suspected dependencies should be examined. Those dependencies or correlations identified as having a significant effect must be accounted for in later analyses.

Conducting a systematic sensitivity study may not be a trivial undertaking, involving significant effort on the part of the risk assessor. Risk assessors should exercise great care not to prematurely or unjustifiably eliminate pathways or parameters from full probabilistic treatment. Any parameter or pathway eliminated from full probabilistic treatment should be identified and the reasons for its elimination thoroughly discussed.

2. Restrict the use of probabilistic assessment to significant pathways and parameters.

Although specifying distributions for all or most variables in a Monte Carlo analysis is useful for exploring and characterizing the full range of variability and uncertainty, it is often unnecessary and not cost effective. If a systematic preliminary sensitivity analysis (that includes examining the effects of various assumptions about distributions) was undertaken and documented, and exposure pathways and parameters that contribute little to the assessment endpoint and its overall uncertainty and variability were identified, the risk assessor may simplify the Monte Carlo analysis by focusing on those pathways and parameters identified as significant. From a computational standpoint, a Monte Carlo analysis can include a mix of point estimates and distributions for the input parameters to the exposure model. However, the risk assessor and risk manager should continually review the basis for "fixing" certain parameters as point values to avoid the perception that these are indeed constants that are not subject to change.

3. Use data to inform the choice of input distributions for model parameters .

The choice of input distribution should always be based on all information (both qualitative and quantitative) available for a parameter. In selecting a distributional form, the risk assessor should consider the quality of the information in the database and ask a series of questions including (but not limited to):

- *Is there any 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?*

When data for an important parameter are limited, it may be useful to define plausible alternative scenarios to incorporate some information on the impact of that variable in the overall assessment (as done in the sensitivity analysis). In doing this, the risk assessor should select the widest distributional family consistent with the state of knowledge and should, for important parameters, test the sensitivity of the findings and conclusions to changes in distributional shape.

4. Surrogate data can be used to develop distributions when they can be appropriately justified.

The risk assessor should always seek representative data of the highest quality available. However, the question of how representative the available data are is often a serious issue. Many times, the available data do not represent conditions (e.g., temporal and spatial scales) in the population being assessed. The assessor should identify and evaluate the factors that introduce uncertainty into the assessment. In particular, attention should be given to potential biases that may exist in surrogate data and their implications for the representativeness of the fitted distributions.

When alternative surrogate data sets are available, care must be taken when selecting or combining sets. The risk assessor should use accepted statistical practices and techniques when combining data, consulting with the appropriate experts as needed.

Whenever possible, collect site or case specific data (even in limited quantities) to help justify the use of the distribution based on surrogate data. The use of surrogate data to develop distributions can be made more defensible when case-specific data are obtained to check the reasonableness of the distribution.

5. When obtaining empirical data to develop input distributions for exposure model parameters, the basic tenets of environmental sampling should be followed. Further,

particular attention should be given to the quality of information at the tails of the distribution.

As a general rule, the development of data for use in distributions should be carried out using the basic principles employed for exposure assessments. For example,

- *Receptor-based sampling in which data are obtained on the receptor or on the exposure fields relative to the receptor;*
- *Sampling at appropriate spatial or temporal scales using an appropriate stratified random sampling methodology;*
- *Using two-stage sampling to determine and evaluate the degree of error, statistical power, and subsequent sampling needs; and*
- *Establishing data quality objectives.*

In addition, the quality of information at the tails of input distributions often is not as good as the central values. The assessor should pay particular attention to this issue when devising data collection strategies.

6. **Depending on the objectives of the assessment, expert [1] judgment can be included either within the computational analysis by developing distributions using various methods or by using judgments to select and separately analyze alternate, but plausible, scenarios. When expert judgment is employed, the analyst should be very explicit about its use.**

Expert judgment is used, to some extent, throughout all exposure assessments. However, debatable issues arise when applying expert opinions to input distributions for Monte Carlo analyses. Using expert judgment to derive a distribution for an input parameter can reflect bounds on the state of knowledge and provide insights into the overall uncertainty. This may be particularly useful during the sensitivity analysis to help identify important variables for which additional data may be needed. However, distributions based exclusively or primarily on expert judgment reflect the opinion of individuals or groups and, therefore, may be subject to considerable bias. Further, without explicit documentation of the use of expert opinions, the

[1] According to NCRP (1996), an expert has (1) training and experience in the subject area resulting in superior knowledge in the field, (2) access to relevant information, (3) an ability to process and effectively use the information, and (4) is recognized by his or her peers or those conducting the study as qualified to provide judgments about assumptions, models, and model parameters at the level of detail required.

distributions based on these judgments might be erroneously viewed as equivalent to those based on hard data. When distributions based on expert judgement have an appreciable effect on the outcome of an analysis, it is critical to highlight this in the uncertainty characterization.

Evaluating Variability and Uncertainty

7. The concepts of variability and uncertainty are distinct. They can be tracked and evaluated separately during an analysis, or they can be analyzed within the same computational framework. Separating variability and uncertainty is necessary to provide greater accountability and transparency. The decision about how to track them separately must be made on a case-by-case basis for each variable.

Variability represents the true heterogeneity or diversity inherent in a well-characterized population. As such, it is not reducible through further study. Uncertainty represents a lack of knowledge about the population. It is sometimes reducible through further study. Therefore, separating variability and uncertainty during the analysis is necessary to identify parameters for which additional data are needed. There can be uncertainty about the variability within a population. For example, if only a subset of the population is measured or if the population is otherwise under-sampled, the resulting measure of variability may differ from the true population variability. This situation may also indicate the need for additional data collection.

8. There are methodological differences regarding how variability and uncertainty are addressed in a Monte Carlo analysis.

There are formal approaches for distinguishing between and evaluating variability and uncertainty. When deciding on methods for evaluating variability and uncertainty, the assessor should consider the following issues.

- *Variability depends on the averaging time, averaging space, or other dimensions in which the data are aggregated.*
- *Standard data analysis tends to understate uncertainty by focusing solely on random error within a data set. Conversely, standard data analysis tends to overstate variability by implicitly including measurement errors.*
- *Various types of model errors can represent important sources of uncertainty. Alternative conceptual or mathematical models are a potentially important source of uncertainty. A major threat to the accuracy of a variability analysis is a lack of representativeness of the data.*

9. **Methods should investigate the numerical stability of the moments and the tails of the distributions.**

For the purposes of these principles, numerical stability refers to observed numerical changes in the characteristics (i.e., mean, variance, percentiles) of the Monte Carlo simulation output distribution as the number of simulations increases. Depending on the algebraic structure of the model and the exact distributional forms used to characterize the input parameters, some outputs will stabilize quickly, that is, the output mean and variance tend to reach more or less constant values after relatively few sampling iterations and exhibit only relatively minor fluctuations as the number of simulations increases. On the other hand, some model outputs may take longer to stabilize. The risk assessor should take care to be aware of these behaviors. Risk assessors should always use more simulations than they think necessary. Ideally, Monte Carlo simulations should be repeated using several non-overlapping subsequences to check for stability and repeatability. Random number seeds should always be recorded. In cases where the tails of the output distribution do not stabilize, the assessor should consider the quality of information in the tails of the input distributions. Typically, the analyst has the least information about the input tails. This suggest two points.

- *Data gathering efforts should be structured to provide adequate coverage at the tails of the input distributions.*
- *The assessment should include a narrative and qualitative discussion of the quality of information at the tails of the input distributions.*

10. **There are limits to the assessor's ability to account for and characterize all sources of uncertainty. The analyst should identify areas of uncertainty and include them in the analysis, either quantitatively or qualitatively.**

Accounting for the important sources of uncertainty should be a key objective in Monte Carlo analysis. However, it is not possible to characterize all the uncertainties associated with the models and data. The analyst should attempt to identify the full range of types of uncertainty impinging on an analysis and clearly disclose what set of uncertainties the analysis attempts to represent and what it does not. Qualitative evaluations of uncertainty including relative ranking of the sources of uncertainty may be an acceptable approach to uncertainty evaluation, especially when objective quantitative measures are not available. Bayesian methods may sometimes be

useful for incorporating subjective information into variability and uncertainty analyses in a manner that is consistent with distinguishing variability from uncertainty.

Presenting the Results of a Monte Carlo Analysis

11. Provide a complete and thorough description of the exposure model and its equations (including a discussion of the limitations of the methods and the results).

Consistent with the Exposure Assessment Guidelines, Model Selection Guidance, and other relevant Agency guidance, provide a detailed discussion of the exposure model(s) and pathways selected to address specific assessment endpoints. Show all the formulas used. Define all terms. Provide complete references. If external modeling was necessary (e.g., fate and transport modeling used to provide estimates of the distribution of environmental concentrations), identify the model (including version) and its input parameters. Qualitatively describe the major advantages and limitations of the models used.

The objectives are transparency and reproducibility - to provide a complete enough description so that the assessment might be independently duplicated and verified.

12. Provide detailed information on the input distributions selected. This information should identify whether the input represents largely variability, largely uncertainty, or some combination of both. Further, information on goodness-of-fit statistics should be discussed.

It is important to document thoroughly and convey critical data and methods that provide an important context for understanding and interpreting the results of the assessment. This detailed information should distinguish between variability and uncertainty and should include graphs and charts to visually convey written information.

The probability density function (PDF) and cumulative distribution function (CDF) graphs provide different, but equally important insights. A plot of a PDF shows possible values of a random variable on the horizontal axis and their respective probabilities (technically, their densities) on the vertical axis. This plot is useful for displaying:

- *the relative probability of values;*
- *the most likely values (e.g., modes);*
- *the shape of the distribution (e.g., skewness, kurtosis); and*

- *small changes in probability density.*

A plot of the cumulative distribution function shows the probability that the value of a random variable is less than a specific value. These plots are good for displaying:

- *fractiles, including the median;*
- *probability intervals, including confidence intervals;*
- *stochastic dominance; and*
- *mixed, continuous, and discrete distributions.*

Goodness-of-fit tests are formal statistical tests of the hypothesis that a specific set of sampled observations are an independent sample from the assumed distribution. Common tests include the chi-square test, the Kolmogorov-Smirnov test, and the Anderson-Darling test. Goodness-of-fit tests for normality and lognormality include Lilliefors' test, the Shapiro-Wilks' test, and D'Agostino's test.

Risk assessors should never depend solely on the results of goodness-of-fit tests to select the analytic form for a distribution. Goodness-of-fit tests have low discriminatory power and are generally best for rejecting poor distribution fits rather than for identifying good fits. For small to medium sample sizes, goodness-of-fit tests are not very sensitive to small differences between the observed and fitted distributions. On the other hand, for large data sets, even small and unimportant differences between the observed and fitted distributions may lead to rejection of the null hypothesis. For small to medium sample sizes, goodness-of-fit tests should best be viewed as a systematic approach to detecting gross differences. The risk assessor should never let differences in goodness-of-fit test results be the sole factor for determining the analytic form of a distribution.

Graphical methods for assessing fit provide visual comparisons between the experimental data and the fitted distribution. Despite the fact that they are non-quantitative, graphical methods often can be most persuasive in supporting the selection of a particular distribution or in rejecting the fit of a distribution. This persuasive power derives from the inherent weaknesses in numerical goodness-of-fit tests. Such graphical methods as probability-probability (P-P) and quantile-quantile (Q-Q) plots can provide clear and intuitive indications of goodness-of-fit.

Having selected and justified the selection of specific distributions, the assessor should provide plots of both the PDF and CDF, with one above the other on the same page and using identical horizontal scales. The location of the mean should be clearly indicated on both curves [See Figure 1]. These graphs should be accompanied by a summary table of the relevant data.

13. Provide detailed information and graphs for each output distribution.

In a fashion similar to that for the input distributions, the risk assessor should provide plots of both the PDF and CDF for each output distribution, with one above the other on the same page, using identical horizontal scales. The location of the mean should clearly be indicated on both curves. Graphs should be accompanied by a summary table of the relevant data.

14. Discuss the presence or absence of dependencies and correlations.

Covariance among the input variables can significantly affect the analysis output. It is important to consider covariance among the model's most sensitive variables. It is particularly important to consider covariance when the focus of the analysis is on the high end (i.e., upper end) of the distribution.

When covariance among specific parameters is suspected but cannot be determined due to lack of data, the sensitivity of the findings to a range of different assumed dependencies should be evaluated and reported.

15. Calculate and present point estimates.

Traditional deterministic (point) estimates should be calculated using established protocols. Clearly identify the mathematical model used as well as the values used for each input parameter in this calculation. Indicate in the discussion (and graphically) where the point estimate falls on the distribution generated by the Monte Carlo analysis. Discuss the model and parameter assumptions that have the most influence on the point estimate's position in the distribution. The most important issue in comparing point estimates and Monte Carlo results is whether the data and exposure methods employed in the two are comparable. Usually, when a major difference between point estimates and Monte Carlo results is observed, there has been a fundamental change in data or methods. Comparisons need to call attention to such differences and determine their impact.

In some cases, additional point estimates could be calculated to address specific risk management questions or to meet the information needs of the audience for the assessment. Point estimates can often assist in communicating assessment results to certain groups by providing a

scenario-based perspective. For example, if point estimates are prepared for scenarios with which the audience can identify, the significance of presented distributions may become clearer. This may also be a way to help the audience identify important risks.

16. A tiered presentation style, in which briefing materials are assembled at various levels of detail, may be helpful. Presentations should be tailored to address the questions and information needs of the audience.

Entirely different types of reports are needed for scientific and nonscientific audiences. Scientists generally will want more detail than non-scientists. Risk managers may need more detail than the public. Reports for the scientific community are usually very detailed. Descriptive, less detailed summary presentations and key statistics with their uncertainty intervals (e.g., box and whisker plots) are generally more appropriate for non-scientists.

To handle the different levels of sophistication and detail needed for different audiences, it may be useful to design a presentation in a tiered format where the level of detail increases with each successive tier. For example, the first tier could be a one-page summary that might include a graph or other numerical presentation as well as a couple of paragraphs outlining what was done. This tier alone might be sufficient for some audiences. The next tier could be an executive summary, and the third tier could be a full detailed report. For further information consult Bloom et al., 1993.

Graphical techniques can play an indispensable role in communicating the findings from a Monte Carlo analysis. It is important that the risk assessor select a clear and uncluttered graphical style in an easily understood format. Equally important is deciding which information to display. Displaying too much data or inappropriate data will weaken the effectiveness of the effort. Having decided which information to display, the risk assessor should carefully tailor a graphical presentation to the informational needs and sophistication of specific audiences. The performance of a graphical display of quantitative information depends on the information the risk assessor is trying to convey to the audience and on how well the graph is constructed (Cleveland, 1994). The following are some recommendations that may prove useful for effective graphic presentation:

- Avoid excessively complicated graphs. Keep graphs intended for a glance (e.g., overhead or slide presentations) relatively simple and uncluttered. Graphs intended for publication can include more complexity.

- Avoid pie charts, perspective charts (3-dimensional bar and pie charts, ribbon charts), pseudo-perspective charts (2-dimensional bar or line charts).

- Color and shading can create visual biases and are very difficult to use effectively. Use color or shading only when necessary and then, only very carefully. Consult references on the use of color and shading in graphics.

- When possible in publications and reports, graphs should be accompanied by a table of the relevant data.

- If probability density or cumulative probability plots are presented, present both, with one above the other on the same page, with identical horizontal scales and with the location of the mean clearly indicated on both curves with a solid point.

- Do not depend on the audience to correctly interpret any visual display of data. Always provide a narrative in the report interpreting the important aspects of the graph.

- Descriptive statistics and box plots generally serve the less technically-oriented audience well. Probability density and cumulative probability plots are generally more meaningful to risk assessors and uncertainty analysts.

Appendix: Probability Distribution Selection Issues

Surrogate Data, Fitting Distributions, Default Distributions Subjective Distributions

Identification of relevant and valid data to represent an exposure variable is prerequisite to selecting a probability distribution However, often the data available are not a direct measure of the exposure variable of interest. The risk assessor is often faced with using data taken in spatial or temporal scales that are significantly different from the scale of the problem under consideration. The question becomes whether or not or how to use marginally representative or surrogate data to represent a particular exposure variable. While there can be no hard and fast rules on how to make that judgment, there are a number of questions risk assessors need to ask when the surrogate data are the only data available.

Is there Prior Knowledge about Mechanisms? Ideally, the selection of candidate probability distributions should be based on consideration of the underlying physical processes or mechanisms thought to be key in giving rise to the observed variability. For example, if the exposure variable is the result of the product of a large number of other random variables, it would make sense to select a lognormal distribution for testing. As another example, the exponential distribution would be a reasonable candidate if the stochastic variable represents a process akin to inter-arrival times of events that occur at a constant rate. As a final example, a gamma distribution would be a reasonable candidate if the random variable of interest was the sum of independent exponential random variables.

Threshold Question - Are the surrogate data of acceptable quality and representativeness to support reliable exposure estimates?

What uncertainties and biases are likely to be introduced by using surrogate data? For example, if the data have been collected in a different geographic region, the contribution of factors such as soil type, rainfall, ambient temperature, growing season, natural sources of exposure, population density, and local industry may have a significant effect on the exposure concentrations and activity patterns. If the data are collected from volunteers or from hot spots, they will probably not represent the distribution of values in the population of interest. Each difference between the survey data and the population being assessed should be noted. The effects of these differences on the desired distribution should be discussed if possible.

How are the biases likely to affect the analysis and can the biases be corrected? The risk assessor may be able to state with a high degree of certainty that the available data over-estimates or under-estimates the parameter of interest. Use of ambient air data on arsenic collected near smelters will almost certainly over-estimate average arsenic exposures in the United States. However, the smelter data can probably be used to produce an estimate of inhalation exposures that falls within the high end. In other cases, the assessor may be unsure how unrepresentative data will affect the estimate as in the case when data collected by a particular State are used in a

national assessment. In most cases, correction of suspected biases will be difficult or not possible. If only hot spot data are available for example, only bounding or high end estimates may be possible. Unsupported assumptions about biases should be avoided. Information regarding the direction and extent of biases should be included in the uncertainty analysis.

How should any uncertainty introduced by the surrogate data be represented?

In identifying plausible distributions to represent variability, the risk assessor should examine the following characteristics of the variable:

1. ***Nature of the variable.***
Can the variable only take on discrete values (e.g., either on or off; either heads or tails) or is the variable continuous over some range (e.g., pollutant concentration; body weight; drinking water consumption rate)? Is the variable correlated with or dependent on another variable?

2. ***Bounds of the variable.***
What is the physical or plausible range of the variable (e.g., takes on only positive values; bounded by the interval [a,b]). Are physical measurements of the variable censored due to limits of detection or some aspect of the experimental design?

3. ***Symmetry of the Distribution.***
Is distribution of the variable known to be or thought to be 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? Is the shape of the distribution likely to be dictated by physical/biological properties (e.g., logistic growth rates) or other mechanisms?

4. ***Summary Statistics.***
Summary statistics can sometimes be useful in discriminating among candidate distributions. For example, frequently the range of the variable can be used to eliminate inappropriate distributions; it would not be reasonable to select a lognormal distribution for an absorption coefficient since the range of the lognormal distribution is $(0,\infty)$ while the range of the absorption coefficient is (0,1). If the coefficient of variation is near 1.0, then an exponential distribution might be appropriate. Information on skewness can also be useful. For symmetric distributions, skewness = 0; for distributions skewed to the right, skewness > 0; for distributions skewed to the left, skewness < 0.

5. ***Graphical Methods to Explore the Data.***
The risk assessor can often gain important insights by using a number of simple graphical techniques to explore the data prior to numerical analysis. A wide variety of graphical methods have been developed to aid in this exploration including frequency histograms for continuous distributions, stem and leaf plots, dot plots, line plots for discrete distributions, box and whisker plots, scatter plots, star representations, glyphs, Chernoff faces, etc. [Tukey (1977); Conover (1980); du Toit *et al.* (1986); Morgan and Henrion, (1990)]. These graphical methods are all

intended to permit visual inspection of the density function corresponding to the distribution of the data. They can assist the assessor in examining the data for skewness, behavior in the tails, rounding biases, presence of multi-modal behavior, and data outliers.

Frequency histograms can be compared to the fundamental shapes associated with standard analytic distributions (e.g., normal, lognormal, gamma, Weibull). Law and Kelton (1991) and Evans et al. (1993) have prepared a useful set of figures which plot many of the standard analytic distributions for a range of parameter values. Frequency histograms should be plotted on both linear and logarithmic scales and plotted over a range of frequency bin widths (class intervals) to avoid too much jaggedness or too much smoothing (i.e., too little or too much data aggregation). The data can be sorted and plotted on probability paper to check for normality (or log-normality). Most of the statistical packages available for personal computers include histogram and probability plotting features, as do most of the spreadsheet programs. Some statistical packages include stem and leaf, and box and whisker plotting features.

After having explored the above characteristics of the variable, the risk assessor has three basic techniques for representing the data in the analysis. In the first method, the assessor can attempt to fit a theoretical or parametric distribution to the data using standard statistical techniques. As a second option, the assessor can use the data to define an empirical distribution function (EDF). Finally, the assessor can use the data directly in the analysis utilizing random resampling techniques (i.e., bootstrapping). Each of these three techniques has its own benefits. However, there is no consensus among researchers (authors) as to which method is generally superior. For example, Law and Kelton (1991) observe that EDFs may contain irregularities, especially when the data are limited and that when an EDF is used in the typical manner, values outside the range of the observed data cannot be generated. Consequently, when the data are representative of the exposure variable and the fit is good, some prefer to use parametric distributions. On the other hand, some authors prefer EDFs (Bratley, Fox and Schrage, 1987) arguing that the smoothing which necessarily takes place in the fitting process distorts real information. In addition, when data are limited, accurate estimation of the upper end (tail) is difficult. Ultimately, the technique selected will be a matter of the risk assessor's comfort with the techniques and the quality and quantity of the data under evaluation.

The following discussion focuses primarily on parametric techniques. For a discussion of the other methods, the reader is referred to Efron and Tibshirani (1993), Law & Kelton (1991), and Bratley *et al (1987)*.

Having selected parametric distributions, it is necessary to estimate numerical values for the intrinsic parameters which characterize each of the analytic distributions and assess the quality of the resulting fit.

Parameter Estimation. Parameter estimation is generally accomplished using conventional statistical methods, the most popular of which include the method of maximum likelihood, method of least squares, and the method of moments. See Johnson and Kotz (1970), Law and

Kelton (1991), Kendall and Stewart (1979), Evans et al. (1993), Ang and Tang (1975), Gilbert (1987), and Meyer (1975).

Assessing the Representativeness of the Fitted Distribution. Having estimated the parameters of the candidate distributions, it is necessary to evaluate the "quality of the fit" and, if more than one distribution was selected, to select the "best" distribution from among the candidates. Unfortunately, there is no single, unambiguous measure of what constitutes best fit. Ultimately, the risk assessor must judge whether or not the fit is acceptable.

Graphical Methods for Assessing Fit. Graphical methods provide visual comparisons between the experimental data and the fitted distribution. Despite the fact that they are non-quantitative, graphical methods often can be most persuasive in supporting the selection of a particular distribution or in rejecting the fit of a distribution. This persuasive power derives from the inherent weaknesses in numerical goodness-of-fit tests. Commonly used graphical methods include: *frequency comparisons* which compare a histogram of the experimental data with the density function of the fitted data; *probability plots* compare the observed cumulative density function with the fitted cumulative density function. Probability plots are often based on graphical transformations such that the plotted cumulative density function results in a straight line; *probability-probability plots* (P-P plots) compare the observed probability with the fitted probability. P-P plots tend to emphasize differences in the middle of the predicted and observed cumulative distributions; *quantile-quantile plots* (Q-Q plots) graph the *ith-quantile* of the fitted distribution against the *ith quantile* data. Q-Q plots tend to emphasize differences in the tails of the fitted and observed cumulative distributions; and *box plots* compare a box plot of the observed data with a box plot of the fitted distribution.

Goodness-of-Fit Tests. Goodness-of-fit tests are formal statistical tests of the hypothesis that the set of sampled observations are an independent sample from the assumed distribution. The null hypothesis is that the randomly sampled set of observations are independent, identically distributed random variables with distribution function F. Commonly used goodness-of-fit tests include the chi-square test, Kolmogorov-Smirnov test, and Anderson-Darling test. The chi-square test is based on the difference between the square of the observed and expected frequencies. It is highly dependent on the width and number of intervals chosen and is considered to have low power. It is best used to reject poor fits. The Kolmogorov-Smirnov Test is a non-parametric test based on the maximum absolute difference between the theoretical and sample Cumulative Distribution Functions (CDFs). The Kolmogorov-Smirnov test is most sensitive around the median and less sensitive in the tails and is best at detecting shifts in the empirical CDF relative to the known CDF. It is less proficient at detecting spread but is considered to be more powerful than the chi-square test. The Anderson-Darling test is designed to test goodness-of-fit in the tails of a Probability Density Function (PDF) based on a weighted-average of the squared difference between the observed and expected cumulative densities.

Care must be taken not to over-interpret or over-rely on the findings of goodness-of-fit tests. It is far too tempting to use the power and speed of computers to run goodness-of-fit tests against a generous list of candidate distributions, pick the distribution with the "best" goodness-of-fit statistic, and claim that the distribution that fit "best" was not rejected at some specific level of significance. This practice is statistically incorrect and should be avoided [Bratley *et al.*, 1987, page 134]. Goodness-of-fit tests have notoriously low power and are generally best for rejecting poor distribution fits rather than for identifying good fits. For small to medium sample sizes, goodness-of-fit tests are not very sensitive to small differences between the observed and fitted distributions. On the other hand, for large data sets, even minute differences between the observed and fitted distributions may lead to rejection of the null hypothesis. For small to medium sample sizes, goodness-of-fit tests should best be viewed as a systematic approach to detecting gross differences.

Tests of Choice for Normality and Lognormality. Several tests for normality (and lognormality when log-transformed data are used) which are considered more powerful than either the chi-square or Komolgarov-Smirnoff (K-S) tests have been developed: Lilliefors' test which is based on the K-S test but with "normalized" data values, Shapiro-Wilks test (for sample sizes $\leq$ 50), and D'Agostino's test (for sample sizes $\geq$ 50). The Shapiro-Wilks and D'Agostino tests are the tests of choice when testing for normality or lognormality.

If the data are not well-fit by a theoretical distribution, the risk assessor should consider the Empirical Distribution Function or bootstrapping techniques mentioned above.

For those situations in which the data are not adequately representative of the exposure variable or where the quality or quantity of the data are questionable the following approaches may be considered.

Distributions Based on Surrogate Data. Production of an exposure assessment often requires that dozens of factors be evaluated, including exposure concentrations, intake rates, exposure times, and frequencies. A combination of monitoring, survey, and experimental data, fate and transport modeling, and professional judgment is used to evaluate these factors. Often the only available data are not completely representative of the population being assessed. Some examples are the use of activity pattern data collected in one geographic region to evaluate the duration of activities at a Superfund site in another region; use of national intake data on consumption of a particular food item to estimate regional intake; and use of data collected from volunteers to represent the general population.

In each such case, the question of whether to use the unrepresentative data to estimate the distribution of a variable should be carefully evaluated. Considerations include how to express the possible bias and uncertainty introduced by the unrepresentativeness of the data and alternatives to using the data. In these situations, the risk assessor should carefully evaluate the basis of the distribution (e.g., data used, method) before choosing a particular surrogate or before picking among alternative distributions for the same exposure parameter. The

following table indicates exposure parameters for which surrogate distributions may be reasonable and useful.

Table 1 Examples of exposure parameters for which distributions based on surrogate data might be reasonable

Receptor Physiological Parameters		body weight height total skin surface area exposed skin - hands, forearms, head, upper body
Behavioral	Receptor Time-Activity Patterns	residency periods - age, residency type weekly work hours time since last job change showering duration
	Receptor Contact Rates	soil ingestion rates soil adherence food ingestion - vegetables, freshwater finfish, saltwater finfish, shellfish, beef water intake - total water, tapwater inhalation rates

Rough Characterizations of Ranges and Distributional Forms. In the absence of acceptable representative data or if the study is to be used primarily for screening, crude characterizations of the ranges and distributions of the exposure variable may be adequate. For example, physical plausibility arguments may be used to establish ranges for the parameters. Then, assuming such distributions as the uniform, log-uniform, triangular and log-triangular distributions can be helpful in establishing which input variables have the greatest influence on the output variable. However, the risk assessor should be aware that there is some controversy concerning the use of these types of distributions in the absence of data. Generally, the range of the model output is more dependant on the ranges of the input variables than it is on the actual shapes of the input distributions. Therefore, the risk assessor should be careful to avoid assigning overly-restrictive ranges or unreasonably large ranges to variables. Distributional assumptions can have a large influence on the shapes of the output distribution. When the shape of the output distribution must be estimated accurately, care and attention should be devoted to developing the input distributions.

Distributions Based on Expert Judgment. One method that has seen increasing usage in environmental risk assessment is the method of subjective probabilities in which an expert or experts are asked to estimate various behaviors and likelihoods regarding specific model variables or scenarios. Expert elicitation is divided into two categories: (1) informal elicitation, and (2) formal elicitation. Informal elicitation methods include self assessment, brainstorming, causal elicitation (without structured efforts to control biases), and taped group discussions between the project staff and selected experts.

Formal elicitation methods generally follow the steps identified by the U.S. Nuclear Regulatory Commission (USNRC, 1989; Oritz, 1991; also see Morgan and Henrion, 1990; IAEA, 1989; Helton, 1993; Taylor and Burmaster; 1993) and are considerably more elaborate and expensive than informal methods.

References Cited in Text

A. H-S. Ang and W. H. Tang, *Probability Concepts in Engineering Planning and Design, Volume I, Basic Principles*, John Wiley & Sons, Inc., New York (1975).

D. L. Bloom, et al., Communicating Risk to Senior EPA Policy Makers: A Focus Group Study, U.S. EPA Office of Air Quality Planning and Standards (1993).

P. Bratley, B. L. Fox, L. E. Schrage, *A Guide to Simulation*, Springer-Verlag, New York (1987).

W.S. Cleveland, *The Elements of Graphing Data*, revised edition, Hobart Press, Summit, New Jersey (1994).

W. J. Conover, *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York (1980).

S. H. C. du Toit, A. G. W. Steyn, R.H. Stumpf, *Graphical Exploratory Data Analysis*, Springer-Verlag, New York (1986).

B. Efron and R. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, New York (1993).

M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, John Wiley & Sons, New York (1993).

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

J. C. Helton, "Uncertainty and Sensitivity Analysis Techniques for Use In Performance Assessment for Radioactive Waste Disposal," *Reliability Engineering and System Safety*, Vol. 42, pages 327-367 (1993).

IAEA, *Safety Series 100, Evaluating the Reliability of Predictions Made Using Environmental Transfer Models*, International Atomic Energy Agency, Vienna, Austria (1989).

N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, volumes 1 & 2, John Wiley & Sons, Inc., New York (1970).

M. Kendall and A. Stuart, The *Advanced Theory of Statistics*, Volume I - Distribution Theory; Volume II - Inference and Relationship, Macmillan Publishing Co., Inc., New York (1979).

A. M. Law and W. D. Kelton, *Simulation Modeling & Analysis*, McGraw-Hill, Inc., (1991).

S. L. Meyer, *Data Analysis for Scientists and Engineers*, John Wiley & Sons, Inc., New York (1975).

M. G. Morgan and M. Henrion, *Uncertainty A guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York (1990).

NCRP 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, D.C. (May, 1996).

N. R. Oritz, M. A. Wheeler, R. L. Keeney, S. Hora, M. A. Meyer, and R. L. Keeney, "Use of Expert Judgment in NUREG-1150, *Nuclear Engineering and Design*, 126:313-331 (1991).

A. C. Taylor and D. E. Burmaster, "Using Objective and Subjective Information to Generate Distributions for Probabilistic Exposure Assessment," U.S. Environmental Protection Agency, draft report (1993).

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

USNRC, *Severe Accident Risks: An Assessment for Five U.S. Nuclear power Plants* (second peer review draft), U.S. Nuclear Regulatory Commission, Washington, D.C. (1989).

References for Further Reading

B. F. Baird, *Managerial Decisions Under Uncertainty,* John Wiley and Sons, Inc., New York (1989).

D. E. Burmaster and P. D. Anderson, "Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessments," *Risk Analysis*, Vol. 14(4), pages 477-482 (August, 1994).

R. Clemen, *Making Hard Decisions*, Duxbury Press (1990).

D. C. Cox and P. Baybutt, "Methods for Uncertainty Analysis: A Comparative Survey," *Risk Analysis*, Vol. 1 (4), 251-258 (1981).

R. D'Agostino and M.A. Stephens (eds), *Goodness-of-Fit Techniques*, Marcel Dekker, Inc., New York (1986).

L. Devroye, *Non-Uniform Random Deviate Generation*, Springer-Verlag, (1986).

D. M. Hamby, "A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models," *Environmental Monitoring and Assessment*, Vol. 32, 135-154 (1994).

D. B. Hertz, and H. Thomas, *Risk Analysis and Its Applications,* John Wiley and Sons, New York (1983).

D. B. Hertz, and H. Thomas, *Practical Risk Analysis - An Approach Through Case Studies*, John Wiley and Sons, New York (1984).

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

F. O. Hoffman and J. S. Hammonds, "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability," *Risk Analysis*, Vol. 14 (5), 707-712 (1994).

R. L. Iman and J. C. Helton, "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models," *Risk Analysis*, Vol. 8(1), pages 71-90 (1988).

R. L. Iman and W. J. Conover, "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Commun. Statistics, Communications and Computation*, 11, 311-331 (1982).

R. L. Iman, J. M. Davenport, and D. K. Zeigler, "Latin Hypercube Sampling (A Program Users Guide)," Technical Report SAND 79:1473, Sandia Laboratories, Albuquerque (1980).

M. E. Johnson, *Multivariate Statistical Simulation*, John Wiley & Sons, Inc., New York (1987).

N. L. Johnson, S. Kotz and A. W. Kemp, *Univariate Discrete Distributions*, John Wiley & Sons, Inc., New York (1992).

R. LePage and L. Billard, *Exploring the Limits of Bootstrap*, Wiley, New York (1992).

J. Lipton, et al. "Short Communication: Selecting Input Distributions for Use in Monte Carlo Analysis," *Regulatory Toxicology and Pharmacology*, 21, 192-198 (1995).

W. J. Kennedy, Jr. and J. E. Gentle, *Statistical Computing*, Marcel Dekker, Inc., New York (1980).

T. E. McKone and K. T. Bogen, "Uncertainties in Health Risk Assessment: An Integrated Case Based on Tetrachloroethylene in California Groundwater," *Regulatory Toxicology and Pharmacology*, 15, 86-103 (1992).

R. E. Megill (Editor), *Evaluating and Managing Risk,* Penn Well Books, Tulsa, Oklahoma (1985).

R. E. Megill, *An Introduction to Risk Analysis*, end Ed. Penn Well Books, Tulsa, Oklahoma (1985).

Palisade Corporation, 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, New York (1994).

W. H. Press, B. P. Flannery, S. A. Teulolsky, and W. T. Vetterling, *Numerical Recipes in Pascal: the Art of Scientific Computing*, Cambridge University Press (1989).

W. H. Press, S. A. Teulolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: the Art of Scientific Computing*, Cambridge University Press (1992).

W. H. Press, S. A. Teulolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: the Art of Scientific Computing*, Cambridge University Press (1992).

T. Read and N. Cressie, *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer-Verlag, New York (1988).

V. K. Rohatgi, *Statistical Inference*, John Wiley & Sons, New York (1984).

R. Y. Rubinstein, *Simulation and the Monte Carlo Method,* John Wiley and Sons, New York (1981).

L. Sachs, *Applied Statistics - A Handbook of Techniques*, Spring-Verlag, New York (1984).

A. Saltelli and J. Marivort, "Non-parametric Statistics in Sensitivity Analysis for Model Output: A Comparison of Selected Techniques," *Reliability Engineering and System Safety*, Vol. 28, 229-253 (1990).

H. Schneider, *Truncated and Censored Distributions form Normal Populations*, Marcel Dekker, Inc., New York (1986).

F. A. Seiler and J. L. Alvarez, "On the Selection of Distributions for Stochastic Variables," *Risk Analysis*, Vol. 16 (1), 5-18 (1996).

F.A. Seiler, "Error Propagation for Large Errors," *Risk Analysis*, Vol 7 (4), 509-518 (1987).

W. Slob, "Uncertainty Analysis in Multiplicative Models," *Risk Analysis*, Vol. 14 (4), 571-576 (1994).

A. E. Smith, P.B. Ryan, J. S. Evans, "The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating the Population Distribution of Risk," *Risk Analysis*, Vol. 12 (4), 467-474 (1992).

U.S. Environmental Protection Agency, Guidelines for Carcinogenic Risk Assessment, Federal Register 51(185), 33992-34003 (May 29, 1992).

U.S. Environmental Protection Agency, *Source Assessment: Analysis of Uncertainty - Principles and Applications*, EPA/600/2-79-004 (August, 1978)

U.S. Environmental Protection Agency, Guidelines for Exposure Assessment, Federal Register 57(104), 22888-22938 (May 29, 1992).

U.S. Environmental Protection Agency, Summary Report for the Workshop on Monte Carlo Analysis, EPA/630/R-96/010 (September, 1996).

Figure 1a. Example Monte Carlo Estimate of the PDF for Lifetime Cancer Risk

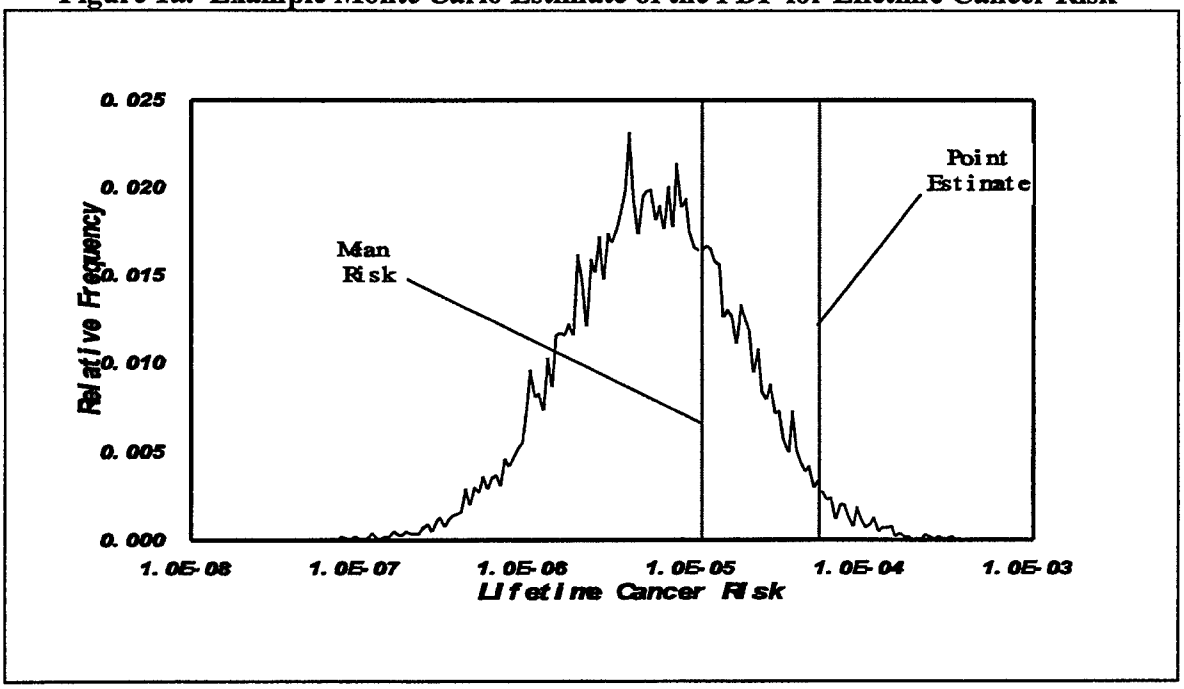


Figure 1b: Example Monte Carlo Estimate of the CDF for Lifetime Cancer Risk

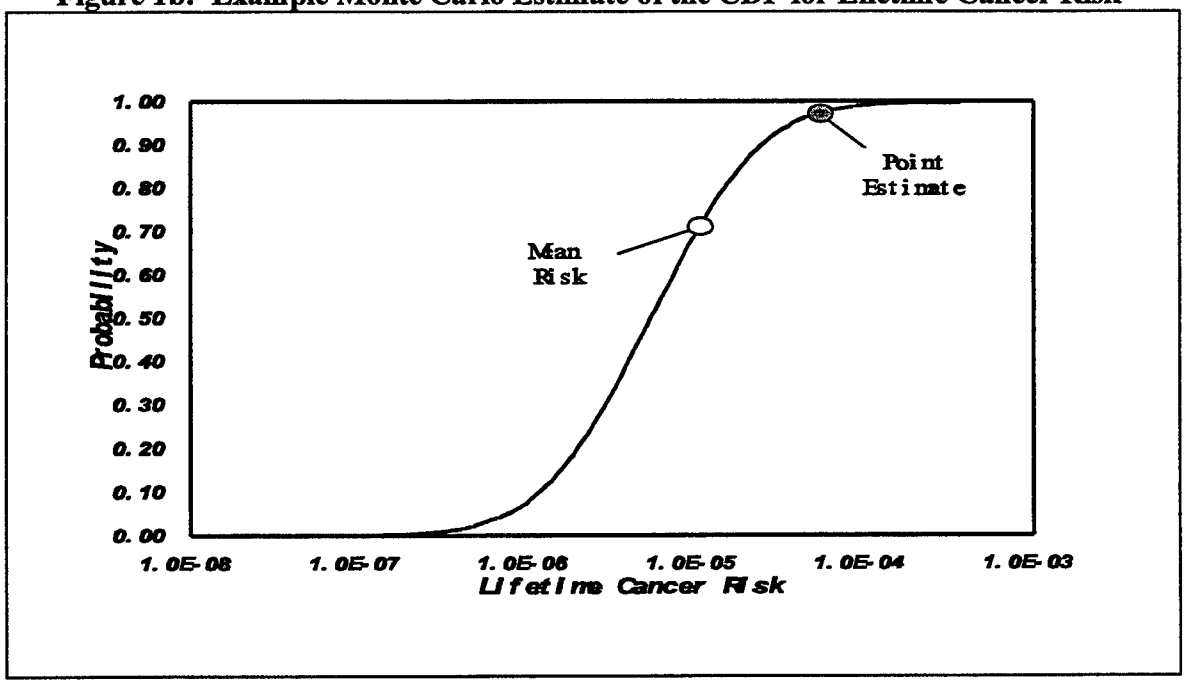
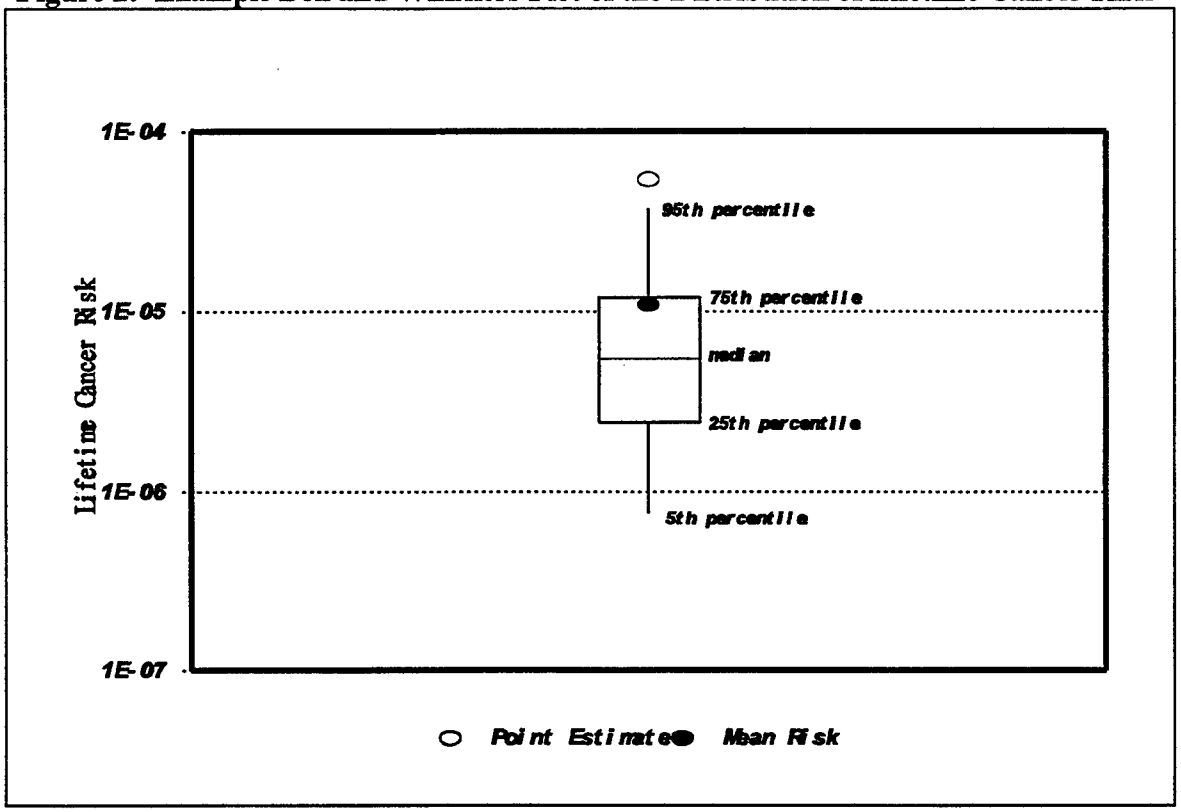


Figure 2: Example Box and Whiskers Plot of the Distribution of Lifetime Cancer Risk



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION VIII (8HWM-SM)
999 18th STREET - SUITE 500
DENVER, COLORADO 80202-2466

Region VIII

REGION 8 SUPERFUND TECHNICAL GUIDANCE

No. RA-10: Monte Carlo Simulation

Sep 1995

Risk Assessment (Short Title / Key Words)

TITLE: *Use of MONTE CARLO Simulation in Risk Assessments*

authors: Region 8 Toxicologists

contact: Dr. Susan Griffin
ph: 303-294-1062

approvals: PSA 9/29/95 Chief, TS /date; Carol L. Campbell 9/29/95 Chief, SM /date; CLC for Dupree 9/29/95 Director, HWM /date
(initials)

HAZARDOUS WASTE MANAGEMENT DIVISION, SUPERFUND MANAGEMENT BRANCH, TECHNICAL SECTION

SUMMARY

EPA's current risk assessment methods express health risks as single numerical values or "single-point" estimates of risk. This technique provides little information about uncertainty and variability surrounding the risk estimate. Recent EPA guidance (EPA, 1992a) recommends developing "multiple descriptors" of risk to provide more complete information to Agency decision-makers and the public. Monte Carlo simulation can be an effective way to produce these multiple risk descriptors. This document recommends guidelines under which Region VIII risk assessors may accept the optional use of Monte Carlo simulation to develop multiple descriptors of risk. *The Region will continue to require RME and Central Tendency single-point risk estimates, prepared under current national guidance, in conjunction with optional Monte Carlo simulations.*

SINGLE RISK ESTIMATES VS. MULTIPLE DESCRIPTORS

EPA designed its human health risk assessment guidance (*e.g.,* EPA, 1991, 1989a and 1989b) to produce protective, rather than best, estimates of risk. EPA is aware that true risks are probably less than its estimates, but has chosen a regulatory policy of giving the benefit of uncertainty surrounding the risk assessment to the exposed public.

These protective risk estimates sometimes create difficulty for Agency decision-makers and the public. Site-specific Regional risk assessments usually present risk as a single number, or single-point estimate, accompanied by a qualitative discussion of uncertainty. The public tends to focus on the single-point estimate and to overlook the uncertainty, which may span several orders of magnitude. EPA risk managers, though aware of the uncertainty, must still justify their decision to either accept or reduce the single-point risk. If the risk is close to the maximum acceptable level, it is likely that different assumptions would have produced a different risk number, leading to a different decision. In this way, single-point risk assessment methods place the risk assessor in an inappropriate risk management role.

Recent EPA guidance on risk characterization (EPA, 1992a) discusses this problem in depth and recommends the use of multiple risk descriptors in addition to protective, single-point risk estimates. Inclusion of these additional risk descriptors provides the public with more complete information on the likelihood of various risk levels and risk managers with multiple risk-based cleanup goals from which to choose. This guidance, based in large part upon that developed for EPA Region III by Dr. Roy L. Smith, concerns the use of Monte Carlo simulation as an effective source of multiple risk descriptors.

MONTE CARLO SIMULATION

Monte Carlo simulation is a statistical technique that can be used to simulate the effects of study variability and informational uncertainty that often accompany many "real world" situations. It is a process whereby an outcome is calculated repeatedly for many "what if" scenarios, using in each iteration randomly selected values, within the probability density function of the variable, for each of the variable or uncertain input values. Information on the range and likelihood of possible values for these parameters is contained in mathematical expressions termed Probability Density (or Distribution) Functions (PDFs). For risk assessment applications of Monte Carlo simulation, PDFs are typically normal, lognormal, uniform or triangular in nature (Smith 1994, Thompson et al. 1992, Whitmyer et al. 1992). Although the simulation process is internally complex, commercial computer software (e.g., @Risk from Palisade Corporation and Crystal Ball from Decisioneering Inc.) performs the calculations as a single operation, presenting results in simple graphs and tables. These results approximate the full range of possible outcomes and the likelihood of each. When Monte Carlo simulation is applied to risk assessment, risk can be displayed as a frequency distribution graph that is similar in appearance to the familiar bell-shaped curve, a form more intuitively understood by nonstatisticians.

When compared with alternative approaches for assessing parameter uncertainty or variability (e.g., analytical uncertainty propagation or classical statistical analysis), the Monte Carlo technique has the advantages of very general applicability, no inherent restrictions on input distributions or input-output relationships, and relatively straight-forward computations (EPA 1992b). In its application to risk assessment, Monte Carlo simulation not only generates results that can be expressed in a more easily understood graphical format, but it also permits the degree of conservativeness to be specified quantitatively (i.e., the Reasonable Maximum Exposure (RME) can be defined as the upper 90th or 95th percentile of output results). Furthermore, risk results are more easily justified statistically, and the uncertainty underlying them can be discussed quantitatively. In general, this approach can satisfactorily address the goals of uncertainty analysis outlined in recent EPA guidance (EPA 1992a).

Monte Carlo-type distributional simulations also have some limitations that have to date restrained EPA from accepting it as a preferred risk assessment tool. These concerns are briefly noted below, as are a number of relevant comments intended to provide additional perspective.

1. Available software cannot distinguish between variability and uncertainty. Some factors, such as body weight and rates of food and water ingestion, evidence natural inter-individual differences that may be described by relatively well known population distributions, and are referred to as "variability" or "stoichasticity." Other factors, such as frequency and duration of trespassing or the true mean contaminant concentrations to which individuals are exposed, can often simply be unknown. This lack of knowledge is called "uncertainty." In specific instances some parameters may reflect both variability and uncertainty. Current Monte Carlo software treats uncertainty as if it were

variability, which is of concern because it could lead to misleading conclusions. Other types of mathematical models are available, such as Multiple Objective Decision Support Systems (MODSS) to deal with subjective uncertainty.

Knowledge uncertainty is potentially reducible by additional research and measurement, and does not reflect an inherent population characteristic of variability that, in this case, contributes to the true distribution of risks experienced by population members. One way of at least partially managing this problem is to employ a "two-phase Monte Carlo sampling structure", whereby true parameter variability is modeled repetitively under different scenarios of uncertainty (MacIntosh et al. 1994). This approach generates a family of distributions, one for each uncertainty scenario (e.g., one each at the 5th, 50th and 95th percentiles of total estimated uncertainty), and thus conceptually and visually segregates stochastic variability from knowledge uncertainty. In the context of portraying risk distributions as a part of site-specific, clean-up decision making, EPA Region VIII prefers that variability and uncertainty be treated separately to the extent possible.

2. Ignoring correlations among exposure variables can bias or distort Monte Carlo calculations (e.g., an iteration using both a very high body weight and a very low body surface area is not realistic), yet information on possible correlations is often lacking.

   To some degree this problem can be ameliorated by using site-specific data and the associated correlations that become evident. Specified correlations can then be taken into account by the computer software package. The adverse consequences of neglecting parameter correlations also may be somewhat mitigated by the presence of other random-input parameters that have similar or greater levels of uncertainty/ variability (Smith et al. 1992). Furthermore, practices such as using multiple age groups to represent chronic or lifetime exposures can substantially reduce the negative effect of not determining the degrees of correlation among interdependent parameters (Finley and Paustenbach 1994).

3. Exposure factors developed from short-term studies with large populations may not accurately represent long-term conditions in small populations. It is difficult to generalize about the degree to which this concern could impact a given risk assessment, but maximizing the collection and use of site-specific data is one tactic that should lessen the problem. Furthermore, standard RME calculations are potentially susceptible to the same type of difficulties.

4. The tails of Monte Carlo risk distributions, which are of the greatest regulatory interest, are the most sensitive to the particular "shape" of the input distributions.

   This presents a significant concern in some cases; in others it does not. The exact shape of the PDF has been reported to exert only a "minimal" effect on the tail values (e.g., the 90th or 95th percentiles) when the mean and variance of the PDF's data were held constant (Finley and Paustenbach 1994, Hoffman and Hammonds 1992). It has been EPA Region VIII's experience that the PDF shape can significantly affect the distribution of risk outcome when the variable in question is a substantial "driver" of risk variability. Such variables can be identified by performing a sensitivity analysis of the risk equation after PDF parameters have been defined. In the occupational human health examples discussed in Appendix 2, negligible changes in risk distri-bution were observed when the shape of the body-weight PDF was changed from lognormal to normal. This is not surprising, given that sensitivity analysis indicated that the body-weight parameter contri-buted only 0.6 percent to risk variability under the example conditions (Simulation 2). In contrast, changing the shape of the PDF for contaminant concentration in soil (a 44 to 76 percent contributor to risk variability, Simulations 2 to 4) resulted in significant alterations in risk distribution in the 90 to 100th percentile range. Furthermore, in this example, while the mean risk level was affected by less than 10 percent, the median (50th percentile) risk changed by a factor of over 11 (Simulation 2 versus Simulation 4). Further discussion and comparative data are provided in Appendix 2.

Because of reservations concerning these practical issues, Region VIII cannot at this time recommend Monte Carlo simulation as the sole, or even principal, risk assessment method. Nevertheless, Monte Carlo simulation provides certain advantages over the qualitative procedures currently used to analyze uncertainty and variability. For baseline risk assessments at NPL sites, Region VIII recommends that uncertainty and variability surrounding single-

point risk estimates rely on multiple descriptors of risk (EPA 1992a). Monte Carlo simulation can be an acceptable method for developing these multiple descriptors.

GUIDELINES FOR USING MONTE CARLO SIMULATION

Region VIII risk assessors believe that Monte Carlo simulation requires more development before it can serve as the primary risk assessment method, for reasons described above. However, the technique has clear advantages over the qualitative analyses of uncertainty and variability currently in use. Region VIII will accept Monte Carlo simulations submitted as uncertainty/variability analyses in risk assessments, under the following guidelines:

1. Submit a work plan for EPA review before doing the Monte Carlo simulation, to ensure the work will be acceptable to EPA. The workplan should describe the software to be used, the exposure routes and models, and input probability distributions and their sources. An issue often of particular concern is how to most appropriately model contaminant concentrations in soil. EPA expects that peer-reviewed literature and site-specific data will be used whenever possible. A partial list of references, not necessarily reviewed or endorsed by EPA Region VIII, that provide example PDFs, PDF selection guidance and related information is presented in Appendix 1. When established PDFs are not available or appropriate, professional judgement and consideration of all relevant information may be used to select candidate PDFs for EPA review. Describe how correlations among input variables will be handled. If reasonable PDFs cannot be produced then default or site-specific point estimates for inputs should be used.

2. Include only exposure variables in the Monte Carlo simulation. Enter reference doses and slope factors as single numbers, except for specific contaminants for which the EPA Office of Research and Development has approved frequency distributions.

3. Include only significant exposure scenarios and contaminants in the Monte Carlo simulation. First, calculate RME risks for all exposure routes under current guidance. In general, Monte Carlo simulation is most appropriate when total RME cancer risk exceeds 1E-06, the noncarcinogenic hazard index exceeds 1.0, or for environmental receptors, when exposures exceed relevant toxicity reference values. Only those exposure routes that contribute significantly to the total risk or hazard under consideration need be modeled. Include only contaminants that contribute 1% or more of the total RME risk or hazard index.

4. Use Monte Carlo simulation only to analyze uncertainty and variability, as a "multiple descriptor" of risk. Remember to discuss with EPA Region VIII the complex issue of variability versus uncertainty before performing Monte Carlo analysis, and where possible to model the two separately. Include standard RME and average exposure risk estimates in all graphs and tables of Monte Carlo results. Generate deterministic risks using current EPA national guidance (EPA 1992a, 1991, 1989a and 1989b).

5. Include graphs and tables showing and describing each input distribution, distributions of risk for each exposure route, and distributions of total risk (summed across exposure pathways and age groups, as appropriate under current guidance). Sensitivity analyses should also be included to help focus on the major contaminants and pathways of exposure.

6. Although the use of Monte Carlo simulation in risk assessment has received greater attention with respect to human receptors, under appropriate conditions of sufficient data and valid PDFs, this guidance may also be applied to environmental receptors (see example in Appendix 3).

SUMMARY

Region VIII will accept Monte Carlo simulations that conform to the guidelines in this document (illustrated in the flow chart in Figure 2), as part of baseline human health or environmental risk assessments. The most important guideline is that all risk assessments must include single-point RME and average exposure risk estimates prepared under current EPA national guidance. The Region will accept Monte Carlo simulation only as an optional addition to, not a substitute for, current risk assessment methods.

REFERENCES

EPA. 1992a. U. S. Environmental Protection Agency. Guidance on risk characterization for risk managers and risk assessors. Washington, DC: U.S. Environmental Protection Agency, Office of the Administrator, memorandum from F. Henry Habicht on 26 February 1992.

EPA. 1992b. U.S. Environmental Protection Agency. Guidelines for exposure assessment (final). Fed. Reg. 57(104)22887-22938.

EPA. 1991. Standard default exposure factors, risk assessment guidance for Superfund, Volume I: Human health evaluation manual supplemental guidance. Washington, DC: U.S. Environmental Protection Agency Office of Solid Waste and Emergency Response, Toxics Integration Branch, OSWER Directive 9285.6-03.

EPA. 1989a. U. S. Environmental Protection Agency. Risk assessment guidance for Superfund, Volume I: Human health evaluation manual (Part A). Washington, DC: U.S. Environmental Protection Agency Office of Solid Waste and Emergency Response, Toxics Integration Branch, EPA/540/1-89/002.

EPA. 1989b. U. S. Environmental Protection Agency. Exposure factors handbook. Washington, DC: U.S. Environmental Protection Agency, Office of Health and Environmental Assessment, EPA/600/8-89/043.

Finley B, Paustenbach D. 1994. The benefits of probabilistic exposure assessment: three case studies involving contaminated air, water, and soil. Risk Analysis 14:53-73.

Hoffman FO, Hammonds JS. 1992. An introductory guide to uncertainty analysis in environmental and health risk assessment. Environmental Sciences Division, Oak Ridge National Laboratory, ESD Publication 3920. Prepared for the U.S. Department of Energy, Washington, DC.

MacIntosh DL, Suter II GW, Hoffman FO. 1994. Uses of probabilistic exposure models in ecological risk assessments of contaminated sites. Risk Analysis 14:405-419.

Smith RL. 1994. Use of Monte Carlo simulation for human exposure assessment at a Superfund site. Risk Analysis 14:433-439.

Smith AE, Ryan PB, Evans JS. 1992. The effect of neglecting correlations when propagating uncertainty and estimating population distribution of risk. Risk Analysis 12:467-474.

Thompson KM, Burmaster DE, Crouch AC. 1992. Monte Carlo techniques for qualitative uncertainty analysis in public health risk assessments. Risk Analysis 12:53-63.

Whitmyer GK, Driver JH, Ginevan ME, et al. 1992. Human exposure assessment I: Understanding the uncertainties. Toxicol. Indust. Health 8:297-320.

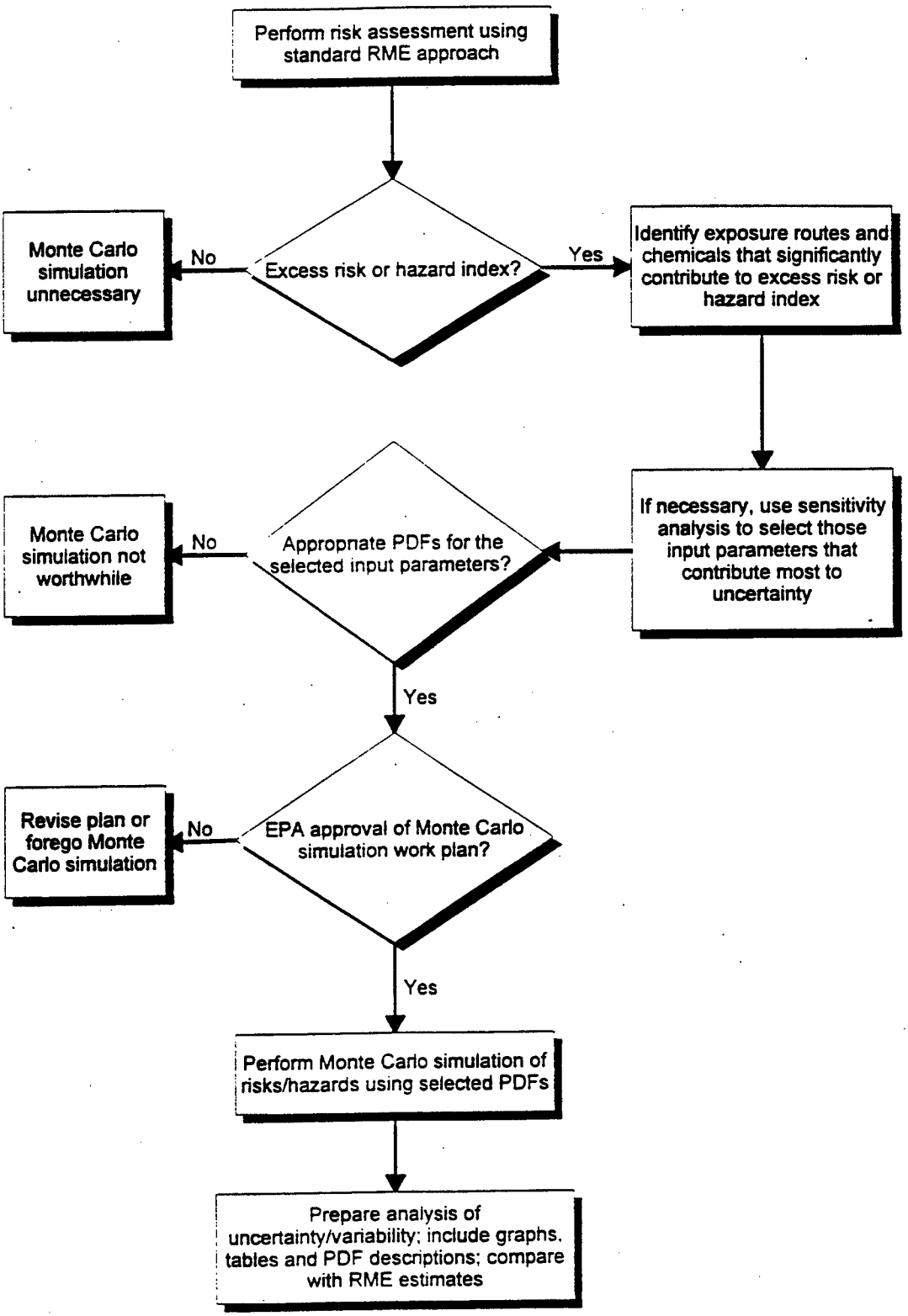


FIGURE 1 FLOW CHART FOR THE USE OF MONTE CARLO SIMULATION TO CHARACTERIZE UNCERTAINTY/VARIABILITY IN RISK ASSESSMENT ESTIMATES

APPENDIX 1
SAMPLE REFERENCES CONTAINING SOME COMMONLY USED OR PROPOSED PDFs AND RELATED INFORMATION[a]

American Industrial Health Council. 1994. Exposure factors sourcebook. Washington, DC: American Industrial Health Council.

Boyce CP, Evans CG, Schoof RA. 1992. Poster presented at the 31st Annual Meeting of the Society of Toxicology, Seattle, WA, February 23-27, 1992. Abstract #933.

Brainard J, Burmaster DE. 1992. Bivariate distributions for height and weight of men and women in the United States. Risk Analysis 12:267-275.

EPA. 1993. U.S. Environmental Protection Agency. Wildlife Exposure Factors Handbook. Office of Health and Environmental Assessment. Office of Research and Development. EPA/600/R-93/187a.

EPA. 1991. U.S. Environmental Protection Agency. Office of Solid Waste and Emergency Response. Supplemental guidance; standard default exposure factors. Interim final. Washington, DC: U.S. Environmental Protection Agency, OSWER Directive 9285.6-03.

EPA. 1989. U.S. Environmental Protection Agency. Exposure factors handbook. Exposure Assessment Group, Office of Health and Environmental Assessment, U.S. Environmental Protection Agency. EPA/600/8-89/043.

Finley B and Paustenbach D. 1994. The benefits of propabilistic exposure assessment: three case studies involving contaminated air, water, and soil. Risk Analysis 14:53-73.

Finley B, Proctor D, Scott P, Harrington N, Paustenbach D, Price P. 1994. Recommended distributions for exposure factors frequently used in health risk assessment. Risk Analysis 14:533-553.

Finley B, Scott PK, Mayhall DA. 1994. Development of a standard soil-to-skin adherence probability density function for use in Monte Carlo analyses of dermal exposure. Risk Analysis 14:555-569.

Hattis D, Silver K. 1994. Human interindividual variability--a major source of uncertainty in assessing risks for noncancer health effects. Risk Analysis 14:421-431.

MacIntosh DL, Suter II GW, Hoffman FO. 1994. Uses of probabilistic exposure models in ecological risk assessments of contaminated sites. Risk Analysis 14:405-419.

McKone TE. 1994. Uncertainty and variability in human exposures to soil contaminants through home-grown food: a Monte Carlo assessment. Risk Analysis 14:449-463.

McKone TE. 1991. The precision of a fugacity-based model for estimating dermal uptake of chemicals from soil. In: Proceedings of hydrocarbon contaminated soils, Vol. 1, Lewis Publishers, Chelsea, MI, pp. 555-574.

McKone TE, Bogen KT. 1991. Predicting uncertainties in risk assessment. Environ. Sci. Technol. 25:)1674-1681.

Meeks YJ, Salhotra AM. 1990. Monte Carlo approach to exposure assessment. In: Proceedings of the 1990 Specialty Conference, National Conference on Environmental Engineering, ASCE, New York, NY, pp. 775-782.

Murray DM, Burmaster DE. 1994. Estimated distributions for average daily consumption of total and self-caught fish for adults in Michigan angler houeholds. Risk Analysis 14:513-519.

Paustenbach DJ, Wenning RJ, Lau V, et al. 1992. Recent developments on the hazards posed by 2,3,7,8-tetrachlorodibenzo-p-dioxin in soil: Implications for setting risk-based cleanup levels at residential and industrial sites. J. Toxicol. Environ. Health 36:103-149.

Roseberry AM, Burmaster DE. 1992. Lognormal

(a) These are example references only, and have not necessarily been reviewed or endorsed by EPA Region VIII. All distributions to be used in Monte Carlo simulations must be agreed to in advance by EPA
Region VIII.

distribution for water intake by children and adults. Risk Analysis 12:99-104.

Ruffle B, Burmaster DE, Anderson PD, Gordon HD. 1994. Lognormal distributions for fish consumption by the general U.S. population. Risk Analysis 14:395-404.

Smith RL. 1994. Use of Monte Carlo simulation for human exposure assessment at a Superfund site. Risk Analysis 14:433-439.

Thompson KM, Burmaster DE, Crouch AC. 1992. Monte Carlo techniques for qualitative uncertainty analysis in public health risk assessments. Risk Analysis 12:53-63.

Thompson KM, Burmaster DE. 1991. Parametric distributions of soil ingestion by children. Risk Analysis 11:339-342.

Whitmyer GK, Driver JH, Ginevan ME, et al. 1992a. Human exposure assessment I: Understanding the uncertainties. Toxicol. Indust. Health 8:297-320.

Whitmyer GK, Driver JH, Ginevan ME, et al. 1992b. Human exposure assessment II: Quantifying and reducing the uncertainties. Toxicol. Indust. Health 8:321-342.

APPENDIX 2
MONTE CARLO SIMULATION: DETAILED HUMAN HEALTH EXAMPLE

To illustrate the use of Monte Carlo simulation in human health risk assessment, as compared with the standard deterministic approach (Risk Assessment Guidance for Superfund (RAGS) Part A, EPA, 1989), a simplified excess cancer risk assessment calculation is used here as an example. There are seven input parameters needed to solve for risk, as shown in the equation below:

$$\text{Risk} = \frac{\text{Cs} \times \text{IR} \times \text{1E-06} \times \text{EF} \times \text{ED} \times \text{SF}}{\text{BW} \times \text{AT}} \quad (1)$$

Where:

- Cs = Concentration of benzene in soil (mg benzene/kg soil)
- IR = Ingestion rate (mg soil/day)
- EF = Exposure frequency (days/year)
- ED = Exposure duration (years)
- SF = Slope factor for benzene (mg benzene/kg BW-day)$^{-1}$
- BW = Body weight (kg)
- AT = Averaging time (days)
- 1E-06 = Mg-to-kg conversion factor

In this example, a future adult worker is assumed to be potentially exposed to a single chemical (benzene), contained in a single medium (soil), by a single route (ingestion). It is further assumed that the future worker's potential exposure will be confined to a site having a relatively small area, all portions of which will be contacted equally by the worker over time. Under such assumptions, which are essentially the same as those typically assumed for residential scenarios, it is appropriate to utilize a determinate value based on the mean of soil sample data. To retain a certain degree of conservativism (especially when sample size is small or contaminant concentration is highly variable), it is recommended that the UCL95 of the arithmetic mean generally be used (assuming a lognormal distribution. the same approach used for the RME scenario) . When this value is associated with a high degree of uncertainty, collection of additional samples should be considered by the risk manager. When a site is very large and/or the contact behavior of individual members of the (potentially) exposed population is such that a given individual's contact with soil is not equal across the site when integrated over time, then it is reasonable to consider distributing the Cs term through the use of an appropriate PDF (see the alternate simulations discussed below). This issue of how, or whether or not to, distribute the "Cs" term should be discussed in advance with EPA Region VIII.

For the sake of simplicity, a number of other parameters used in estimating risk (for example, bioavailability of benzene from soil, fraction of soil ingested which is contaminated, etc.) have been omitted.

Calculation of Risk Using a Deterministic Approach

Selected standard or typical RME values for the input parameters in Equation No. 1 are listed below.

- Cs = 1,000 mg/kg soil (based on the following hypothetical set of sample concentrations (mg benzene/kg soil): 0.1, 0.1, 0.1, 0.5, 1.0, 5.0, 10.0, 15.0, 20.0, 23.0, 25.0, 27.0, 28.0, 30.0, 35.0, 40.0, 50.0, 75.0, 150, 400 and 1,000; sample arithmetic mean = 96.7, and the upper 95th confidence limit (UCL95) of this mean = 4,587 (assuming a lognormal distribution); since this value exceeds the maximum hit, the maximum hit value of 1,000 was selected as the RME concentration value).
- IR = 100 mg soil/day
- EF = 250 days/year (assuming 2 weeks vacation and 5 work days/week)
- ED = 25 years (upper 95th percentile value for years worked at the same location, as reported by the U.S. Bureau of Labor Statistics 1990)
- SF = 2.9E-02 (mg/kg-day)$^{-1}$ (upper 95th confidence limit of the dose-response curve)
- BW = 70 kg (mean value)
- AT = 70 years or 25,550 days

Based on these RME assumptions, excess cancer risk from the ingestion of soil contaminated with benzene is estimated to be 1.0E-05. As directed by EPA (1992a), a central tendency (AVG) risk descriptor is also calculated for this example. The values for Cs, IR, EF and ED can be represented by more average values as follows:

- Cs = 96.7 mg/kg (sample arithmetic mean)
- IR = 50 mg soil/day
- EF = 219 days/year (average of both full- and part-time workers, U.S. Bureau of Labor Statistics, 1991)
- ED = 5 years

The resulting excess cancer risk estimate is 8.6E-08.

However, there are several limitations inherent in this approach. For example, by using upper percentiles or confidence limits for at least three of the seven input parameters, the resulting estimate may be overly conservative. Burmaster and Harris (1993) have noted that the result generated from a calculation containing three values at their respective upper 95th confidence limits is itself at the upper 99.78th confidence limit. Also, by using a somewhat arbitrary mixture of mean and upper confidence limit/percentile values in the calculation, uncertainty and variability in the risk estimate cannot easily be discussed quantitatively.

Calculation of Risk Using Monte Carlo Simulation

As previously noted, when compared with alternative approaches for assessing parameter uncertainty (e.g., analytical uncertainty propagation or classical statistical analysis), the Monte Carlo technique has the advantages of very general applicability, no restrictions on input distributions or input-output relationships and relatively straightforward computations (EPA, 1992b).

The same equation used to calculate risk by the standard approach (Equation 1) is used in Monte Carlo simulation. The major difference in this approach is that four of the seven input parameters are now defined by PDFs instead of discrete values. There are several sources of PDFs, including site-specific data, published information such as that included in Appendix 1, and as a last resort, estimation based on professional judgment. The PDFs selected for this example as input into Equation 1, expressed in terms of "distribution type (mean, standard deviation), restrictions," are listed as follows:

- Cs = 1,000 mg/kg soil (the same as the determinate RME Cs assumption)
- IR = Lognormal (50, 50) (the "standard" central tendency value of 50 mg/kg was taken as the mean, and professional judgment was used to select the lognormal shape; the selected standard deviation results in 90th and 100th percentile values of approximately 100 and 1,000 mg/kg, respectively)
- EF = Triangular (minimum = 125 (half-time, or 5 days/week, 25 weeks/year), best = 219 (as above); maximum = 300 (6 days/week, 50 weeks/year)
- ED = Lognormal (7.3, 12.0), truncated at 0 and 50 years (arithmetic mean taken from Finley et al. 1994, citing U.S. Bureau of Statistics 1992; professional judgment was used to select the lognormal shape, and the standard deviation was selected to approximate 50th and 95th percentile values from the same U.S. Bureau of Statistics data)
- SF = Determinate value (2.9E-02 $[mg/kg\text{-}day]^{-1}$), upper 95th % confidence limit of the dose-response curve
- BW = Lognormal (71.0, 15.9), truncated at 40 and 150 kg (derived from Brainard and Burmaster 1992)
- AT = Determinate value (25,550 days), assumed lifetime in SF calculations

Based on these assumptions, the mean risk was calculated (rounded to two significant figures) to be 1.2E-06.[(a)]

This value is about 14-fold higher than the central tendency (AVG) estimate calculated previously, but still significantly below the standard RME risk. In this example, the 90th and 95th percentile risk estimates (2.8 and 4.4E-06, respectively) were also somewhat below the calculated RME risk of 1.0E-05, which was in fact virtually equivalent to the 99th percentile estimate of 1.1E-05. The Monte Carlo risk distribution is highly skewed, with the obviously improbable 100th percentile scenario having an estimated risk of 5.8E-05.

To illustrate the effects that precise PDF shape can have on output distribution, the above example (simulation 1) was run three additional times under somewhat varying conditions. The mean and several benchmark percentile values for each of these simulations are presented in the table shown below. In simulation 2, Cs was represented by a lognormal PDF having a mean of 282 mg/kg soil and a standard deviation of 5,315 mg/kg soil, with truncations at 0 and 10,000 mg/kg soil (based upon the same soil concentration data set used above); as would be expected, most values are lower than in the original simulation. In simulation 3, the BW distribution was changed from lognormal to normal (as is suggested in some references), while in simulation 4 the Cs distribution was changed from lognormal to normal (with the BW PDF returned to lognormal). The mean and several benchmark percentile risk values for all four simulations are compared below:

(a) Monte Carlo simulation was conducted using Crystal Ball from Decisioneering, Inc. Results were based on 10,000 iterations using Latin hypercube sampling (100 divisions) and a burst mode option of 5.

Percentile	Simulation 1	2	3	4
Mean	1.16E-06	2.19E-07	2.16E-07	2.64E-07
50th	4.46E-07	6.56E-09	6.48E-09	7.46E-08
90th	2.81E-06	2.28E-07	2.25E-07	6.18E-07
95th	4.41E-06	5.88E-07	5.78E-07	1.08E-06
≈99th	1.07E-05	3.85E-06	3.79E-06	3.03E-06
100th	5.78E-05	1.08E-04	1.05E-04	2.34E-05

As can be seen, changing the "non-driver" body-weight variable (a 0.6 percent contributor to risk variability in Simulation 2, see page A2-8) from a lognormal to a normal PDF had only negligible effect. However, when the same PDF shape change was tried with the concentration-in-soil (Cs) variable, which is a major contributor to risk variability (approximately 44 to 76 percent in Simulations 2 to 4), the resulting changes in risk distribution were much more substantial (Simulation 2 versus Simulation 4). While alterations in "RME-representative" 90th and 95th percentile risks were significant (2 to 3 fold), changes in the median and upper bound (50th and 100th percentiles, respectively) risks were in this case more dramatic (5 to 11 fold). The results illustrate the importance of giving attention to the selection of PDF shape, especially for those variables that are drivers of risk distribution. Significant changes in risk output may be expected in many cases with shifts from uniform to triangular to normal/lognormal PDF shapes, particularly when the input data have widely dispersed values.

A typical Crystal Ball output report (minus the actual frequency counts) of this example (Simulation 1) is presented on the following pages for illustration purposes.[a] This includes a sensitivity analysis, a summary description and statistics of the simulation, an output frequency distribution chart of the calculated excess cancer risk probabilities, risk values for specified percentiles and the numerical descriptors and distribution shapes for each of the variable input parameters. Results of the RME approach can be added to these Crystal Ball output figures and tables for comparative purposes.

References

Brainard J, Burmaster DE. 1992. Bivariate distributions for height and weight of men and women in the United States. Risk Analysis 12:267-275.

Burmaster DE, Harris RH. 1993. The magnitude of compounding conservatism in superfund risk assessments. Risk Analysis 13:131-134.

EPA. 1992a. U. S. Environmental Protection Agency. Guidance on risk characterization for risk managers and risk assessors. Washington, DC: U. S. Environmental Protection Agency, Office of the Administrator, memorandum from F. Henry Habicht on 26 February 1992.

EPA. 1992b. U.S. Environmental Protection Agency. Guidelines for exposure assessment (final). Fed. Reg. 57(104)22887-22938.

EPA. 1989. U.S. Environmental Protection Agency. Office of Emergency and Remedial Response. Risk assessment guidance for Superfund. Volume I. Human health evaluation manual (Part A). Interim final. Washington, DC: U.S. Environmental Protection Agency. EPA/540/1-89/002.

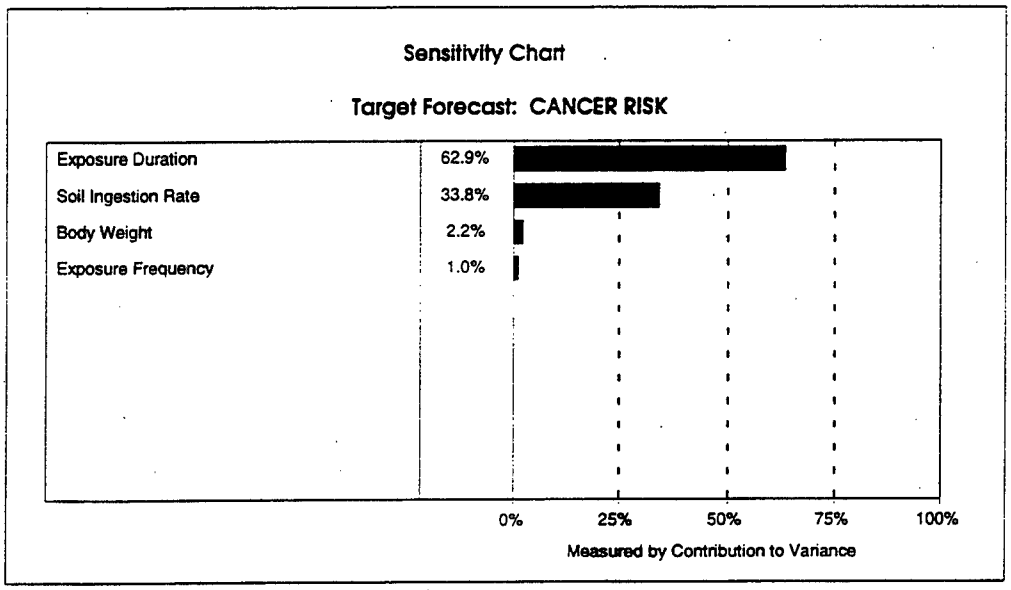
Finley B, Proctor D, Scott P, Harrington N, Paustenbach D, Price P. 1994. Recommended distributions for exposure factors frequently used in health risk assessment. Risk Analysis 14:533-553.

U.S. Bureau of Statistics. 1992. Employee tenure and occupational mobility in the early 1990s. U.S. Department of Labor, Bureau of Labor Statistics, Washington, DC, USDL92-386.

(a) The report presents results calculated by Crystal Ball using its "exact statistics" option. This takes about 12% longer (under these run conditions) on an IBM PS/2 Model 70 20 MHZ computer than using the "approximate statistics" option. Use of a 16 MHZ computer with a math coprocessor appeared to speed up exact calculations by about 14-18%, but approximate calculations by only 2 to 3%. Run conditions can substantially affect computation time (e.g., on this system, using a "burst mode" size of 50 required twice as much time as a burst mode of only 5). This 10,000 iteration run on the coprocessor machine required about 25 minutes. Employing the "sensitivity analysis" option did not appear to significantly increase run time. Several runs were performed on a 486DX66 MHZ computer, and run times were drastically reduced to just over 3 minutes.

OCCUPATIONAL HUMAN HEALTH EXAMPLE
(SIMULATION 1)

Crystal Ball Report
Simulation started on 9/10/95 at 21:54:53
Simulation stopped on 9/10/95 at 22:00:53



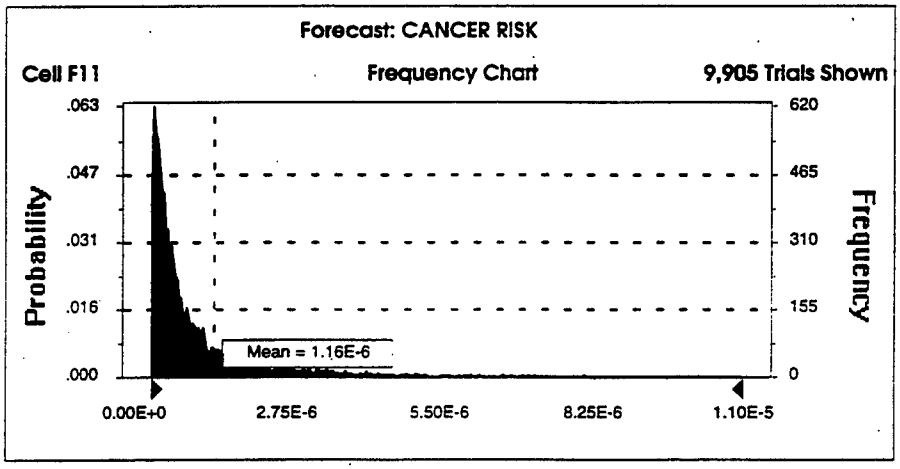
OCCUPATIONAL HUMAN HEALTH EXAMPLE
(SIMULATION 1)

Forecast: CANCER RISK

Summary:

Display Range is from 0.00E+0 to 1.10E-5
Entire Range is from 2.33E-9 to 5.78E-5
After 10,000 Trials, the Std. Error of the Mean is 2.34E-8

Statistics:	Value
Trials	10000
Mean	1.16E-06
Median	4.46E-07
Mode	---
Standard Deviation	2.34E-06
Variance	5.48E-12
Skewness	7.78
Kurtosis	111.57
Coeff. of Variability	2.02
Range Minimum	2.33E-09
Range Maximum	5.78E-05
Range Width	5.78E-05
Mean Std. Error	2.34E-08



Forecast: CANCER RISK (cont'd)

Percentiles:

Percentile	Value
0%	2.33E-09
5%	4.17E-08
10%	7.19E-08
15%	1.05E-07
20%	1.39E-07
25%	1.76E-07
30%	2.16E-07
35%	2.60E-07
40%	3.16E-07
45%	3.81E-07
50%	4.46E-07
55%	5.31E-07
60%	6.43E-07
65%	7.75E-07
70%	9.43E-07
75%	1.17E-06
80%	1.50E-06
85%	1.97E-06
90%	2.81E-06
95%	4.41E-06
100%	5.78E-05

End of Forecast

OCCUPATIONAL HUMAN HEALTH EXAMPLE
(SIMULATION 1)

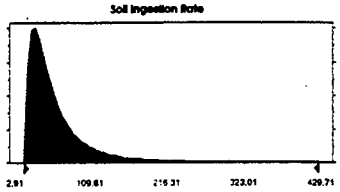
Assumptions

Assumption: Soil Ingestion Rate

Lognormal distribution with parameters:

Mean	50.00
Standard Dev.	50.00

Selected range is from 0.00 to 1,000.00
Mean value in simulation was 49.88

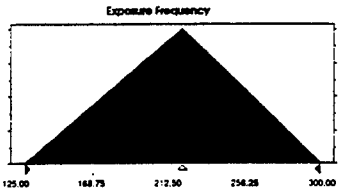


Assumption: Exposure Frequency

Triangular distribution with parameters:

Minimum	125.00
Likeliest	219.00
Maximum	300.00

Selected range is from 125.00 to 300.00
Mean value in simulation was 214.67

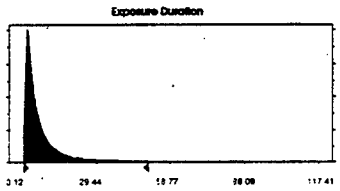


Assumption: Exposure Duration

Lognormal distribution with parameters:

Mean	7.30
Standard Dev.	12.00

Selected range is from 0.00 to 50.00
Mean value in simulation was 6.40

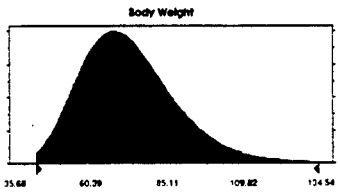


Assumption: Body Weight

Lognormal distribution with parameters:

Mean	71.00
Standard Dev.	15.90

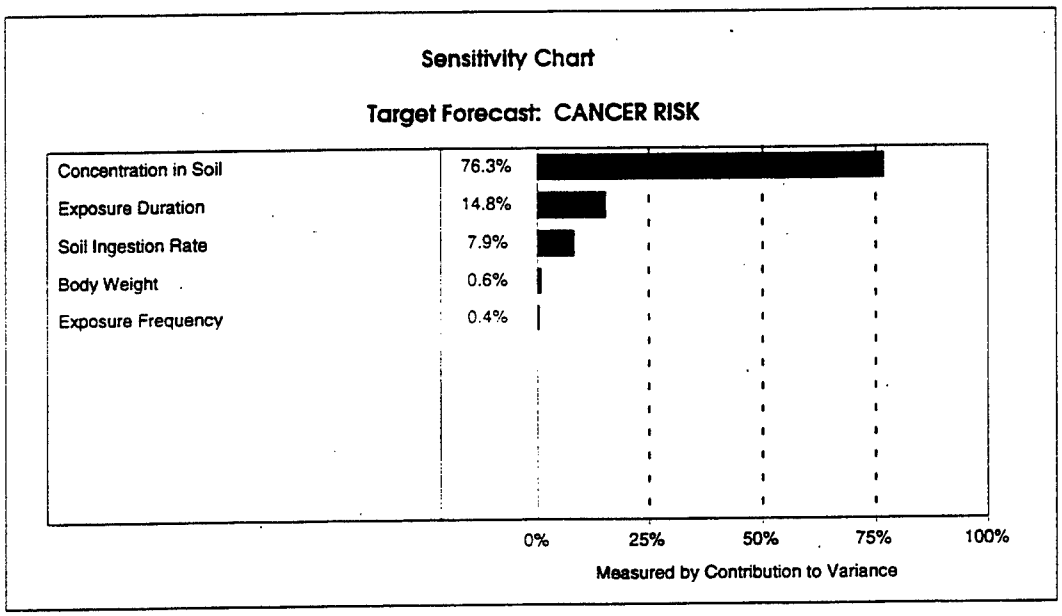
Selected range is from 40.00 to 150.00
Mean value in simulation was 71.19



End of Assumptions

OCCUPATIONAL HUMAN HEALTH EXAMPLE
(SIMULATION 2)

Crystal Ball Report
Simulation started on 2/22/95 at 23:46:27
Simulation stopped on 2/22/95 at 23:48:42



Note: In this alternate simulation, Cs (concentration in soil) was changed from a determinate value of 1,000 mg/kg to a lognormal PDF having a mean of 282 mg/kg, a standard deviation of 5,315 mg/kg and a truncated range of 0 to 10,000 mg/kg. The Cs data set is the same as that used for the primary simulation and the RME/AUG calculations.

APPENDIX 3
MONTE CARLO SIMULATION: ECOLOGICAL EXAMPLE

A simple food chain transfer from a prairie dog to a bald eagle was used to illustrate the use of Monte Carlo simulation in an environmental or ecological evaluation. For illustration purposes, a hazard quotient (HQ) equation was used as follows:

$$HQ = \frac{DI}{TRV} \quad (1)$$

Where:

DI = Daily Intake of a specific chemical contaminant (mg/kg-day)
TRV = Toxicity Reference Value (mg/kg-day)

The toxicity reference value (TRV) is a species-specific (i.e., bald eagle) and chemical-specific toxicity factor analogous to a reference dose (RfD) for wildlife. The daily intake (DI) (i.e., the average amount of the chemical at the body's exchange boundary) is described by the following equation:

$$DI = \frac{C \times CR \times EF \times FI \times FPD \times ED}{BW \times AT} \quad (2)$$

Where:

C = Contaminant Concentration (mg/kg Prairie Dog tissue)
CR = Contact Rate (kg food/day), the amount of food ingested on a daily basis by a bald eagle
EF = Exposure Frequency (365 days/year)
FI = Fraction Ingested (unitless), the portion of the diet that comes from the contaminated area
FPD = Fraction of a bald eagle's diet that is prairie dog (kg Prairie Dog tissue/kg food)
ED = Exposure Duration (years or days)
BW = Bald eagle's Body Weight (kg)
AT = Averaging Time (years or days)

In this example the concentration term (C) reflects the level of contaminant (dieldrin) in prairie dog.

Data available for this evaluation included information on a bald eagle's feed rate (FR), defined as the amount of food ingested daily by a bald eagle per unit of body weight:

$$FR = \frac{CR \times EF}{BW} \quad (3)$$

Available data for this exercise also included the fraction of a bald eagle's diet that was comprised of medium mammals (FMM) and the portion of this medium mammal diet that was prairie dog (PPD). These variables define the FPD, as follows:

$$FPD = FMM \times PPD \quad (4)$$

Where:

FMM = kg medium mammals/kg food
PPD = kg Prairie Dog/kg medium mammals

Additionally, it was assumed that the exposure duration equaled the averaging time (i.e., ED = AT). Equation No. 2 therefore was modified to allow for these available input variables and assumptions, as follows:

$$HQ = \frac{C \times FR \times FI \times FMM \times PPD}{TRV} \quad (5)$$

Calculation of a Hazard Quotient Using a Deterministic Approach

Point estimate hazard quotients were calculated for both average (AVG) and reasonable maximum exposures (RME), analogous to the risk descriptors for human health evaluations. Assumed values for the input parameters in Equation No. 5 are listed below:

C = 3.74E-01 mg dieldrin/kg prairie dog (upper 95th confidence limit of the arithmetic mean of 126 measured values, calculated using one-half the reported detection limit for nondetects and the measured value for detected results) (ESE 1988).
FR = 12.3E-02 kg/kg (for RME, maximum value, equal to the 93rd percentile of data cited in USFWS 1992); 8.91E-02 (for AVG, arithmetic mean of data cited in USFWS 1992)

FI = 1.0 (for RME, assumes all prairie dogs ingested are from the contaminated area); 0.5 (for AVG, assumes one-half of all prairie dogs ingested are from the contaminated area)
FMM = 0.936 (value cited in USFWS 1992)
PPD = 0.706 (calculated from data in USFWS 1992)
TRV = 1.10E-03 (Ford et al. 1992)

Based on these assumptions, the hazard quotients for a bald eagle due to the ingestion of a prairie dog food source contaminated with dieldrin are estimated to be 2.8E+01 (RME) and 1.00E+01 (AVG).

Calculation of Hazard Quotient Using Monte Carlo Simulation

The same equation (Equation 5) used to calculate a point estimate of the hazard quotient is used in Monte Carlo simulation. The major difference is that most of the six input parameters can be defined by probability distribution functions (PDFs) instead of discrete values. For this example, PDFs were readily available for two variables in Equation No. 5, the concentration and feed rate terms. For the remaining input parameters, PDFs were estimated or assumed or discrete values were used. Each input variable selected for this example, expressed in terms of "distribution type (mean, standard deviation), restrictions," is listed below.

C = Lognormal (3.634E-01, 1.368E+00), truncated at 0 mg/kg (the concentration of dieldrin cannot be a negative number; arithmetic mean and standard deviation were calculated from 126 measured values, using one-half the reported detection limit for nondetects and the measured value for detected results)
FR = Normal (8.913E-02, 2.689E-02) (values cited in USFWS 1992), truncated at 4.45E-02 kg/kg-bw/day (professional judgment) because it was assumed that a bald eagle consuming less food than one-half the arithmetic mean would either starve to death or become so weakened as to become vulnerable to predation
FI = Uniform: minimum = 0.01, maximum = 1.0 (professional judgment)
FMM = Determinate value: 0.936 (value cited in USFWS 1992)
PPD = Determinate value: 0.706 (calculated from data in USFWS 1992)
TRV = Determinate value: 1.10E-03 (Ford et al. 1992)

Based on these assumptions, the mean HQ was calculated to be 1.04E+01[(a)]. An HQ of less than 1.0E+00 indicates a nonhazardous situation. The Monte Carlo evaluation provides additional statistical information. For example, examination of the distribution of hazard quotients calculated by the Monte Carlo simulation indicates that approximately 60 to 65% of the bald eagle population would exceed the benchmark value of 1.0E+00. Additionally, the results predict that 50% of the bald eagle population would exhibit a HQ greater than 2.05E+00 and for 15 to 20% of the bald eagle population the HQ would be expected to exceed the calculated mean HQ value. The 90th and 95th percentiles for the hazard estimate were 2.06E+01 and 3.86E+01, respectively. The 100th percentile of this highly skewed distribution is 2.01E+03, which represents a very unlikely scenario.

A typical Crystal Ball output report of this example problem is presented on the following pages for illustration purposes.[(b)] This includes a summary description and statistics of the simulation, an output frequency distribution chart of the calculated hazard quotient probabilities, hazard quotient values for specified percentiles, and the numerical descriptors and distribution shapes for each of the variable input parameters.

References

ESE. 1988. Litigation technical support services, Rocky Mountain Arsenal, biota assessment, bald eagle study, winters 1986-1987, 1987-1988, final report. Contract DAAK11-84-D-0016. September 1988.

(a) Monte Carlo simulation was conducted using Crystal Ball from Decisioneering, Inc. Results were based on 10,000 iterations using Latin hypercube sampling (100 divisions) and a burst mode option of 5.
(b) The report presents results calculated by Crystal Ball using its "exact statistics" option. This takes about 12% longer (under these run conditions) on an IBM PS/2 Model 70 20 MHZ computer than using the "approximate statistics" option. Use of a math coprocessor appeared to speed up exact calculations by about 14-18%, but approximate calculations by only 2 to 3%. Run conditions can substantially affect computation time (e.g., on this system, using a "burst mode" size of 50 required twice as much time as a burst mode of only 5). This 10,000 iteration run on the coprocessor machine required about 25 minutes.

Ford KL, Applehans FM, Ober R. 1992. Development of toxicity reference values for terrestrial wildlife. Hazardous Materials Control Resources Institute Superfund '92 Proceedings. Washington, DC: December 1-3, 1992.

USFWS. 1992. U.S. Fish and Wildlife Service. The potential effects of Rocky Mountain Arsenal cleanup and Denver metropolitan transportation development on bald eagles, final report. U.S. Fish and Wildlife Service. Denver, CO.

Crystal Ball Report
Simulation started on 11/29/94 at 15:04:23
Simulation stopped on 11/29/94 at 15:28:38

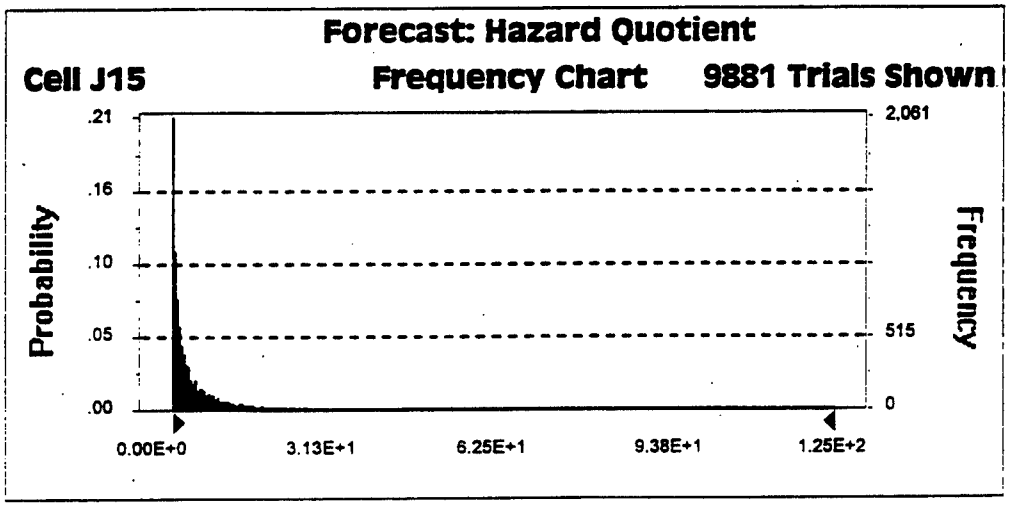
Forecast: Hazard Quotient

Summary:

Certainty Level is 98.81% based on Entire Range
Certainty Range is from -Infinity to 1.25E+2
Display Range is from 0.00E+0 to 1.25E+2
Entire Range is from 1.66E-3 to 2.01E+3
After 10,000 Trials, the Std. Error of the Mean is 0.44

Statistics:

	Display Range	Entire Range
Trials	9881	10000
Mean	7.09E+00	1.04E+01
Median (exact)	1.99E+00	2.05E+00
Mode (exact)	1.66E-03	1.66E-03
Standard Deviation	1.41E+01	4.36E+01
Variance	2.00E+02	1.90E+03
Skewness	4.01	(unavailable)
Kurtosis	22.59	(unavailable)
Coeff. of Variability	1.99	4.19
Range Minimum	0.00E+00	1.66E-03
Range Maximum	1.25E+02	2.01E+03
Range Width	1.25E+02	2.01E+03
Mean Std. Error	1.42E-01	4.36E-01



Forecast: Hazard Quotient (cont'd)

Percentiles for Entire Range:

Percentile	Hazard Quotient (exact)
0%	1.66E-03
5%	7.41E-02
10%	1.60E-01
15%	2.69E-01
20%	3.98E-01
25%	5.53E-01
30%	7.52E-01
35%	9.87E-01
40%	1.26E+00
45%	1.60E+00
50%	2.05E+00
55%	2.56E+00
60%	3.23E+00
65%	4.21E+00
70%	5.45E+00
75%	7.07E+00
80%	9.30E+00
85%	1.32E+01
90%	2.06E+01
95%	3.86E+01
100%	2.01E+03

End of Forecast

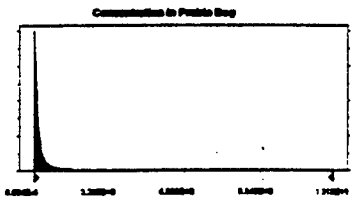
Assumptions

Assumption: Concentration in Prairie Dog

Lognormal distribution with parameters:

Mean	3.634E-01
Standard Dev.	1.368E+00

Selected range is from 0.000E+0 to +Infinity
Mean value in simulation vas 3.663E-1

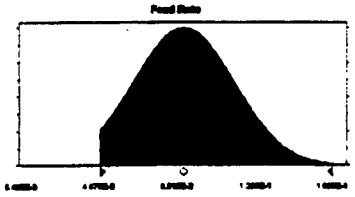


Assumption: Feed Rate

Normal distribution with parameters:

Mean	8.913E-02
Standard Dev.	2.689E-02

Selected range is from 4.450E-2 to +Infinity
Mean value in simulation was 9.197E-2

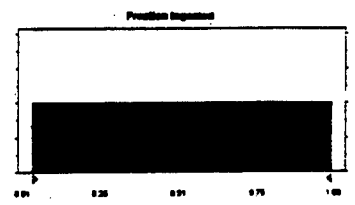


Assumption: Fraction Ingested

Uniform distribution with parameters:

Minimum	0.01
Maximum	1.00

Selected range is from 0.01 to 1 00
Mean value in simulation was 0.50



End of Assumptions

Does the use of Monte Carlo Methods in risk assessment add value? A case study of the Sangamo-Weston/Lake Hartwell site.

by Ted W. Simon, Ph.D., D.A.B.T., Toxicologist
U.S. Environmental Protection Agency

Probabilistic methods in risk assessment such as Monte Carlo have been touted in many quarters as a preferred alternative to the presently used deterministic methods. Many believe these deterministic methods lead to cleanup levels that are excessively health protective and more costly to achieve. The health protectiveness or "conservative" nature of present methods in risk assessment stem from an appropriate desire by regulators in the face of often considerable uncertainty to give the benefit of the doubt to the potentially exposed public. This article presents a case study of a Superfund site at which a Monte Carlo analysis was used in the risk assessment.

As will be seen, the Monte Carlo risk assessment formed a small but important part of the set of decision criteria used at the Sangamo site. Although the Monte Carlo risk assessment did not drive the decision, its value was to provide additional information or perspective for risk managers.

The Sangamo-Weston Site, Pickens, South Carolina

The Sangamo plant manufactured capacitors that used polychlorinated biphenyls (PCBs) as their dielectric substance from 1955 to 1977. During that time, waste water containing PCBs was discharged into Town Creek. The PCBs migrated downstream where they were deposited into the sediment of Twelve Mile Creek and Lake Hartwell. Sediment samples taken from Lake Hartwell in 1991 and 1992 showed that over 700 acres of lake bottom were contaminated with PCB-containing sediments. Fish tissue collected from Lake Hartwell by the South Carolina Department of Health and Environmental Control (SCDHEC) in 1991 and 1992 revealed concentrations of PCBs in largemouth and hybrid bass, channel catfish and bluegill.

Lake Hartwell is renowned for the quality of its fishing throughout the south. The larger migratory hybrid bass are prized as game fish. Many people living in northeast Georgia and western South Carolina depend on the lake for their living - fishing guides, bait shop operators, etc. Fishing restrictions at the lake affect them profoundly.

Fish Tissue Sampling Data and the Distribution in Game Fish

PCBs were found in 100% of the fish sampled. In largemouth and hybrid bass, the concentrations ranged from 0.5 to 19.7 mg/kg with a mean of 5.4 mg/kg. The acceptable level for the Food and Drug Administration for PCBs in fish tissue for occasional consumption (e.g. fish bought in the grocery store) is 2 mg/kg. For the Monte Carlo risk assessment, a lognormal distribution was used to model the uncertainty in fish tissue concentrations in fish from Lake Hartwell. The corresponding

deterministic Exposure Point Concentration for fish in the Twelve Mile Creek arm was 7.0 mg/kg, calculated as the 95th percent upper confidence limit on the mean.

Exposure Assessment

Exposure distributions were obtained from the Lake Hartwell and Twelve Mile Creek Recreational Angler Survey performed in 1992 by the South Carolina Wildlife and Marine Resources Department (SCWMRD). Almost 900 survey responses were available and were used to evaluate risk associated with fish consumption lake wide. Survey respondents answered questions regarding typical meal sizes and the number of individuals who might share the angler-caught fish, i.e. Ingestion Rate (IR). The number of meals and individual would eat in a month, i.e. the Exposure Frequency (EF) was also determined.

One would expect meal sizes to be correlated with age, body weight and gender; however, the information needed to effect these correlations in the Monte Carlo simulation was not included in the survey. The results of the survey for meal size may be biased upwards. The geometric mean of this survey was more than two times higher than the geometric mean of a USDA survey (120 g/meal) performed in the '70's.

Exposure Duration (ED) was assumed to be equal to enure of residence. Body weights (BW) were taken as the mean value for males and females within ten year age spans from age 25 to age 75 with an additional age group of 18-24 years.

The distribution of lifetime PCB intakes was determined by random sampling from the concentration, EF and ED distributions using Latin Hypercube sampling. Once complete, the distribution of intakes was multiplied by the upper bound cancer slope factor for PCBs to obtain a distribution of risks.

Distribution of Risks from Fish Consumption at Sangamo-Weston/Lake Hartwell

The deterministic risk from fish consumption calculated using reasonable maximum exposure (RME) assumptions to be $1x10^{-2}$. From the Monte Carlo risk assessment, the 90th percentile risk was $7x10^{-3}$, and the 95th percentile was $1x10^{-2}$, the same as the RME risk.

Cleanup Goals for Fish in Lake Hartwell

The FDA tolerance level of 2 ppm was selected as the preliminary remediation goal. Using EPA's risk assessment methods, a cancer risk of $1x10^{-4}$ is associated with a fish tissue concentration of 0.036 ppm. This level was based on a receptor consuming 357 grams of fish per meal, 60 meals per year over a 30 year period. Achieving this low concentration in fish tissue was determined to be technically impracticable.

Sedimentation Study

As a part of the Remedial Investigation, a modeling study of sediment deposition was performed for the Twelve Mile Creek arm. The study showed that sediment is continually being transported from upstream locations and deposited in Lake Hartwell. The study also showed that the PCB-contaminated sediment would be completely covered by upstream sediment within 30 years.

Time	Risk
Present	$1x10^{-2}$
1 year	$4x10^{-4}$
10 years	$6x10^{-5}$
20 years	$6x10^{-6}$
27 years	$1x10^{-6}$

Risk Reduction in time due to sediment deposition

Combining the sediment deposition model with a model for accumulation of PCBs in fish from sediment showed that the risks from fish consumption would continue

to decrease. The table shows the modeled reduction in risk with time assuming that clean sediment was being deposited in the lake.

Remedial Alternatives

The nine evaluation criteria for Superfund remedies are specified in the National Contingency Plan 40 CFR 300.430(e)(9). These nine criteria are:

1) Overall protection of human health and the environment
2) Compliance with Applicable or Relevant and Appropriate Requirements
3) Long-term Effectiveness and Permanence
4) Reduction of Toxicity, Mobility or Volume
5) Short-term Effectiveness
6) Implementability
7) Cost
8) State Acceptance
9) Community Acceptance

There were five proposed remedial alternatives: (1) No action; (2) Institutional Controls, including fish and sediment monitoring and public education; (3) Fishery Isolation, consisting of a semi-permanent "fish fence" at a bridge to prevent migratory hybrid and striped bass from entering the Twelve Mile Creek arm; (4) Capping with a Sediment Control Structure, consisting of placing clean sediment over the PCB contaminated sediment in the lake and constructing a weir to reduce the flow of contaminated sediment from upstream locations; and (5) Confined Disposal, a plan to rechannel the Twelve Mile Creek arm, dredging approximately 1.3 million cubic yards of sediment into a near shore landfill and stabilization of the dredge spoils with cement-based additives.

EPA chose Fishery Isolation as its preferred alternative. The community around Lake Hartwell did not accept this alternative (criterion 9). Instead, the community preferred Institutional Controls. It was felt that the "fish fence" would interfere with boating. The cost for alternatives 4 and 5 was significantly greater than any of the others with alternative 4 priced at between 30 and 50 million dollars and alternative 5 at about 600 million dollars.

The exact role of the Monte Carlo assessment was equivocal. The point estimate of risk for fish consumption in the Twelve Mile Creek arm is $1x10^{-2}$ as shown in the table. This risk level is far above the upper end of the risk range of $1x10^{-4}$ discussed in the NCP. However, the possible upward bias in the fish ingestion estimates indicated that the risks estimated from the distribution could be overestimates. The magnitude of overestimation of risk is unknown. Despite the fact the both the high end risk estimate from the Monte Carlo analysis and the RME point estimates were outside the acceptable risk range, the sedimentation study, cost and community acceptance were the basis for the remedial decision.

Percentile	Risk Level
RME (point Est.)	$1x10^{-2}$
90 (High End)	$7x10^{-3}$
50 (Central Tendency)	$1x10^{-3}$
20 (Approx. Mode)	$5x10^{-4}$

Risks from fish consumption at Lake Hartwell in 1992

You may contact Ted W. Simon for information at simon.ted@epamail.epa.gov.