♻EPA

Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000)

Technical Support Document
Volume 1: Risk Assessment

Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000)

Technical Support Document Volume 1: Risk Assessment

Final

Office of Science and Technology
Office of Water
U.S. Environmental Protection Agency
Washington, DC 20460

LIST OF ACRONYMS

AEL	Adverse-Effect Level
AWQC	Ambient Water Quality Criteria
BAF	Bioaccumulation Factor
BCF	Bioconcentration Factor
BMC	Benchmark Concentration
BMD	Benchmark Dose
BMR	Benchmark Response
BW	Body Weight
CFR	Code of Federal Regulations
CR	Consumption Rate
CWA	Clean Water Act
D	Dose
DI	Drinking Water Intake
DNA	Deoxyribonucleic Acid
ED_{10}	Dose Associated with a 10 Percent Extra Risk
EPA	Environmental Protection Agency
ER	Extra Risk
FEL	Frank Effect Level
FI	Fish Intake
GI	Gastrointestinal
HA	Health Advisory
IARC	International Agency for Research on Cancer
ILSI	International Life Sciences Institute
IRIS	Integration Risk Information System
kg	kilogram
L	Liter
LC_{50}	Lethal concentration to 50 percent of the population
LD_{50}	Lethal dose to 50 percent of the population
LED_{10}	The Lower 95 Percent Confidence Limit on a Dose Associated with a 10 Percent Extra Risk
LMS	Linear Multistage Model
LOAEL	Lowest Observed Adverse Effect Level
LR	Lifetime Risk
MF	Modifying Factor
mg	Milligrams
ml	Milliliters
MLE	Maximum Likelihood Estimate
MOA	Mode of Action
MOE	Margin of Exposure
NOAEL	No-Observed-Adverse-Effect Level
NOEL	No-Observed-Effect Level
NTIS	National Technical Information Service

OSTP	Office of Science and Technology Policy
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
POD	Point of Departure
$q_1{}^*$	Cancer Potency Factor
RfC	Reference Concentration
RfD	Reference Dose
RfD_{DT}	Developmental Toxicity Reference Dose
RPF	Relative Potency Factor
RSC	Relative Source Contribution
RSD	Risk-Specific Dose
SAB	Science Advisory Board
TEF	Toxicity Equivalency Factor
TSD	Technical Support Document
UF	Uncertainty Factor
USEPA	U.S. Environmental Protection Agency

METHODOLOGY FOR DERIVING AMBIENT WATER QUALITY CRITERIA FOR THE PROTECTION OF HUMAN HEALTH (2000)

TECHNICAL SUPPORT DOCUMENT
VOLUME 1
RISK ASSESSMENT

NOTICE

The policies and procedures set forth in this document are intended solely to describe EPA methods for developing or revising ambient water quality criteria to protect human health, pursuant to Section 304(a) of the Clean Water Act, and to serve as guidance to States and authorized Tribes for developing their own water quality criteria. This guidance does not substitute for the Clean Water Act or EPA's regulations; nor is it a regulation itself. Thus, it does not impose legally-binding requirements on EPA, States, Tribes or the regulated community, and may not apply to a particular situation based upon the circumstances.

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.

1. INTRODUCTION

This document provides technical support concerning cancer and noncancer risk assessment methods used in the *Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000)* (USEPA, 2000a; hereafter the "2000 Human Health Methodology").

1.1 BACKGROUND

Ambient water quality criteria (AWQC) developed under Section 304(a) of the Clean Water Act (hereafter the "CWA" or "the Act") are based solely on data and scientific judgments on the relationship between pollutant concentrations and environmental and human health effects. The 304(a) criteria do not reflect consideration of economic impacts or the technological feasibility of meeting the chemical concentrations in ambient water. As discussed below, 304(a) criteria are used by States and authorized Tribes to establish water quality standards, and ultimately provide a basis for controlling discharges or releases of pollutants.

The U.S. Environmental Protection Agency (EPA) published the availability of AWQC documents for 64 toxic pollutants and pollutant categories identified in Section 307(a) of the CWA in the *Federal Register* on November 28, 1980 (USEPA, 1980). The November 1980 *Federal Register* notice (hereafter the "1980 Methodology") also summarized the criteria documents and discussed in detail the methods used to derive the AWQC for those pollutants. The AWQC for those 64 pollutants and pollutant categories were published pursuant to Section 304(a)(1) of the CWA:

> *The Administrator, . . . shall develop and publish, . . . , (and from time to time thereafter revise) criteria for water quality accurately reflecting the latest scientific knowledge (A) on the kind and extent of all identifiable effects on health and welfare including, but not limited to, plankton, fish, shellfish, wildlife, plant life, shorelines, beaches, esthetics, and recreation which may be expected from the presence of pollutants in any body of water, including ground water; (B) on the concentration and dispersal of pollutants, or their byproducts, through biological, physical, and chemical processes; and (c) on the effects of pollutants on the biological community diversity, productivity, and stability, including information on the factors affecting rates of eutrophication and rates of organic and inorganic sedimentation for varying types of receiving waters.*

The 1980 Methodology provided two essential types of information: (1) discussions of available scientific data on the effects of the pollutants on public health and welfare, aquatic life, and recreation; and (2) quantitative concentrations or qualitative assessments of the levels of pollutants in water which, if not exceeded, will generally ensure adequate water quality for a specified water use.

The 1980 AWQC were derived using guidelines and methodologies developed by the Agency for calculating the impact of waterborne pollutants on aquatic organisms and on human health. Those guidelines and methodologies consisted of systematic procedures for assessing valid and appropriate data concerning a pollutant's acute and chronic adverse effects on aquatic organisms, nonhuman mammals, and humans. The guidelines and methodologies were fully described in Appendix B (for protection of aquatic life and its uses) and Appendix C (for protection of human health) of the November 1980 *Federal Register* Notice.

1.2 NEED FOR REVISION OF THE 1980 HUMAN HEALTH METHODOLOGY FOR DERIVING AWQC

l.2.1 Scientific Advances Since 1980

Since 1980, EPA risk assessment practices have evolved significantly in the areas of cancer and noncancer risk assessments, exposure assessments, and bioaccumulation assessment.

In cancer risk assessment, there have been advances on the use of mode of action (MOA) information to support both the identification of carcinogens and the selection of procedures to characterize risk at low, environmentally relevant exposure levels. Related to this is the development of new procedures for quantifying cancer risk at low doses to replace the current default linear multistage model (LMS).

In noncancer risk assessment, the Agency is moving toward the use of statistical models, such as the benchmark dose approach and categorical regression, to derive reference doses (RfDs) in place of the traditional NOAEL-(no-observed-adverse-effect level)-based method.

In exposure analysis, several new studies have addressed water consumption and fish consumption. These exposure studies provide a more current and comprehensive description of national, regional, and special-population consumption patterns; these are reflected in the 2000 Human Health Methodology (USEPA, 2000). In addition, more formalized procedures are now available to account for human exposure from multiple sources when setting health goals that address only one exposure source.

With respect to bioaccumulation, the Agency has moved toward the use of a bioaccumulation factor (BAF) to reflect the uptake of a contaminant by fish from all sources rather than just from the water column as reflected by the use of a bioconcentration factor (BCF) in the 1980 Methodology. The Agency has also developed detailed procedures and guidelines for estimating BAF values.

1.2.2 EPA Risk Assessment Guidelines Development Since 1980

When the 1980 Methodology was developed, EPA had not yet developed formal cancer or noncancer risk assessment guidelines. Since then, EPA has published several risk assessment guidelines documents. *Guidelines for Carcinogen Risk Assessment* were published in 1986

(USEPA, 1986a) (hereafter the "1986 cancer guidelines") as were *Guidelines for Mutagenicity Risk Assessment* (USEPA, 1986b). In 1996, the Agency published *Proposed Guidelines for Carcinogen Risk Assessment* (USEPA, 1996a) (hereafter the "1996 proposed cancer guidelines"), which were subsequently revised in July 1999 following extensive external review (USEPA, 1999a, hereafter the "1999 draft revised cancer guidelines"). When final guidelines are published, they will replace the current *Guidelines for Carcinogen Risk Assessment* published in 1986 (USEPA, 1986a) (hereafter the "1986 cancer guidelines").

With respect to noncancer risk assessment, the Agency published *Guidelines for Developmental Toxicity Risk Assessment* in 1991 (USEPA, 1991) and *Guidelines for Reproductive Toxicity Risk Assessment* in 1996 (USEPA, 1996b). In 1998, EPA published final *Guidelines for Neurotoxicity Risk Assessment* (USEPA, 1998), and in 1999, it issued draft *Guidance for Conducting Health Risk Assessment of Chemical Mixtures* (USEPA, 1999b). In addition, the Agency is developing a framework for cumulative risk assessment and the Office of Pesticide Programs has developed draft guidance for assessing cumulative risk of common mechanism pesticides and other substances.

1.3   PURPOSE OF THIS DOCUMENT

This *Risk Assessment Technical Support Document* (TSD) (hereafter the "Risk Assessment TSD") provides additional technical detail on the principles and recommendations presented in the 2000 Human Health Methodology for risk assessments to be used in deriving AWQC. Also included are illustrative examples to explain the thought process behind many of the new risk assessment directions being taken by EPA. For instance, there is an example of how to apply principles of the 1999 draft revised cancer guidelines to a chemical for which the MOA is considered to be a threshold process.[1] For noncancer assessment, an example is included on how to use the benchmark dose (BMD) approach.

The focus of the 2000 Human Health Methodology, which this document accompanies, is the development of AWQC to protect human health. The Agency intends to use the 2000 Human Health Methodology both to develop new AWQC for additional chemicals and to revise existing AWQC.

It is important to emphasize that the 2000 Human Health Methodology is also intended to provide States and authorized Tribes flexibility in setting water quality standards by providing scientifically valid options for developing their own water quality criteria that consider local conditions. States and authorized Tribes are encouraged to use the Methodology to derive their own AWQC. The 2000 Human Health Methodology also defines the default factors EPA will use in evaluating and determining consistency of State and Tribal water quality standards with the requirements of the CWA and the implementing federal regulation (40 CFR 131). These

---

[1] Throughout this document, the term "risk level" regarding a cancer assessment using linear approach refers to an upper bound estimate of excess lifetime cancer risk.

default factors will also be used by the Agency to calculate 304(a) criteria when promulgating water quality standards for a State or Tribe under Section 303(c) of the Act.

1.4   CRITERIA EQUATIONS

The following equations for deriving AWQC include toxicological parameters which are derived from scientific analysis, science policy, and risk management decisions. An example of an empirically measured, science-based value is a point of departure (POD) from an animal study [in the form of a lowest-observed-adverse-effect level (LOAEL)/no-observed-adverse-effect level (NOAEL)/ lower 95 percent confidence limit on a dose associated with a 10 percent extra risk (LED_{10})]. The decision to use animal effects as a surrogate for human effects involves judgment on the part of the EPA (and other agencies) as to the best practice to follow when human data are lacking. Such a decision is a matter of science policy. On the other hand, the choice to base AWQC on protection of the 90th percentile of the general population's water consumption rate is a risk management decision. In many cases, the Agency has selected parameters using its best judgment of the overall protection afforded by the resulting AWQC when all parameters are combined. This issue is discussed further in the 2000 Human Health Methodology document, along with further details on risk characterization as related to this Methodology with emphasis placed on explaining the uncertainties in the overall risk assessment.

The generalized equations for deriving AWQC based on noncancer and cancer effects are:

Noncancer Effects

$$\text{AWQC} = \text{RfD} \cdot \text{RSC} \cdot \left(\frac{\text{BW}}{\text{DI} + \sum_{i=2}^{4} (\text{FI}_i \cdot \text{BAF}_i)} \right)$$

(Equation 1-1)

Cancer Effects: Nonlinear Low-Dose Extrapolation

$$\text{AWQC} = \frac{\text{POD}}{\text{UF}} \cdot \text{RSC} \cdot \left(\frac{\text{BW}}{\text{DI} + \sum_{i=2}^{4} (\text{FI}_i \cdot \text{BAF}_i)} \right)$$

(Equation 1-2)

Cancer Effects: Linear Low-Dose Extrapolation

$$AWQC = RSD \cdot \left(\frac{BW}{DI + \sum\limits_{i=2}^{4} (FI_i \cdot BAF_i)} \right)$$

(Equation 1-3)

where:

AWQC	=	Ambient Water Quality Criterion (mg/L, or milligrams/Liter)
RfD	=	Reference dose for noncancer effects (mg/kg-day, or milligram/kilogram-day)
POD	=	Point of departure for carcinogens based on a nonlinear low-dose extrapolation (mg/kg-day), usually a LOAEL, NOAEL, or LED_{10}
UF	=	Uncertainty Factor for carcinogens based on a nonlinear low-dose extrapolation carcinogens (unitless)
RSD	=	Risk-specific dose for carcinogens based on a linear low-dose extrapolation (mg/kg-day) (Dose associated with a target risk, such as 10^{-6})
RSC	=	Relative source contribution factor to account for non-water sources of exposure. (Not used for carcinogens based on a linear low-dose extrapolation) May be either a percentage (multiplied) or amount subtracted, depending on whether multiple criteria are relevant to the chemical.
BW	=	Human body weight (default = 70 kg for adults)
DI	=	Drinking water intake (default = 2 L/day for adults)
FI	=	Fish intake (defaults = 0.0175 kg/day for general population and sport anglers, and 0.142 kg/day for subsistence fishers)
BAF	=	Bioaccumulation factor, lipid normalized (L/kg)

1.5   REFERENCES

USEPA (U.S. Environmental Protection Agency). 1980. Guidelines and methodology used in the preparation of health effect assessment chapters of the consent decree water criteria documents. *Federal Register* 45: 79347.

USEPA (U.S. Environmental Protection Agency). 1986a. Guidelines for carcinogen risk assessment. *Federal Register* 51: 33992-34003. September 24.

USEPA (U.S. Environmental Protection Agency). 1986b. Guidelines for mutagenicity risk assessment. *Federal Register* 51: 34006-34012. September 24.

USEPA (U.S. Environmental Protection Agency). 1986c. Guidelines for the health risk assessment of chemical mixtures. *Federal Register* 51: 33992-34003. September 24.

USEPA (U.S. Environmental Protection Agency). 1991. Guidelines for developmental toxicity risk assessment. *Federal Register* 56: 63798-63826.

USEPA (U.S. Environmental Protection Agency). 1996. Proposed guidelines for carcinogen risk assessment. *Federal Register* 61: 17960.

USEPA (U.S. Environmental Protection Agency). 1998. Guidelines for neurotoxicity risk assessment. *Federal Register* 63: 26926.

USEPA (U.S. Environmental Protection Agency). 1999a. *1999 Guidelines for Carcinogen Risk Assessment. Review Draft.* Risk Assessment Forum. Washington, DC. EPA/NCEA-F-0644. July.

USEPA (U.S. Environmental Protection Agency). 1999b. Guidance for conducting health risk assessment of chemical mixtures. *Federal Register* 64: 23833.

USEPA (U.S. Environmental Protection Agency). 2000. *Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000).* Office of Science and Technology, Office of Water. Washington, DC. EPA-822-B-00-004. August.

2. CANCER EFFECTS

This section provides a discussion of the current status of the cancer risk assessment methodology employed by EPA which is based on recent scientific developments and the Agency's experience in this field. A discussion is provided of:

C   Background information on the current cancer risk assessment methods in the 1986 *Guidelines for Carcinogen Risk Assessment* (USEPA, 1986; hereafter "1986 cancer guidelines"); and

C   New principles recommended in the *Guidelines for Carcinogen Risk Assessment. Review Draft* (USEPA, 1999a; hereafter "1999 draft revised cancer guidelines").[2] When final guidelines are published, they will replace the 1986 cancer guidelines, including their application in the Methodology for deriving AWQC for carcinogens.

2.1   1986 EPA GUIDELINES FOR CARCINOGENIC RISK ASSESSMENT

In 1986, EPA published its *Guidelines for Carcinogenic Risk Assessment* (hereafter "1986 cancer guidelines"). These guidelines were based on the publication by the Office of Science and Technology Policy (OSTP, 1985) that provided a summary of the state of knowledge in the field of carcinogenesis and a statement of broad scientific principles of carcinogen risk assessment on behalf of the federal government.

The 1986 cancer guidelines established a classification scheme to describe the nature of the cancer database and evidence supporting the carcinogenicity of an agent. This classification system is based on a similar scheme used at the time by the International Agency for Research on Cancer (IARC). This scheme is described briefly below. More detailed information can be obtained from the 1986 cancer guidelines.

The classification scheme utilizes several alpha-numerical groups for classifying chemicals with respect to the evidence available regarding their likely carcinogenic potential for humans:

Group A: Human carcinogen; sufficient evidence from epidemiological studies.

Group B: Probable human carcinogen; sufficient evidence in animals or limited evidence in humans.

Group C: Possible human carcinogen; limited evidence of carcinogenicity in animals in the absence of adequate human data.

---

[2] This is a revision of the *Proposed Guidelines for Carcinogen Risk Assessment* published in 1996 (USEPA, 1996).

Group D: Not classifiable; inadequate data or no data.

Group E: Evidence of noncarcinogenicity for humans; no evidence of carcinogenicity in adequate studies in at least two species or in both epidemiological and animal studies.

Within Group B there are two subgroups: B1 and B2. Group B1 is reserved for agents for which there is limited evidence of carcinogenicity from epidemiologic studies. Group B2 is generally for agents for which there is sufficient evidence from animal studies and for which there is inadequate evidence or no data from epidemiologic studies.

The 1986 cancer guidelines also include guidance on the definition of sufficient or limited evidence. The evidence from human studies is evaluated as "sufficient" when a causal relationship is indicated by the study. Human evidence is considered "limited" when a causal interpretation is credible, but alternative explanations are not sufficiently excluded.

When animal studies are used in the evaluation of carcinogenicity, "sufficient" evidence includes agents which have been demonstrated to cause:

C   An increased incidence of malignant tumors; or combined malignant and benign tumors;
    1)   in multiple species or strains;
    2)   in multiple experiments (e.g., with different routes of administration or using different dose levels); or
    3)   to an unusual degree in a single experiment with regard to high incidence, unusual site or type of tumor; or
C   An early age at onset.

For quantitative cancer risk estimation, the 1986 cancer guidelines recommended the use of the linearized multistage model (LMS) as the default approach based on the default assumption that chemical carcinogens cause DNA mutations. The 1986 cancer guidelines also stated that low-dose extrapolation models and approaches other than the LMS model might be considered more appropriate based on biological information showing mechanisms of action other than mutagenesis. However, no guidance was given in choosing other approaches; thus, departures from the LMS procedure have been rare in practice. The 1986 cancer guidelines recommended the use of body weight raised to the two/thirds power ($BW^{2/3}$) as a dose scaling factor between species based on the idea that dose would scale as a function of surface area of the body.

2.2   REVISIONS TO EPA'S CARCINOGEN RISK ASSESSMENT GUIDELINES

In 1996, EPA published *Proposed Guidelines for Carcinogen Risk Assessment* (USEPA, 1996; hereafter the "1996 proposed cancer guidelines"). EPA developed its 1999 draft revised cancer guidelines in response to the February 1997 and January 1999 USEPA Science Advisory Board (SAB) reviews of the proposal. When final guidelines are published, they will replace the 1986 cancer guidelines. These revisions are designed to ensure that the Agency's cancer risk

assessment methods reflect the most current scientific information and advances in risk assessment methodology.

In the meanwhile, the 1986 cancer guidelines are used and extended with principles discussed in the 1999 draft revised cancer guidelines. These principles arise from scientific discoveries concerning cancer made in the last 15 years and from EPA policy of recent years supporting full characterization of hazard and risk both for the general population and potentially sensitive groups such as children. These principles are incorporated in recent and ongoing assessments such as the reassessment of dioxin, consistent with the 1986 guidelines. Until final guidelines are published, information is presented to describe risk under both the old 1986 guidelines and 1999 draft revisions.

The 1999 draft revised cancer guidelines require the full use of all relevant information to convey the circumstances or conditions under which a particular hazard is expressed (e.g., route, duration, pattern, or magnitude of exposure). The 1999 draft revised cancer guidelines emphasize the understanding of mode of action (MOA) whereby the agent induces tumors. The MOA underlies the hazard assessment and provides the rationale for dose-response assessments.

The key principles in the 1999 draft revised cancer guidelines include:

a)   Hazard assessment is based on the analysis of all biological information rather than just tumor findings.

b)   An agent's MOA in causing tumors is emphasized to reduce the uncertainty in describing the likelihood of harm and in determining the dose-response approach(es).

c)   The 1999 draft revised cancer guidelines emphasize the conditions under which the hazard may be expressed (e.g., route, pattern, duration and magnitude of exposure). Further, these guidelines require a hazard characterization to integrate the analysis of all relevant studies into a weight-of-evidence narrative, and to develop a working conclusion regarding the agent's mode of action in leading to tumor development.

d)   A weight-of-evidence narrative with accompanying descriptors (listed in Section 2.3.1 below) replaces the current alphanumeric classification system. The weight-of-evidence narrative is a summary of the key evidence for carcinogenicity. It describes the agent's MOA, characterizes the conditions of hazard expression including route of exposure and any anticipated disproportionate effects on sensitive subgroups, and recommends appropriate dose-response approach(es). Significant strengths, weaknesses, and uncertainties of contributing evidence are also highlighted.

e)   Biologically based extrapolation models are the preferred approach for quantifying risk. These models integrate events in the carcinogenic process throughout the dose-response range from high to low doses. It is anticipated, however, that the necessary data for the parameters used in such models will not be available for most chemicals. The 1999 draft

revised cancer guidelines allow for alternative quantitative methods, including several default approaches.

f)   Dose-response assessment is a two-step process when a biologically based model is not used. The first step is the assessment of observed data to derive a point of departure (POD), and the second step is the extrapolation below the range of observation. In addition to modeling tumor data, the 1999 draft revised cancer guidelines call for the use and modeling of other kinds of responses if they are considered to be more informed measures of carcinogenic risk that reflect key events in the carcinogenic process (see Section 2.3.3).

For the second extrapolation step, three default approaches are provided–linear, nonlinear, or both. The standard POD for animal studies is the effective dose (ED) corresponding to the lower 95 percent limit on a dose associated with 10 percent extra risk[3] (LED_{10}). A lower POD may be used for human studies of large populations. The choice of extrapolation approach is based on conclusions about an agent's MOA as described in Section 2.3.3.2 below.

Linear. The linear default is a straight line extrapolation from the POD to the origin (zero dose, zero extra risk).

Nonlinear. The nonlinear default begins with the identified POD and provides a margin of exposure (MOE) analysis rather than estimating the probability of effects at low doses. The MOE analysis is used to determine the appropriate margin between the POD and the exposure level of interest (in this Methodology, the AWQC). The goal is to provide information about the risk reduction that accompanies lowering of exposure and the adequacy of an MOE. Factors considered for MOE analysis include the nature of the

---

[3]   "Two risk measures of increased response for quantal data have been proposed in the literature, additional risk and extra risk (Crump, 1984).

Additional risk is defined as P(d) - P(0) and extra risk as [P(d) - P(0)[/[1 - P(0)], where P(d) is the probability of response at dose d, and P(0) is the probability of response at dose 0 (no exposure). Thus, extra risk is additional risk divided by the proportion of individuals that will not respond in the absence of exposure, i.e. additional risk and extra risk differ quantitatively in the way they account for background response.

If the spontaneous incidence of a tumor is zero (or close to zero), then the tumor incidence observed reflects the risk of the tumor from exposure to the chemical agent. In this case, the estimate of extra risk and additional risk are the same. If the spontaneous tumor incidence is greater than zero, then the risk of developing a tumor due to exposure to a specific dose of a chemical agent will not be the incidence of the tumor at that dose *per se*, but will be the incidence of the tumor at that dose corrected for the spontaneous incidence.

Additional risk is the proportion of individuals with tumors in the exposed groups beyond that in the control group, and extra risk is the proportion of individuals responding that would not otherwise have responded. This assumes that the processes leading to tumors in the unexposed individuals are independent of the processes that lead to tumors in the exposed animals. The greater the background incidence, the greater the difference between extra and additional risk. If there are no tumors in the control group [P(0) = 0], there is no difference between extra and additional risk.

Extra risk provides an expanded measure of the incidence of adverse effects when the background incidence is high, with the effect becoming more marked as the background incidence increases. In effect it provides a more sensitive measure of tumor response to a chemical agent when the spontaneous incidence of tumors is high."

response, slope of the observed dose-response curve, human sensitivity compared with experimental animals, and nature and extent of human variability in sensitivity. For more detail about MOE analysis, see Section 2.3.3.2.

Linear and Nonlinear. Both approaches can be used when different modes of action are thought to be responsible for different tumor or other key event responses.

g)   The approach used to calculate an oral human equivalent dose when assessments are based on animal bioassays has been refined and includes a change in the default assumption for interspecies dose scaling. The 1999 draft revised cancer guidelines use body weight raised to the 3/4 power ($BW^{3/4}$).

EPA modeling approaches for the observed range of cancer and noncancer assessments are being consolidated. The modeling of observed response data to identify the POD in a standard way for both kinds of response will be based on the benchmark dose (BMD) modeling approach described briefly in Section 3.6 below.

Until new cancer guidelines are published, the 1986 guidelines will be used along with principles of the 1999 draft revised cancer guidelines. The 1986 cancer guidelines are the basis for IRIS risk numbers which were used to derive the current AWQC. Each new assessment applying the principles of the 1999 draft revised cancer guidelines will be subject to peer review before being used as the basis of AWQC.

Section 2.3 describes the methodology for deriving numerical AWQC for carcinogens applying the principles of the 1999 draft revised cancer guidelines. This discussion of the revised methodology for carcinogens focuses primarily on the quantitative aspects of deriving numerical AWQC values. It is important to note that the cancer risk assessment process outlined in the 1999 draft revised cancer guidelines is not limited to the quantitative aspects. A numerical AWQC value derived for a carcinogen is to be based on appropriate hazard characterization and accompanied by risk characterization information.

2.3   DESCRIPTION OF THE METHODOLOGY FOR DERIVING AWQC BASED ON THE REVISED CARCINOGEN RISK ASSESSMENT

Following the publication of the *Draft Water Quality Criteria Methodology: Human Health* (USEPA, 1998a) and the accompanying TSD (USEPA, 1998b), EPA received comments from the public. EPA also held an external peer review of the draft Methodology, including the cancer methodology. Both the peer reviewers and the public recommended that EPA incorporate the new cancer risk assessment approaches into the AWQC Methodology.

The 2000 Human Health Methodology for deriving numerical AWQC for carcinogens is consistent with the 1986 cancer guidelines and principles included in the 1999 draft revised cancer guidelines. This discussion of applying the 2000 Human Health Methodology to carcinogens focuses primarily on the quantitative aspects of deriving numerical AWQC values, but also

emphasizes the importance of qualitative information as critical to the cancer risk evaluation process.

This section contains a discussion of the weight-of-evidence narrative, describing information relevant to a cancer risk evaluation and characterization. It also includes a discussion of general considerations and a framework of analysis for the MOA. These topics are followed by discussions of the quantitative aspects of deriving numerical AWQC values for carcinogens. It is assumed that data from an appropriately conducted animal bioassay or human epidemiological study provide the underlying basis for deriving the AWQC value. The discussion of quantitative risk estimation focuses on the following topics:

- Dose estimation;

- Characterizing dose-response relationships in the range of observation and at low, environmentally relevant doses;

- Calculating the AWQC value;

- Risk characterization; and

- Use of Toxicity Equivalent Factors (TEF) and Relative Potency Estimates.

2.3.1   Weight-of-Evidence Narrative

The 1999 draft revised cancer guidelines include a weight-of-evidence narrative that is based on an overall judgment of biological, chemical, and physical considerations. The hazard assessment emphasizes analysis of all relevant information rather than just tumor findings. The weight-of-evidence narrative lays out key evidence and includes a discussion of tumor data, information on the MOA, its implications for human hazard including sensitive subgroups, and dose-response evaluation. The narrative emphasizes route and level of exposure and relevance to humans. In addition, a discussion of the strengths and weaknesses of the database is included.

The weight-of-evidence narrative is written in nontechnical language. It provides the key data with conclusions, as well as the conditions for hazard expression. Conclusions about potential human carcinogenicity are presented by route of exposure. Contained within this narrative are simple likelihood descriptors that essentially distinguish whether there is enough evidence to make a projection about human hazard (i.e., carcinogenic to humans; likely to be carcinogenic to humans; suggestive evidence of carcinogenicity but not sufficient to assess human carcinogenic potential; data are inadequate for an assessment of human carcinogenic potential; and not likely to be carcinogenic to humans). Because one encounters a variety of data sets on agents, these descriptors are not meant to stand alone; rather, the context of the weight-of-evidence narrative is intended to provide a transparent explanation of the biological evidence and how the conclusions were derived. Moreover, these descriptors should not be viewed as classification categories (like the alphanumeric system), which often obscure key scientific

differences among chemicals. The new weight-of-evidence narrative also presents conclusions about how the agent induces tumors and the relevance of the MOA to humans including sensitive subgroups, and recommends a dose-response approach based on an understanding of the MOA.

2.3.1.1 Mode of Action: General Considerations and Framework for Analysis

An MOA is a description of key events and processes starting with the interaction of an agent with a cell, through operational and anatomical changes, and resulting in cancer formation. "Mode" of action is contrasted with "mechanism" of action, which implies a more detailed, molecular description of events than is meant by MOA.

Mode of action conclusions are used to address the question of human relevance of animal tumor responses, to address differences in anticipated response among humans such as between children and adults or men and women, and as the basis of decisions about the anticipated shape of the dose-response relationship.

Mode of action analysis is based on physical, chemical, and biological information that helps to explain key events[4] in an agent's influence on development of tumors.

There are many examples of possible modes of carcinogenic action such as mutagenicity, mitogenesis, inhibition of cell death, cytotoxicity with reparative cell proliferation, and immune suppression. All pertinent studies are reviewed in analyzing an MOA, and an overall weighing of evidence is performed, laying out the strengths, weaknesses, and uncertainties of the case as well as potential alternative positions and rationales. Identifying data gaps and research needs is also part of the assessment.

2.3.1.2 Framework for Evaluating a Postulated Carcinogenic Mode(s) of Action

The framework is intended to be an analytic tool for judging whether available data support a mode of carcinogenic action postulated for an agent, and includes nine elements:

1.   Summary description of postulated MOA
2.   Identification of key events
3.   Strength, consistency, specificity of association
4.   Dose-response relationship
5.   Temporal relationship
6.   Biological plausibility and coherence
7.   Other modes of action
8.   Conclusion
9.   Human relevance, including subpopulations

---

[4]A "key event" is an empirically observable, precursor step that is itself a necessary element of the mode of action, or is a marker for such an element.

In reaching conclusions, the question of "general acceptance" of an MOA will be tested as part of the independent peer review that EPA obtains for its assessment and conclusions.

2.3.2   Dose Estimation by the Oral Route

2.3.2.1 Determining the Human Equivalent Dose

An important objective in the dose-response assessment is to use a measure of internal or delivered dose at the target site when sufficient data are available. This is particularly important in those cases where the carcinogenic response information is being extrapolated to humans from animal studies. Generally, the measure of dose provided in the underlying human studies and animal bioassays is the applied dose, typically given in terms of mg/kg-day. When animal bioassay data are used, it is necessary to make adjustments to the applied oral dose values to account for differences in toxicokinetics between animals and humans that affect the relationship between applied dose and delivered dose at the target organ and estimate a human equivalent dose.

In the estimation of a human-equivalent dose, the 1999 draft revised cancer guidelines recommend that when toxicokinetic data are available, they are used to convert the doses used in animal studies to equivalent human doses. However, in most cases, there are insufficient data available to compare dose between species. In these cases, the estimate of a human-equivalent dose is based on science policy default assumptions. In the past, a standard surface area conversion was used; the surrogate for surface area was body weight raised to the 2/3 power ($BW^{2/3}$). To derive an equivalent human dose from animal data, the new default procedure is to scale daily applied oral doses experienced over a lifetime in proportion to $BW^{3/4}$.

The $BW^{3/4}$ adjustment factor is used because metabolic rates, as well as most rates of physiological processes that determine the disposition of a dose, scale this way. Thus, the rationale for this factor rests on the empirical observation that rates of physiological processes consistently tend to maintain proportionality with body weight raised to 3/4 power. Based on this assumption, the "human equivalent" of the applied oral dose in an animal study is obtained from the following algorithm where the doses are in mg/kg-day:

$$\text{Human Equivalent Dose} =$$

$$\text{Animal Dose} \times \left(\frac{\text{Animal BW}}{\text{Animal BW}^{3/4}} \right) \times \left(\frac{\text{Human BW}^{3/4}}{\text{Human BW}} \right)$$

(Equation 2-1)

This equation can be simplified to:

$$\text{Human Equivalent Dose} = (\text{Animal Dose})[(\text{Animal BW})/(\text{Human BW})]^{1/4}$$

(Equation 2-2)

A more extensive discussion of the rationale and data supporting the Agency's change in scaling factors from $(BW)^{2/3}$ to $(BW)^{3/4}$ is in USEPA (1992b) and the 1999 draft revised cancer guidelines.

2.3.3   Dose-Response Analysis

Dose-response analysis addresses the relationship of dose to the degree of response observed in an animal or human study. Extrapolations are necessary when environmental exposures are outside of the range of study observations. Past observations of response have focused on the observation of tumors. The 1999 draft revised cancer guidelines suggest that responses may also include tumor precursors or other effects related to carcinogenicity. These effects may include: changes in DNA, chromosomes, or other key macromolecules; effects on growth signal transduction; induction of physiological or hormonal changes; effects on cell proliferation; or other effects that play a role in the carcinogenic process. Non-tumor effects are referred to as "precursor data" in the following discussion.

Specific guidance regarding the use of animal data, presentation of study results, and selection of the optimal data for use in a dose-response analysis is discussed in detail in the 1999 draft revised cancer guidelines.

2.3.3.1 Characterizing Dose-Response Relationships in the Range of Observation

The first quantitative component in the derivation of AWQC for carcinogens is the dose-response assessment in the range of observation. The objective of this component is to identify a POD for low-dose extrapolation. Two options are available for the assessment in the observed range:

C   Development of a biologically-based model or
C   Curve-fitting of the tumor or precursor data.

If data are extensive and sufficient to quantitatively relate specific key events in the cancer process to neoplasia and the purpose of the assessment is such as to justify investing the necessary resources, a biologically based model can be used for both the observed tumor and related response data and for extrapolation below the range of observed data in either animal or human studies. Extensive data are required to both build the model and to estimate how well it conforms

with observed tumor development specific to the agent. There are not sufficient data to utilize these types of models for most agents.

In the absence of adequate data to generate a biologically based model, dose-response relationships in the observed range can be addressed through curve-fitting procedures for tumor or precursor data. The models should be appropriate to the type of response data in the observed range (see Internet site http://www.epa.gov/ncea/bmds.htm).

The 1999 draft revised cancer guidelines call for modeling not only tumor data in the observable range, but also other responses thought to be important events preceding tumor development (e.g., DNA adducts, cellular proliferation, receptor binding, hormonal changes). The modeling of those data is intended to better inform the dose-response assessment by providing insights into the relationships of exposure (or dose) below the observable range for tumor response. These non-tumor response data can only play a role in the dose-response assessment if the agent's carcinogenic mode of action is reasonably understood, as well as the role of that precursor event.

The 1999 draft revised cancer guidelines recommend calculating the lower 95 percent confidence limit on a dose associated with an estimated 10 percent increased tumor or relevant non-tumor response (LED_{10}) for quantitative modeling of dose-response relationships in the observed range. The estimate of the LED_{10} is used as the POD for low-dose extrapolations discussed below. This standard point of departure (LED_{10}) is adopted as a matter of science policy to remain as consistent and comparable from case to case as possible. It is also a convenient comparison point for noncancer endpoints. The rationale supporting use of the LED_{10} is that a 10 percent response is at or just below the limit of sensitivity for discerning a statistically significant tumor response in most long-term rodent studies and is within the observed range for other toxicity studies. Use of lower limit takes experimental variability and sample size into account. The ED_{10} (central estimate) is also presented as a reference for comparison uses, especially for use in relative hazard/potency ranking among agents for priority setting.

For some data sets, a choice of the POD other than the LED_{10} may be appropriate. The objective is to determine the lowest reliable part of the dose-response curve for the beginning of the second step of the dose-response assessment—determine the extrapolation range. Therefore, if the observed response is below the LED_{10}, then a lower point may be a better choice (e.g., LED_{5}). Human studies more often support a lower POD than animal studies because of greater sample size.

The POD may be a NOAEL when a MOE analysis is the nonlinear dose-response approach. The kinds of data available and the circumstances of the assessment both contribute to deciding to use a NOAEL or LOAEL, which is not as rigorous or as ideal as curve fitting, but can be appropriate. If several data sets for key events and tumor response are available for an agent, and they are a mixture of continuous and incidence data, the most practicable way to assess them together is often through a NOAEL/LOAEL approach.

When a POD is estimated from animal data, it is adjusted to the human equivalent dose using an interspecies dose adjustment or toxicokinetic analysis.

Analysis of human studies in the observed range is designed on a case-by-case basis depending on the type of study and how dose and response are measured in the study. In some cases, the analysis may incorporate consideration of an agent's interactive effects with other agents.

2.3.3.2 Extrapolation to Low, Environmentally Relevant Doses

In most cases, the derivation of an AWQC will require an evaluation of carcinogenic risk at environmental exposure levels substantially lower than those used in the underlying study. Various approaches are used to extrapolate risk outside the range of observed experimental data. In the 1999 draft revised cancer guidelines, the choice of extrapolation method is largely dependent on the mode of action. It should be noted that the term "mode of action" (MOA) is deliberately chosen in the 1999 draft revised cancer guidelines in lieu of the term "mechanism" to indicate using knowledge that is sufficient to draw a reasonable working conclusion without having to know the processes in detail as the term mechanism might imply. The 1999 draft revised cancer guidelines favor the choice of a biologically based model, if the parameters of such models can be calculated from data sources independent of tumor data. It is anticipated that the necessary data for such parameters will not be available for most chemicals. Thus, the 1999 draft revised cancer guidelines allow for several default extrapolation approaches (low-dose linear, nonlinear, or both).

A. Biologically Based Modeling Approaches

If a biologically based model has been used to characterize the dose-response relationships in the observed range, and the confidence in the model is high, it may be used to extrapolate the dose-response relationship outside the observed data range. Although biologically based approaches are appropriate both for characterizing observed dose-response relationships and extrapolating to environmentally relevant doses, it is not expected that adequate data will be available to support such approaches for most substances. In the absence of such data, the default linear approach, the nonlinear (or MOE) approach, or both linear and nonlinear approaches are used.

B. Default Linear Extrapolation Approach

The default linear approach replaces the LMS approach that has served as the default for EPA cancer risk assessments. Any of the following conclusions leads to selection of a linear dose-response assessment approach:

- The chemical has direct DNA mutagenic reactivity or other indications of DNA effects that are consistent with linearity.

- Mode of action analysis does not support direct DNA effects, but the dose-response relationship is expected to be linear (e.g., certain receptor-mediated effects).

- Human exposure or body burden is high and near doses associated with key events in the carcinogenic process (e.g., 2,3,7,8-tetrachlorodibenzo-p-dioxin).

- There is an absence of sufficient tumor MOA information.

The procedures for implementing the default linear approach begin with the estimation of a POD as described above. The point of departure, LED_{10}, reflects the interspecies conversion to the human equivalent dose and the other adjustments for less-than-lifetime experimental duration. In most cases, the extrapolation for estimating response rates at low, environmentally relevant exposures is accomplished by drawing a straight line between the POD and the origin (i.e., zero dose, zero extra risk). This is mathematically represented as:

$$y = mx + b$$
$$b = 0$$

(Equation 2-3)

where:

y	=	Response or incidence
m	=	Slope of the line (cancer potency factor) $= {}^a y / {}^a x$
x	=	Dose
b	=	Slope intercept

The slope of the line, "m" (i.e., ${}^a y / {}^a x$, the estimated cancer potency factor at low doses), is computed as:

$$m = \frac{0.10}{LED_{10}}$$

(Equation 2-4)

When an LED_{10} isn't used, the standard equation for the slope of a line may be used:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

(Equation 2-5)

where:

2-12

$$
\begin{array}{ll}
y_2 & = \quad \text{Response at the POD} \\
y_1 & = \quad \text{Response at the origin (zero)} \\
x_2 & = \quad \text{Dose at the POD} \\
x_1 & = \quad \text{Dose at the origin (zero)}
\end{array}
$$

Due to the use of the origin for y_1 and x_1, the equation simplifies to:

$$
\frac{y_2}{x_2}
$$

(Equation 2-6)

The risk-specific dose (RSD) is then calculated for a specific incremental targeted lifetime cancer risk (in the range of 10^{-6} to 10^{-4}) as:

$$
\text{RSD} = \frac{\text{Target Incremental Cancer Risk}}{\text{m}}
$$

(Equation 2-7)

where:

$$
\begin{array}{lll}
\text{RSD} & = & \text{Risk-specific dose (mg/kg-day)} \\
\text{Target Risk}[7] & = & \text{Value typically in the range of } 10^{-4} \text{ to } 10^{-6} \\
\text{m} & = & \text{Cancer potency factor (mg/kg-day)}^{-1}
\end{array}
$$

The use of the RSD to compute the AWQC is described below in the Section 2.3.4, AWQC Calculation.

C. Default Nonlinear Approach

As discussed in the 1999 draft revised cancer guidelines, any of the following conclusions leads to a selection of a nonlinear (MOE) approach to dose-response assessment:

- A tumor MOA supporting nonlinearity applies (e.g., some cytotoxic and hormonal agents such as disruptors of hormonal homeostasis), and the chemical does not demonstrate mutagenic effects consistent with linearity.

---

[7] In 1980, the target lifetime cancer risk range was set at 10^{-7} to 10^{-5}. However, both the expert panel for the AWQC workshop (USEPA, 1992a) and SAB recommended that EPA change the risk range to 10^{-6} to 10^{-4}, to be consistent with drinking water.

- A MOA supporting nonlinearity has been demonstrated, and the chemical has some indication of mutagenic activity, but it is judged not to play a significant role in tumor causation.

A default assumption of nonlinearity is appropriate when there is no evidence for linearity and sufficient evidence to support an assumption of nonlinearity. The MOA may lead to a dose-response relationship that is nonlinear, with response falling much more quickly than linearly with dose or with response being most influenced by individual differences in sensitivity. Alternatively, the MOA may theoretically have a threshold (e.g., the carcinogenicity may be a secondary effect of toxicity or of an induced physiological change that is itself a threshold phenomenon) (see Appendix C, Example 5, or Appendix D, Example 2 in USEPA, 1999a). The EPA does not generally try to distinguish between modes of action that might imply a "true threshold" from others with a nonlinear dose-response relationship. Except in unusual cases where extensive information is available, it is not possible to distinguish between these empirically.

As a matter of science policy under this analysis, nonlinear probability functions are not fit to the response data to extrapolate quantitative low-dose risk estimates. This is because different models can lead to a very wide range of results, and there is currently no basis, generally, to choose among them. Thus, the default procedure for nonlinear extrapolation is to conduct an MOE analysis to evaluate concern for levels of exposure.

An MOE is defined as the POD divided by the environmental exposure of interest. The environmental exposures of interest, for which MOEs are estimated, may be actual or projected exposure levels. An acceptable MOE is estimated.

MOE analysis is applicable if data are sufficient to presume a nonlinear dose-response function containing a significant change in slope. An RfD[8] or RfC-like value may be estimated and considered based on a precursor event that is key to the cancer process.

To support a risk manager's consideration of the MOE, all of the pertinent hazard, dose-response, and human exposure information is characterized to provide insights about the scientific community's current understanding of the phenomena that may be occurring as dose (exposure) decreases substantially below the observed data. The goal is to provide as much information as possible about the risk reduction that accompanies lowering of exposure and the adequacy of an MOE based on scientific input.

Operationally, there are two main steps in the MOE approach:

- The first step is the selection of a POD that is a "minimum effect dose level." The POD would ideally be the dose where the key events in tumor development would not occur in

---

[8] A reference dose (RfD) or reference concentration (RfC) for noncancer toxicity is an estimate with uncertainty spanning perhaps an order of magnitude of daily exposure to the human population (including sensitive subgroups) that is anticipated to be without appreciable deleterious effects during a lifetime. It is arrived at by dividing empirical data on effects by uncertainty factors that consider inter- and intraspecies variability, extent of data on all important chronic exposure toxicity endpoints, and availability of chronic as opposed to subchronic data.

a heterogeneous human population, thus representing an actual "no-effect level". As noted above, the POD may be the LED_{10} [9] for tumor incidence or a precursor. In some cases, it may also be appropriate to use a NOAEL or LOAEL value from a precursor. When animal data are used, the POD is a human equivalent dose or concentration arrived at by interspecies dose adjustment (as discussed above) or toxicokinetic analysis.

- The second step in using MOE analysis to establish an AWQC is the selection of an appropriate margin or UF to apply to the POD. This is supported by analysis in the MOE discussion provided in the risk assessment. The Agency will develop more specific guidance on the MOE approach, as recommended by the Agency's SAB in its January, 1999 review. The guidance will be peer reviewed and published separately as part of the Agency's implementation activity of these guidelines. The general principles and major elements to be considered in an MOE analysis are listed below.

  - The nature of the response used for the dose-response assessment, for instance, whether it is a precursor effect or a tumor response. The latter may support a greater MOE.

  - The slope of the observed dose-response relationship at the POD and its uncertainties and implications for risk reduction associated with exposure reduction. A steeper slope implies a greater reduction in risk as exposure decreases. This may support a smaller MOE.

  - Human sensitivity compared with that of experimental animals. How sensitive is the *human population* compared with the tested animals? For this comparison, all doses should have already been converted to equivalent human doses, using either a toxicokinetic model or the default cross-species scaling factor. These dose conversions reflect interspecies differences in toxicokinetics, not toxicodynamics. When information is not sufficient to quantify human sensitivity with regard to the toxicodynamics compared with the tested animals, this uncertainty needs to be taken into account in the discussion of an adequate MOE. As with noncancer assessment, the default assumption is that the most sensitive humans are more sensitive than the test animals. Depending on the data available on the sensitivity of the test species to the agent and the endpoint of concern compared with humans, the MOE decision may need to incorporate more or less conservatism.

  - Nature and extent of human variability and sensitivity. Is there information on *sensitive individuals* that would be part of a heterogeneous human population? Pertinent information would come from human studies, since animal studies, particularly those using homogeneous animal strains, do not provide information

---

[9] The LED_{10} is adopted as the standard POD for non tumor key event or toxicity incidence data in order to harmonize curve-fitting procedures between cancer and non cancer toxicity assessments. Because the NOAEL in study protocols for non tumor toxicity can range from about a 5% to a 30% effect level, adopting the 10% effect level as the standard POD will accommodate most of these data sets without departing the range of observation. The LED_{10} can be regarded as an improved and harmonized estimate of the NOAEL (USEPA, 1999a).

about human variability. When information is not sufficient to quantify the extent of human variability in sensitivity, this uncertainty should be reflected in the discussion of an adequate MOE (also see discussion below on human exposure).

- Human exposure. The MOE evaluation also takes into account the magnitude, frequency, and duration of exposure. If the population exposed in a particular scenario is wholly or largely composed of a subpopulation of special concern (e.g., children) for whom evidence indicates a special sensitivity to the agent's MOA, an adequate MOE would be larger than for general population exposure.

Considering the toxicity and other data presented in the weight-of-evidence narrative and the MOE analysis provided in the risk assessment for the chemical, a UF is selected [10] on a case-by-case basis, with full explanation of the rationale.

The UF is used to modify the POD in the final equation. This is shown below in Section 2.3.4 on AWQC calculation.

D. Both Linear and Nonlinear Approaches

Any of the following conclusions leads to selection of both a linear and nonlinear approach to dose-response assessment. Relative support for each dose-response method and advice on the use of that information needs to be presented. In some cases, evidence for one MOA is stronger that for the other, allowing emphasis to be placed on that dose-response approach. In other cases, both modes of action are equally possible, and both dose-response approaches should be emphasized.

- Modes of action for a single tumor type support both linear and nonlinear dose response in different parts of the dose-response curve (e.g., 4,4' methylene chloride).

- A tumor mode of action supports different approaches at high and low doses; e.g., at high dose, nonlinearity, but, at low dose, linearity (e.g., formaldehyde).

- The agent is not DNA-reactive and all plausible modes of action are consistent with nonlinearity, but a key event is not fully established.

- Modes of action for different tumor types support differing approaches, e.g., nonlinear for one tumor type and linear for another due to lack of MOA information (e.g., trichloroethylene).

---

[10] EPA will develop more specific guidance on the margin of exposure approach, as recommended by the Agency's SAB in 1999. The guidance will be peer reviewed and published separately as part of the Agency's implementation of the Final Revised Cancer Guidelines.

2.3.4   AWQC Calculation

2.3.4.1 Linear Approach

The following equation is used for the calculation of the AWQC for carcinogens where an RSD is obtained from the linear approach:

$$AWQC = RSD \times \left(\frac{BW}{DI + (FI \times BAF)} \right)$$

(Equation 2-8)

where:

AWQC	=	Ambient water quality criterion (mg/L)
RSD	=	Risk-specific dose (mg/kg-day)
BW	=	Human body weight (kg)
DI	=	Drinking water intake (L/day)
FI	=	Fish intake (kg/day)
BAF	=	Bioaccumulation factor (L/kg)

The AWQC calculation shown above is appropriate for water bodies that are used as sources of drinking water (and for other uses).

2.3.4.2 Nonlinear Approach

In those cases where the nonlinear, MOE approach is used, a similar equation is used to calculate the AWQC:

$$AWQC = \frac{POD}{UF} \times RSC \times \left(\frac{BW}{DI+(FI \times BAF)} \right)$$

(Equation 2-9)

where:

AWQC	=	Ambient water quality criterion (mg/L)
RSD	=	Risk-specific dose (mg/kg-day)
POD	=	Point of departure (mg/kg-day)
UF	=	Uncertainty factor (unitless)
BW	=	Human body weight (kg)
DI	=	Drinking water intake (L/day)

FI   =   Fish intake (kg/day)
BAF   =   Bioaccumulation factor (L/kg)
RSC   =   Relative source contribution (percentage or subtraction)

As noted above for the linear approach, the AWQC calculation shown above is appropriate for water bodies that are used as sources of drinking water (and for other uses).

A difference between the AWQC values obtained using the linear and nonlinear approaches is that the AWQC value obtained using the default linear approach corresponds to a specific estimated incremental lifetime cancer risk level in the range of 10^{-4} to 10^{-6}. In contrast, the AWQC value obtained using the nonlinear approach does not describe or imply a specific cancer risk.

The actual AWQC chosen is based on a review of all relevant information, including cancer, noncancer, ecological, and other critical data. The AWQC might not utilize the value obtained from the cancer analysis if it is less protective than that derived from the noncancer endpoint.

2.3.5   Risk Characterization

Risk characterization information accompanies the numerical AWQC value and addresses the major strengths and weaknesses of the assessment arising from the availability of data and the current limits of understanding of the process of cancer causation. Key issues relating to the confidence in the hazard assessment and the dose-response analysis (including the low dose extrapolation procedure used) are discussed.

Whenever more than one interpretation of the weight of evidence for carcinogenicity or the dose-response characterization can be supported, and when choosing among them is difficult, the alternative views are provided along with the rationale for the interpretation chosen in the derivation of the AWQC value. Where possible, quantitative uncertainty analyses of the data are provided; at a minimum, a qualitative discussion of the important uncertainties is presented.

Important features of the risk characterization include significant scientific issues, significant science and science policy choices that were made when alternative interpretations of data exist, and the constraints of the data and the state of knowledge. The assessments of hazard, dose-response, and exposure are summarized to generate risk estimates for the exposure scenarios of interest.

The 1999 draft revised cancer guidelines contain more detailed guidance regarding the development of risk characterization summaries and analyses.

2.3.6 Use of Toxicity Equivalence Factors and Relative Potency Estimates

The 1999 draft revised cancer guidelines state:

A Toxicity equivalence factor (TEF) procedure is one used to derive quantitative dose-response estimates for agents that are members of a category or class of agents. TEFs are based on shared characteristics that can be used to rank or order the class members by carcinogenic potency when cancer bioassay data are inadequate for this purpose. The ordering is by reference to the characteristics and potency of a well-studied member or members of the class. Other class members are indexed to the reference agent(s) by one or more shared characteristics to generate their TEFs.

In addition, the 1999 draft revised cancer guidelines state that TEFs are generated and used for the limited purpose of assessment of agents or mixtures of agents in environmental media when better data are not available. When better data become available for an agent, the TEF should be replaced or revised. To date, adequate data to support use of TEFs has been found only for dibenzofurans (dioxins) and coplanar polychlorinated biphenyls (PCBs) (USEPA, 1989, 1999b).

The uncertainties associated with TEFs must be discussed when this approach is used. This is a default approach to be used when tumor data are not available for individual components in a mixture. Relative potency factors (RPFs) can be similarly derived and used for agents with carcinogenicity or other supporting data. The RPFs are conceptually similar to TEFs, but do not have the same level of data to support them. TEFs and RPFs are used only when there is no better alternative. When they are used, uncertainties associated with them must be discussed. As of today, there are only three classes of compounds for which relative potency approaches have been examined by EPA: dioxins, PCBs, and polycyclic aromatic hydrocarbons (PAHs). There are limitations to the use of TEF and RFP approaches, and caution should be exercised when using them. More guidance can be found in the *Draft Guidance for Conducting Health Risk Assessment of Chemical Mixtures* (USEPA, 1999b).

2.4 CASE STUDY (COMPOUND Z, A RODENT BLADDER CARCINOGEN)

This section illustrates an application of the nonlinear method (MOE) for a rodent bladder carcinogen (Compound Z). A brief summary of the data set is provided below with conclusions regarding the weight of evidence *"Likely/Not Likely Human Carcinogen"-Range of Dose Limited, Margin-of-Exposure Extrapolation*. For more details in the hazard evaluation and in the mode of action evaluation of this chemical, see Appendices A and B, respectively, which are selected from the case studies in the 1999 draft revised cancer guidelines. The AWQC obtained using the default linear and LMS approaches are included for purposes of comparison only and would not be used for agents with the characteristics described for Compound Z.

2.4.1 Background and Evaluation for Compound Z

Compound Z is a metal organophosphonate which has been tested in acute, subchronic, chronic, reproductive, mutagenic and carcinogenic assays in multiple species. Tumors were observed only in rat studies. No human data are available. Based on a review of the toxicity, mechanistic, metabolic, and other data summarized below for this agent, it was concluded that a nonlinear approach is most appropriate for establishing AWQC based on carcinogenicity. (See Appendices A and B for more detail.)

Lifetime cancer bioassays of Compound Z identified bladder tumors and hyperplasia in male rats at doses of 1500 mg/kg-day and higher in the diet. These effects were not observed at 100 and 400 mg/kg-day. In a 90-day study designed to evaluate the mechanisms of tumor induction, the following sequence was identified as critical to bladder tumor formation in rats:

1)   large doses of Compound Z produce urinary calcium/potassium imbalance followed by
2)   diuresis, a sharp drop in urine pH, formation of urinary calculi, and
3)   appearance of transitional cell hyperplasia in the renal pelvis, ureter, and urinary bladder.

These effects occurred within two weeks of exposure onset, persisted to the end of exposure, and were reversible upon cessation of the 90-day exposure.

The pathological events caused by Compound Z are believed to result from prolonged mechanical irritation of the bladder by calculi that developed in response to the exposure. At high but not lower subchronic doses in the male rat, Compound Z leads to elevated blood phosphorus levels; the body responds by releasing excess calcium into the urine. The calcium and phosphorus combine in the urine and precipitate into multiple stones in the bladder. The stones are very irritating to the bladder; the bladder lining is eroded, and cell proliferation occurs to compensate for the loss of the lining. This leads to development of hyperplasia, with subsequent tumor formation. A prolonged increase in the rate of proliferation of cells of the urinary bladder has been proposed to be an important step in the induction of urinary bladder tumors (Cohen and Ellwein, 1990, 1991). Thus, the association of cell proliferation, hyperplasia, and subsequent cancer induction as a result of urinary stone formations due to exposure to Compound Z is proposed as one mode of action which may justify, after a review of all relevant data, the use of a nonlinear approach, such as the MOE approach.

Studies of the effects of separated components of this agent (i.e., the metal and the organophosphate components) yield no evidence of carcinogenicity in the bladder. In metabolic studies in animals, the metallic component in isolation from the parent molecule was not absorbed to a significant extent from the gastrointestinal tract.

Compound Z has been assessed via a battery of mutagenicity assays that have yielded negative results, and a review of the chemical structure does not suggest potential genotoxicity. The metabolites of Compound Z have also yielded negative results in mutagenicity assays and yielded no evidence of carcinogenicity. The negative genotoxicity results for Compound Z and

structurally related agents provide further support for the use of a nonlinear approach, such as the MOE approach, to establish AWQC.

2.4.2   Conclusion and Use of the MOE Approach for Compound Z

Compound Z, a metal aliphatic phosphonate, is likely to be carcinogenic to humans only under high-exposure conditions following oral and inhalation exposure that lead to bladder stone formation, but is not likely to be carcinogenic under low-exposure conditions. It is not likely to be a human carcinogen via the dermal route, given that the compound is a metal conjugate that is readily ionized, and its dermal absorption is not anticipated. The weight of evidence is based on: (1) bladder tumors only in male rats at high exposure; (2) the absence of tumors at any other site in rats or mice; (3) the formation of calcium-phosphorus-containing bladder stones in male rats at high, but not low, exposure. The bladder stones erode bladder epithelium and result in profound increases in cell proliferation and cancer; and (4) the absence of carcinogenic structural analogues or mutagenic activity.

There is a strong mode of action basis for the requirements of high doses of Compound Z, which leads to excess calcium and increased acidity in the urine, resulting in the precipitation of bladder stones and subsequent increase in cell proliferation and tumor hazard potential. Lower doses fail to perturb urinary constituents, lead to stones, produce toxicity, or give rise to tumors. Therefore, dose-response assessment should assume nonlinearity.

A major uncertainty is whether the profound effects of Compound Z may be unique to the rat. Even if Compound Z produced stones in humans, there is only limited evidence that humans with bladder stones develop cancer.

Based on the progression of pathology leading to tumors, in which hyperplasia is an early critical step, hyperplasia was selected as the sentinel precursor effect which was used as the basis for the calculation of AWQC using the MOE approach. Hyperplasia incidence data from a lifetime rat study are available for Compound Z. Tumor data from the same lifetime rat study were used to calculate AWQC using the default linear and LMS approaches for purposes of comparison. The data used for all three approaches are summarized in Table 2-1 below.

2.4.2.1 Identification of the Point of Departure for Compound Z

The POD chosen for the MOE calculations was 400 mg/kg-day, which is the maximum animal dose yielding no observable hyperplastic effects (the NOAEL shown in Table 2-1).[11] The study found males to be more sensitive than females, and the hyperplasia results in male rats were used for AWQC calculations. The human equivalent dose for the NOAEL of 106.4 mg/kg-day was calculated using the new scaling factor of body weight raised to the 3/4 power (as shown in Equation 2-1).

---

[11]This is based on a dietary conversion factor for rats from ppm to mg/kg-day of 0.05.

Table 2-1. Study Results from a Lifetime Exposure of Male Rats to Compound Z

Animal Dose in mg/kg-day (scaled human equivalent doses)	Number in Group	Number Responding	
		tumors (combined papilloma & carcinoma)	hyperplasia
0	73	3	5
400 $(BW^{3/4} = 106.4)^a$ $(BW^{2/3} = 68.4)^b$	78	2	5
1500 $(BW^{3/4} = 398.9)^a$ $(BW^{2/3} = 256.5)^b$	78	21*	29*

a.   The $(BW)^{3/4}$ scaling factor is based on the 1999 draft revised cancer guidelines.

b.   The $(BW)^{2/3}$ scaling factor is based on the 1986 cancer guidelines and is used with the LMS method later in this section for comparative purposes.

*   There were statistically significant ($p<0.05$) increases in both tumor incidence and hyperplasia in the treated group compared with the control group.

2.4.2.2 Discussion of the Points Affecting Selection of the UF for Compound Z

•   The Nature of the Response. The response used for the dose-response assessment is hyperplasia, which is a precursor effect. Therefore, a smaller UF is needed.

•   Slope of the Dose-Response Relationship. The data available indicate a steep slope at the point of departure (at 400 mg/kg-day animal dose). This would suggest a rapid reduction in risk with lower doses, or a smaller UF.

•   Intraspecies Variability. There is variability within the human population in responses to xenobiotic agents which may result from a variety of factors including health status, diet, age, and genetic composition. Research on Compound Z did not identify a common health or genetic condition which would yield a subpopulation who are particularly susceptible to the carcinogenic effects of Compound Z nor did it indicate an exceptionally high or low level of intraspecies variability.

- Interspecies Variability. Animals and humans may vary widely in their responses to agents due to their differing physiology and metabolism. A review of human case studies and epidemiological studies indicate that humans may be significantly less susceptible to the influence of bladder irritation, stone formation, and subsequent tumor formation than male rodents. This would suggest a smaller UF for interspecies variability.

- Human Exposure. This exposure scenario is chronic, so there is no need to apply an additional UF.

After considering all the issues together, a decision is made on the margin of safety (MOS) exposure or the UF. The size of the UF is a matter of policy and is selected on a case-by-case basis, considering the weight of evidence and the MOE analysis provided in the risk assessment.[12]

In summary, an overall UF of 30 is used in the MOE calculation. The selection of the UF is based on a consideration of all the factors discussed above, such as intraspecies variability (10), interspecies variability (3 is used here to account for toxicokinetic differences, a scaling factor of body weight raised to 3/4 power has already applied to adjust for toxicokinetic differences).   In addition, the database for this chemical is very extensive, as described in detail in Appendix B (selected from the case study of the 1999 draft revised cancer guidelines). Further, the duration of the key study used for quantification is chronic. Thus, this factor of 30 is considered to be sufficient for human health protection. The risk may decline considerably with doses lower than the POD; the male rat is a very sensitive model (mice do not respond). Physiological phenomena are likely to fall off sharply with dose as shown by the dose-response curve. Further, bladder stone and subsequent tumor formation is not a common phenomenon in humans.

2.4.2.3 AWQC Calculations for Compound Z

Equation 2-9 shown in Section 2.3.4.2 was used to calculate the AWQC for Compound Z:

$$\text{AWQC} = \frac{\text{POD}}{\text{UF}} \text{ x} \left(\frac{\text{BW}}{\text{DI+(FI x BAF)}} \right) \text{ x RSC\%}$$

(Equation 2-9)

The following input parameters were used:

POD   =   Point of departure (106.4 mg/kg-day (NOAEL))

---

[12] EPA will provide specific guidance on the margin of exposure approach. The guidance will be peer reviewed and published separately as part of the Agency's implementation activity of the Draft Revised Cancer Guidelines.

UF	=	Uncertainty factor of 30
BW	=	Body weight for adult (70 kg)
DI	=	Drinking water intake (2 L/day)
FI	=	Fish intake (0.0175 kg/day)
BAF	=	Assumed bioaccumulation factor (BAF) (300 L/kg)
RSC	=	Relative source contribution (20% assumed)

This calculation yields an AWQC of 6.7 mg/L. The body weight, water intake, fish intake, and RSC percentage values used in the above calculation are the current default values for adults. The BAF, which accounts for the accumulation of Compound Z from water through the food chain and into fish tissue, has been arbitrarily chosen for purposes of this case study.

The AWQC calculations shown above is appropriate for water bodies that are used as sources of drinking water (and for other uses).

2.4.3   Use of the Default Linear Approach for Compound Z

This section is provided for purposes of illustrating the use of the default linear approach for deriving AWQC based on carcinogenicity and to compare the resulting AWQC to that obtained above using the MOE approach. As discussed in Section 2.4.1 above, it is important to note that the default linear method would most likely not, in practice, be recommended as an approach for quantifying the risk and deriving the AWQC for Compound Z given the hazard characteristics described for this substance.

2.4.3.1 <u>Computing the Human Equivalent Dose for Compound Z</u>

The doses used in the study were adjusted to obtain a human equivalent dose, as shown in Table 2-1. In the absence of toxicokinetic data, this was done using a scaling factor of $BW^{3/4}$, with a male rat weight of 0.35 kg and a human weight of 70 kg (as shown in Equation 2-1).

2.4.3.2 <u>Calculation of AWQC for Compound Z</u>

To describe the dose-response of tumor incidence data in the observed range, a curve-fitting model such as the multistage or other approach appropriate for the data can be used. In the case of Compound Z, three data points (at doses of 0, 400, and 1500 mg/kg-day) were used in the multistage model (GLOBAL86) to calculate the LED_{10} (the 95 percent lower confidence limit on a dose associated with a 10 percent increase in response). The value obtained for the LED_{10} is 204 mg/kg-day.

The cancer slope factor (m) is calculated by dividing 0.1 by the LED_{10} using Equation 2-4:

$$m = \frac{0.10}{LED_{10}}$$

(Equation 2-4)

This yields an estimated cancer slope factor of 4.9×10^{-4} per mg/kg-day. The cancer slope factor is then used in Equation 2-7 with a specified risk level (in this case 10^{-6}) to calculate an RSD:

$$RSD = \frac{Target\ Incremental\ Cancer\ Risk}{m}$$

(Equation 2-7)

This yields an RSD of 2.0×10^{-3} mg/kg-day.

The RSD is used in Equation 2-8 with the same input parameters (body weight, drinking water intake, fish intake, and BAF) as those used for the MOE approach:

$$AWQC = RSD \times \left(\frac{BW}{DI + (FI \times BAF)} \right)$$

(Equation 2-8)

This yields an AWQC of 0.019 mg/L (rounded from 0.0189 mg/L) for a target risk of $10^{-6.}$

2.4.4 Use of the LMS Approach for Compound Z

This section is provided strictly for purposes of comparing the use of the MOE approach with the traditional LMS method for deriving AWQC for carcinogens. As discussed above, the LMS approach would not be used in practice to quantify risk and derive the AWQC for Compound Z given the hazard characteristics described for this substance.

First, the LMS approach was used to fit the male rat tumor data shown in Table 2-1 using the computer program GLOBAL86. This program calculates the 95th percentile upper confidence limit on the linear slope (i.e., the q_1^*) in the low dose range. A human equivalent dose was calculated using the $BW^{2/3}$ interspecies dose scaling factor for purposes of illustrating the results obtained applying the 1980 Methodology. The human equivalent doses obtained using this scaling factor are shown in Table 2-1 above. (The same data set, using differently scaled doses, was employed for both the new linear and LMS approaches.) The q_1^* value obtained using the LMS approach is 6×10^{-4} (mg/kg-day)$^{-1}$.

2-25

Equation 2-7 was used with a reference incremental cancer risk of 10^{-6} to calculate an RSD of 1.7×10^{-3}. Equation 2-8 was then used to calculate the AWQC with the same input parameters (body weight, drinking water intake, fish intake, and BAF) as those used for the MOE approach. The AWQC was calculated to be 0.016 mg/L and was rounded from 0.0157 mg/L.

2.4.5 Comparison of Approaches and Results for Compound Z

The results of the three approaches used for Compound Z are summarized in Table 2-2. The AWQC calculated using the MOE approach is substantially higher than that obtained using the default linear and LMS approaches. If larger or smaller UFs were used in the MOE calculations, the AWQC obtained using the MOE approach would decrease or increase accordingly. The quantitative relationship between AWQC derived using different methods will vary depending on the nature of the data set and the UFs and POD selected for use in the MOE approach.

**Table 2-2. Comparison of AWQC Obtained for Compound Z
Using the MOE, Default Linear, and LMS Approaches**

Method	AWQC (mg/L)
MOE: Using hyperplasia as a precursor for determining the POD and a UF of 30.	6.7
Default Linear: Using linear extrapolation - straight line drawn from the LED_{10} to the origin with a 10^{-6} target risk level and an interspecies scaling factor based on $BW^{3/4}$.	0.019
LMS: Using the linearized multistage approach with a 10^{-6} risk level and an interspecies scaling factor based on $BW^{2/3}$.	0.016

2.5   REFERENCES

Barnes, D.G., G.P Daston, J.S. Evans, A.M. Jarabek, R.J. Kavlock, C.A. Kimmel, C. Park, and H.L. Spitzer. 1995. Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. *Regul. Toxicol. Pharmacol.* 21:296-306.

Chen, CW. and G. Oberdorster. 1996. Selection of models for assessing dose-response relationship for particle-induced lung cancer. *Inhalation Toxicol.* 8:259-278.

Cohen, S.W. and L.B. Ellwein. 1990. Cell proliferation in carcinogenesis. *Science* 249:1007-1011.

Cohen, S.W. and L.B. Ellwein. 1991. Genetic errors, cell proliferation and carcinogenesis. *Cancer Res.* 51:6493-6505.

Crump, K. 1984. A new method for determining allowable daily intakes. *Fund. Appl. Toxicol.* 4:854-891.

OSTP (Office of Science and Technology Policy). 1985. Chemical carcinogens: Review of the science and its associated principles. *Federal Register* 50: 10372-10442.

USEPA (U.S. Environmental Protection Agency). 1986. Guidelines for carcinogen risk assessment. *Federal Register* 51:33992-34003.

USEPA (U.S. Environmental Protection Agency). 1989. *Interim Procedures for Estimating Risks Associated with Exposures to Mixtures of Chlorinated Dibenzo-p-dioxins and Dibenzofurans (CDDs and CDFs) and 1989 Update*. Risk Assessment Forum. Washington, DC. EPA/625/3-89/016.

USEPA (U.S. Environmental Protection Agency). 1991. *Workshop Report on Toxicity Equivalency Factors for Polychlorinated Biphenyl Congeners.* Risk Assessment Forum. Washington, DC. EPA/625/3-91/020.

USEPA (U.S. Environmental Protection Agency). 1992a. *Report of the National Workshop on Revision of the Methods for Deriving National Ambient Water Quality Criteria for the Protection of Human Health.* Office of Water. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1992b. Draft report: A cross-species scaling factor for carcinogen risk assessment based on equivalence of $Mg/Kg^{3/4}/day$. *Federal Register* 57: 24152-24173.

USEPA (U.S. Environmental Protection Agency). 1996. *Proposed Guidelines for Carcinogen Risk Assessment.* Office of Research and Development. Washington, DC. EPA/600/P-92/003C. (*Federal Register* 61:17960)

USEPA (U.S. Environmental Protection Agency). 1998a. *Draft Water Quality Criteria Methodology: Human Health. Federal Register Notice*. Office of Water. Washington, DC. EPA-822-Z-98-001.

USEPA (U.S. Environmental Protection Agency). 1998b. *Ambient Water Quality Criteria Derivation Methodology - Human Health. Technical Support Document. Final Draft.* Office of Water. Washington, DC. EPA

USEPA (U.S. Environmental Protection Agency). 1999a. *Guidelines for Carcinogen Risk Assessment. Review Draft.* Risk Assessment Forum. Washington, DC. NCEA-F-0644. July.

USEPA (U.S. Environmental Protection Agency). 1999b. Draft guidance for conducting health risk assessment of chemical mixtures. *Federal Register* 64:23833-23834.

3. NONCANCER EFFECTS

3.1 INTRODUCTION

The evaluation of risks from noncarcinogenic chemicals traditionally has been based on the assumption that noncarcinogens have a dose or level below which no adverse effects are expected to occur. The risk estimate developed by EPA for noncarcinogens is the reference dose (RfD). The Integrated Risk Information System (IRIS) Background Document entitled *Reference Dose (RfD): Description and Use in Health Risk Assessments* (USEPA, 1988; hereafter the "1988 RfD background document") defines an RfD as "an estimate (with uncertainty spanning approximately an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects over a lifetime." The RfD is acknowledged to be an estimate and, thus, may not be completely protective of every individual within a highly variable population; conversely, exposures above the RfD are not necessarily unsafe. Some individuals may have better adaptive or protective capacities than others, and responses may vary with age and state of health; thus, individuals respond differently to toxicant exposure (Barnes and Dourson, 1988).

The key step in deriving water quality criteria for the protection of human health from noncancer effects is the determination of the RfD. As described in Section 1, the RfD is used in concert with additional information regarding exposure and the bioaccumulation potential of the substance to derive an AWQC for noncancer effects. The procedures presented in USEPA's 1988 RfD background document for deriving the RfD using an experimentally derived NOAEL/LOAEL approach are incorporated into this chapter. The Agency is also investigating alternative methods for estimating the RfD. Thus, this guidance document contains information on two alternative methods: BMD and categorical regression approaches. The Agency continues to conduct research on the utility of both of these methods in the noncancer risk assessment process and recommends their application in circumstances where the data are sufficient. The Agency used the BMD approach to derive a RfD for methylmercury as described in *Reference Dose (RfD) for Oral Exposure for Methylmercury* (USEPA, 1994a).

This section begins with a discussion of hazard identification and dose-response characterization. This is followed by a description of factors to be considered in the selection of critical data sets for use in the risk assessment evaluation. The procedures for deriving an RfD for a substance using the traditional NOAEL/LOAEL approach are presented as the accepted current risk assessment practice used by EPA. Next, the BMD method for deriving an RfD is discussed, and an example of its application is provided for illustrative purposes. A brief discussion of categorical regression is also included, with references to the relevant literature. The chapter concludes with specific sections on several issues relevant to noncancer risk assessment, including practical nonthreshold effects and risks from short-term exposures and mixtures.

While the intent of this guidance is to provide sufficient information to apply methods for deriving RfDs, this document does not detail all relevant issues and underlying theory associated with these methods. For further information, the reader is referred to the sources cited in the reference list (in particular, USEPA, 1988; Crump et al., 1995; and Hertzberg and Miller, 1985).

3.2   HAZARD IDENTIFICATION

The first step in the risk assessment involves preparing a hazard identification, based on a review of data available to characterize the health effects associated with chemical exposure. The 1988 RfD background document outlines considerations for choosing data upon which to base a hazard identification for noncancer health effects.[13] Assessors should prepare a hazard identification document that describes the nature of exposure, the type and severity of effects observed, and the quality and relevance of data to humans. Well-conducted human studies are considered the best for establishing a link between exposure to an agent and manifestation of an adverse effect. In the absence of adequate human data, the Agency relies primarily on animal studies. In such cases, the principle studies are drawn from experiments conducted on laboratory mammals, most often rat, mouse, rabbit, guinea pig, dog, monkey, or hamster. Well-designed animal studies offer the benefit of controlled chemical exposures and definitive toxicological analysis. Supporting evidence provides additional information for dose-response assessment and may come from a wide variety of sources, such as metabolic and pharmacokinetic studies. *In vitro* studies seldom provide definitive hazard identification data, but they can often provide insight into the compound's potential for human toxicity.

Important to the hazard identification is consideration of the biological and statistical significance of observed effects. The determination of whether an effect is adverse requires professional judgment. Generally, adverse health effects are considered to be those deleterious effects which are or may become debilitating, harmful, or toxic to the normal functions of an organism, including reproductive and developmental effects. Adverse effects do not include such effects as tissue discoloration without histological or biochemical effects, or the induction of the enzymes involved in the metabolism of the substance. Guidelines for defining the severity of adverse effects have been suggested by Hartung and Durkin (1986). EPA has also developed guidelines for the ranking of observed effects (USEPA, 1995) and a ranking scheme for slight to severe effects. Distinguishing slight effects such as reversible enzyme induction and reversible subcellular change from more serious effects is critical in distinguishing between a NOAEL and LOAEL.

It is also important to evaluate the reversibility of an effect. Reversibility refers to whether or not a change will return to normal or within normal limits either during the course of or following exposure. However, even a reversible effect may be adverse to an organism. In

---

[13]The Agency has also developed guidelines that explain the process of hazard identification for developmental (USEPA, 1991a) and reproductive (USEPA, 1994b) effects. Please refer to these EPA documents for guidance in these areas.

performing a hazard identification, irreversible effects should be distinguished from less serious, but still adverse, reversible changes.

The exposure conditions for toxicity tests, including the route (e.g., inhaled versus ingested), source (e.g., water versus food), and duration, should be discussed in the hazard identification. The hazard identification should also include an evaluation of the quality of studies. Elements that affect the quality of studies include the soundness of the study protocol, the adequacy of data analysis, the characterization of the study compound, the types of species used, the number of individuals per study group, the number of study groups, dose spacing, the types of observations recorded, sex and age of animals, and the route and duration of exposure (USEPA, 1988).

The hazard identification should conclude with a weight-of-evidence discussion. In general, the discussion should review the results of different studies and develop an overall picture of the chemical's toxicity. Evidence for possible toxicity in humans is supported by similar results across species and across investigators. A plausible mechanism of action for the effect, as well as similar toxic activity in chemicals of similar structure, also add to the weight of evidence.

3.3   DOSE-RESPONSE ASSESSMENT

The dose-response assessment involves the evaluation of toxicity data to identify doses at which statistically and/or biologically significant effects occur, and to identify NOAEL and/or LOAEL values. The effects data are also evaluated to see if there is a quantitative relationship between dose and the magnitude of the effect. Dose-response relationships can be linear, curvilinear, or U-shaped. The RfD is traditionally estimated by identifying the most appropriate NOAEL for the critical effect. The LOAEL may be used to estimate the RfD if no appropriate NOAELs have been identified.

3.4   SELECTION OF CRITICAL DATA

3.4.1   Critical Study

Ideally, the scientific data for noncancer effects should include sufficient information to characterize quantitatively the incidence and severity of response as dose increases. However, complete data are frequently lacking. Instead, the Agency bases the derivation of the RfD on the NOAEL or LOAEL from a critical study or collection of critical studies. The choice of the critical study or studies to use in the derivation of the chronic RfD requires professional judgment concerning the quality of the studies, the definition of adverse effects and their level of occurrence. As part of the hazard identification, all relevant toxicity data on a chemical should be evaluated to support the establishment of the RfD. Those studies representing the best quality

and most appropriate data should be considered for defining adverse effects and their level of occurrence.

In choosing a study on which to base the RfD, the Agency recommends a hierarchy of acceptable data. Most preferable is a well-conducted epidemiologic study that demonstrates a positive association between a quantifiable exposure to a chemical and human disease. Use of acceptable human studies avoids the problems of interspecies extrapolation, and thus, confidence in the estimate is often greater. At present, however, human data adequate to serve as a basis for quantitative risk assessment are available for only a few chemicals. Most often, inference of adverse health effects for humans must be drawn from toxicity information gained through animal experiments with human data serving qualitatively as supporting evidence. Under this condition, health effects data must be available from well-conducted animal studies and be relevant to humans based on a defensible biological rationale (e.g., similar metabolic pathways). In the absence of data from a more "relevant" species, data from the most sensitive animal species tested (i.e., the species demonstrating an adverse health effect at the lowest administered dose via a relevant route of exposure), shall generally be used as the critical study.

The route of administration must be considered when choosing the critical study from among quality toxicity tests. The vehicle in which the chemical is administered is also relevant. For example, within the oral route of exposure, the bioavailability of a chemical ingested from one source (e.g., food) may differ from when it is ingested from another source (e.g., water). Usually, the toxicity database does not provide data on all possible routes, sources, and/or durations of administration. In general, the preferred exposure route is that which is considered most relevant to environmental exposure. For example, when developing drinking water standards, the Agency has placed greater weight on oral studies in experimental animals, especially those studies in which the contaminant is administered via water. However, in the absence of data on the exposure route and/or source of concern, it is the Agency's view that the potential for the toxicity manifested by one route and/or source of exposure may be relevant to other exposure routes and/or sources. EPA's *Interim Methods for the Development of Inhalation Reference Doses* (USEPA, 1989) discusses specific issues relevant to route-to-route extrapolation. These include issues of portal-of-entry effects, available toxicokinetic data for the routes of interest, measurements of absorption efficiency by each route of interest, comparative excretion data when the associated metabolic pathways are equivalent by each route of interest, and comparative systemic toxicity data when such data indicate equivalent effects by each route of interest.

Preference should be given to studies involving exposure over a significant portion of the animal's lifespan since this is anticipated to reflect the most relevant environmental exposure. Studies with shorter time frames can miss important effects. In selected cases, studies of less than 90 days can be used for quantification, but the study must be of exceptionally high quality. In general short-term tests should not be used for anything other than interim RfDs or for developmental RfDs. However, developmental effects can sometimes be the critical effect and serve as the basis of an RfD. The duration of a developmental study is generally less than 15 days.

3.4.2   Critical Data and Endpoint

The experimental exposure level representing the highest dosage level tested at which no adverse effects were demonstrated in any of the species evaluated should be used for criteria development. By basing criteria on the critical toxic effect, it is assumed that all toxic effects are prevented (USEPA, 1988). In the absence of such data, the lowest LOAEL dosage may be used for criteria development and an additional uncertainty factor for LOAEL to NOAEL extrapolation is applied. When two or more studies of equal quality and relevance exist, the geometric means of the NOAELs or LOAELs may be used.

Often a chemical may elicit multiple effects, each with a different NOAEL and LOAEL. From among these effects, the Agency selects a critical endpoint. The critical endpoint is generally the effect that exhibits the lowest LOAEL (USEPA, 1988).

3.5   DERIVING RFDS USING THE NOAEL/LOAEL APPROACH

The 1988 RfD background document describes methods used to derive an RfD for a given chemical and criteria for selection of the critical NOAEL or LOAEL. Appropriate UFs and modifying factors (MF) are then applied to the selected endpoint to derive the RfD.

The general equation for deriving the RfD is (USEPA, 1988):

$$\text{RfD (mg/kg-day)} = \frac{\text{NOAEL}}{\text{UF} * \text{MF}} \text{ or } \frac{\text{LOAEL}}{\text{UF} * \text{MF}}$$

(Equation 3-1)

where:

NOAEL   =   An exposure level at which there are no statistically or biologically significant increases in the frequency or severity of observed adverse effects between the exposed population and its appropriate control; *some effects may be produced at this level, but they are not considered as adverse, nor precursors to specific adverse effects.*

LOAEL   =   The lowest experimental exposure level at which there are statistically or biologically significant increases in frequency or severity of observed adverse effects between the exposed

population and its appropriate control group. The LOAEL may be used if the NOAEL cannot be determined.

UF   =   An uncertainty factor which reduces the dose to account for several areas of scientific uncertainty inherent in most toxicity databases. Standard UFs are used to account for variation in sensitivity among humans, extrapolation from animal studies to humans, and extrapolation from less than chronic NOAELs to chronic NOAELs. An additional UF may be employed if a LOAEL is used to define the RfD.

MF   =   A modifying factor, to be determined using professional judgment. The MF provides for additional uncertainty not explicitly included in UF, such as completeness of the overall database and the number of species tested. (The value for MF must be greater than zero and less than or equal to 10; the default value for the MF is 1).

The RfD is generally expressed in units of milligrams per kilogram of body weight per day (mg/kg-day).

3.5.1   Selection of Uncertainty Factors and Modifying Factors

The choice of appropriate UFs and MFs must be a case-by-case judgment by experts and should account for each of the applicable areas for uncertainty and nuances in the available data that impact uncertainty. Several reports describe the underlying basis of UFs (Zielhuis and van der Kreek, 1979; Dourson and Stara, 1983) and research into this area (Calabrese, 1985; Hattis et al., 1987; Hartley and Ohanian, 1988; Lewis et al., 1990; Dourson et al., 1992).

The UFs summarized in Table 3-1 account for five areas of scientific uncertainty inherent in most toxicity databases: inter-human variability (H) (to account for variation in sensitivity among the members of the human population); experimental animal-to-human extrapolation (A); subchronic to chronic extrapolation (S) (to account for uncertainty in extrapolating from less-than-chronic NOAELs (or LOAELs) to chronic NOAELs); LOAEL-to-NOAEL extrapolation (L); and database completeness (D) (to account for the inability of any single study to adequately address all possible adverse outcomes). Each of these five areas is generally addressed by the Agency with a factor of 1, 3, or 10. The default value is 10.

Table 3-1. Uncertainty Factors and the Modifying Factor

Uncertainty Factor	Definition

UF_H — Use a 1-, 3-, or 10-fold factor when extrapolating from valid data in studies using long-term exposure to average healthy humans. This factor is intended to account for the variation in sensitivity (intraspecies variation) among the members of the human population.

UF_A — Use an additional 1-, 3-, or 10-fold factor when extrapolating from valid results of long-term studies on experimental animals when results of studies of human exposure are not available or are inadequate. This factor is intended to account for the uncertainty involved in extrapolating from animal data to humans (interspecies variation).

UF_S — Use an additional 1-, 3-, or 10-fold factor when extrapolating from less-than-chronic results on experimental animals when there are no useful long-term human data. This factor is intended to account for the uncertainty involved in extrapolating from less-than-chronic NOAELs to chronic NOAELs.

UF_L — Use an additional 3- or 10-fold factor when deriving an RfD from a LOAEL, instead of a NOAEL. This factor is intended to account for the uncertainty involved in extrapolating from LOAELs to NOAELs.

UF_D — Use an additional 1-, 3-, or 10-fold factor when deriving an RfD from an "incomplete" database. Missing studies, e.g., reproductive, are often encountered with chemicals. This factor is meant to account for the inability of any study to consider all toxic endpoints. The intermediate factor of 3 (½ log unit) is often used when there is a single data gap exclusive of chronic data. It is often designated as UF_D.

Modifying Factor

Use professional judgment to determine the MF, which is an additional uncertainty factor that is greater than zero and less than or equal to 10. The magnitude of the MF depends upon the professional assessment of scientific uncertainties of the study and database not explicitly treated above (e.g., the number of species tested). The default value for the MF is 1.

Note: With each UF or MF assignment, it is recognized that professional scientific judgment must be used.

In addition, an MF may be used to account for areas of uncertainty that are not explicitly considered using the standard UF. This value of the MF is greater than zero and less than or equal to 10, but it should generally be used on a log 10 basis (i.e., 0.3, 1, 3, 10) as are the standard UFs. The default value for this factor is 1.

The Agency's reasoning in its use of the MF is that the areas of scientific uncertainty labeled H, A, S, L, or D do not represent all of the uncertainties in the estimation of an RfD. For

example, the fewer the number of animals used in a dosing group, the more likely it is that no adverse effect will be observed at a dose point which may have had an effect in a larger population. Such a case might argue for modifying the usual 10-fold factors—a 100-fold UF might be raised to 250 if too few animals were used in a chronic study. While this increase is scientifically reasonable, it introduces two difficulties: the adjustments applied could differ between risk assessors, and the applied precision of the result might not be justified by the data. For example, a UF of 250 has an implied precision of 2 digits and is not appropriate in relation to the variability of the biological response. The Agency intends to avoid these difficulties through limiting the options for the modifying factor (1, 3, 10).

In practice, the magnitude of the overall UF is dependent on professional judgment as to the total uncertainty in all areas. When uncertainties exist in one, two or three areas, the Agency generally uses 10-, 100-, and 1,000-fold UF respectively. When uncertainties exist in four areas, the Agency generally uses an UF no greater than 3,000. It is the Agency's opinion that toxicity databases that are weaker and would result in UFs in excess of 3,000 are too uncertain as a basis for quantification. In such cases, the Agency does not estimate an RfD, and additional toxicity data are sought or awaited. For a few chemicals, an UF of 10,000 was applied. However, in such cases, the risk assessment was completed before current policies for the maximum UF were in place.

The Agency occasionally uses a factor of less than 10 or even a factor of 1, if the existing data reduce or obviate the need to account for a particular area of uncertainty. For example, the use of a 1-year rat study as the basis of an RfD may suggest the use of a 3-fold, rather than 10-fold, factor to account for subchronic to chronic extrapolation, since it can be empirically demonstrated that 1-year rat NOAELs are generally closer in magnitude to chronic values than are 3-month NOAELs (Swartout, 1990). Lewis et al. (1990) more fully investigate this concept of variable uncertainty factors through an analysis of expected values.

The modification of UFs from their standard values should follow the general guidelines for composite UFs and the overall precision of one digit for UFs. The composite uncertainty factor to use with a given database is again strictly a case-by-case judgment by experts. It should be flexible enough to account for each of the applicable five areas of uncertainty and any nuances in the available data that might change the magnitude of any factor. The Agency describes its choice for the composite UF and sub-components for individual RfDs on its IRIS. Table 3-2 presents examples of the UFs employed for several chemicals recently accepted into IRIS through the consensus process. Because of the high degree of judgment involved in the selection of UFs and MFs, the risk assessment justification should include a detailed discussion of the selection of these factors, along with the data to which they are applied.

**Table 3-2. Examples of Uncertainty Factors and Modifying Factors
from IRIS Risk Assessments**

Chemical	Total UF	MF	Rationale
Barium	3	1	The RfD is based on NOAELS from two human studies that were supported by NOAELS and LOAELS from two well designed animal studies. The UF of 3 was applied because of inadequate data on differences between adults and children with regard to the critical effect (hypertension) and incomplete data on possible developmental effects.
Beryllium	300	1	The RfD is based on a BMD_{10} from a dietary study in dogs. The UF includes 10-fold values for intraspecies and interspecies variability and a 3 to accommodate database deficiencies regarding human effects via the oral route, reproductive/developmental, and immunological effects.
Chromium VI	300	3	The RfD was based on a NOAEL from an animal study. The UF includes 10-fold values for intraspecies and interspecies variability and a 3 to accommodate the less-than-lifetime exposure in the principle study. A modifying factor of 3 was added because of concerns for acute gastrointestinal effects in humans with reported exposures to about 20 mg/L in drinking water.
Naphthalene	3000	1	The RfD is based on a duration-adjusted NOAEL from a subchronic animal study. The UF includes 10-fold values for intraspecies and interspecies variability and the less-than-chronic duration of the study. An additional 3-fold factor was added because of the lack of a two-generation reproductive study.

3.5.2   Confidence in NOAEL/LOAEL-Based RfD

As stated previously, when available, adequate data from acceptable human studies should be used as the basis for the RfD. Use of good epidemiology studies generally give the highest confidence in RfDs. In the absence of such data, RfDs are estimated from studies in experimental animals.

The Agency generally considers a "complete" database for calculating a chronic RfD for noncancer health effects to include the following:

C Two adequate mammalian chronic toxicity studies, by the appropriate route in different species, one of which must be a rodent.

C One adequate mammalian multi-generation reproductive toxicity study by an appropriate route.

C Two adequate mammalian developmental toxicity studies by an appropriate route in different species.

For a "complete" database, the likelihood that additional toxicity data may change the RfD is low. Thus, the Agency usually has confidence in such an RfD because additional toxicity data are not likely to change the value.

The Agency considers a NOAEL from a well-conducted, mammalian subchronic (90-day) study by the appropriate route as a minimum database for estimating an RfD. However, for such a database, additional toxicity data may change the RfD. Thus the Agency generally has less confidence in such an RfD.

For some chemicals, an acute health hazard is the critical effect of concern. These could include neurotoxic, portal of entry, or immunotoxic effects of acute exposures at environmental levels of contaminant. In such cases, longer term studies (subchronic or chronic) that would typically be included in a review of the toxicity literature may not capture the critical endpoint. Under such circumstances, greater emphasis should be placed on characterizing the acute threshold as opposed to the potential chronic effects.

Developmental toxicity data, if they constitute the sole source of information, are not considered an adequate basis for chronic RfD estimation. This is because such data are often generated from short-term chemical exposures, and, thus, are of limited relevance in predicting possible adverse effects from chronic exposures. However, if a developmental toxicity endpoint is the critical effect established from a "complete" database, a chronic RfD can be derived from such data, applying the uncertainty and MFs normally required. Developmental data are the basis for developmental reference doses (RfD_{DT}).[14] The term RfD_{DT} is used to distinguish the developmental value from the chronic RfD which refers to chronic exposure situations. Uncertainty factors for developmental toxicity include a 10-fold factor for interspecies variation and a 10-fold factor for intraspecies variation; in general, an uncertainty factor is not applied to

---

[14]A RfD for developmental toxicity (RfD_{DT}) is discussed in USEPA (1991a).

account for duration of exposure. In some cases, additional factors may be applied due to a variety of uncertainties that exist in the database. For example, the standard study design for developmental toxicity study calls for a low dose that demonstrates a NOAEL, but there may be circumstances where a risk assessment must be based on the results of a study in which a NOAEL for developmental toxicity was not identified. For details regarding risk assessment for developmental toxicants, refer to EPA's *Final Guidelines for Developmental Toxicity Risk Assessment* (USEPA, 1991b).

3.5.3   Presenting the RfD as a Single Point or as a Range

Although the RfD has traditionally been presented and used as a single point estimate, its definition contains the phrase ". . . an estimate (with uncertainty spanning perhaps an order of magnitude) . . ." (USEPA, 1988). Underlying this concept is the reasoning that during the derivation of the RfD, the selection of the critical effect and of the total uncertainty factor is based on the "best" scientific judgment of the Agency Work Group and that other groups of competent scientists examining the same database would reach a similar conclusion, within an order of magnitude.

Presenting the RfD as a range may be more appropriate than expressing it as a point estimate because rarely are sufficient data available to precisely determine a lifetime threshold for a human. Even when there are good, reliable data, the variability of response in the human population argues for expressing the RfD as a range. However, although EPA supports the use of a range that spans one order of magnitude for most RfDs, there are a number of potential interpretations of the term "order of magnitude" as described below:

C   Range = x to 10x. (where point estimate of RfD=x). This view is supported by those who believe that the risk assessment process is so inherently conservative that the RfD should be considered to be the lowest estimate, with the range of imprecision all resting <u>above</u> this point estimate.

C   Range = 0.3x to 3x. This view is held by many EPA scientists who have developed RfDs. The RfD point estimate, x, is the midpoint of a range that spans an order of magnitude.

C   Range = 0.1x to x. This is the view held by many risk managers. Regulatory decisions (e.g., setting of standards or cleanup levels) are made based on the assumption that standards or cleanup levels are protective as long as they do not exceed the RfD.

C   Range = 0.1x to 10x. This range represents the assumption that the order of magnitude range could be on <u>either</u> side of the point estimate x.

The Agency is proposing a risk management approach where the upper and lower bounds of the range are correlated to the uncertainty. Because the uncertainty around the dose response relationship increases as extrapolation below the observed data increases, the use of an alternative point within the range for the RfD may be more appropriate in characterizing risk than the calculated point estimate. Therefore, as a matter of risk management policy, it is proposed that if the product of the UFs and MF used to derive the RfD is 100 or less, there can be no consideration of a range. When greater than 100 but less than 1,000, a range can be established which is one half of a $\log_{10}$ (3-fold) or a number ranging from the point estimate divided by 1.5 to the point estimate multiplied by 1.5. With a UF of 1,000 and above, the range can span a number ranging from the point estimate divided by 3 to the point estimate multiplied by 3 (a 10-fold range). A risk assessor can then select a single point within the defined range to use as an alternate to the calculated RfD.

The use of an alternative value within the range defined by the uncertainty must be justified. As used in this document, justification means that there are scientific data which indicate that some value in the range other than the point estimate may be more appropriate than the point estimate, based on human health or environmental fate considerations. One example of a situation where a point other than the calculated RfD might be applied would be where there is a lower bioavailability of the contaminant in fish than in water. In such an instance, the decreased bioavailability from fish tissues could be used to support selection of an RfD value greater than the calculated value if the critical study were one where the contaminant was administered through drinking water. For example, most inorganic contaminants, particularly divalent cations, have bioavailability values of 20 percent or less from a food matrix, but are much more available (about 80 percent or higher) from drinking water. Accordingly, the external dose necessary to produce a toxic internal dose would likely be higher for a study where the exposure occurred through the diet rather than the drinking water. As a result, the RfD from a dietary study would likely be higher than that for the drinking water study if equivalent external doses had been used.

The exposures considered in the derivation of the AWQC include fish (food) and water. Thus, one might be able to justify an alternative value to the RfD estimate that was slightly higher than the RfD estimate in cases where the NOAEL that was the basis for the RfD came from a drinking water study, but slightly lower than the RfD estimate if the NOAEL was from a dietary study.

Another situation where a point from the lower end of the range could be selected is one where there is a well-defined sensitive population, such as women in the first trimester of pregnancy. In this situation, the presence of the contaminant in both water and fish and average body weights for women of reproductive age that are less than the 70 kg default may justify an alternative value from the low end of the range about the RfD estimate.

Table 3-3 gives examples of some factors to consider when determining whether to use the point estimate of the RfD or values higher or lower than the point estimate. The factors presented in Table 3-3 should be considered in making the decision as to whether or not to use a value other than the point estimate. EPA advocates the use of the point estimate of the RfD as the default to derive the AWQC.

**Table 3-3. Some Scientific Factors to Consider
When Using the RfD Range**

Use point estimate RfD	- Default position - Total UF/MF product is 100 or less - Essential nutrient
Use point from the lower range of RfD	- Increased bioavailability from the exposure vehicle verses the experimental conditions used in the RfD study - The seriousness of the effect and whether or not it is reversible - A shallow dose-response curve in the range of observation - Exposed group contains a sensitive population (e.g., children or fetuses)
Use point from the upper range of RfD	- Decreased bioavailability for humans as compared to experimental animals - RfD based on minimal LOAEL and a UF/MF of 1,000 or greater - A steep dose-response curve in the range of observation - No sensitive populations identified

There are many factors that can affect the uncertainty in the RfD, and thereby affect the selection of an alternative value within a range. The completeness of the database plays a major role. Observing the same effects in several animal species, including humans, can increase confidence in the RfD point estimate and thereby narrow the range of uncertainty. Other factors that can affect the uncertainty are the slope of the dose-response curve, seriousness of the observed effect, spacing of doses, and the route of exposure. For example, a steep dose-response curve indicates that relatively large differences in effect occur with a small change in dose; thus, there will be a greater chance that the data will allow scientists to distinguish clearly (i.e., statistically) between doses that produce an effect and those that do not. For a situation where the RfD is derived from a LOAEL for a serious effect, an additional uncertainty factor is often used in the RfD derivation to protect against less serious effects that could have occurred at lower doses had lower doses been evaluated. Dose spacing and the size of the study groups used in the experiment can also affect the confidence in the RfD. The "true" NOAEL is not identified by a standard toxicology study. The wider the dose spacing, and the smaller the number of animals studied, the greater the margin of uncertainty about where the "true" NOAEL may fall. Finally, for some RfDs, the route of exposure in the experiment may not match the route of exposure for

humans, and interroute extrapolation or toxicokinetic modeling may be considered using assumptions about differences in absorption rates between routes.

There are cases where an alternative value within a range should not be used. For example, the RfD for zinc (USEPA, 1992) is based on consideration of nutritional data, a minimal LOAEL, and a UF of 3. If the factor of 3 were used to bound the RfD for zinc, then the upper-bound level would approach the minimal LOAEL. This situation must be avoided, since it is unacceptable to set a standard at levels that may cause an adverse effect. The risk manager must be informed of those specific cases when it is not scientifically correct to use the RfD range. Table 3-3 provides managers with guidelines on the scientific basis for using the range.

3.6   DERIVING AN RFD USING A BENCHMARK DOSE APPROACH

A number of issues have been raised regarding the development of the RfD based on the traditional NOAEL/LOAEL approach. These concerns include the following:

C   The traditional approach does not incorporate information on the shape of the dose-response curve, but focuses only on a single point (the NOAEL or LOAEL).

C   The value of the NOAEL depends on the number of doses and spacing of the doses in the experiment. The possible NOAEL values are limited to the discrete values of the experimental doses. Theoretically, the experimental no adverse effect level could be any value between the experimental NOAEL and the LOAEL, and sometimes the true NOAEL is below the observed NOAEL, especially in studies with a limited number of animals in each dose group.

C   Data variability is not directly taken into account. For example, studies based on a larger number of animals may detect effects at lower doses than studies with fewer animals; as a result, the NOAEL from a small study may be higher than the NOAEL from a similar but larger study in the same species. The traditional approach does not have a mechanism to account for such data variability.

C   The determination of the NOAEL is dependent on the background incidence of the effect in control animals; therefore, statistically significant differences between the dose groups and the control group are more difficult to detect if background incidence is relatively high, even if biologically significant effects occur.

C   In conjunction with exposure data, the NOAEL-based RfD can be used to estimate the size of the population at risk, but not the magnitude of the risk.

In response to these concerns, alternative approaches have been developed that attempt to address some of these shortcomings. One such alternative, the BMD approach, has been the subject of extensive research over the past decade (Crump 1984, 1995; Gaylor, 1983, 1989; Dourson et al., 1985; Brown and Erdreich, 1989; Kimmel, 1990; Faustman et al., 1994; Allen et al., 1994a, 1994b). The EPA Risk Assessment Forum is in the process of developing guidance on procedures and models to be used in the calculation of BMDs. The following discussion presents the general methods for calculation of a RfD using the BMD approach; for more extensive discussion, the reader is referred to Crump et al. (1995). To date, the Agency has used the BMD approach for deriving the RfD for methylmercury (USEPA, 1994a) and the RfC for several compounds.

3.6.1 Overview of the Benchmark Dose Approach

A BMD or benchmark concentration (BMC) is defined as a statistical lower confidence limit on the dose producing a predetermined level of change in response (the benchmark response, or BMR) relative to controls. The BMD/BMC is intended to be used as an alternative to the NOAEL in deriving a point of departure for low dose extrapolations. The BMD/BMC is a dose corresponding to some change in the level of response relative to background and is not dependent on the doses used in the study. The BMR is based on a biologically significant level of response or on the response level at the lower end of the observable range for a particular endpoint. The BMD/BMC approach does not reduce uncertainty inherent in extrapolating from animal data to humans (except for that in the LOAEL to NOAEL extrapolation), and does not require that a study identify a NOAEL. The BMD/BMC approach requires only that at least one dose be near the range of the response level for the BMD/BMC.

Modeling of dose and response is central to the BMD approach. The modeling process is limited to the experimental range and no attempt is made to extrapolate to doses far below the experimental range. Generally, the models used in the BMD approach are statistical rather than biologically-based models; thus, they cannot be reliably used to extrapolate to low doses without incorporating detailed information on the mechanisms through which the toxic agent causes the particular effect being modeled.

Once a mathematical dose-response curve and its corresponding curve of confidence limits are established, the assessor selects a point on the lower confidence dose curve corresponding to the chosen BMR. This point on the lower confidence curve is the lower confidence bound of the effective dose for that BMR (denoted as the BMD) (see Exhibit 3-1). A BMD may be calculated for each endpoint for which there is an adequate database.

Exhibit 3-1

Derivation of RfD Using BMD Approach
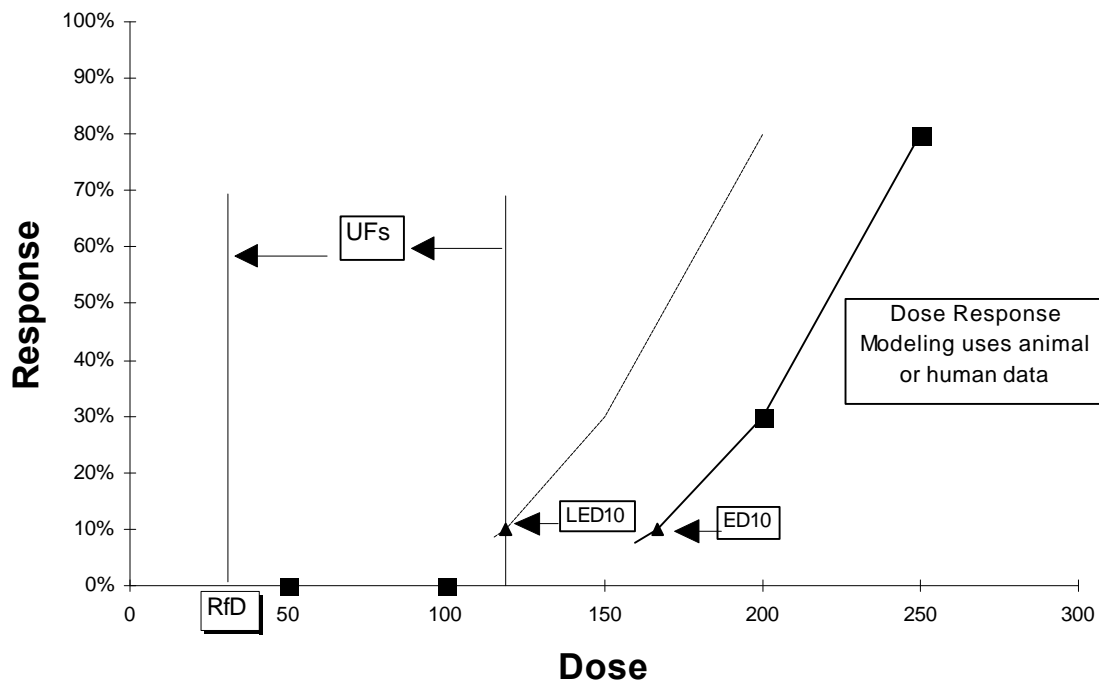
The BMD approach offers a number of advantages over the traditional approach for deriving the RfD from the NOAEL/LOAEL divided by uncertainty factors. Some of the advantages of the BMD approach are the following:

- it considers the dose-response curve, including its shape;

- it better accounts for statistical variability in the data; and

- it is not overly sensitive to dose spacing and, thus, is not limited to experimental doses for determining the effect level. However, the data requirements for using the BMD approach are more extensive than those for the NOAEL/LOAEL approach.

Studies with small group sizes and evaluation of a limited number of endpoints will tend to yield lower BMD values because the confidence bands will be wider. Therefore, the BMD approach provides an incentive to conduct more robust studies, since better studies give narrower confidence bands.

3.6.2   Calculation of the RfD Using the Benchmark Dose Method

The determination of an RfD using the BMD approach involves four basic steps. The first step involves the selection of the experiments and responses that will be used for modeling the BMD. Second, BMDs are calculated for the selected responses; BMD values should be calculated for all endpoints that have the potential for yielding the critical BMD. Third, a single BMD is selected from among those calculated. Finally, the RfD is calculated by dividing the chosen BMD by appropriate UFs. The decision points associated with these steps are outlined in Table 3-4. The discussion that follows summarizes the critical issues unique to the BMD approach and is based largely on information from Crump et al. (1995).

Table 3-4. Steps and Decisions Required in the BMD Approach

Step	Decisions
1. Selection Study/Response	1. Experiments to include 2. Responses to model
2. Model dose-response	1. Format of data 2. Mathematical model(s) 3. Handling model fit 4. Measure of altered response
3. Select BMD(s)	1. Critical BMR 2. Confidence limit calculation
4. Calculate RfD	1. Uncertainty factors

Source: Crump et al., 1995

3.6.2.1 <u>Selection of Response Data to Model</u>

The selection of experiments and responses suitable for BMD modeling involves considerations similar to those for identifying the appropriate studies upon which to base a NOAEL. There may be several appropriate studies and relevant health effects that could be modeled for a chemical. Ideally, BMD calculations would be performed for the complete set of relevant effects. However, utilizing all relevant responses for the calculation of BMDs may be

resource-intensive. Further, it is difficult to interpret results from a large number of dose-response analyses. When selecting the data to model it is considered appropriate to limit attention to those responses for which there is evidence of a dose-response relationship. Statistically, such a relationship may be indicated by significant trends (either increasing or decreasing) in the response as dose level increases. Considerations of biological significance may also be warranted. Another alternative is to focus efforts on modeling the most critical effects as seen at the LOAEL. However, limiting the number of responses modeled may potentially misrepresent the minimum BMD.[15]

3.6.2.2 Use of Categorical Versus Continuous Data

A central issue in the selection of data to model concerns the form of the data used. Categorical data, particularly quantal data, are relatively straightforward to use in the BMD approach, since the data are expressed as the number (or percent) of subjects exhibiting a defined response at a given dose. Data may also be of the continuous form, where results are expressed as the measure of a continuous biological endpoint, such as a change in organ weight or serum enzyme level. With continuous data the results are generally presented in terms of means and standard deviation for dose groups but are most valuable when data for individual animals are available. To perform dose-response modeling of such data, the bounds for a normal response as opposed to an adverse response must be decided. Continuous data can be modeled by looking at the mean response for each dose group as a fraction of the mean response of the control group or as the percentage of animals showing an adverse response at each dose level (Gaylor and Slikker, 1990; Crump, 1995). Such approaches take advantage of the continuous nature of the response data, but express the results in terms that are directly comparable to those derived from analysis of categorical data, i.e., in terms of additional or extra risk, rather than in terms of changes in mean response. Crump (1995) provided options for handling continuous data that can be applied to the same models used for analysis of quantal endpoints. Such developments have enhanced the consistency of results across different endpoints for any particular chemical. In any case, application of the BMD approach to continuous data requires professional judgment in order to determine what level or category of response constitutes an abnormal (adverse) effect. The BMD approach is not recommended for routine use but may be used when data are available and justify the extensive analyses required.

3.6.2.3 Choice of Mathematical Model

Various mathematical approaches have been proposed for determining the BMD. Table 3-5 shows a number of dose-response models that may be used for estimating the BMD with quantal or continuous data. The EPA Benchmark Dose modeling program (Version 1.2) includes

---

[15]This is due to the fact that an effect seen only at doses above the LOAEL but having a shallow dose-response could produce a lower BMD than an effect seen at the LOAEL, which has a steeper dose-response.

Table 3-5. Dose-Response Models Proposed for Estimating BMDs

Model	Formula
Quantal Data	
Quantal linear regression (QLR)	$P(d) = c + (1-c)\{1-\exp[-q_1(d-d_0)]\}$
Quantal quadratic regression (QQR)	$P(d) = c + (1-c)\{1-\exp[-q_1(d-d_o)^2]\}$
Quantal polynomial regression (QPR)	$P(d) = c + (1-c)\{1-\exp[-q_1d_1-...q_kd^k]\}$
Quantal Weibull (QW)	$P(d) = c + (1-c)\{1-\exp[-q_1d^k]\}$
Log-normal (LN)	$P(d) = c + (1-c)N(a+b \log d)$
Continuous data	
Continuous linear regression (CLR)	$m(d) = c + q_1(d-d_0)$
Continuous quadratic regression (CQR)	$m(d) = c + q_1(d-d_0)^2$
Continuous linear-quadratic regression (CLQR)	$m(d) = c + q_1d+q_2d^2$
Continuous polynomial regression (CPR)	$m(d) = c + q_1d+...+q_kd^k$
Continuous power (CP)	$m(d) = c + q_1(d)^k$

Notes: $P(d)$ is the probability of a response at the dose, d; $m(d)$ is the mean response at the dose, d. In all models, c, $q_1,...q_k$, and d_o are parameters estimated from the data. For the quantal models, $0\#c\#1$ and $q_i\$0$. For the CPR model proposed by Crump (1984), all the q_i have the same sign. In the CLQR model discussed by Gaylor and Slikker (1990), q_1 and q_2 were not constrained to have the same sign. For all models, $d_0\$0$, $k\$1$. $N(x)$ denotes the normal cumulative distribution function.

SOURCE: Crump et al., 1995.

option for using gamma, logistic, multistage, probit, quantal-linear, quantal-quadratic and Weibull models for quantal data. Linear, polynomial, power and Hill models are available for use with continuous data. The EPA software for benchmark modeling can be downloaded from http://www.epa.gov/ncea/bmds.htm. The Agency is also developing guidance for use of BMD model results.

Information generally required for application of dose-response models for categorical (including quantal) data includes the experimental doses, the total number of animals in each dose group, and the number of these whose responses are in each of the categories of response. For continuous data, the experimental doses, number of animals in each dose group, mean response in each group, and sample variance of response in each dose group are needed.

The BMD approach should not be applied to data sets with only two experimental groups (a control and one positive dose). In such cases, much of the advantage of the BMD approach with respect to consideration of the dose-response shape will be lost; such data supply little information about the shape of the dose-response curve. The more doses available, especially at lower doses, the greater the expected benefit of the BMD approach as compared to the NOAEL-based approach.

3.6.2.4 Handling Model Fit

Fitting the models to experimental data gives estimates of the parameters that help determine the model which has the best fit to the data. This fitting, usually accomplished through maximum likelihood methods, estimates the probability of response (for quantal data) or the mean response (for continuous data) for each dose level. Goodness-of-fit tests can be used to determine if a model adequately describes the dose-response data. The experimental data should be plotted against the model projection, thereby providing a visual representation of fit.

In many cases, several models may appear to fit the data well. In these cases, other considerations can be used to select an appropriate model. For example, the statistical assumptions underlying the model should be reasonable for the given data. Quantal results, for example, are assumed to follow a binomial distribution around a dose-dependent expected value. This assumption requires that each subject responds independently and that all have an equal probability of responding. Continuous responses for each dose level are assumed to follow a normal distribution and are also assumed to be independent. When biological factors may be important (e.g., intralitter correlation for developmental toxicity data) they may also be used to select appropriate models. Another biological consideration may be whether or not a threshold is assumed to exist. If a threshold is expected for the given effect, then a model that allows for a threshold dose may be chosen for modeling. The biological plausibility of the dose-response curve shape should always be a consideration in model selection.

Even with these considerations, several different models may often adequately describe the data. In such cases it is important to examine fit about the BMR. Models that have similar fit to the entire data set may differ with respect to their predictions near the BMR. It may be possible to select one model over another on the basis of that more local behavior.

In certain data sets, none of the standard models may provide a reasonable fit to the data. Fit is assessed statistically by comparing the model predictions to the observations. Goodness-of-fit statistics formalize that comparison and provide p-values, ranging between 0 and 1, as a measure of fit. When using a P^2 statistic, larger p-values are indicative of good fit; smaller p-values of poorer fit. Sufficiently small p-values (e.g., less than 0.01 or 0.05) are typically viewed as an indication that the model was not adequate for describing the observed dose-response pattern.

Poor fit is often due to reduced responses at higher doses that are inconsistent with the dose-response trend for lower doses, perhaps due to competing toxic processes or saturation of metabolic systems related to the toxic response of interest. Several procedures can be used to adjust the modeling process in these circumstances. For example, responses at the highest doses could be eliminated, since those doses are usually least informative of responses in the lower dose region of interest. In the case of saturated metabolic pathways, toxicokinetic data can be used to estimate delivered dose to the organ of interest. The BMD modeling can then be conducted on the delivered dose (Andersen et al., 1987, 1993; Gehring et al., 1978).

Visual (graphical) examination of the model predictions in relation to the observations is an essential exercise with respect to all of these fit issues. This supplements the formal statistical assessment of fit and may, in fact, be equally or more informative. Biological plausibility is another critical factor to consider when selecting the best BMD from among several options.

3.6.2.5 Measure of Altered Response

Crump (1984) proposed two measures of increased response for quantal data. These are additional risk and extra risk. Additional risk is the probability of response at dose d, P(d), minus the probability of response at zero dose (control response), P(0). It describes the additional proportion of animals that respond in the presence of a dose. Extra risk is additional risk divided by [1-P(0)]. It describes the additional proportion of animals that respond in the presence of a dose, divided by the proportion of animals that would not respond under control conditions. These measures are distinguished in the way they account for control responses. For example, if a dose increases a response from 0 to 1 percent, both the additional risk and the extra risk is 1 percent. However, if a dose increases risk from 90 to 91 percent, the additional risk is still 1 percent, but the extra risk is 10 percent. The choice of extra risk versus additional risk is based to some extent on assumptions about whether an agent is adding to the background risk. Extra risk is viewed as the default because it is more conservative.

Analogous measures of risk have been proposed for continuous data (Crump, 1984). First, altered response can be expressed as the difference between the mean response to dose d minus the mean control response. The second measure is simply the difference between dose and control means divided by (i.e., normalized by) the control mean response. The second measure expresses change as a fraction of the control response rather than as an absolute change.

More recent consideration of BMDs for continuous endpoints have suggested other alternatives. Allen et al. (1994a, 1994b) and Kavlock et al. (1995) determined that normalizing changes in mean responses by a multiple of the background standard deviation produced BMDs that were comparable, on average to NOAELs. For the developmental endpoints that those investigators studied, the preferred multiple for the standard deviation was 0.5.

It is not clear when measures of risk expressed relative to the background (e.g., extra risk) are preferable to measures expressed as absolute changes. Additional research is required to provide guidance regarding the measure of altered response that is most appropriate in particular circumstances.

3.6.2.6 Selection of the BMR

A critical decision for deriving the BMD is the selection of the Benchmark Response (BMR). Since the BMD is used like a NOAEL in the derivation of the RfD, the BMR should be selected near the low end of the range where effects were detected in a study. The dose predicted to cause a 10 percent increase in the incidence of the effect in the test population (ED_{10}) is frequently chosen as the BMR. For some data, it may be possible to adequately estimate the ED_{05} or ED_{01}, which are closer to a true no-effect dose. However, in many cases the ED_{10} is the lowest level of risk that can be estimated from standard toxicity studies (Crump, 1984).

During a BMD Workshop, sponsored by EPA, participants generally agreed that the appropriate BMR should either be 5 percent or 10 percent, but acknowledged that future research might demonstrate the advisability of selecting one value over another (ILSI, 1993). Research by Allen et al. (1994a, 1994b) and Faustman et al. (1994) indicates that BMDs defined in terms of 10 percent increases in probability of response tend to be, on average, similar to corresponding NOAELs for quantal developmental toxicity studies. For the purposes of water quality criteria derivation, EPA recommends the use of the ED_{05} or ED_{10} when deriving a BMD.

3.6.2.7 Calculating the Confidence Interval

The BMD is defined to be the lower confidence bound on the dose corresponding to the selected BMR. A statistical lower confidence limit is used rather than a maximum likelihood estimate (MLE) for several reasons. The use of confidence limits accounts for population variability. Most biological responses are normally distributed within a population. Accordingly,

if one were to randomly select two groups of animals from the population to study, the lowest response-level responders from one study group might differ from that for the second group exposed to identical experimental conditions. Use of the lower bound confidence interval increases the confidence that the results from a study of a small group of animals can be extrapolated to the entire population.

To calculate the upper confidence bound on response and the lower bound on effective dose, one must select a procedure for calculating confidence limits and the size of the confidence limits. The recommended method used to calculate the confidence bounds on the curve relies on maximum likelihood theory. This approach is the same one used by EPA in the computer program for cancer dose-response modeling. The approach can be applied to BMD modeling using the EPA Benchmark software as well as other commercially available benchmark programs. A detailed explanation of theory supporting this approach is found in Crump (1984).

By convention, the size of the statistical confidence limits can range from 90 to 99 percent. The methods of confidence limit calculation and choice of confidence limits are critical. The Agency recommends the use of one-sided 95th percentile confidence limits for BMD modeling. This is consistent with the size of the confidence limits used in cancer dose-response modeling.

3.6.2.8 Selection of the BMD as the Basis for the RfD

An important decision is the choice of the appropriate BMD to use in the RfD calculation when multiple BMDs are calculated. Multiple BMDs can be calculated when different models fit the response data for a single study, when more than one response is modeled in a single study, and when there are different BMDs from different studies. When multiple BMDs are calculated because several models fit a single data set, the analyst may select the smallest BMD or combine BMDs by using the geometric mean. When multiple BMDs are calculated from different responses or different studies that examine the same endpoint, the choice among BMDs may also involve the selection of the "critical effect" and the most appropriate species, sex, or other relevant feature of experimental design. Graphic representations of the model output and experimental data, as well as an understanding of the biological mode of action, may help in the selection of the BMD.

3.6.2.9 Use of Uncertainty Factors with BMD Approach

Once a single or averaged BMD is selected, the RfD can be calculated by dividing the BMD by one or more uncertainty factors. As a default, all applicable uncertainty factors used in the traditional NOAEL-based RfD approach, except for the LOAEL-NOAEL extrapolation factor, should be considered. Other factors, such as the size of the BMR and confidence bounds, biological considerations (such as the possibility of a threshold), severity of the modeled effect,

and the slope of the dose-response curve, may affect the choice and magnitude of uncertainty factors (see Crump et al., 1995, for more detailed discussion).

3.6.3   Limitations of the BMD Approach

The BMD approach has been proposed as an alternative procedure that can be used until biologically motivated approaches are available for some or all effects. It provides specific improvements over NOAEL-based approaches, but by no means does it resolve all issues or difficulties associated with noncancer risk assessment. The BMD approach allows for objective extrapolation of animal response data to human exposures across the different study designs encountered in noncancer risk assessment.

3.6.4   Example of the Application of the BMD Approach

The following provides a simple example of the application of the BMD approach to quantal toxicity data. The example given is taken from Crump et al. (1995) for acrylamide. The purpose of presenting this example is to illustrate the method only; no actual risk value nor AWQC for acrylamide is derived.

3.6.4.1 Selection of Data to Model

This example takes the approach of identifying a critical study rather than modeling all endpoints seen in valid studies. For this example, a 2-year drinking water study of chronic effects in rats is used as the critical study for acrylamide (Johnson et al., 1986). The endpoint examined was tibial nerve degeneration in male rats. The researchers recorded the occurrence of nerve degeneration in two categories: none or mild; and moderate or severe. Since mild nerve degeneration occurs spontaneously in older rats, and because mild degeneration showed no dose-response relationship, only moderate and severe degeneration were recorded as responses. The data are presented in quantal form, with no or mild degeneration considered "no response," and moderate to severe degeneration recorded as a response. The dose levels and number of animals responding in each dose group are shown in Table 3-6.

3.6.4.2 Choice of Mathematical Model

From Table 3-5, we can select from among the various models available for quantal data. Fitting is accomplished through the use of maximum likelihood estimation to estimate the probability of a response at each dose level. The actual fitting exercise is done through the use of computer software.

Table 3-6. Rats Experiencing Moderate or Severe Nerve Degeneration in Response to Acrylamide Dose

Dose (mg/kg-day)	Number affected	Number tested
0	9	60
0.01	6	60
0.1	12	60
0.5	13	60
2.0	16	60

3.6.4.3 Results of Information Above

All of the models can be tried to see which achieves the best fit. The following Exhibits illustrate the best-fit modeling of the study data for the Weibull model (Exhibit 3-2) and the quadratic model (Exhibit 3-3). Table 3-7 provides the best-fit model parameters for the two equations.

Note that in example given here, the measure of altered response is extra risk (ER), which is defined as:

$$ER(d) = [P(d)-(P(0)]/[1-P(0)]$$

(Equation 3.2)

where:

ER = Extra risk
d = Dose
P = Probability

Extra risk is the fraction of animals that respond when exposed to a dose, d, among animals who otherwise would not respond.

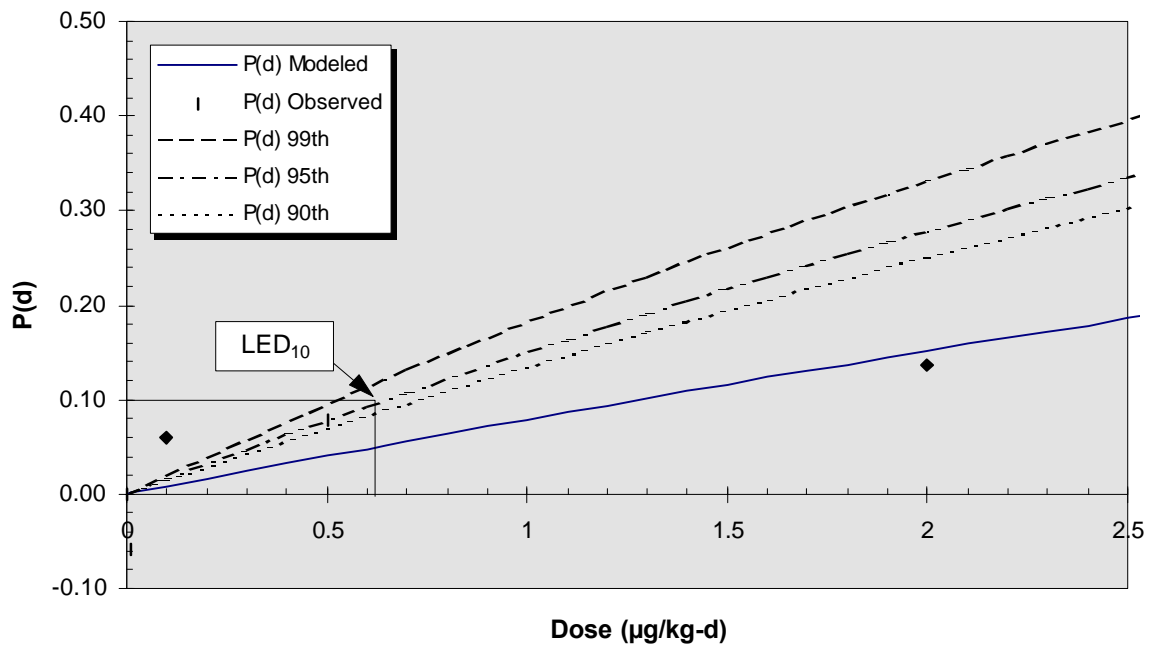**Exhibit 3-2. Quantal Weibull Regression - Extra Risk**
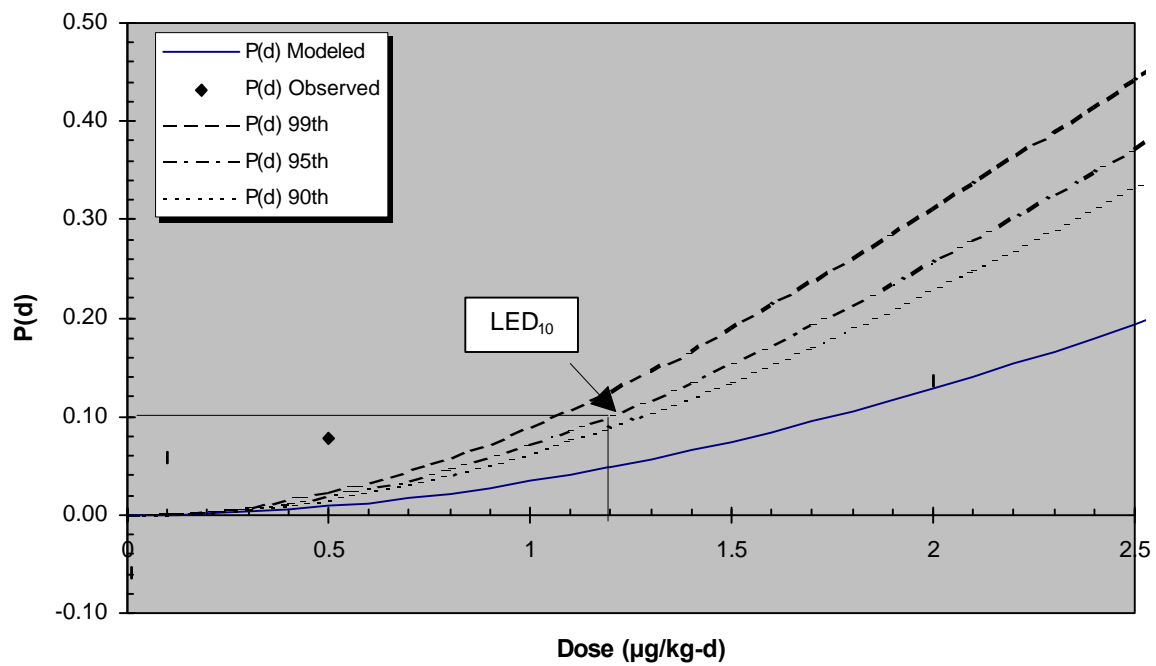

Exhibit 3-3. Quantal Quadratic Regression - Extra Risk

Table 3-7. Best-Fit Model Parameters from Modeling of the Acrylamide Data

Model	Background rate	q1	k	Chi-square p value
Quantal Weibull	0.15	0.08	1	0.48
Quantal quadratic	0.16	0.034	--	0.34

Both models fit the data adequately as shown in Table 3-7. In both cases the chi-squared goodness of fit yields P-values greater than 0.05. Therefore, either model can be used for derivation of BMD. Neither model, as fitted to this data set, suggests a threshold for this response. However, both models do indicate a background rate in the absence of exposure to acrylamide.

3.6.4.4 Selection of the BMR

For the data set discussed above, the BMDs were calculated using the quantal Weibull and the quantal quadratic models for 1, 5, and 10 percent extra risk (Table 3-8 estimates are in units of mg/kg-day):

Table 3-8. BMD Values Calculated Using Quantal Weibull and Quadratic Models

Model	BMR	BMD (mg/kg-day) for Confidence Limit:		
		90th percentile	95th percentile	99th percentile
Weibull	10	0.73	0.64	0.52
	5	0.35	0.31	0.25
	1	0.07	0.06	0.05
Quadratic	10	1.28	1.19	1.06
	5	0.89	0.83	0.74
	1	0.39	0.37	0.33

The calculated BMDs are about a factor of two apart for the BMD_{10} values, but are about a factor of six apart for the BMD_1. This demonstrates the model dependence of the BMD values when low BMR levels are selected.

3.6.4.5 Calculating the Confidence Interval

As shown in Table 3-8, the BMDs were calculated for 90th, 95th, and 99th percentile confidence limits. The effect of the confidence limit on the estimated BMD was slightly less for the quantal quadratic than for the quantal Weibull. Model results were most comparable for the 90th and 95th percentile confidence limits and least comparable for the 99th percentile confidence limits. These results demonstrate that the BMD tends to be more model-dependent for wider (higher percentile) confidence intervals. For the remainder of the example, the 95th percentile confidence limit estimate is used.

3.6.4.6 Selection of the BMD as the Basis for the RfD

The example above yields different 95th percentile BMD_{10} values based on the two models. Since there is no basis upon which to eliminate one of the BMDs (i.e., goodness of fit, statistical assumptions and biological considerations), both must be considered. Either the smaller estimate or a geometric average may be used. In this case, the selection of which BMD to use is a risk management decision. In the example, the lower of the two BMDs (0.64) was chosen for the RfD calculation since it is the more conservative value.

3.6.4.7 Use of Uncertainty Factors with BMD Approach

Once the BMD is chosen, the RfD is derived by dividing the BMD by UFs. The same UFs applied to a NOAEL are used. In this case, a factor of 10 was selected for interspecies extrapolation and a factor of 10 for human interspecies variability. Using a total UF of 100 and applying it to the 95th percentile confidence limit BMD for 10 percent response (derived with the quantal Weibull model) yields an RfD of 0.006 mg/kg-day.

3.7 CATEGORICAL REGRESSION

3.7.1 Summary of the Method

Categorical regression is another method under investigation to estimate risks associated with systemic toxicity (Dourson et al., 1997; Guth et al., 1997). In this approach, health effects are grouped into ordered severity categories (ranging from no effect to severe effect). This simplification allows for both quantal and continuous data to be utilized, as well as data that are reported qualitatively rather than quantitatively. Furthermore, information on many health effects can be considered together. Logistic regression analysis techniques are then applied to the data: the cumulative odds of falling into severity categories is the dependent variable, and the independent variables are exposure concentration, exposure duration, and other parameters. Using the regression results, the RfD is then specified as the dose at which the probability of adverse effects is sufficiently small at some level of confidence, modified, as in the NOAEL and

BMD approaches, by appropriate uncertainty factors. For example, the dose of interest, D, might be defined as that dose for which one could conclude with 95 percent certainty that the probability of an adverse effect was less than 0.01. The value D would then be adjusted by uncertainty factors to derive the RfD.[16]

3.7.2   Steps in Applying Categorical Regression

The categorical regression approach begins with a review of the toxicological database for the chemical. For each valid study, the responses observed are assigned to one of several ordered severity categories, based on biological and statistical considerations. For example, responses may be grouped into four categories: (1) no effect; (2) no adverse effect; (3) mild-to-moderate adverse effect; and (4) severe or lethal effect. These correspond to the dose categories used in setting the RfD, namely the No Observed Effect Level (NOEL), NOAEL, LOAEL, and Frank Effect Level (FEL), respectively. Judgment is required to define the types of effects that correspond to the severity categories.

Since all response data are used in categorical regression analysis, there is no need to specify the lowest dose showing "mild-to-moderate" adverse effects. Accordingly, a more general term, adverse-effect level (AEL), is generally used in categorical regression in place of the term LOAEL to describe mild-to-moderate effects.

The probability of observing a response in a category at a given dose level is estimated by dividing the number of responses observed in that severity category divided by the total number of observations recorded for that dose level. Sufficient numbers of dose groups in each of several categories are required for the categorical regression.

The log odds for each dose and severity level is calculated, and then regressed against dose. The resulting regression equation can be used to calculate the probability of an effect of given severity for any dose.

Several model structures (logistic, Weibull, or others) may be used to perform the categorical regression. Logistic regression on the ordered categories (Harrell, 1986; Hertzberg, 1989) allows the dependent variable (e.g., severity parameter) to be categorical and the independent variables to be either categorical or continuous.

The goodness of the fit of the model to the data can be judged using several statistical measures: the overall P^2 value; model parameter standard errors and their P^2 significance levels;

---

[16]Note that the logistic regression could be used to estimate the response to exposures greater than the RfD. BMD models could be used similarly, but caution is warranted when doing so in either case.

concordance statistics and correlation coefficients for the overall model; and the model covariates (Hertzberg and Wymer, 1991). A variety of criteria for judging goodness of fit are currently being investigated.

Some advantages of using the categorical regression to derive the RfD are that data on more than one health effect can be incorporated and likely responses to exposures above the RfD can be evaluated. Predictions for responses above the RfD can incorporate effects other than the critical effect, a limitation for both the NOAEL/UF and BMD approaches.

3.8   CHRONIC, PRACTICAL NONTHRESHOLD EFFECTS

Noncarcinogenic effects are generally assumed to exhibit a threshold below which adverse effects are unlikely to occur. There are, however, exceptions to this general rule. Of particular concern are teratogenic and reproductive toxicants that may act through genetic mechanisms. EPA has recognized the potential for genotoxic teratogens and germline mutagens and discussed this issue in the 1991 *Amendments to Agency Guidelines for Health Assessments of Suspect Developmental Toxicants* (USEPA, 1991a) and in the 1986 *Guidelines for Mutagenicity Risk Assessment* (USEPA, 1986). These risk assessment guidelines raise concern for the potential for future generations inheriting chemically induced germline mutations or suffering from mutational events occurring *in utero*. At this time, genotoxic teratogens and germline mutagens should be considered an exception to the threshold assumption. In the absence of adequate data to support a genetic or mutational basis for developmental or reproductive effects, the default becomes an threshold approach. For such chemicals, this guidance recommends the procedures described above for noncarcinogens assumed to have a threshold. A nonthreshold approach should only be applied when there are substantial scientific data supportive of a non-threshold mechanism of toxicity, as is the situation for the neuro-developmental effects of lead. Ideally, a proposed mode of action would be available and would support the no-threshold hypothesis.

Where evidence for a genetic or mutational basis does exist, a nonthreshold mechanism shall be assumed for genotoxic teratogens and germline mutagens. Since there is no well established mechanism for calculating criteria protective of human health from the effects of these agents, criteria will be established on a case-by-case basis.

3.9   ACUTE, SHORT-TERM EFFECTS

States may choose to derive criteria that correspond to acute or short-term exposures. These criteria should correspond to a level of exposure that is "without appreciable risk of deleterious effects during some relatively short period of time" (USEPA, 1991c). The derivation of such values follows the same general approaches described above for criteria based on chronic effects. The primary difference lies in the type of toxicity data used as the basis for the

evaluation. Generally, studies that mimic the exposure pattern and duration of interest will be considered more relevant to the development of acute or short-term criteria. This is especially important where acute or short-term effects are of a substantially different nature than low-level chronic effects.

The Office of Water has established procedures for deriving Health Advisories (HAs) for one day, ten days, and longer-term. In general, HAs are developed by using NOAELs or LOAELs from studies with similar duration to the exposure period of concern, though there is some flexibility in this regard. Studies used for HAs should provide information on the critical endpoint. Studies that identify only frank toxic responses should not be used since these levels are far above the protective level targeted by HAs. More information on the derivation of HAs is given in Ware (1988).

Data from short-term studies should not be used when determing the longer-term or lifetime HAs. In instances where the database is inadequate to support a longer-term or lifetime HA no value is calculated. The Agency does not use data from less-than-90-day studies purely because they are the only available data. Factors such as the toxicokinetics, potential recovery periods, and potential for bioaccumulation should be considered in judging the relevance of the data to the HA derivation.

3.10   MIXTURES

Exposures to multiple contaminants may occur simultaneously. Possible interactions among chemicals in a mixture are usually placed in one of three categories:

•   Antagonistic, where the chemical mixture exhibits less toxicity than is suggested by the sum of the toxic effects of the components.

•   Synergistic, where the chemical mixture exhibits greater toxicity than is suggested by the sum of the toxic effects of the components.

•   Additive, where the toxicity of the chemical mixture is equal to the sum of the toxicities of the components.

Approaches to conducting a risk assessment for a mixture are presented in the 1999 draft *Guidelines for Health Risk Assessment of Chemical Mixtures* (USEPA, 1999). In only a few instances have the interactive effects of chemical mixtures been specifically studied. Where data on the effects of chemical mixtures exist, they should be used to characterize risk. Using the available data is especially important in cases where the resulting toxic effect from the mixture HAs been demonstrated to be greater than the sum of the individual effects. Certain categories of

contaminants, in particular, persistent organic pollutants that share a common mode of action and/or target tissue, are of elevated concern when they co-occur in the fish and drinking water.

Where specific data are not available on the interactive effects of particular chemical mixtures or on similar mixtures, the methods described below can be used by states to characterize risks from chemicals in a mixture. When risks from multiple chemicals are added, the quality of experimental evidence that supports the assumption of dose addition should be stated clearly (USEPA, 1999) and the approach should only be applied when data on the same or a similar mixture are not available.

In cases where the chemicals in the mixture induce the same effect by similar modes of action, contaminants may be assumed to contribute additively to risk (USEPA, 1999), unless specific data indicate otherwise. To characterize risks from multiple chemical exposure to noncarcinogens, the dose for each chemical with a similar effect first is expressed as a fraction of its RfD. These ratios are added for all chemicals to obtain the chemical mixture hazard index:

$$HI_{mix} = \sum_{m=1}^{n} \frac{E_m}{RfD_m}$$

(Equation 3.3)

where:

HI_{mix} = hazard index of the mixture (unitless)
E_m = the exposure to chemical m
RfD_m = the reference dose for chemical m
n = the number of chemicals in the mixture.

A hazard index greater than one implies an increased risk for non-carcinogenic effects from the mixture. However, the numerical value of the hazard index does not indicate the magnitude and severity of the risk (USEPA, 1999). Mode of action is an important consideration. Two chemicals with the same target tissue but totally different modes of action may or may not increase risk in an additive fashion.

Some chemical mixtures may contain chemicals that cause dissimilar health effects. Methods currently do not exist for combining dissimilar health effects to characterize overall health concerns from chemical mixtures. Instead, States should characterize and present the risks from these contaminants separately.

3.11 REFERENCES

Allen, B.C., R.J. Kavlock, C.A. Kimmel and E.M. Faustman. 1994a. Dose response assessments for developmental toxicity: II. Comparison of generic benchmark dose estimates with NOAELs. *Fundam. Appl. Toxicol.* 23:487-495.

Allen, B.C., R.J. Kavlock, C.A. Kimmel and E.M. Faustman. 1994b. Dose response assessments for developmental toxicity: III. Statistical models. *Fundam. Appl. Toxicol.* 23:496-509.

Andersen M., H. Clewell, M. Gargas, F.A. Smith, and R.H. Ritz. 1987. Physiologically based pharmacokinetics and risk assessment process for methylene chloride. *Toxicol. Appl. Pharmacol.* 87:185-205.

Andersen, M., J. Mills, M. Gargas, L. Kedderis, L. Birnbaum, D. Neubert, and W. Greenlee. 1993. Modeling receptor-mediated process with dioxin: Implications for pharmacokinetics and risk assessment. *Risk Analysis* 13:25-26.

Barnes, D.G., and M. Dourson. 1988. Reference Dose (RfD): Description and use in health risk assessments. *Regul. Toxicol. Pharmacol.* 8:471-486.

Brown, K.G. and L.S. Erdreich. 1989. Statistical uncertainty in the no-observed-adverse-effect level. *Fund. Appl. Toxicol.* 13(2): 235-244.

Calabrese, E. 1985. Uncertainty factors and interindividual variation. *Regul. Toxicol. Pharmacol.* 5:190-196.

Crump, K.S. 1984. A new method for determining allowable daily intakes. *Fund. Appl. Toxicol.* 4:854-871.

Crump, K. 1995. Calculation of benchmark doses from continuous data. *Risk Analysis* 15:79-89.

Crump, K.S., B. Allen, and E. Faustman. 1995. *The Use of the Benchmark Dose Approach in Health Risk Assessment*. Prepared for USEPA Risk Assessment Forum. EPA/630/R-94/007.

Dourson, M.L. and J. Stara. 1983. Regulatory history and experimental support of uncertainty (safety) factors. *Regul. Toxicol. Pharmacol.* 3:224-239.

Dourson, M.L., R.C. Hertzberg, R. Hartung and K. Blackburn. 1985. Novel approaches for the estimation of acceptable daily intake. *Toxicol. Ind. Health* 1:23-41.

Dourson, M.L., L.A. Knauf, and J.C. Swartout. 1992. On reference dose (RfD) and its underlying toxicity database. *Toxicol. Ind. Health* 8(3):171-189.

Dourson, M.L., L.K. Teuschler, P.R. Durkin, and W.M. Stiteler. 1997. Categorical regression of toxicity data, a case study using Aldicarb. *Regulatory Toxicity and Pharmacology* 25:121-129.

Faustman, E.M., B.C. Allen, R.J. Kavlock, and C.A. Kimmel. 1994. Dose response assessment for developmental toxicity: I. Characterization of database and determination of NOAELs. *Fundam. Appl. Toxicol*. 23:478-486.

Gaylor, D.W. 1983. The use of safety factors for controlling risk. *J. Toxicol. Environ. Health* 11:329-336.

Gaylor, D.W. 1989. Quantitative risk analysis for quantal reproductive and developmental effects. *Environ. Health Perspect*. 79:243-246.

Gaylor, D.W. and W. Slikker, Jr. 1990. Risk assessment for neurotoxic effects. *Neurotoxicology* 11:211-218.

Gehring, P.J., P.J. Watanabe, and C.N. Park. 1978. Resolution of dose-response toxicity data for chemicals requiring metabolic activation: Example -- vinyl chloride. *Toxicol. Appl. Pharmacol.* 44:581-591.

Guth, D.J., R.J. Carroll, D.G. Simpson, and H. Zhou. 1997. Categorical regression analysis of acute exposure to tetrachloroethylene. *Risk Analysis* 17(3):321-332.

Harrell, F. 1986. The logist procedure. *SUGI Supplemental Library Users Guide,* Ver. 5th. Ed. SAS Institute. Cary, NC.

Hartley, W.R. and E.V. Ohanian. 1988. The use of short-term toxicity data for prediction of long-term health effects. In: *Trace Substances in Environmental Health* - XXII. D.D. Hemphil, (ed). University of Missouri. May 23-26. Pp. 3-12.

Hartung, R. and P.R. Durkin. 1986. Ranking the severity of toxic effects: Potential applications to risk assessment. *Comments on Toxicology* 1:49-63.

Hattis, D., L. Erdreich and M. Ballew. 1987. Human variability in susceptibility to toxic chemicals -- A preliminary analysis of pharmacokinetics data from normal volunteers. *Risk Analysis* 7(4):415-426.

Hertzberg, R.C. 1989. Fitting a model to categorical response data with applications to species extrapolation of toxicity. *Health Physics* 57: 405-409.

Hertzberg, R.C. and M.E. Miller. 1985. A statistical model for species extrapolation using categorical response data. *Toxicol. Ind. Health* 1(4):43-63.

Hertzberg, R.C. and L. Wymer. 1991. Modeling the severity of toxic effects. Presentation at the *84th Annual Meeting of the Air and Waste Management Association*. June 16-21, 1991.

ILSI (International Life Sciences Institute). 1993. *Report on the Benchmark Dose Workshop*. International Life Sciences Institute, Risk Science Institute. Washington, DC.

Johnson, K.A., S.J. Gorzinski, K.M. Bodner, R.A. Campbell, C.H. Wolf, M.A. Friedman, and R.W. Mast. 1986. Chronic toxicity and oncogenicity study on acrylamide incorporated in the drinking water of Fischer 344 rats. *Toxicol. Appl. Pharmacol.* 85:154-168.

Kavlock, R.J., B.C. Allen, E.M. Faustman, and C.A. Kimmel. 1995. Dose response assessment for developmental toxicity: IV. Benchmark doses for fetal weight changes. *Fundam. Appl. Toxicol.* 26:211-222.

Kimmel, C.A. 1990. Quantitative approaches to human risk assessment for noncancer health effects. *Neurotoxicology* 11:189-198.

Lewis, S.C., J.R. Lynch, and A.I. Nikiforov. 1990. A new approach for deriving community exposure guidelines from no-observed-adverse-effect levels. *Reg. Toxicol. Pharmacol.* 11:314-330.

Swartout. 1990. Personal Communication to M.L. Dourson of the Office of Technology Transfer and Regulatory Support on January 12 . Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1986. Guidelines for mutagenicity risk assessment. *Federal Register* 51:34006-34012.

USEPA (U.S. Environmental Protection Agency). 1988. Reference dose (RfD): Description and use in health risk assessments. *Integrated Risk Information System* (IRIS). Online. Intra-

Agency Reference Dose (RfD) Work Group. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office. Cincinnati, OH. February.

USEPA (U.S. Environmental Protection Agency). 1989. *Interim Methods for Development of Inhalation Reference Doses*. Office of Health and Environmental Assessment. Washington, DC.   EPA/600/8-88-066F.

USEPA (U.S. Environmental Protection Agency). 1991a. Amendments to agency guidelines for health assessments of suspect developmental toxicants. *Federal Register* 56:63798-63826. December 5.

USEPA (U.S. Environmental Protection Agency). 1991b. Final guidelines for developmental toxicity risk assessment. *Federal Register* 56:63798-63826. December 5.

USEPA (U.S. Environmental Protection Agency). 1991c. *General Quantitative Risk Assessment Guidelines for Noncancer Health Effects*. Second External Review Draft. Technical Panel for Development of Risk Assessment Guidelines for Noncancer Health Effects. Cincinnati, OH. ECAO CIN-538.

USEPA (U.S. Environmental Protection Agency). 1992. Reference dose (RfD) for oral exposure for inorganic zinc. *Integrated Risk Information System* (IRIS). Online. (Verification date 10/1/92). Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office. Cincinnati, OH.

USEPA (U.S. Environmental Protection Agency). 1993. Reference dose (RfD) for oral exposure for inorganic arsenic. *Integrated Risk Information System* (IRIS). Online. (Verification date 2/01/93). Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office. Cincinnati, OH.

USEPA (U.S. Environmental Protection Agency). 1994a. Reference dose (RfD) for oral exposure for methylmercury. *Integrated Risk Information System* (IRIS). Online. (Verification date 11/23/94). Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office. Cincinnati, OH.

USEPA (U.S. Environmental Protection Agency). 1994b. *Guidelines for Reproductive Toxicity Risk Assessment*. External Review Draft. Risk Assessment Forum. Washington, DC. EPA/600/AP-94/001. February.

USEPA (U.S. Environmental Protection Agency). 1995. *RQ Document for Solid Waste. Report on the Benchmark Dose Peer Consultation Workshop: Risk Assessment Forum*. Office of Research and Development. Washington, DC.   EPA/630/R-96/011. November.

USEPA (U.S. Environmental Protection Agency). 1999. *Guidelines for the Health Risk Assessment of Chemical Mixtures*. External Peer Review Draft. Risk Assessment Forum. Washington, DC. NCEA-C-0148. April.

Ware, G.W. (ed). 1988. Reviews of Environmental Contamination and Toxicology: U.S. Environmental Protection Agency Office of Drinking Water Health Advisories. Vol. 104. Springer-Verlag, Inc. New York, NY.

Zielhuis, R.L. and F.W. van der Kreek. 1979. The use of a safety factor in setting health based permissible levels for occupational exposure. *Int. Arch. Occup. Environ. Health* 42:191-201.

APPENDIX A

CASE STUDY EXAMPLE
HAZARD EVALUATION FOR COMPOUND Z

A.1   HUMAN DATA

Compound Z is a metal-conjugated phosphonate. No human tumor or toxicity data exist on this chemical.

A.2   ANIMAL DATA

Compound Z caused a statistically significant increase in the incidence of urinary bladder tumors in male, but not female, rats at 30,000 ppm (3%, 1500 mg/kg/day) in the diet in a long-term study. Some of these animals had accompanying urinary tract stones and toxicity. No bladder tumors or adverse urinary tract effects were seen in two lower dose groups (2,000 and 8,000 ppm equivalent to 100 and 400 mg/kg/day) in the same study. A chronic dietary study in mice at doses comparable to those in the rat study showed no tumor response or urinary tract effects. A 2-year study in dogs at doses up to 40,000 ppm showed no adverse urinary tract effects.

A.3   OTHER KEY DATA

Subchronic dosing of rats confirmed that there was profound development of stones in the male bladder at doses comparable to those causing cancer in the chronic study, but not at lower doses. Sloughing of the epithelium of the urinary tract accompanied the stones.

There was a lack of mutagenicity relevant to carcinogenicity. In addition, there is nothing about the chemical structure of Compound Z to indicate DNA reactivity or carcinogenicity.

Compound Z is composed of a metal, an ethanol, and a simple phosphorus-oxygen-containing component. The metal is not absorbed from the gut, whereas the other two components are absorbed. At high doses, ethanol is metabolized to carbon dioxide, which makes the urine more acidic; the phosphorus level in the blood and calcium in the urine are increased. Chronic testing of the phosphorus-oxygen-containing component alone in rats did not show any tumors or adverse effects on the urinary tract.

Because Compound Z is a metal complex, it is not likely to be readily absorbed from the skin.

A.4   EVALUATION

Compound Z produced cancer of the bladder and urinary tract toxicity in male rats, but not in female rats or mice, and dogs also failed to show the toxicity noted in male rats. The mode of action developed from the other key data to account for the toxicity and tumors in the male rats is the production of bladder stones. At high, but not lower, subchronic doses in the male rat, Compound Z leads to elevated blood phosphorus levels; the body responds by releasing excess calcium into the urine. The calcium and phosphorus combine in the urine and precipitate into multiple stones in the bladder. The stones are very irritating to the bladder; the bladder lining is eroded and cell proliferation occurs to compensate for the loss of the lining. Cell layers pile up, and finally, tumors develop. Stone formation does not involve the chemical per se but is secondary to the effects of its constituents on the blood and, ultimately, the urine. Bladder stones, regardless of their cause, commonly produce bladder tumors in rodents, especially the male rat.

A.5   CONCLUSION

Compound Z: "Likely/Not Likely Human Carcinogen" ! Range of Dose Limited, Margin-of-Exposure Extrapolation

Compound Z, a metal aliphatic phosphonate, is *likely to be carcinogenic to humans* only under high-exposure conditions following *oral and inhalation exposure* that lead to bladder stone formation, but is *not likely* to be carcinogenic under low-exposure conditions. It is *not likely to be a human carcinogen* via the *dermal* route, given that the compound is a metal conjugate that is readily ionized and its dermal absorption is not anticipated. The weight of evidence is based on (a) bladder tumors only in male rats; (b) the absence of tumors at any other site in rats or mice; (c) the formation of calcium-phosphorus-containing bladder stones in male rats at high, but not low, exposures that erode bladder epithelium and result in profound increases in cell proliferation and cancer; and (d) the absence of structural alerts or mutagenic activity.

There is a strong mode-of-action basis for the requirements of (a) high doses of Compound Z, (b) which lead to excess calcium and increased acidity in the urine, (c) which result in the precipitation of stones, and (d) the necessity of stones for toxic effects and tumor hazard potential. Lower doses fail to perturb urinary constituents, lead to stones, produce toxicity, or give rise to tumors. Therefore, dose-response assessment should assume nonlinearity.

A major uncertainty is whether the profound effects of Compound Z may be unique to the rat. Even if Compound Z produced stones in humans, there is only limited evidence that humans with bladder stones develop cancer. Most often human bladder stones are either passed in the urine or lead to symptoms resulting in their removal. However, since one cannot totally dismiss

the male rat findings, some hazard potential may exist in humans following intense exposures. Additional research would be needed to reduce this uncertainty.

APPENDIX B

CASE STUDY EXAMPLE
MODE OF ACTION EVALUATION: COMPOUND Z (BLADDER TUMOR)

B.1   HAZARD DATA SUMMARY

B.1.1 Data Availability

Data include a rat chronic/carcinogenicity feeding study, an 18 month CD-1 mouse carcinogenicity study, a three-generation reproduction study in the rat, and a 2-year feeding study in dogs. There are no data on the effects in humans of exposure to compound Z.

A 13-week feeding study in rats included interim sacrifices at 2, 4, and 8 weeks and establishment of 16-week recovery groups at 8 weeks and a 21-week recovery group at 13 weeks.

B.1.2 Tumor Observations

B.1.2.1 <u>Tumor Response</u>

Rats. Administration of compound Z in the diet to male Sprague-Dawley rats at dose levels of 30,000 ppm or more for 2 years resulted in an increase in bladder urothelial tumors in male rats. Statistically significant increases ($p<0.05$) were noted at the high dose only (40,000/30,000 ppm) in the incidences of transitional cell papillomas, carcinomas, combined papillomas and carcinomas, and hyperplasia in the 2-year SD rat bioassay (Table B-1). Bladder calculi were observed in some animals but correlation between stones and tumors was not evident at final sacrifice.

Mice. No increase in tumor incidences was observed in an 18-month bioassay with mice.

Dogs. When administered to dogs at dose levels up to 40,000 ppm in the diet for up to 2 years, the compound produced no tumors.

B.1.3 Mutagenicity

Compound Z has not shown mutagenic activity in *Salmonella sp.* or micronucleus assays. No evidence exists that the chemical produces effects on DNA synthesis, nor does it appear to be clastogenic. There are no structural attributes that suggest mutagenic potential for the chemical.

Table B-1. Incidence of Transitional Cell Lesions and Stones in the Bladder of Males from a 2-Year Sprague-Dawley Rat Study

Parameter	Dose (ppm)			
	0	2000	8000	40,000/30,000
N	73	75	78	78
Lesion				
Papilloma	1	1	1	5
Carcinoma	2	2	1	16
Combined	3	3	2	21
Hyperplasia	5	7	5	29
Stones	0	0	0	5

B.1.4 Toxicity, Uroliths, and Hyperplasia

There was a strong association among disruptions in urinary physiology, toxicity, uroliths, and hyperplasia in the 13-week study in mid-dose and high-dose animals (30,000 and 50,000 ppm respectively; $p < 0.05$). In the control and 8,000 ppm group, no animals had stones and no animals had hyperplasia (see Table B-2).

B.1.4.1 Thirteen-Week Study

Urothelial toxicity and disruptions in urinary physiology and urothelial toxicity appeared early in the study. Early changes in urinary physiology (decreased pH and increased cation concentration) were observed following 2 weeks of treatment and persisted throughout the duration of the study. Urothelial toxicity was expressed as edema, cystitis, and hyperplasia; hyperplasia (simple and papillary transitional cell combined) increased in overall incidence with continued treatment. It was present in 70% of mid-dose (30,000 ppm) animals and 80% of high-dose (50,000 ppm) animals following 2 weeks of exposure, and in 70% of the mid-dose group and 100% of the high-dose group at 13 weeks. There was some indication of a decrease in

Table B-2. Incidence of Bladder Hyperplasia and Stones in
Male Sprague-Dawley Rats Treated up to 13 Weeks

Parameter	2 weeks				8 weeks				13 weeks			
Dose[a]	1	2	3	4	1	2	3	4	1	2	3	4
N	10	10	10	10	10	10	10	9	10	10	10	6
Papillary hyperplasia	0	0	7	8	0	0	9	7	0	0	5	6
Simple hyperplasia									0	0	2	0
Stones	0	0	3	4	0	0	9	8	0	0	7	6

[a]Dose (ppm): 1 = control, 2 = 8000, 3 = 30,000, 4 = 50,000.

severity of hyperplasia at 13 weeks when compared with earlier time periods, as there was an apparent shift from the incidence of papillary hyperplasia to simple hyperplasia and a decrease in the combined incidence of hyperplasia in the 30,000 ppm group of animals.

Uroliths were found to be present as early as 2 weeks (0%, 0%, 30%, and 40% in the four dose groups, respectively) and the incidence increased over the period of the study. The incidence of uroliths at termination of the 13-week study was 0%, 0%, 70%, and 100% in the four dose groups, respectively, but there was a decrease in size and number of stones per animal at 13 weeks.

B.1.4.2 Three-Generation Reproduction Study in Rats

High dose levels (>20,000 ppm in the diet) led to formation of lesions in the urinary tract of males and females of the F1, F2, and F3 generations. The lesions included hemorrhage of the bladder wall, increased pelvic dilation, and papillary necrosis. In the F3 generation, additional effects noted in renal tissue were hyperplasia of the transitional epithelium and desquamation of cells in the lumen of the urinary tract. The changes were associated with crystalline or calcareous deposits.

B.1.5 Reversibility of Effects

There was strong evidence of reversibility of bladder stones and bladder hyperplasia. When animals that had been treated for 8 weeks were returned to basal diet for 16 weeks, uroliths were found in 30% of 30,000 ppm animals and 25% of high-dose animals. Bladder hyperplasia (papillary and transitional cell combined) was reduced to 25% and 30% in each of these two dose

groups (Table B-3). An analysis of individual animal data revealed a strong correlation between the incidence of uroliths and hyperplasia at the termination of the recovery period.

Table B-3. Reversal of incidence of bladder hyperplasia
and stones following 8 weeks treatment and 16 weeks recovery

Parameter	Dose (ppm)			
	0	**8000**	**30,000**	**50,000**
N	10	10	10	8
Papillary hyperplasia	0	0	2	1
Simple hyperplasia	0	0	1	1
Stones	0	0	3	2

B.1.6 Blood and Urine Chemistry

Compound Z administration resulted in increases in blood phosphorus and carbon dioxide (data not shown). Urinalyses (Table B-4) showed elevated calcium levels, reduced urinary phosphorus, and a profound lowering of urinary pH (5.0), which began at 2 weeks and persisted throughout the 13-week study in the 30,000 and 50,000 ppm group of rats. These changes occurred in the presence of bladder stones, which were reported to consist of 33% calcium and 23% phosphorus.

Table B-4. Clinical Chemistry Values (Urine) in Male Sprague-Dawley Rats
Treated up to 13 weeks

Parameter	2 weeks				8 weeks				13 weeks			
Dose	1	2	3	4	1	2	3	4	1	2	3	4
N	10	10	10	10	10	10	10	9	10	10	10	6
Calcium - g/dL	6	11	56[b]	36[c]	11	11	18	65[b]	5	7	14[b]	58[b]
Phosphorus - mg/dL	90	62	2[b]	13[c]	109	90	19	1[b]	57	67	26	1[b]
pH	7	6.5	5[b]	5[b]	7.4	6.9	5.8[b]	5.0[b]	7.2	6.7	6.0[b]	5.0[b]
Stones	0	0	3	4	0	0	9	8	0	0	7	6

[a]Dose (ppm): 1 = control, 2 = 8000, 3 = 30,000, 4 = 50,000.
[b]$p<0.01$; [c]$p<0.05$

B.1.7 Metabolism

Upon ingestion by rats, the ethyl moiety of compound Z is rapidly absorbed, hydrolyzed to a phosphite, and oxidized via acetaldehyde and acetate to carbon dioxide and water. Absorption of the phosphite moiety leads to increased blood phosphorus levels. There is also an increase in blood calcium load, which leads to increased excretion of calcium via the urine. Ethyl phosphite moieties and carbon dioxide are also eliminated via the urine. A marked depression of urinary pH (5.0) results from acidification of the urine by carbon dioxide. An aluminum moiety of the parent chemical is poorly absorbed, and most is eliminated in the feces. The phosphite metabolite, the major urinary metabolite, was not shown to express carcinogenic potential when administered to Sprague-Dawley rats at dose levels up to 32,000 ppm. It also does not express any mutagenic potential and does not have any structural alerts.

B.1.8 Structure-Activity Relationships

There are no data on structurally related chemicals.

B.2 MODE OF ACTION ANALYSIS

B.2.1 Summary Description of Postulated Mode of Action

Compound Z produces transitional cell tumors in male Sprague-Dawley rats. The mode of action includes disruption in urinary physiology, including precipitation of calcium and phosphorus and formation of bladder calculi. The stones irritate the urothelium of the bladder, followed by transitional cell hyperplasia and bladder tumor formation. Disruption of urinary physiology is a consequence of a metabolic sequence involving (1) absorption and metabolism of the ethyl moiety to carbon dioxide, resulting in a reduction in urinary pH; and (2) absorbtion of the phosphite moiety, which leads to increased blood phosphorus levels and increased release of calcium into the urine. Increases in water consumption followed by increased urinary volume may contribute to bladder toxicity, but a precise role of increased urinary volume has not been established.

The mode of action for compound Z is consistent with other data that demonstrate that solid masses in the rodent bladder, regardless of their origin **!** insertion of solid materials, including inert pellets, precipitation of administered chemicals (e.g., melamine) or disruption of urinary physiology (e.g., diethylene glycol) **!** lead to urothelial toxicity and the formation of tumors.

B.2.2 Key Events

The key precursor events associated with bladder tumor formation following administration of compound Z to rats include increased blood phosphorus and carbon dioxide, elevated urinary calcium and volume, decreased urinary pH and phosphorus, formation of bladder stones, and irritation and hyperplasia of the urothelium.

B.2.3 Strength, Consistency, and Specificity of Association of Tumor Response with Key Events

The only tumor response seen in animal studies is bladder tumors in male Sprague-Dawley rats. Studies in dogs and mice showed no effect on the bladder. The rat tumor response was seen only at high doses that lead to key precursor effects: altered urinary physiology (volume, calcium, pH) results in stones and produces toxicity and hyperplasia of the urothelium. The high-dose changes were noted in a rat chronic, a rat subchronic, and a three-generation reproduction study in rats. The key events, including hyperplasia, were observed to be reversible in subchronic stop/recovery studies. Administration of the major metabolite of compound Z, monosodium phosphite, fails to reduce urinary pH, increase urinary volume, or produce nonneoplastic or neoplastic lesions of the bladder. The database on compound Z is sufficient to evaluate the proposed mode of action despite the absence of more complete information on the composition of the stones and questions regarding the absence of toxicity following the administration of monosodium phosphite. There is a high degree of confidence that the findings accurately reflect the effects associated with administration of the chemical. No data gaps were identified that would substantially alter the evaluation of the proposed mode of action.

B.2.4 Dose-Response Relationships

The 2-year bioassay showed urothelial hyperplasia, transitional cell papillomas, and transitional cell carcinomas and a few bladder stones at 40,000/30,000 ppm. Of 78 high-dose animals, 37% showed bladder tumors. Tumors, hyperplasia, and stones were not increased at 8000 ppm. A special 13-week feeding study demonstrated that key events ! increased urinary calcium levels, decreased urinary phosphorus levels, decreased pH, bladder stones, irritation, edema, and hyperplasia ! occurred consistently only at dose levels of 30,000 ppm or greater. A strong dose-response correlation was shown between calculus formation and hypercalciuria, acidic urine, and bladder hyperplasia. In a rat reproduction study, bladder effects were noted at 24,000 ppm but not at 12,000 ppm.

B.2.5 Temporal Association

A subchronic rat study with serial sacrifices at 2, 4, 8, and 13 weeks, including evaluation of 16-week recovery groups after 8 weeks and a 21-week recovery group after 13 weeks, was

performed. By 2 weeks of administration, compound Z produced stones that filled the bladder and resulted in advanced papillary hyperplasia. The number and size of stones was greatest at two weeks and there was a progressive decrease over the 13 week period. Early changes in urinary physiology (decreased urinary pH, increased calcium concentration, and decreased phosphorus concentration) were observed following 2 weeks of treatment and persisted throughout the duration of the study. Observation of the 8-week treatment/16-week recovery groups showed that incidence of both stones and hyperplasia significantly decreased as compared with incidence in animals sacrificed at 8 weeks. Also, upon cessation of dosing at 13 weeks, the incidence of animals with stones, the incidence of papillary hyperplasia, and the severity of hyperplasia decreased significantly by the end of a 21-week recovery period (data not shown). The changes noted within 2 weeks of dosing appear to have set in motion a series of events beginning with increased urinary calcium concentrations, followed or accompanied by stone formation, irritation of the bladder urothelium, hyperplasia and, eventually, neoplasia.

B.2.6 Biological Plausibility and Coherence of the Database

Long-term and subchronic studies with compound Z have demonstrated a dose correlation between development of stones and bladder tumor formation in male rats. Data from the 13-week study indicate a rapid onset of effects (changes in urinary parameters, formation of stones, and hyperplasia within 2 weeks of dosing) and adaptation of treated animals to compound Z exposure by 13 weeks (decreased numbers and size of stones per animal, decreased severity of hyperplasia). Tumors were observed only at doses at which key events were observed.

Additional bioassay data provide support for the association of tumors in rats with the key events in rats and the absence of both tumors and similar key events in other species treated with compound Z. Treatment of rats in a three-generation reproduction study at high dose levels (>20,000 ppm in the diet) led to formation of lesions in the urinary tract of males and females. When administered to dogs at dose levels up to 40,000 ppm in the diet for up to 2 years, the chemical produced minimal toxic effects overall, no effects on the urinary tract, and no tumors. Compound Z produced no effects in mice when administered up to a dose level of 20,000/30,000 ppm in the diet for 2 years.

Observations with compound Z are in keeping with those observed in many other experimental settings. Stones, regardless of their chemical makeup, are irritating to the rodent bladder, causing irritation, hyperplasia, and eventually neoplasia.

There are some uncertainties regarding the role of certain findings following compound Z administration. Generally, an increase in urinary pH is associated with the precipitation of calcium and phosphorus-containing stones in rats. However, stones are formed in the presence of a low urinary pH in rats administered compound Z. It is also unclear whether or not the acidic environment of the urine (most likely a consequence of the conversion of the ethyl moiety to carbon dioxide in the blood) contributes to or enhances any effects noted in bladder tissue in rats.

There was a paucity of stones in high-dose animals at termination of the 2-year study but a higher incidence of bladder tumors, which suggests that bladder stones may not be the causative factor involved in bladder tumor formation. Other considerations discount this presumption. First, a number of the high-dose animals showed hydronephrosis or dilation of the ureters, presumptive indications of past urinary tract obstruction. Second, the 13-week study provided evidence that bladder calculi develop rapidly (within 2 weeks), but then decreased in frequency and size. The decrease in size and number of bladder calculi was accompanied by a decrease in severity of bladder hyperplasia in animals treated with 30,000 ppm of compound Z. Third, it is recognized that a constant ppm of an agent in the diet results in a reduction in dose per unit body weight as an animal grows. Finally, the increased urinary volume or decreased urinary pH may have led to a dissolution of stones over time.

The absence of bladder stones and urothelial toxicity following administration of the major metabolite, monosodium phosphite, is puzzling, as one might expect administration to rats would lead to similar bladder effects as with compound Z. However, the metabolite when administered to rats, leads to an increase in blood levels of phosphorus but does not alter urinary volume or pH as would be expected with an increase in sodium consumption. Considering the high dose-level of metabolite administered to rats (32,000 ppm), it is unlikely an additional bioassay using higher dose-levels would provide useful information.

B.2.7 Other Modes of Action

Compound Z is not mutagenic in short-term tests and it does not have a structure suggesting biological reactivity. No other modes of action, apart from that postulated, are in evidence. The fact that bladder tumors were the sole tumors seen in rats and that no other species showed tumors or other toxicities like those in the rat make it less likely that the agent has another generalized mode of action.

B.2.8 Conclusion

The available bioassay data on compound Z are sufficient to support the postulated mode of action that the chemical, which lacks mutagenic potential, leads to bladder tumor formation in male rats through a sequence of key events involving perturbations in urinary physiology, especially increased calcium concentration, calculus formation, urothelial irritation, hyperplasia, and neoplasia.

B.3   RELEVANCE OF THE MODE OF ACTION TO HUMANS

Bacterial infection, urinary stones or a combination of the two may be risk factors for human urinary tract cancer (Burin et al., 1995; Davis et al., 1984; Gonzalez et al., 1991; Kawai et al., 1994; Hiatt et al., 1982). Infection of the bladder with *Schistosoma haematobium* leads to

bladder tumors, and part of its action may be associated with stone formation (IARC, 1994). A significant relationship has also been shown between spinal cord injury and bladder cancer; chronic infection and stones are found in individuals so affected (Bickel et al., 1991; Broecker et al., 1981; Dolin et al., 1994; El-Marsi and Fellows, 1981; Stonehill et al., 1996). Case control epidemiologic studies (relative risks less than three) suggest associations between bladder cancer and urinary tract stones (Burin et al., 1995; Gonzalez et al. 1991). A large cohort study supports the association shown between bladder stones and bladder cancer (Chow et al., 1997). Taken as a whole, stones may play some role (particularly, along with infection) in bladder cancer formation. Bladder cancer is a disease of advancing age, with about 2/3 of all cases occurring among persons aged 65 years or older (Hankey et al., 1993).

Stones occur much more frequently in the upper urinary tract than in the bladder of humans (about 10% of urinary stones are found in the bladder), presumably because the upright posture of humans predisposes them to expelling stones through the urethra once a stone passes from the kidney to the bladder (Hiatt et al., 1982; Johnson et al. 1979; DeSesso, 1995). This characteristic, as well as the pain which accompanies such stones and leads to their surgical removal. Stones in the rodent bladder tend to be retained, because of their horizontal position. These findings suggest that there may be a lower susceptibility of humans compared to rodents to the development of urinary tract tumors associated with stones.

Precipitation of chemicals in the urinary tract with the formation of stones is a common finding, with about 12% of males and 5% of females having a history over a lifetime of at least one stone (Johnson et al., 1979). Compared to adults, urinary stone formation in children is an uncommon occurrence except in individuals with a predisposing condition, such as, various inborn errors of metabolism (e.g., cystinuria) and congenital malformations (Gearhart et al., 1991). The prevalence of urinary stones in children is about 1 case per 20,000 per year (0.005%) (Khoory et al., 1998). Only about 5% of stones are initially manifest during the first 20 years of life (Johnson et al., 1979). Causes of urinary stones in children are remarkably similar to those of adults (Khoory et al., 1998; Stapleton, 1996). Like with adults, the urine of children varies in pH and osmolality, particularly in response to diet and physiologic stressors (e.g., exercise, heat). Urinary excretion of chemicals occurs throughout life, although there may be quantitative differences associated with a number of factors including disease states and nutritional status. Stones used to be more common in children in developed countries than they are now, largely due to malnutrition, which is still a problem in developing nations today (Trinchieri, 1996).

Compound Z is converted to metabolic derivatives through simple hydrolysis, a chemical conversion that does not depend on enzymatic activity. It is not plausible that differences in levels of enzymatic activity, such as detoxification via hepatic metabolism or metabolism in other tissues will alter, qualitatively, responses in population subgroups such as the aged, the infirm, or infants and children who may be exposed to Compound Z.

In summary, the potential human carcinogenic hazard of the chemical cannot be dismissed for Compound Z. Compound Z poses a carcinogenic hazard to humans only under conditions that would lead to the formation of bladder stones. It is reasonable to conclude that the mode of action involving stone formation for Compound Z that has been developed for adult animals may be applicable to young animals and to children. Information suggests that effects in the young may not be any greater than in adults and, in fact, the young may be less susceptible unless there are rare extenuating factors.

B.4   REFERENCES

Bickel, A., Culkin, D.J., Wheeler, J.S. 1991. Bladder cancer in spinal cord injury patients. *J. Urol*. 146:1240-1242.

Broecker, B.H., Klein, F.A., Hackler, R.H. 1981. Cancer of the bladder in spinal cord injury patients. *J. Urol*. 125:196-7.

Burin, G.J., Gibb, H.J., Hill, R.N. 1995. Human bladder cancer: Evidence for a potential irritation-induced mechanism. Fd. Chem. Toxicol. 33:785-795.

Chow, W-H., Lindbald, P., Gridley, G., Nyren, O., McLaughlin, J.K., Linet, M.S., Pennello, G.A., Adami, H-O., Fraumeni, J.F. Jr. 1997. Risk of urinary tract cancers following kidney or ureter stones. *J. Natl. Cancer Inst*. 89: 1453-1457.

Davis, C.P., Cohen, M.S., Gruber, M.B., Anderson, M.D., Warren, M.M. 1984. Urothelial hyperplasia and neoplasia: A response to chronic urinary tract infections in rats. *J. Urol.* 132:1025-1031.

DeSesso, J.M. 1995. Anatomical relationships of urinary bladders compared: their potential role in the development of bladder tumours in humans and rats. *Food Chem Toxicol.* 33:705-714.

Dolin, P.J., Darby, S.C., Beral, V. 1994. Paraplegia and squamous cell carcinoma of the bladder in young women: findings from a case-control study. *Br. J. Cancer* 70:167-168.

El-Masri, W.S., Fellows, G. 1981. Bladder cancer after spinal cord injury. *Paraplegia* 19:265-70.

Gearhart, J.P., Herzberg, G.Z., Jeffs, R.D. 1991. Childhood urolithiasis: Experiences and advances. *Pediatrics* 87:445-450.

Gonzalez, C.A., Errezola, M., Izarzugaza, I., Lopez-Abente, G., Escolar, A., Nebot, M., Riboli, E. 1991. Urinary infection, renal lithiasis and bladder cancer in Spain. *Eur. J. Cancer* 27:498-500.

Hankey, B.F., Silverman, D.T., Kaplan, R. 1993. Urinary bladder. In Miller, B.A., Ries, L.A.G., Hankey, B.F. et al., eds. *SEER Cancer Statistics Review: 1973-1990.* NIH Pub. No. 93-789. Bethesda, MD: National Cancer Institute: XXXVI.1-17.

Hiatt, R.A., Dales, L.G., Friedman, G.D., Hunkeler, E.M. 1982. Frequency of urolithiasis in a prepaid medical program. *Amer. J. Epidemiol*. 115: 255-265.

IARC. 1994. Some Industrial Chemicals. In *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon, France, 60: 13-33.

Johnson, C.M., Wilson, D.M., O'Fallon, W.M., Malek, R.S., Kurland, L.T. 1979. Renal stone epidemiology: A 25-year study in Rochester, Minnesota. *Kid. Internat*. 16:624-631.

Kawai, K., Kawamata, H., Kemeyama, S., Rademaker, A., Oyasu, R. 1994. Persistence of carcinogen-altered cell population in rat urothelium which can be promoted to tumors by chronic inflammatory stimulus. *Cancer Res*. 54:2630-2632.

Khoory, B.J., Pedrolli, A., Vecchni, S., Benini, D., Fanos, V. 1998. Renal caculosis in pediatrics. *Pediatr. Med. Chir*. 20:367-376.

Stapelton, F.B. 1996. Clinical approach to children with urolithiasis. *Semin. Nephrol*. 16:389-397.

Stonehill, W.H., Dmochowski, R.R., Patterson, A.L., Cox, C.E. 1996. Risk factors for bladder tumors in spinal cord injury patients. *J. Urol*. 155:1248-1250.

Trinchieri, A. 1996. Epidemiology of urolithiasis. *Arch. Ital. Urol. Androl*. 68:203-249.

APPENDIX C

EVALUATION OF THE QUALITY OF DATA SET(S)
FOR USE IN DERIVING AN RfD

The derivation of RfDs begins with a thorough review and assessment of the toxicological database to identify the type and magnitude of possible adverse health effects associated with a chemical. This evaluation should include an examination of the full range of possible health effects, including acute, short-term (14 to 28 days), subchronic, reproductive/developmental, and chronic effects.

To be useful for supporting the derivation of an RfD, a study must meet certain standards with regard to experimental design, conduct and data reporting. This appendix provides general guidance on criteria for appropriate study design for a variety of types of toxicity studies. These guidelines provide the assessor with a means to evaluate the quality and adequacy of data. Appropriate studies are used both for the evaluation of potential hazard of the chemical and for the derivation of the RfD.

C.1   ACUTE TOXICITY DETERMINATION

Studies of acute exposure (one dose or multiple dose exposure occurring within a short time (e.g. less than 24 hours)) are widely available for many chemicals. Acute toxicity [often expressed in terms of the lethal dose (or concentration) to 50 percent of the population (LD_{50} or LC_{50})] is usually the initial step in experimental assessment and evaluation of a chemical's toxic characteristics. Such studies are used in establishing a dosage regimen in subchronic and other studies and may provide initial information on the mode of toxic action of a substance. Because LD_{50} or LC_{50} studies are of short duration, inexpensive and easy to conduct, they are commonly used in hazard classification systems.

Acute lethality studies are of limited use, however, in the derivation of chronic criteria, since the establishment of chronic criteria should never be based on exposures that approach acutely lethal levels. However, the data from such studies do provide information on health hazards likely to arise from individual short-term exposures. Such studies provide high dose effects data from which to evaluate potential effects from exposures which may temporarily exceed the acceptable chronic exposure level. An evaluation of the data should include the incidence and severity of all abnormalities, the reversibility of abnormalities observed other than lethality, gross lesions, body weight changes, effects on mortality, and any other toxic effects.

In recent years guidelines have been established to improve quality and provide uniformity in test conditions. Unfortunately, many published LD_{50} or LC_{50} tests were not

conducted in accordance with current EPA or Organization for Economic Cooperation and Development (OECD) guidelines (USEPA, 1985; OECD, 1987) since they were conducted prior to establishment of those guidelines. For this reason, it becomes necessary to examine each test or study to determine if the study was conducted in an adequate manner.

The following is a list of ideal conditions compiled from various testing guidelines which may be used for determination of adequacy of acute toxicity data. Many published studies do not report details of test conditions making such determinations difficult. However, test conditions guidelines that might be considered ideal may include:

General:

- Animal age and species identified.

- Minimum of 5 animals per sex per dose group (both sexes should be used).

- 14-day or longer observation period following dosing.

- Minimum of 3 dose levels appropriately spaced (most statistical methods require at least 3 dose levels).

- Identification of purity or grade of test material used (particularly important in older studies).

- If a vehicle used, the selected vehicle is known to be non-toxic.

- Gross necropsy results for test animals.

- Acclimation period for test animals before initiating study.

Specific conditions for oral LD_{50}:

- Dosing by gavage or capsule.

- Total volume of vehicle plus test material remain constant for all dose levels.

- Animals were fasted before dosing.

Specific conditions for dermal LD_{10}:

- Exposure on intact, clipped skin and involve approximately 10 percent of body surface.

- Animals prevented from oral access to test material by restraining or covering test site.

Specific conditions for inhalation LC_{50}:

- Duration of exposure at least 4 hours.

- If an aerosol (mist or particulate), the particle size (median diameter and deviation) should be reported.

Although the above listed conditions would be included in an ideally conducted study, not all of these conditions need to be included in an adequately conducted study. Therefore, some discretion is required on the part of the individual reviewing these studies (USEPA, 1985; OECD, 1987).

C.2   SHORT-TERM TOXICITY STUDIES (14-DAY OR 28-DAY REPEATED DOSE TOXICITY)

Short-term exposure generally refers to multiple or continuous exposure usually occurring over a 14-day to 28-day time period. The purpose of short-term repeated dose studies is to provide information on possible adverse health effects from repeated exposures over a limited time period.

The following guidelines were derived using the OECD *Guidelines for Testing of Chemicals* (OECD, 1987) for determining the design and quality of a repeated dose short-term toxicity study:

- Minimum of 3 dose levels administered and an adequate control group used.

- Minimum of 10 animals per sex, per dose group (both sexes should be used).

- The highest dose level should ideally elicit some signs of toxicity without inducing excessive lethality and the lowest dose should ideally produce no signs of toxicity.

- Ideal dosing regimes include 7 days per week for a period of 14 days or 28 days.

- All animals should be dosed by the same method during the entire experiment period.

- Animals should be observed daily for signs of toxicity during the treatment period (i.e., 14 or 28 days). Animals that die during the study are necropsied and all survivors in the treatment groups are sacrificed and necropsied at the end of the study period.

- All observed results, quantitative and incidental, should be evaluated by an appropriate statistical method.

- Clinical examinations should include hematology and clinical biochemistry, urinalysis may be required when expected to provide an indication of toxicity. Pathological examination should include gross necropsy and histopathology.

The findings of short-term repeated dose toxicity studies should be considered in terms of the observed toxic effects and the necropsy and histopathological findings. The evaluation will include the incidence and severity of abnormalities, gross lesions, body weight changes, effects on mortality, and other general or specific toxic effects (OECD, 1987).

These guidelines represent ideal conditions and studies will not be expected to meet all standards in order to be considered to be adequate. For example, the National Toxicology Program's cancer bioassay program has generated a substantial database of short-term repeated dose studies. The study periods for these range from 14 days to 20 days with 12 to 15 doses administered generally for 5 dose levels and a control. Since the quality of this data is good, it is desirable to consider these study results even though they do not always identically follow the protocol.

C.3   SUBCHRONIC AND CHRONIC TOXICITY

Studies involving subchronic exposure (occurring usually over 3 months) and chronic exposure (those involving an extended period of time, or a significant fraction of the subject's lifetime) are designed to permit a determination of no-observed-effect levels (NOEL) and toxic effects associated with continuous or repeated exposure to a chemical. Subchronic studies provide information on health hazards likely to arise from repeated exposure over a limited period of time. They provide information on target organs, the possibilities of accumulation, and, with the appropriate uncertainty factors, may be used in establishing water quality criteria for human health. Chronic studies provide information on potential effects following prolonged arid repeated exposure. Such effects might require a long latency period or are cumulative in nature before manifesting disease. The design and conduct of such tests should allow for detection of general

toxic effects including neurological, physiological, biochemical, and hematological effects and exposure-related pathological effects.

The following guidelines were derived using the EPA *Health Effects Testing Guidelines* (USEPA, 1985), for determining the quality of a subchronic or chronic (long term) study. Additional detailed guidance may be found in that document. These guidelines represent ideal conditions and studies will not be expected to meet all standards in order to be considered for use as the basis for RfD derivation. Ideally, a subchronic/chronic study should include:

- Minimum of 3 dose levels administered and an adequate control group used.

- Minimum of 10 animals for subchronic, 20 animals for chronic studies per sex, per dose group (both sexes should be used).

- The highest dose level should elicit some signs of toxicity without inducing excessive lethality and the lowest dose should ideally produce no signs of toxicity.

- Ideal dosing regimes include dosing for 5-7 days per week for 13 weeks or greater (90 days or greater) for subchronic, and at least 12 months or greater for chronic studies in rodents. For other species repeated dosing should ideally occur over 10 percent or greater of animal's lifespan for subchronic studies and 50 percent or greater of the animal's lifespan for chronic studies.

- All animals should be dosed by the same method during the entire experimental period.

- Animals should be observed daily during the treatment period (i.e., 90 days or greater).

- Animals that die during the study are necropsied and, at the conclusion of the study, surviving animals are sacrificed and necropsied and appropriate histopathological examinations carried out.

- Results should be evaluated by an appropriate statistical method selected during experimental design.

- Such toxicity tests should evaluate the relationship between the dose of the test substance and the presence, incidence and severity of abnormalities (including behavioral and clinical abnormalities), gross lesions, identified target organs, body weight changes, effects on mortality, and any other toxic effects noted in USEPA (1985).

C.4   DEVELOPMENTAL TOXICITY

Guidelines for reproductive and developmental toxicity studies have been developed by EPA (USEPA, 1985 and OECD, 1987). Developmental toxicity can be evaluated via a relatively short-term study in which the compound is administered during the period of organogenesis. Based on the EPA *Health Effects Testing Guidelines* (USEPA, 1985), ideal studies should include:

- Minimum of 20 young, adult, pregnant rats, mice, or hamsters or 12 young, adult, pregnant rabbits recommended per dose group.

- Minimum of 3 dose levels with an adequate control group used.

- The highest dose should induce some slight maternal toxicity but no more than 10 percent mortality. The lowest dose should not produce grossly observable effects in dams or fetuses. The middle dose level, in an ideal situation, will produce minimal observable toxic effects.

- Dose period should cover the major period of organogenesis (days 6 to 15 gestation for rat and mouse, 6 to 14 for hamster, and 6 to 18 for rabbit).

- Dams should be observed daily; weekly food consumption and body weight measurements should be taken.

- Necropsy should include both gross and microscopic examination of the dams; the uterus should be examined so that the number of embryonic or fetal deaths and the number of viable fetuses can be counted; fetuses should be weighted.

- One-third to one-half of each litter should be prepared and examined for skeletal anomalies and the remaining animals prepared and examined for soft tissue anomalies.

As with any other type of study, the appropriate statistical analyses must be performed on the data for a study to qualify as a good quality study. In addition, developmental studies are unique in the sense that they yield two potential experimental units for statistical analysis, the litter and the individual fetus. The EPA testing guidelines do not provide any recommendation on which unit to use, but the *Guidelines for the Developmental Toxicity Risk Assessment* (USEPA, 1991) states that "since the litter is generally considered the experimental unit in most developmental toxicity studies…, the statistical analyses should be designed to analyze the relevant data based on incidence per litter or on the number of litters with a particular endpoint."

Others have also identified the litter as the preferred experimental unit (Palmer, 1981 and Madson et al., 1982).

Information on maternal toxicity is very important when evaluating developmental effects because it helps determine if differential susceptibility exists for the offspring and mothers. Since the conceptus relies on its mother for certain physiological processes, interruption of maternal homeostasis could result in abnormal prenatal development. Substances which affect prenatal development without compromising the dam are considered to be a greater developmental hazard than chemicals which cause developmental effects at maternally toxic doses. Unfortunately, maternal toxicity information has not been routinely presented in earlier studies and has become a standard practice in studies only recently. In an attempt to use whatever data are available, maternal toxicity information may not be required if developmental effects are serious enough to warrant consideration regardless of the presence of maternal toxicity.

C.5   REPRODUCTIVE TOXICITY

The EPA *Health Effects Testing Guidelines* (USEPA, 1985) include guidelines for both reproduction and fertility studies and developmental studies. These EPA guidelines can serve as the ideal experimental situation with which to compare study quality. Studies being evaluated do not need to match precisely but rather should be similar enough that one can be assured that the chemical was adequately tested and that the results are a reliable estimate of the true reproductive or developmental toxicity of the chemical.

These guidelines also recommend a two-generation reproduction study to provide information on the ability of a chemical to impact gonadal function, conception, parturition and the growth and development of the offspring. Additional information concerning the effects of a test compound on neonatal morbidity, mortality, and developmental toxicity may also be provided. The recommendations for reproductive testing are lengthy and quite detailed and may be reviewed further in the EPA Health Effects Testing Guidelines. In general, the test compound is administered to the parental (P) animals (at least 20 males and enough females to yield 20 pregnant females) at least 10 weeks before mating, through the resulting pregnancies and through weaning of their offspring (F1 or first generation). The compound is then administered to the Fl generation similarly through the production of the second generation (or F2) offspring until weaning. Recommendations for numbers of dose groups and dose levels are similar to those reported for developmental studies. Details should also be provided on mating procedures, standardization of litter sizes (if possible, 4 males and 4 females from each litter are randomly selected), observation, gross necropsy and histopathology. Full histopathology is recommended on the following organs of all high dose and control P and Fl animals used in mating: vagina, uterus, testes, epididymides, seminal vesicles, prostate, pituitary gland, and target organs. Organs of animals from other dose groups should be examined when pathology has been demonstrated in high dose animals (USEPA, 1985).

C.6   REFERENCES

Madson, J.M. et al. 1982. Teratology test methods for laboratory animals. In: *Principles and Methods of Toxicology*. Hayes, A.W. (ed). Raven Press. New York, NY.

OECD (Organization for Economic Cooperation and Development). 1987. *Guidelines for Testing of Chemicals*. Paris, France.

Palmer, AK. 1981. Regulatory requirements for reproductive toxicology: Theory and practice. In: *Developmental Toxicology*. Kimmel, C.A. and J. Buelke-Sam (eds). Raven Press. New York, NY.

USEPA (U.S. Environmental Protection Agency). 1985. Health Effects Testing Guidelines. 40 CFR Part 798.

USEPA (U.S. Environmental Protection Agency). 1991. Final guidelines for development toxicity risk assessment. *Federal Register* 56: 63798-63826. December 5.