Abt Associates Inc.

Cambridge, MA
Bethesda, MD
Atlanta, GA
Durham, NC

# Evaluating Methods for Quantifying Human Noncancer Health Risks: Case Study Application

## Contract # EP-W-05-022 (Work Assignment #77)

***REVISED DRAFT***

March 1, 2009

*Prepared for*
Mr. Ronn Dexter (1809T)
U.S. EPA
National Center for Environmental Economics
1200 Pennsylvania Ave. NW
Washington, DC 20460

*Prepared by*
Sue Greco
Jay Smith
Brian Segal
Ellen Post
Meghan Lynch
Dale Hattis

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

# Table of Contents

# Tables

## Figures

# 1. Introduction

"Traditional noncancer risk assessment"[1] methods have limited application in benefits assessments. While exposure to chemicals below the reference levels (e.g., RfD or RfC) is likely to be "without an appreciable risk of deleterious effects" (U.S., EPA, 2007), there is not a standard method to quantify the actual risk. The two case studies presented in this report begin to explore the merits of using three alternative noncancer risk assessment methods in setting reference doses and concentrations, as well as in estimating the potential human health benefits (such as those resulting from the application of a new rule to, e.g., restrict or ban on emissions of a certain compound). While we stop short of monetizing the benefits, we do estimate risk at various exposure levels, and we expect the monetized benefits to vary linearly with risk. The two case studies illustrate inhalation and ingestion routes of exposure, and highlight differences in inputs and outputs across the three models. While the statistical framework for two of the models (BMDS and CatReg) is similar, the third model is probabilistic in nature, was recently espoused in an NRC report, and merits additional work beyond this report.

## 1.1 Objectives

Abt Associates was tasked with analyzing the capabilities of current quantitative noncancer risk assessment methods and software packages to develop dose-response curves to aid in benefits assessment. This report evaluates three noncancer risk assessment methods using a case study framework. The chemicals chosen for the case studies are 1,2,4,5-tetrachlorobenzene (TCB) and carbonyl sulfide (COS)[2]. The methods evaluated are: Benchmark Dose software (BMDS), Categorical Regression (CatReg), and Straw Man[3]. Our approach does not include a full-scale benefit-cost analysis, but focuses on the differences across the three models as they apply to human health risk and thus benefits assessment. As a result of limitations in extrapolating information from animal studies to humans, the change in the incidence of the health effect in humans and the monetized social costs of these health effects were not estimated, but are expected to vary in a linear manner with the model output (e.g., risk).

The specific objectives of this report are to (first for COS, then for TCB):

1. Apply three quantitative models (BMDS, CatReg, Straw Man) to the same compound and assess model applicability and fit.
2. Estimate and compare reference doses and concentrations (RfD and RfC) from the traditional noncancer risk assessment method to those derived from each of the three quantitative models considered.

[1] In this report, we use the term, "traditional noncancer risk assessment", to refer to a reference dose or concentration that has been established by identifying a NOAEL or LOAEL (from a toxicological study), then dividing this "point of departure" by uncertainty factors.

2 The chemical selection process is described in Section 2.2.

[3] These models were chosen after examination of several models in an earlier work assignment.

3. Evaluate and compare the risk estimates provided by each of the three models at doses near the RfD/RfC and point of departure.
4. Qualitatively evaluate the capabilities of the different models/software packages with regard to their potential application to benefits assessment and potential improvements over the traditional noncancer risk assessment method.

## 1.2 Background

Benefit-cost analysis evaluates the favorable effects of policy actions and the associated opportunity costs of these actions on a common dollar basis (U.S. EPA, 2000). In evaluating proposed or existing EPA policy actions, monetized human health benefits from reduced exposure to environmental contaminants generally comprise the majority of the benefits. To estimate these benefits, policy analysts estimate (Axelrad et al., 2005):

- the adverse health effects associated with the exposure to the contaminant in question,
- the probability of adverse effects occurring in the exposed population for different levels of potential exposure (i.e., the exposure-response function),
- the corresponding change in the incidence of the health effect that may result from a particular policy option relative to some policy baseline case, and,
- the economic value of each adverse health effect avoided (i.e., the unit value).

The first two items above are examined in detail in this report using a risk assessment framework. Risk assessment involves identifying the hazard, conducting an exposure assessment, identifying an appropriate dose-response function, and characterizing the human health risk from exposure to the pollutant. Risk assessment methods for cancer and noncancer health outcomes differ in that a threshold or safe value is presumed for most noncancer health outcomes (such as birth defects, respiratory effects, and hepatotoxicity).

Benefits for noncancer health effects, are difficult to quantify for two main reasons. First most information about the toxicity (mode of action, health impact) of environmental contaminants arises from well-controlled laboratory experiments conducted on animals. There are well-known and documented issues with extrapolating experimental information from animals to humans. For example, there are intraspecies differences whereby different strains of rats may respond differently to contaminants, and these differences are only amplified when comparing a strain of rats to humans (i.e., interspecies differences). Furthermore, in toxicological studies animals are often exposed to defined doses at levels which are orders of magnitude higher than the doses to which humans are typically exposed. The high-to-low dose and animal-to-human projections can be problematic since effects seen in animal experiments at the relatively high doses used in animal toxicology studies may not correspond to those seen in humans at lower doses. As a result, substantial uncertainties exist about the shape of the dose-response function at typical levels of human exposure, the variability of human response to exposures, and the characteristics of any adverse health effect in humans.

Second, the traditional noncancer risk assessment method – as we have defined it in this report- does not attempt to estimate a dose-response function, and so does not lend itself to benefits analysis. The traditional method selects a point of departure – a NOAEL[4] or LOAEL[5] - from a critical effect toxicological study and divides it by appropriate uncertainty factors to estimates threshold values for ingestion (the reference dose, RfD) and inhalation (the reference concentration, RfC). The RfD is "an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime" (U.S. EPA, 2007). Because the traditional approach does not estimate a dose-response function that is relevant to humans, this approach presents significant challenges in quantifying health benefits in terms of the number of cases of the adverse health effect that can be potentially prevented as a result of a specific policy option. At best, the number of people who may be at some (not precisely defined) risk for adverse effects because of exposures to the chemical above the RfD or RfD can be quantified.

The inability to quantify the noncancer risk posed by chemicals has been overcome, to some extent, by the availability of several methods (and models), including the benchmark dose method (e.g., BMDS), which fits a function to a toxicological data set; categorical regression (e.g., CatReg), that can combine information from several toxicological studies into a summary function; and probabilistic methods (e.g., Straw Man), that explicitly account for the some of the uncertainties associated with the use of data from animal experiments described earlier. However, each of these approaches was developed for a different purpose and utilizes specific analytic assumptions, statistical methods, and input data. Consequently, application of these models to the same chemical may lead to substantially different dose-response functions and estimates of risk for a given exposure dose. This report describes these differences in two chemical case studies. None of the models fully address the issues of extrapolating high-dose animal study information to low-dose human health endpoints. Nevertheless, moving from a threshold value to risk estimates is an important step in benefits assessment.

## 1.3 Report Overview

In the subsequent chapters, we use two chemical case studies to illustrate the impact of different noncancer risk assessment methods on human health benefits assessment. We compare the derivation of reference levels across the traditional (NOAEL/LOAEL) and quantitative (BMDS, CatReg, Straw Man) methods and estimate risk associated with each of the quantitative methods. We examine the potential impact of different analytic assumptions such as deterministic versus probabilistic approaches and incorporation of the uncertainty associated with human interindividual variability in response to pollutants. The report is organized into several sections.

[4] No Observed Adverse Effects Level (NOAEL) determined from critical effect study.

[5] Lowest Observed Adverse Effect Level (LOAEL) determined from critical effect study.

- Methods (**Section 2**)
    - Analytical steps
    - Selection of Case Study compounds
    - Three alternative noncancer risk assessment model descriptions
        - BMDS
        - CatReg
        - Straw Man
- Case Studies
    - Carbonyl Sulfide Case Study (**Section 3**)
    - 1,2,4,5-Tetrachlorobenzene Case Study (**Section 4**)
        - Hazard Assessment
        - Model Results
            - RfC or RfD
            - Risk
        - Comparison across models
- Discussion (**Section 5**)
- References (**Section 6**)
- Appendices
    - Straw Man Inputs (**Appendix A**)
    - CatReg: Additional COS analysis and model inputs (**Appendixes B to D**)

# 2. Methods

In this section, we first describe the overall framework for conducting the case study comparisons. Next, we describe why these two compounds were selected. Finally, we provide an overview of each quantitative noncancer risk assessment model. Only the model frameworks are described in the Methods section; relevant model inputs are described within the case studies (for BMDS) or in Appendices (for CatReg and Straw Man).

## 2.1 Analytical Steps

The analytic approach for the case studies consisted of four major steps to identify the health outcome associated with the contaminant exposure; construct dose-response functions and estimate a reference dose or concentration (RfD/RfC) using each of the three models; estimate risk levels around the RfD/RfC; and discuss model differences. These steps are described in more detail below.

**Step 1: Identify Health Outcome**
The first step of the analysis identified the health outcomes and mode of action associated with exposure to the chemical of interest. For both case study chemicals, a literature review was conducted using TOXNET and the U.S. EPA's Integrated Risk Information System (IRIS). If an RfC/RfD was established within IRIS then the supporting critical effect study was obtained and used as the model input data source. If no critical effect study was identified in IRIS, an extensive review of the toxicology literature was conducted leading to a hypothesis about the mode of action and the critical effect associated with the chemical. General information was obtained on the chemical's use, ambient concentrations, and potential pathways of human exposure. (See **Section 3.1** and **4.1**.)

**Step 2: Dose-Response function, Point of Departure, and RfD/RfC**
The second step of the analysis involved the development of three dose-response functions (for both chemicals) using the alternative quantitative models, as well as the identification of the point of departure and the subsequent calculation of the RfD/RfC. Input datasets were constructed for each model using the critical effect study and health endpoint identified in Step 1. Model settings and data inputs were standardized to improve comparability among models. (For example, each of the software packages used a log-probit function to model the relationship between the dose and the incidence of the quantal[6] response parameter.) Since each model has different data input configurations, care was taken to ensure inputs were equivalent with respect to health endpoint, animal sex, concentration scaling, and exposure duration. The final model fit was assessed. Then, the point of departure from each

[6] Quantal: A dichotomous classification where an individual or animal is placed in one of two categories, e.g., dead or alive, with or without a particular type of tumor, normal or abnormal level of hormone, or other effect at a specific level of severity (U.S.EPA 2007).

model was identified and divided by relevant uncertainty factors in order to calculate the RfC or RfD. Note: the Straw Man model does not identify a point of departure, but estimates the RfC or RfD using probabilistic means described in Section 2.3.3. (See **Sections 3.2 to 3.5** and **4.2 to 4.5**.)

**Step 3: Human Health Risk**
The third step was to develop estimates of risk from the three models at exposure levels near the RfD/RfC and point of departure and to examine how these risk values differed from each other across the different exposure levels. The fitted models were used to estimate the human health risk associated with inhalation or ingestion of the case study compounds. Particular emphasis was placed on the estimation of human health risks at the low contaminant concentrations expected for human inhalation or exposure rather than at the high exposure levels normally used in toxicological animal studies. Each of the methods attempt to account for the animal-to-human extrapolation and human interindividual variability to some extent by applying uncertainty factors that are based on either accepted conventions or probability distributions derived from known information. However, it must be acknowledged that neither of these uncertainty factors is fully characterized when we rely on toxicological studies as the basis for estimating human health risks to environmental contaminants. (See **Sections 3.2 to 3.5** and **4.2 to 4.5**.)

**Step 4: Relative Model Merits**
The final step is a qualitative discussion of important differences and relative merits of the alternative noncancer risk assessment methods and software packages applied in each case study. We evaluate the statistical and scientific validity of the alternative methods used in the two case studies. We discuss the advantages (such as minimal data input needs) and disadvantages (such as lack of transparency) of each alternative method over the traditional noncancer human health risk assessment method. We also examined practical considerations such as the quantity and quality of input data necessary for the alternative method, the level of effort, and the amount of user expertise/judgment required to yield valid outputs. (See **Sections 3.6 and 4.6**, as well as **Section 5**.)

## 2.2 Case Study Compound Selection

The chemicals for the case studies were selected in order to illustrate the level of effort required to obtain the input data for a compound that had been studied in IRIS (compared to one that had not), and to examine the results for both inhalation and ingestion pathways. The peer-reviewed literature available for the compound needed to be adequate to establish a point of departure. The case study compound's critical effect study necessitated a quantal (yes or no) health endpoint since not all of the models readily handle continuous response data. To improve efficiency, one compound previously examined using the Straw Man model was selected for analysis. Based on these criteria, the compounds selected were carbonyl sulfide (COS) for inhalation and 1,2,4,5-tetrachlorobenzene (TCB) for ingestion. All case study selection parameters are summarized in Table 2- 1.

**Table 2- 1 Comparison of Case Study Compounds Across Selection Parameters**

| **Compound** | carbonyl sulfide | 1,2,4,5-tetrachlorobenzene |
|---|---|---|
| **Abbreviation** | COS | TCB |
| **Route of Exposure** | Inhalation | Ingestion |
| **Type of reference value estimated** | RfC | RfD |
| **Quantal Response** | Yes | Yes |
| **Previously examined in Straw Man?** | No | Yes |
| **Noncancer reference value available in IRIS?** | No | Yes |
| **Sufficient information in the peer-reviewed literature to develop an RfC/RfD?** | Yes[7] | Yes |

## 2.3 Noncancer Risk Assessment Models

We selected three methods/models to examine noncancer human health risk benefits assessment in the case studies. They are Benchmark Dose Software (BMDS), Categorical Regression (CatReg), and an approach outlined by Hattis et al. 2002 (Straw Man). These three models have appeared extensively in the literature, are well documented, and are publicly accessible. Each has advantages: BMDS contains a large number of mathematical models available for fitting curves to toxicological data; CatReg can accommodate multiple endpoints and exposure durations; and the probabilistic Straw Man approach uses available information about uncertainty factors to create a distribution of RfD/RfC values, rather than point estimates.

These quantitative noncancer risk assessment methods might refine benefits assessment because of their abilities to incorporate more information from the critical effect toxicological experiment (the traditional method uses just one data point – the NOAEL or LOAEL – as the point of departure) and to generate dose-response functions

[7] Based on the literature search conducted in March 2007 by Abt Associates Inc.

(which can be combined with population exposures to estimate animal and then human health risks). The Straw Man method further allows for estimating a distribution of RfD/RfC's and incorporating variability in human response into the estimates.

The three models differ in their conceptual and statistical approaches. Conceptually, the way CatReg and BMDS are used to establish an RfC or RfD (i.e., by establishing a point of departure, POD) differs from the Straw Man approach. While all approaches generate dose-response functions, the CatReg or BMDS-established POD does not account for the uncertainty about the estimated dose-response function or the uncertainty around the "uncertainty factors" they subsequently use, whereas Straw Man explicitly attempts to quantify these sources of uncertainty. CatReg and BMDS fit mathematical dose-response functions to the animal data, which can then be used to develop single point estimates of risk for a given health effect at a given human equivalent dose (exposure level). One of the primary criticisms of deterministic approaches is that they convey an impression of undue precision by expressing the RfD/RfC as a single value, sometimes presented as the result of "the risk assessment".

The Straw Man model, on the other hand, produces a *distribution* of values for the risk at the RfC or RfD defined by the traditional approach, rather than a single "no risk" number. It yields a *distribution* of concentrations or doses associated with a pre-defined risk level (e.g., 1/100,000). Probabilistic analyses, such as Straw Man, are commonly implemented using a Monte Carlo approach where distributional inputs can be combined to generate a variability and/or uncertainty distribution of the outputs. Furthermore, the uncertainty associated with the inputs can be characterized and used as a basis for assessing the uncertainty distribution for the outputs. The derivation and use of probabilistic risk estimates can help correct the misconception that an estimated RfD/RfC represents a level of exposure that poses zero risk for a population by recognizing that individual thresholds for response to a given chemical can vary greatly, and that there is uncertainty about the extent of this variation. This explicit treatment of uncertainty in the variation in human response to exposure represents another important consideration in the Straw Man model.

#### 2.3.1 Benchmark Dose Software (BMDS)

Benchmark Dose Software (BMDS) was developed by the EPA for estimation of the benchmark dose (BMD) – the dose associated with a specified percentage increase in health risk of, for example, 10% - and its lower confidence limit ($BMD_L$) or the analogous inhalation values (BMC[8] and $BMC_L$[9]). This approach is currently used to evaluate a point of departure that is used in conjunction uncertainty factors to set reference doses and concentrations in the Integrated Risk Information System (IRIS) database. From a review of the IRIS database[10], we observed that approximately 4% (of 359) of the RfD

[8] BMC: An exposure due to a dose or concentration of substance associated with a specified low incidence of risk, generally in the range of 1% to 10%, of a health effect: or the dose associated with a specified measure or change of a biological effect.

[9] $BMC_L$: A lower one-sided confidence limit on the BMC.

[10] Review was conducted by Abt Associates, Inc. in July 2007

values and 23% (of 70) of the RfC values in IRIS were established using the BMDS-generated points of departure for a total of 31 compounds with established reference values in IRIS using BMDS. (The majority of the entries were set using a NOAEL or LOAEL as the point of departure.) To derive the benchmark dose, the software fits a dose-response function to the toxicological data from user-specified mathematical functions (such as probit, logistic, linear, or Weibull distribution). This resulting dose-response function can be used in benefits assessment with human equivalent dose inputs. The software output includes the $BMC_L$ or $BMD_L$ (point of departure), model parameter statistics, and an assessment of the model fit to the input data. To facilitate comparisons to the Straw Man approach we modeled the data using a dichotomous model type and a log-probit dose-response relationship.[11] The log-probit model for the probability of adverse health effect, P[Response], is a 3-parameter model:

$$P[\text{Response}] = \gamma + (1 - \gamma)\,\Phi\{\alpha + \beta \log(\text{Dose})\} \qquad (1)$$

where

- $\alpha$ = intercept
- $\beta$ = slope
- $\gamma$ = background
- $\Phi$ = cumulative standard normal distribution, with mean ($X$) of 0 and standard deviation ($\sigma$) of 1.

Note that when the background equals zero ($\gamma = 0$), the log-probit model reduces to a 2-parameter model:

$$P[\text{Response}] = \Phi\{\alpha + \beta \log(\text{Dose})\} \qquad (2)$$

This is similar to the probit model, where linear dose is employed.

$$P[\text{Response}] = \Phi\{\alpha + \beta\,\text{Dose}\} \qquad (3)$$

Equations (1) and (2), which employ log-transformed doses, essentially assume that there is a lognormal distribution of individual thresholds in the human population to which the model is applied. Equation (3), with linear dose, assumes a normal distribution. Models which take the logarithm of the dose have traditionally been chosen by toxicologists for dose-response modeling in a wide variety of contexts.

11 The log probit model is used because it supports the assumption that the logarithm of the dose associated with a threshold response in individual animals or people has a standard normal distribution. This model has a long history of use in toxicology for a wide variety of effects that are thought to occur by overwhelming homeostatic systems. A lognormal distribution of individual thresholds would be expected if there are many factors that cause individuals to vary from one another in their sensitivity, and if these factors tend to act multiplicatively in affecting the dose needed to cause effects in different subjects.

### 2.3.2 Categorical Regression (CatReg)

CatReg is another EPA software program and is used for conducting categorical regression. The software was developed for meta-analysis of toxicological data for acute inhalation exposures (U.S. EPA 2006a). The software can estimate probabilities of different levels of user-defined response severity, given the input concentration and duration of exposure. A further advantage is that information from multiple studies can be combined. A disadvantage, however, is that input data requirements may be more extensive and stricter (e.g., more input data needed, the data must be in a particular form).

Similar to the BMDS, CatReg provides an estimate of the concentration corresponding to a 10% extra risk (ERC10). The point of departure can be selected as the lower bound of the confidence interval on the ERC10, called the ERC10-L. The equivalent ingestion (dose) terms are ERD10 and ERD10-L.

CatReg output is similar to that of BMDS, and includes a dose-response function, which can also be used to estimate benefits using human equivalent doses.

The general form of the model, incorporating both concentration and duration, is:

$$\Pr(Y \geq s|C,T) = H[\alpha_s + \beta_1 * f_1(C+\gamma) + \beta_2 * f_2(T)] \quad (4)$$

where:

- Pr = probability of adverse health effect
- Y = adverse health effect
- s = severity level (an ordinal score corresponding to 0 (no response) or 1 (response) (additional categories can be added)
- C = concentration (in ppm)
- T = time or duration (in weeks)
- H = is a probability function taking values between 0 and 1, and is based on the standard normal distribution
- $\alpha_s$ = intercept for severity level
- $\beta$ = slope parameters for concentration (1) and time (2)
  f(C), f(T) = functions of concentration and time (log-transformed)
- $\gamma$ = hypothetical background concentration needed to produce background response (in ppm)

### 2.3.3 Hattis et al. 2002 Approach (Straw Man)

In contrast to the deterministic models BMDS and CatReg, the Straw Man model (Hattis et al., 2002; Hattis and Lynch, 2007) is a probabilistic model that generates a distribution of possible risks in a general human population for any specified dose, or alternatively a distribution of doses for any specified level of risk (i.e., a “risk specific dose”). In this section, we describe the Straw Man model in six steps, both using text and mathematical

equations. The implementation of probabilistic modeling techniques was recommended in a recent NRC report as part of a general effort to unify approaches for risk assessment for cancer and noncancer effects and to make the outputs of risk assessments more helpful for decision-making[12].

> "Scientific and risk-management considerations both support unification of cancer and noncancer dose-response assessment approaches. The committee therefore recommends a consistent, unified approach for dose-response modeling that includes formal, systematic assessment of background disease processes and exposures, possible vulnerable populations, and modes of action that may affect a chemical's dose-response relationship in humans. That approach redefines the RfD or RfC as a risk specific dose that provides information on the percentage of the population that can be expected to be above or below a defined acceptable risk with a specific degree of confidence. The risk-specific dose will allow risk managers to weigh alternative risk options with respect to that percentage of the population. It will also permit a quantitative estimate of benefits for different risk-management options. For example, a risk manager could consider various population risks associated with exposures resulting from different control strategies for a pollution source and the benefits associated with each strategy. The committee acknowledges the widespread applications and public-health utility of the RfD; the redefined RfD can still be used as the RfD has been to aid risk-management decisions." (NRC, 2008)

For illustrative purposes, the initial Straw Man publications have used a risk specific dose that would generate a 1/100,000 excess risk of adverse outcome in the general population or 1/1,000 in sensitive subpopulations and a confidence level of 95% for comparisons with existing RfDs. Choosing a set of incidence and confidence levels for replacing existing RfDs is a risk management decision that would have to be made by an appropriate governing body.

The Straw Man model estimates risk specific doses by drawing on information from other chemicals that is relevant to each of the traditional uncertainty factors used in traditional noncancer risk assessments. As with traditional single-point uncertainty factors, the uncertainty distributions depend on the nature of the chemical-specific toxicity information available for the chemical under study and therefore the data deficiencies that need to be offset by supplemental information from other sources—e.g. subchronic/chronic, animal/human, incomplete/complete data set for risk projection, and human interindividual variability for pharmacokinetic and pharmacodynamic steps in the causal pathway to production of adverse effects.

Once the uncertainty factor distributions are assembled, a large number of random draws are made from each distribution. An expected risk level as a function of dose is generated from each set of random draws by applying a log-probit function. A log-probit dose-response model is expected to result when:

[12] Committee on Improving Risk Analysis Approaches Used by the U.S. EPA, NRC (2008). Science and Decisions: Advancing Risk Assessment. Page 8 of the summary.

(a) The mode of action is one of a broad family of homeostasis overwhelming processes that produce an "individual threshold" dose-response relationship. In other words, the fraction of people responding (at a given level of severity) at each dose depends on the fraction who have personal thresholds below that dose.
(b) Each individual's threshold is determined by a large number of factors (e.g. individual uptake rates, activating and detoxifying metabolism rates, excretion rates, and pharmacodynamic factors) that act multiplicatively to affect individual susceptibility.
(c) Interacting background factors that produce the same response in the absence of exposure to the agent being assessed are relatively unimportant.

#### *2.3.3.1 General Straw Man Framework*

The general structure of the Straw Man can be broken down into six steps:

1) Estimate an animal $ED_{50}$ or $EC_{50.}$ Ideally this is done by fitting dose-response models (such as those available in the BMD system) to the original toxicological data. However if only LOAEL's or NOAEL's are available for the chemical and response being evaluated, the framework provides distributions to represent the uncertainty in translating these into $ED_{50}$'s or $EC_{50}$'s. For this type of risk projection, LOAEL's are preferred to NOAEL's because LOAEL's represent doses that are directly observed to produce adverse effects.

2) If the best data available for assessing toxicity are from an animal subchronic (rather than an animal chronic) toxicity dosing protocol, use an empirical distribution of ratios of chronic-subchronic NOAELs and their uncertainty to transform the subchronic dose that affects 50% of the lab animal populations to a chronic dose that would be expected affect 50% of the animals.

3) Use an empirical distribution of human/animal doses needed to produce putatively equivalent toxicity to convert the estimated animal $ED_{50}$'s to human $ED_{50}$'s.

4) Use an empirical distribution to represent deficiencies in the toxicological datasets.

5) Use observations from human studies of analogous chemicals to derive separate distributions of human pharmacokinetic and pharmacodynamic variability, tuned to the organ system affected (e.g., neurotoxicity versus immunotoxicity) and the severity of the effect (less severe effects tend to show greater human pharmacodynamic variability than more severe effects). Use the combined pharmacokinetic and pharmacodynamic variability to assess risks as a function of dose, using the probit model.

6) Combine input from Steps 1 though 5 into a Monte Carlo simulation. After a large number of trials, sort the doses associated with the target risk

(1/100,000 in the general population, or 1/1,000 in sensitive subpopulations) from the lowest to the highest values and choose the proper percentile value to represent the dose that is expected to produce the target risk or less with 95% confidence.

Mathematically, these six steps are represented in the following manner: (Note that terms representing distributions are indicated with a "hat" above the variable.)

(1) Implementation of a curve fit to estimate the animal $ED_{50}$ distribution, or an extrapolation from the point of departure (POD) using an uncertainty factor:

$$\mathrm{POD} \cdot \hat{\mathrm{U}}\mathrm{F}_{\mathrm{POD}\rightarrow\mathrm{ED}_{50}} \qquad (5)$$

(2) Multiplication by the subchronic to chronic uncertainty factor distribution ($\hat{U}F_S$) if the underlying data consist of subchronic effects

(3) Multiplication by the animal to human uncertainty factor distribution ($\hat{U}F_A$)

(4) Multiplication by the database uncertainty factor distribution ( $\hat{\mathrm{D}}$ )

(5) Aggregation of human interindividual variability from pharmacokinetic and pharmacodynamic differences ($\hat{U}F_H$):

$$\sqrt{\left(\log\hat{\mathrm{GSD}}_{\mathrm{PK}}\right)^2+\left(\log\hat{\mathrm{GSD}}_{\mathrm{PD}}\right)^2} \qquad (6)$$

where $\hat{GSD}_{PK}$ and $\hat{GSD}_{PD}$ represent the geometric standard deviation (GSD) of the pharmacokinetic (PK) and pharmacodynamic (PD) variability, respectively.

Incorporation of the probit model into the human interindividual variability to create a dose-response curve:

$$10^{\sqrt{\left(\log\hat{\mathrm{GSD}}_{\mathrm{PK}}\right)^2+\left(\log\hat{\mathrm{GSD}}_{\mathrm{PD}}\right)^2}\cdot\mathrm{probit}\left(\mathrm{P}_{\mathrm{response}}\right)} \qquad (7)$$

where probit(z) is equivalent to the inverse cumulative standard normal function, or the inverse function of the area under the standard normal curve.

Note that if the GSD's are large, there will be more variation in response than if the GSD's are small, resulting in a larger $\hat{U}F_H$.

(6) When all of these above elements are incorporated into the Straw Man model, the general dose-response function for humans can be represented as:

$$\hat{\text{D}}\text{ose}\left(\text{P}_{\text{response}}\right) = \text{POD} \cdot \hat{\text{UF}}_{\text{POD}\rightarrow\text{ED}_{50}} \cdot \hat{\text{UF}}_{\text{A}} \cdot \hat{\text{UF}}_{\text{S}} \cdot \hat{\text{D}} \cdot 10^{\sqrt{\left(\log\hat{\text{GSD}}_{\text{PK}}\right)^2+\left(\log\hat{\text{GSD}}_{\text{PD}}\right)^2}\cdot\text{probit}\left(\text{P}_{\text{response}}\right)} \quad (8)$$

where,

Variables with hats represent distributions or probabilistic draws from distributions

$\hat{\text{D}}\text{ose}(\text{P}_{\text{response}})$: Distribution of doses at which a given probability of response, $\text{P}_{\text{response}}$, is expected

POD: Point of departure (e.g., NOAEL, LOAEL)

$\hat{\text{UF}}_{\text{POD}\rightarrow\text{ED50}}$: Uncertainty distribution for projecting the animal $\text{ED}_{50}$ from the POD

$\hat{\text{UF}}_{\text{A}}$: Uncertainty distribution for interspecies projection (e.g., animal to human)

$\hat{\text{UF}}_{\text{S}}$: Uncertainty distribution for subchronic to chronic dose projection

$\hat{\text{D}}$: Conversion factor distribution for database deficiency

$\hat{\text{GSD}}_{\text{PK}}$: Geometric standard deviation of human variability due to pharmacokinetic differences

$\hat{\text{GSD}}_{\text{PD}}$: Geometric standard deviation of human variability due to pharmacodynamic differences

$\text{P}_{\text{response}}$: Specified probability of response in a general human population (varies from 0 to 1)

Rearranging this equation so that the probability of response is the dependent variable:

$$\hat{\text{P}}_{\text{response}}\left(\text{Dose}\right) = \Phi\left(\frac{\log\left(\text{Dose}\right) - \log\left(\text{POD} \cdot \hat{\text{UF}}_{\text{POD}\rightarrow\text{ED}_{50}} \cdot \hat{\text{UF}}_{\text{A}} \cdot \hat{\text{UF}}_{\text{S}} \cdot \hat{\text{D}}\right)}{\sqrt{\left(\log\hat{\text{GSD}}_{\text{PK}}\right)^2 + \left(\log\hat{\text{GSD}}_{\text{PD}}\right)^2}}\right) \quad (9)$$

where,

Variables with hats represent distributions or probabilistic draws from distributions

$\Phi$: The standard normal cumulative distribution

Dose: Specified dose level in general population (can also be modeled as a concentration)

The calculation of the dependent variables (e.g., dose or response) in both equations results in distributions. The generation of these distributions is best considered as resulting from a long sequence of random draws from each of the distributional input variables (variables with hats). Thus, in the calculation of the distribution of the risk in the general population (Eq. 9), each iteration results in a single value being chosen from the $\hat{\text{UF}}_{\text{POD}\rightarrow\text{ED50}}$, $\hat{\text{UF}}_{\text{A}}$, $\hat{\text{UF}}_{\text{S}}$, $\hat{\text{D}}$, $\hat{\text{GSD}}_{\text{PK}}$, and $\hat{\text{GSD}}_{\text{PD}}$ distributions, all of which in turn calculate a single risk. After a large number of iterations have been run, the resulting risk distribution can be thought of as precisely carrying through the variance of the input distributions. In other words, the characteristics of the distribution of risk are due to the characteristics of the distributions of the input variables. The empirical data from which

the uncertainty and conversion factor distributions are constructed are documented in **Appendix A**.

***2.3.3.2 Straw Man Framework if Detailed Animal Dose-Response Data are Available***

If the subchronic animal dose-response data are detailed enough to generate an animal $ED_{50}$ with its attendant uncertainty using curve fitting procedures, the first step in the Straw Man model should be replaced by the empirically derived $ED_{50}$, since the following relationship should hold:

$$\text{empirically fit } \hat{\text{E}}\text{D}_{50} \subset \text{POD} \cdot \hat{\text{U}}\text{F}_{\text{POD}\rightarrow\text{ED}_{50}} \qquad (10)$$

The empirically fit animal $ED_{50}$ should, in most cases, be a more precise subset of the distribution generated by multiplying the POD and the uncertainty factor distribution used to turn that POD into an $ED_{50}$. Assuming that the empirically fit animal $ED_{50}$ distribution is a more precise set of estimates of the value, it can be substituted into the dose-response equations as a distribution.

$$\hat{\text{D}}\text{ose}\left(\text{P}_{\text{response}}\right) = \hat{\text{E}}\text{D}_{50} \cdot \hat{\text{U}}\text{F}_{\text{A}} \cdot \hat{\text{U}}\text{F}_{\text{S}} \cdot \hat{\text{D}} \cdot 10^{\sqrt{\left(\log\text{G}\hat{\text{S}}\text{D}_{\text{PK}}\right)^2+\left(\log\text{G}\hat{\text{S}}\text{D}_{\text{PD}}\right)^2}\cdot\text{probit}\left(\text{P}_{\text{response}}\right)} \qquad (11)$$

$$\hat{\text{P}}_{\text{response}}\left(\text{Dose}\right) = \Phi\left(\frac{\log\left(\text{Dose}\right) - \log\left(\hat{\text{E}}\text{D}_{50} \cdot \hat{\text{U}}\text{F}_{\text{A}} \cdot \hat{\text{U}}\text{F}_{\text{S}} \cdot \hat{\text{D}}\right)}{\sqrt{\left(\log\text{G}\hat{\text{S}}\text{D}_{\text{PK}}\right)^2 + \left(\log\text{G}\hat{\text{S}}\text{D}_{\text{PD}}\right)^2}}\right) \qquad (12)$$

Where,

$\hat{E}D_{50}$: Distribution representing the empirical estimate of the animal $ED_{50}$ with measurement uncertainty

***2.3.3.3 Final Straw Man Equations Used in this Report***

Equations 8 and 11 can be used to estimate a distribution of doses corresponding to a specific health risk level in the general population. The risk assessor must define this level. However, because the output of interest from the Straw Man model is a distribution of doses at this response level, the risk assessor must also define a level of confidence in order to produce an RfD or RfC point estimate. As stated above, Hattis et al. recommend the 95% confidence level (i.e., 5$^{th}$ percentile dose) at which a response is seen in 1/100,000 of the general population or in 1/1,000 of a sensitive subpopulation. The Straw Man model is typically run for 15,000 iterations to allow for a sufficiently precise estimate of variance in the resulting distribution, which in this case is dose. The equations below illustrate these calculations for an excess risk of 1 in 100,000 adverse health outcomes in the general population.

$$\hat{R}fD \text{ or } \hat{R}fC = \hat{D}ose\left(\frac{1}{100{,}000}\right) = \hat{E}D_{50} \text{ or } \hat{E}C_{50} \cdot \hat{U}F_A \cdot \hat{U}F_S \cdot \hat{D} \cdot 10^{\sqrt{\left(\log\hat{GSD}_{PK}\right)^2 + \left(\log\hat{GSD}_{PD}\right)^2} \cdot \text{probit}\left(\frac{1}{100{,}000}\right)} \quad (13)$$

Where
Rf^D: distribution of doses corresponding to a 1 in 100,000 risk in the general population
Rf^C: distribution of concentrations corresponding to a 1 in 100,000 risk in the general population

The point estimate RfD or RfC are then defined as the 5$^{th}$ percentile of their corresponding distributions.

The calculations of the distributions of risk at and around a given exposure level (such as the RfD or RfC) are also possible, using the following equation:

$$\hat{P}_{response}\left(RfD \text{ or } RfC\right) = \Phi\left(\frac{\log\left(RfD \text{ or } RfC\right) - \log\left(\hat{E}D_{50} \text{ or } \hat{E}C_{50} \cdot \hat{U}F_A \cdot \hat{U}F_S \cdot \hat{D}\right)}{\sqrt{\left(\log\hat{GSD}_{PK}\right)^2 + \left(\log\hat{GSD}_{PD}\right)^2}}\right) \quad (14)$$

From the risk distribution output of this equation, the risk assessor can extract any desired summary statistics (e.g., mean, median, n$^{th}$ percentile risk).

The Straw Man model relies on uncertainty factor distributions, which must be continually updated to account for new data. While some work has been done to update these distributions, continued efforts are needed on this front. To evaluate the uncertainties associated with the uncertainty factor distributions, older Straw Man outputs can be compared to updated Straw Man outputs. There are, in particular, 3 sources of uncertainty that may have varying effects on Straw Man results (with expected effect in parentheses):

1. Representativeness of the populations (may expand estimates of $GSD_{PK}$ and $GSD_{PD}$, and therefore increase expected risks at low doses)
2. Short and long-term variability and measurement errors (may reduce $GSD_{PK}$ and $GSD_{PD}$ and decrease risk at low doses)
3. Potential interactions between common mechanisms of disease and contaminant-caused disease processes (may increase low dose risks).

There are also other sources of uncertainty that have not yet been quantified, and which will be further assessed in subsequent work. As with any analytical tool based on empirical information, it is advisable to have a continuous process of updating methods and associated uncertainties.

# 3. Case Study Application of BMDS, CatReg, and Straw Man to Carbonyl Sulfide (COS)

To help illustrate the difference between the noncancer risk assessment models, we conduct a case study of carbonyl sulfide (COS). We first describe the literature on this compound (hazard assessment), which is expected to act on the brain via the inhalation route of exposure. Next, we conduct a "traditional" noncancer risk assessment, followed by assessments using BMDS, CatReg and Straw Man. We compare the RfC generated by each method, along with the estimated risk at doses around the RfC across the models.

## 3.1 COS Hazard Assessment

Carbonyl Sulfide (COS) is a colorless, odorless, relatively stable gas (Figure 3- 1). Its largest anthropogenic source is from its primary use as a chemical intermediate; however, it is also released from automobiles, coal-fired power plants, biomass combustion, fish processing, combustion of refuse and plastics, petroleum manufacture, and manufacture of synthetic fibers, starch and rubber (NLM, 2007). Most recently, it has been used as a grain fumigant. Natural emissions of COS from deciduous and coniferous trees, volcanoes, salt marshes, and soil have been estimated to be greater than anthropogenic emissions (NLM, 2007). According to the U.S. EPA Toxics Release Inventory (TRI), over 21 million pounds of COS were released in the United States during 2005, a level that has remained relatively constant since 1988 (U.S. EPA, 2006b).

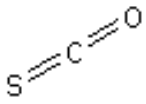



**Figure 3- 1 Chemical structure of Carbonyl Sulfide (COS)**

Carbonyl sulfide has drawn appreciable scrutiny in recent years, in part because of its increasing use as a grain fumigant to replace methyl bromide. It is also known to pose neurotoxic risks at extremely high exposure levels (typically hundreds or thousands of ppm for a duration of minutes to hours).

If released to the atmosphere by point or mobile source emissions, carbonyl sulfide has a long atmospheric residence time, ranging from 200 to 7,300 days. Atmospheric removal of carbonyl sulfide may occur by a slow gas-phase reaction with photochemically produced hydroxyl radicals or oxygen, direct photolysis, and unknown removal processes invoked to balance the sulfur cycle. Direct photolysis is not expected to occur in the troposphere (the first 10 miles of the atmosphere) but it may in the stratosphere (NLM, 2007). When COS is released to moist or dry soils or to water it is expected to rapidly volatilize to the atmosphere. The estimated half-life for volatilization from a model of river contamination is approximately 2 hours. Carbonyl sulfide is not

expected to adsorb to sediment or suspended organic matter, nor is it expected to bioconcentrate in fish and aquatic organisms (NLM, 2007).

People are exposed to carbonyl sulfide through inhalation due to its background presence in air, ingestion/inhalation from natural occurrence in food, through endogenous production, and from direct or secondary cigarette smoke (NLM 2007; Bartholomaeus and Haritos, 2005). COS may be produced in humans as a by-product of aerobic metabolism. Levels of carbonyl sulfide in the atmosphere have been measured in rural and urban sites and over oceans and salt marshes, which are major natural sources. In the United States, the reported concentration of carbonyl sulfide in the air ranged from 0.27 $\mu g/m^3$ to 0.80 $\mu g/m^3$ in rural areas and from 1.17 $\mu g/m^3$ to 1.37 $\mu g/m^3$ in urban areas. (The relative urban and rural levels of COS may now change with its new use as a grain fumigant.) Over the ocean, COS levels were between 14 $\mu g/m^3$ to 19 $\mu g/m^3$ (NLM, 2007). Reported concentrations of COS over salt marshes were even higher, ranging from 60 $\mu g/m^3$ to 180 $\mu g/m^3$. For reference, the range in COS concentration of 0.27 $\mu g/m^3$ to 180 $\mu g/m^3$ corresponds to 0.00011 ppm to 0.073 ppm[13] (or 0.11 ppb to 73 ppb).

While the mechanism(s) for COS-induced neurotoxicity are unknown, the predominant theory for its mode of action is metabolism by carbonic anhydrase to hydrogen sulfide ($H_2S$), followed by inhibition of cytochrome c oxidase —a key mitochondrial enzyme for the production of chemical energy (ATP) in both neurons and other cells (Chengelis and Neal 1980). Unlike its chemical relative, carbon disulfide ($CS_2$) that acts in the peripheral nervous system, carbonyl sulfide (COS) acts primarily in the brain. Decrease in ATP (energy) production in brain neurons can lead to cell death and may eventually result in chronic functional deficits such as those of chronic noise induced hearing loss or Parkinson's disease. Carbonyl sulfide has not yet been shown to pose the same chronic cardiovascular risks that have been documented in workers chronically exposed to $CS_2$ (Chang et al., 2006). However there are presently no chronic lifetime exposure bioassays or human epidemiological studies capable of detecting this.

Cytochrome *c* oxidase (Cyt c Oxidase) is a rate-limiting enzyme in the last step of the electron transport chain. It helps to establish the chemiososmotic potential necessary to drive adenosine triphosphate (ATP) synthesis. ATP is the critical energy source for the majority of cellular functions. Interfering with steps in the electron transport chain can stop the process and lead to cell death. Sills et al (2004) developed a working hypothesis for carbonyl sulfide neurotoxicity whereby COS inhibits mitochondrial cytochrome c oxidase, resulting in ATP depletion, followed by neuronal injury and glial cell response. We adopt their framework for the purpose of this study with modifications to reflect a wider range of clinical endpoints that could result from necrosis of central nervous system neurons (Figure 3- 2).

[13] This conversion assumes a temperature of 25 °C and an atmospheric pressure of 1 atm. The molecular weight of COS is 60 g/mol.

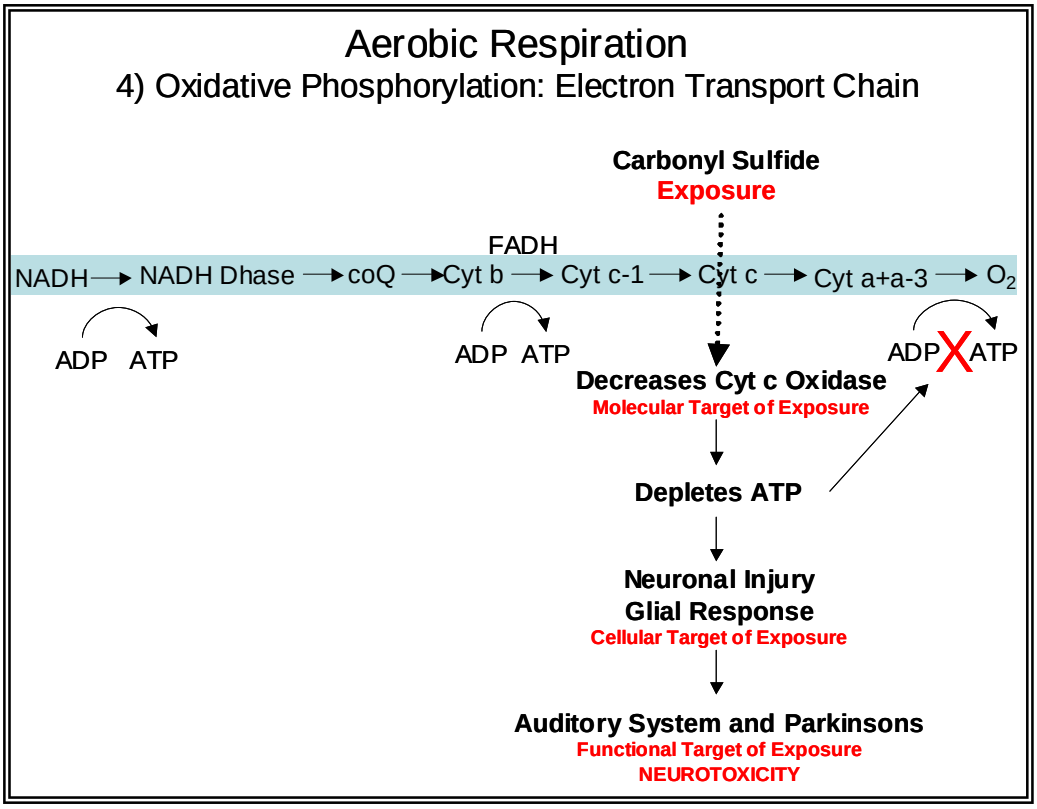



**Figure 3- 2 Possible mode of action for carbonyl sulfide exposure in rats** (Adopted from Sills et al., 2004).

Other researchers reported a concentration-related decrease in cytochrome c oxidase activity in the posterior colliculus and the parietal cortex areas of the brain of male and female rats exposed to COS for 24, 52, and 86 days (Figure 3-3) (Morgan et al., 2004). This supports the mode of action proposed above. Cytochrome oxidase activity in exposed male rats was only 58% of that in controls at day 24 but some apparent recovery of activity was seen at day 52 (Morgan et al., 2004). The inhibition of cytochrome c oxidase in various parts of the brain indicated appreciable (and statistically significant) effects at lower exposure levels (down to 200 ppm) than the brain morphology/histology observations. This effect on a vital energy-related function raises the possibility that these effects of carbonyl sulfide could have important interactions with the actions of other pathological processes that cause the deaths of neurons on a cumulative basis in such conditions as Parkinson's disease, chronic noise-induced hearing loss, and possibly Alzheimer's.

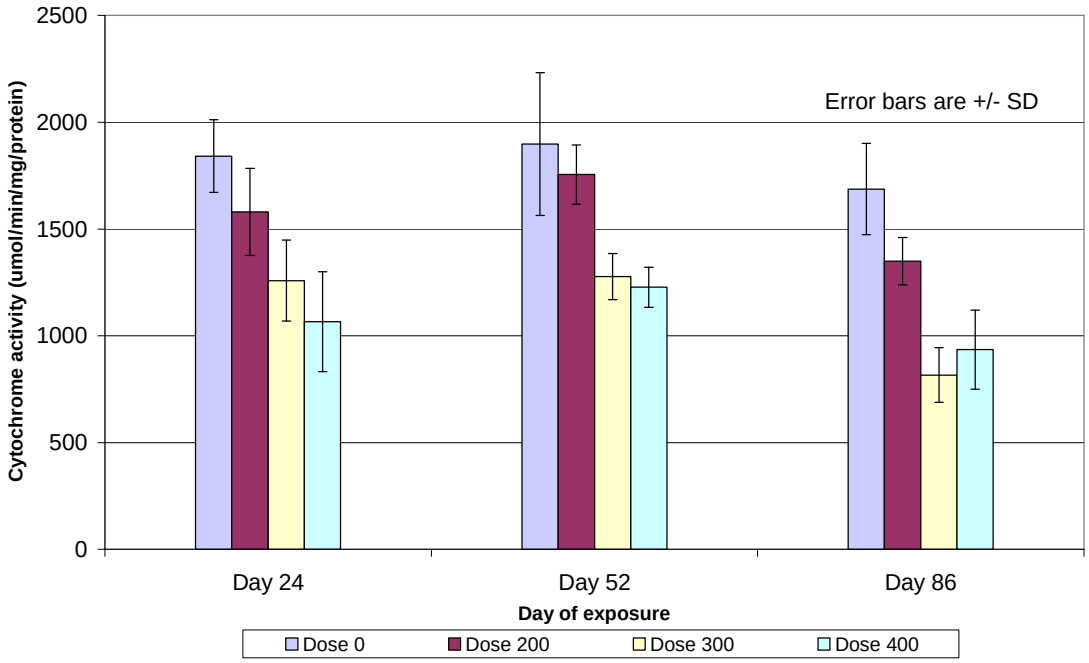



**Figure 3- 3 Cytochrome oxidase activity in parietal cortex of rats exposed to carbonyl sulfide by dose and duration.**
(Source: Morgan et al., 2004; Study dose units in ppm).

### 3.1.1 COS Critical Effect Study Identification

In September 1991, the U.S. EPA RfD/RfC Work Group for the Integrated Risk Information System (IRIS) database reviewed the health effects data for COS and found it to be inadequate for the derivation of an inhalation RfC. The Agency Work Group Review conducted a screening-level literature review in September 2002, but did not identify any critical new studies (U.S. EPA 2007). However, we uncovered several COS studies published since 2002 in our literature review that might support the development of an RfC using neurotoxic effects as the most sensitive health endpoint of COS exposure in rats. Furthermore, we felt it important to evaluate the model capabilities for one case study compound for which there was no IRIS-identified critical effect study or reference level, since this is a common scenario for environmental contaminants, despite the fact that IRIS contains health information for hundreds of compounds.

From all the studies identified, we selected one study and one health endpoint to facilitate comparisons across the three models. The study was published by Morgan et al. in 2004 and examined neurotoxicity associated with COS inhalation in rats. Both male and female F344 rats were dosed with COS according to an occupational dosing schedule of 6 hours per day for 5 days per week for two weeks. Doses ranged from 0 to 500 ppm. There were approximately 10 male and 10 female animals in each dose group. The control group breathed filtered, conditioned air. Another experiment was

performed for 12 weeks with the same number of animals, but COS doses ranged from 0 to 400 ppm, since 500 ppm concentrations were found to be acutely toxic in the 2-week exposure group.

Despite the potential importance of the biochemical observations presented in Morgan et al., 2004, our application of the models was based on the quantal observations of unequivocally adverse morphological changes in the most sensitive of the many areas of the brain where effects were observed. Rats exposed to 400 ppm COS showed lesions in the parietal cortex (and posterior colliculus). Lesions were observed less frequently in the putamen, thalamus, and anterior olivary nucleus (Morgan et al., 2004). The reported dose-response relationships were expressed as number of individual animals with the effect in relation to the number examined. Parietal necrosis results were presented for animals dosed for 2 and 12 weeks. Unfortunately, a quantitative measure regarding the severity or extent of neuronal death necessary for an animal to be classified as having necrosis in a particular brain region was not reported. However the pictures provided by the authors in supplementary information indicated that these classifications often involve extensive destruction of neural cells. Some of those animals classified as not having necrosis may nonetheless have still experienced some destruction of neurons and supporting glial cells.

#### 3.1.2 Data Inputs for COS Case Study

Using the 2- and 12-week, male and female F344 rat parietal necrosis data from Morgan et al. (2004), input data files were created (Table 3-1). The data files included variables for COS dose (ppm), as well as the total number of subjects and the incidence of lesions by gender and dose group. Since the exposure data from Morgan et al. (2004) were based on an occupational dosing schedule of 6 hours per day for 5 days per week, the animals were exposed approximately 18% of the hours in a week. As such, the doses in Table 3-1 have been converted to putatively equivalent 24-hour/7-day continuous doses, the usual basis for expressing RfCs. (A more sensitive endpoint, the inhibition of cytochrome c oxidase as illustrated in Figure 3-3, is examined in **Appendix B**.)

**Table 3-1 Incidence of Brain Lesions in the Parietal Cortex in Male and Female F344 Rats Exposed to COS for 2 and 12 weeks**

| Exposure Duration | Study Dose (ppm) | Continuous Dose (ppm)^ | Incidence of lesions | Total number of subjects | Probability of Effect |
|---|---|---|---|---|---|
| 2 weeks | 0 | 0 | 0 | 20 | 0% |
| | 300[a] | 54 | 1 | 20 | 5% |
| | 400 | 71 | 13 | 20 | 65% |
| | 500 | 89 | 16 | 16 | 100% |
| 12 weeks | 0 | 0 | 0 | 20 | 0% |
| | 300 | 54 | 0 | 20 | 0% |
| | 400 | 71 | 9 | 20 | 45% |

^Dose adjusted to a 24-hour 7 day/week continuous exposure.
(Source: from Morgan et al., 2004 with additions by Abt Associates, Inc.).
[a] The dose of 300 ppm was selected as the NOAEL by Abt Associates Inc.

## 3.2 Application of the Traditional Noncancer Risk Assessment Approach to COS

### 3.2.1 Estimating a Reference Concentration (RfC)

It should be noted that we identify an RfC in this report strictly for the purposes of comparing this value with one that might be estimated using the BMDS, CatReg, and Straw Man models. Using the traditional noncancer risk assessment approach as we have defined it, the RfC is estimated based on the point of departure from the critical effect study, the NOAEL or the LOAEL.

Based on a review of the experiments by Morgan et al. (2004), the no observed adverse effects level (NOAEL) for 2-week exposure to carbonyl sulfide was determined by Abt Associates to be 300 ppm. Using Haber's rule[14], the NOAEL of 300 ppm is converted to a continuous (24 hours/day for 7 days/week) exposure of 54 ppm by noting that the animals were exposed approximately 18% of the time (6 hours/day for 5 days/week).

We estimate the composite uncertainty factor (the product of the relevant uncertainty factors) to be 300— the product of 10 for human interindividual variability ($UF_H$), 10 for

[14] Haber's rule states that equal values of concentration*time are expected to produce equal biological effects. This would be expected, for example, if the agent causes effects that depend on a reaction between the toxicant and a cellular macromolecule that is not ordinarily reversed on the time scale of the exposures. This assumption can either under- or over-state the biologically effective dose and resulting adverse effects that would take place during a continuous exposure delivering the same product of toxicant concentration and time, depending on the details of pharmacokinetic handling and the pharmacodynamic processes leading to cell death or impaired function.

subchronic/chronic exposure uncertainty ($UF_S$), and 3 for uncertainty relating to the absence of potentially salient studies from database (D, in this case a developmental neurotoxicity study, which we believe is indicated because of the apparent central neurotoxic action of this toxicant). No adjustment was made for interspecies extrapolation ($UF_A$), because we assumed a 1:1 concordance of effect between the observed adverse events in the experimental rat and the parallel signs and pathologies in the exposed human. (The slower metabolism rate expected in humans is exactly offset by the slower uptake rate expected in humans relative to the test animals.)

Therefore, the RfC, as estimated using the traditional noncancer risk assessment method is 54 ppm (NOAEL, the point of departure) divided by the composite uncertainty factor of 300, or 0.18 ppm.

$$RfC_{Traditional} = \frac{NOAEL_{continuous}}{UF_{\text{int}erindividual} UF_{subchronic} UF_{animal} UF_{database}} = \frac{54\,ppm}{10x10x1x3} = 0.18\,ppm$$

## 3.3 Application of the BenchMark Dose Software to COS

### 3.3.1 Estimating a Dose-Response Function and Deriving an RfC based on the $BMC_L$

Benchmark Dose Software (BMDS) (Version 1.4.1) was used to develop a dose-response function to estimate the probability of necrosis in the parietal cortex of the laboratory animals as a function of exposure to carbonyl sulfide. The software then uses this estimated dose-response function to estimate the concentration corresponding to a specified percent increase in extra risk[15] (the benchmark concentration, BMC) and the lower confidence limit on the benchmark concentration ($BMC_L$). Using this method, the $BMC_L$ is the point of departure. We employed a 10% increase in extra risk.

There were an insufficient number of data points (3) to generate reliable estimates of model parameters for the 12-week study results (the model over fit the data). This highlights a limitation of the BMDS: as with any model, the number of data points must generally be greater than the number of parameters to be estimated (three in the case of a probit model with a background term). The traditional noncancer risk assessment method may still be used to determine a point of departure in this case. However, we were able to fit several models to the 2-week data.

The BMDS model parameters and fit statistics for the 2-week Morgan et al. data are shown in Table 3-2. The chi-square goodness of fit p-value tests the difference between the full-fitted model (whose number of parameters are sufficient for the model to pass through all data points) and the model with the intercept (or power) and slope indicated in the table. All models, except the Quantal-Linear model provided a good fit to the 2-week data. These models had chi square p-values were greater than 0.1 indicating no

[15] Extra Risk, as defined in the BMDS model, is the additional risk divided by the proportion of animals that will not respond in the absence of exposure, 1 - P(0).

difference between the specified model and the full-fitted model[16]. Akaike's Information Criterion (AIC) is also used to compare model fits using different functions (log-probit, Weibull, linear, etc.) The preferred model is the one with the lowest AIC value. With one exception, the log-probit models provided the best fit to the animal dose-response data. For the 2-week parietal necrosis outcome, the Weibull model provided a somewhat better fit (AIC = 38.0) than the log-probit model (AIC =39.7), but the difference in AIC is so small that it was outweighed by our desire to use the log-probit fit for comparability with the CatReg and Straw Man outputs. We use the log-probit model to determine the point of departure and to estimate risk levels around the RfC.

16 Note that in our case, we have only four dose groups (one of them the control group) and the Chi square statistic may not be an appropriate goodness of fit test.

**Table 3-2 Selected BMDS Model Results for dose-response models of COS Exposure and Risk of Parietal Necrosis**

| Model | BMC (ppm) | $BMC_L$ (ppm) | Background | Intercept /Power[a] | Slope | AIC[b] | Chi Square Goodness of Fit p-value[c] |
|---|---|---|---|---|---|---|---|
| Log-Probit | 55 | 49[d] | 0 | -27.02 | 6.4 | 39.7 | 0.5427 |
| Weibull | 57 | 48 | 0 | 8.75 | 4.8E-17 | 38.0 | 0.9390 |
| logistic | 57 | 49 | 0 | -48.91 | 11.6 | 39.9 | 0.5172 |
| Gamma | 49 | 44 | 0 | 18 | 0.3 | 41.0 | 0.3678 |
| Quantal Linear (Weibull) | 9 | 7 | 0 | n/a | 0.01 | 66.5 | 0.0001 |

Notes:
ppm = parts per million
BMC and $BMC_L$ for all models are based on a 10% increase in extra risk
The natural logarithm of the dose is used in the log-probit model and the gamma model.
[a] The log-probit and logistic models estimate intercepts; the Weibull, gamma, and quantal linear (Weibull) estimate powers.
[b] Lower Akaike's Information Criterion (AIC) indicates a better model fit.
[c] p-value tests whether the full model (that goes through all points) differs from the fitted model, with the indicated intercept (or power) and slope. Generally, $p > 0.1$ indicates an acceptable fit.
[d] Point of departure selected as $BMC_L$ of 49 ppm from the log-probit fit
Log-probit model form: P[response] =
Background + (1-Background)·CumNorm(Intercept+Slope·log(Dose)),
where CumNorm(.) is the cumulative normal distribution function.
Weibull model form: P[response] =
$$\text{Background} + (1-\text{Background})\cdot(1-\exp(-\text{Slope}\cdot\text{Dose})^{\text{Power}})$$
Logistic model form: P[response] =
$$1/(1 + \exp(-\text{Intercept}-\text{Slope}\cdot\text{Dose}))$$
Gamma model form: P[response] =
Background + (1-Background)·(1/Γ(Power))·(integral from 0 to Dose of $t^{\text{Power}-1}\cdot\exp(-t)dt$)
Where Γ(t) is the gamma function, or the integral from 0 to ∞ of $t^{-1}\cdot\exp(-t)dt$
Quantal Linear (Weibull) model form: P[response] =
Background + (1-Background)·(1-exp(-Slope·Dose))

In each model fit, BMDS software estimated the background term as zero ($\gamma = 0$). This reduces the log-probit model form to Equation (2). Using the log-probit results, the $BMC_L$, which is the point of departure, for the 2-week study was 49 ppm, a factor of 1.1 lower than the BMC, indicating that taking the lower bound on the BMC provides a point of departure that is approximately 10% lower than the BMC (using this metric to denote a level of protection was noted in Castorina and Woodruff, 2003).

The fitted BMDS dose-response function is approximately "S-shaped" as shown (dashed line) using the 2-week exposure study data (squares) for necrosis in the parietal cortex in Figure 3- 4. The slope for the BMDS fit is steep between approximately 40 and 100 ppm and nearly flat below 40 ppm and above 100 ppm. Note the estimated RfC (0.18 ppm) is located on the dose-response function where the slope is nearly zero.

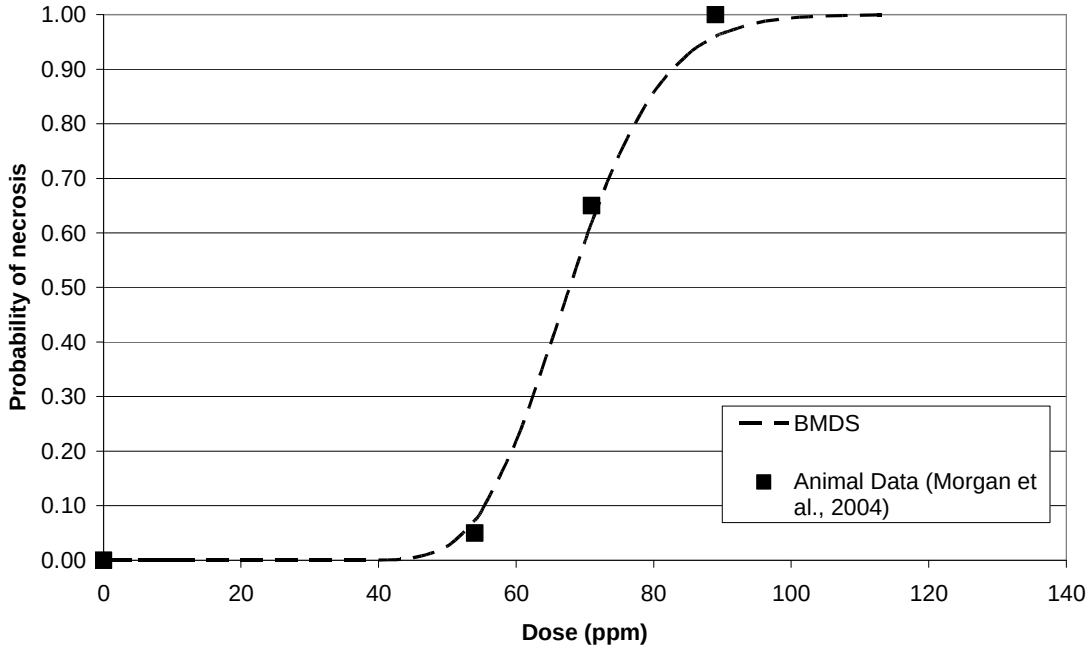



**Figure 3- 4 Comparison of Dataset to BMDS Log-Probit Model results for Probability of Parietal Necrosis due to Two Week Carbonyl Sulfide Exposure**

The reference concentration (RfC) determined using BMDS is calculated by dividing the $BMC_L$ (point of departure) by the appropriate uncertainty factors. As in the traditional noncancer framework, the uncertainty factors selected were 10 for human interindividual variability ($UF_H$), 10 for subchronic to chronic uncertainty ($UF_s$), 1 for animal to human ($UF_A$), and 3 for database uncertainty (D) due to lack of neurodevelopmental studies, for a composite uncertainty factor of 300. This results in an RfC of 0.16 ppm.

$$RfC_{BMDS} = \frac{BMC_L}{UF_{\text{int}erindividual} UF_{subchronic} UF_{animal} UF_{database}} = \frac{49\,ppm}{10x10x1x3} = 0.16\,ppm$$

### 3.3.2 Estimates of COS-Risk near the RfC using the BMDS Model Fit

Using the log-probit model fits, we determined the BMDS-predicted risk levels at concentrations around the point of departure ($BMC_L$) and reference concentration (RfC). For comparability between methods, we use an RfC of 0.18 ppm (as estimated by the traditional noncancer risk assessment method). Risks are computed at the RfC and an

order of magnitude above (1.8 ppm) and below (0.018 ppm) for consistency with Straw Man output, but they also reflect exposures that might be typical for humans and thus the range at which benefits would be estimated. COS concentrations in air are on the order of 0.27 $\mu g/m^3$ (0.00011 ppm) in rural areas and 180 $\mu g/m^3$ (0.073 ppm) over salt marshes. Thus, humans are likely to be exposed to levels well below the RfC and policy changes would be expected to bring about small changes below ambient levels. Using the BMDS models, risk estimates were nearly zero (to 30 decimal places) for the RfC (Table 3-3). This is consistent with the dose-response curve shown in Figure 3-5 that is flat below 40 ppm. At the $BMC_L$ (point of departure), which is approximately 300 times larger than the estimated RfC, the risk was determined to be approximately 2%.

Castorina and Woodruff (2003) recently proposed a method for estimating risk at the RfC/RfD using BMDS output. For sublinear dose-response functions, like log-probit, the authors used the BMDS model to estimate risk at the RfC as we have done. However, for linear dose-response functions, the authors point out that the point of departure ($BMC_L$) results in a benchmark response level of risk (BMR); thus to obtain the level of risk at the RfC (= $BMC_L$/UF), one might divide the BMR by the composite uncertainty factor. In other words, they propose dividing the BMR (in our case, 10%) by the composite uncertainty factor (in our case, 300) to estimate risk at the RfC. Performing this calculation would yield a 3.3E-04 risk at the RfC. Similarly, the risk at 0.1·RfC would be 3.3E-05 and the risk at 10·RfC would be 3.3E-03. Although the dose-response function for COS does not appear to be linear, we provide these risk values as examples of an upper bound level of risk to compare to the BMDS model-predicted negligible risks (Table 3-3).

**Table 3-3 Estimated risks at the point of departure ($BMC_L$) and around the RfC for the 2-week parietal necrosis data as predicted by BMDS log-probit model and linear extrapolation**

| Endpoint | Model type | $BMC_L$ (ppm) | Risk at the $BMC_L$ | Risk at the RfC^ | Risk at the 0.1 RfC^ | Risk at the 10 RfC^ |
|---|---|---|---|---|---|---|
| Parietal necrosis | Log-probit | 49[a] | 2.0E-02 | 0.0E+00 | 0.0E+00 | 0.0E+00 |
| Parietal necrosis | Linear* | 49 | 1.0E-01 | 3.3E-04 | 3.3E-05 | 3.3E-03 |

[a] $BMC_L$ = point of departure = 49 ppm

^ Using an RfC = 0.18 ppm; RfC/10 = 0.018 ppm; RfC*10 = 1.8 ppm

* Using a linear extrapolation method suggested by Castorina and Woodruff (2003) for illustrative purposes only as an example of upper-bound risk levels.

## 3.4 Application of CatReg to COS

### 3.4.1 Estimating a Dose-Response Function and Deriving an RfC based on the Extra Risk Concentration (ERC)

CatReg was used to develop models to estimate the probability of necrosis in the parietal cortex as a function of carbonyl sulfide exposure and duration as detailed in the Table 3 and Table 6 data reported by Morgan et al. (2004). A cumulative odds model with a probit link and log of the exposure dose and duration was used to estimate model parameters (See **Appendix C** for data input file). The background risk was assumed to be zero which is consistent with the data and the BMDS model fit. CatReg was used to explore the impact of COS dose and duration of exposure on a given health endpoint (i.e., necrosis of tissue in the parietal cortex). The software was also used to test for differences in the dose-response relationships by sex (male, female), dose level (0, 300, 400, 500 ppm), and duration (2-weeks versus 12-weeks).

In a CatReg model for necrosis in the parietal cortex that included both 2- and 12-week durations, concentration (i.e., dose), but not duration, was a significant predictor of effect. In fact, the exposure duration parameter was negative, suggesting a decrease in risk with time, which led the model to terminate the run prematurely (no ERC10 or ERC10-L outputs were generated) (Table 3-4). Problems with the run include insufficient data to estimate the number of parameters in the model as well as the inability of CatReg to accommodate health effects that do not get more severe over time (i.e., milder health effects with longer exposure duration.) In the study data (Table 3-1), at 300 ppm, 5% of animals dosed for 2-weeks displayed parietal necrosis compared to none of 12-week animals. Similarly, at 400 ppm, 65% of 2-week animals responded compared to only 45% of the 12-week animals.

**Table 3-4 CatReg parameter estimates for the probability of necrosis in the parietal cortex as a function of COS exposure concentration and duration (2 and 12 weeks)**

| Model | Intercept ($\alpha$) | Concentration Slope ($\beta_1$) | Duration Slope ($\beta_2$) | Chi Square Goodness of fit p-value |
|---|---|---|---|---|
| log-probit | -35.05 | 19.32 | -0.82^ | n/a* |

Model form: $\Pr(Y \geq S | C,T) = H[\alpha_s + \beta_1 * f_1(C) + \beta_2 * f_2(T)]$

*H* is a probability function taking values between 0 and 1, and is based on the Gaussian distribution.

^CatReg has a non-negativity constraint on parameters that leads to termination of the program before ERC10 and ERC10-L values were determined

*The program did not provide analysis of deviance statistics due to early termination.

We next conducted a stratified analysis restricted to the 2-week exposure group. It showed that the risk of adverse responses in both male and female rats varied significantly by concentration level. A comparison of the sex-specific results indicated that the parameter estimates for female rats were overall slightly higher than for male rats, but not significantly so at the 0.05 level (p = 0.08762, data not shown). The final

model to predict necrosis of the parietal cortex as a function of COS exposure was based on the 2-week exposure data, combining the results of both sexes, as was done in the BMDS approach (Table 3-5). CatReg estimates the concentration corresponding to a 10% extra risk[17] in the health effect. In CatReg, the 10% extra risk concentration (ERC10) has a lower bound confidence interval that can be treated as the point of departure on the dose-response curve. For the 2-week parietal cortex model, the lower bound on the 10% extra risk concentration (ERC10-L) was 53 ppm, slightly higher than the point of departure predicted by BMDS.

**Table 3-5 CatReg parameter estimates for the probability of necrosis in the parietal cortex as a function 2-week exposure to Carbonyl Sulfide (cumulative odds model with a probit link).**

| Model | ERC10 (ppm) | ERC10-L (ppm) | Intercept ($\alpha$) | Concentration Slope ($\beta_1$) | Chi Square Goodness of fit p-value^ |
|---|---|---|---|---|---|
| log-probit | 57 | 53[a] | -33.78 | 18.49 | 0.464 |

[a] ERC10-L = point of departure = 53 ppm

Model form: $\Pr(Y \geq s|C) = H[\alpha_s + \beta_1 * f_1(C)]$

*H* is a probability function taking values between 0 and 1, and is based on the Gaussian distribution.
^ Tests whether parameter estimates are significantly different from 0.

To estimate the reference concentration (RfC) using CatReg, we divide the ERC10-L (point of departure) by the previously derived composite uncertainty factor (300, corresponding to the product of 10 for human interindividual variability ($UF_H$), 10 for subchronic to chronic uncertainty ($UF_s$), and 3 for database uncertainty (D) due to lack of developmental studies). No adjustment was made for interspecies extrapolation because we assumed a 1:1 concordance of effect ($UF_A$). The RfC was approximately 0.18 ppm.

$$RfC_{Cat\,Reg} = \frac{ERC10 - L}{UF_{interindividual} UF_{subchronic} UF_{animal} UF_{database}} = \frac{53\,ppm}{10x10x1x3} = 0.18\,ppm$$

### 3.4.2 Estimates of COS Risk near the RfC using CatReg Model fits

Using the 2-week log-probit model fit, the risks at concentrations equal to the lower bound on the 10% extra risk concentration (ERC10-L) and near the traditional RfC (0.18 ppm) were computed (Table 3-6).

[17] Extra risk (ER) adjusts for background incidence rates of the same effects, by estimating risk at dose, d, only among the fraction of the population not expected to respond to the secondary (background) causes: ER = [P(d)-P(0)/1-P(0)]. Note that this is the Abbott correction described in Section 4.3.2.1.

**Table 3-6 Estimated risk predicted by CatReg at the point of departure (ERC10-L) and around the RfC**

| Endpoint | Study Length | ERC 10-L (ppm) | RfC (ppm) | Risk at ERC10-L | Risk at RfC^ | Risk at 0.1 RfC^ | Risk at 10 RfC^ |
|---|---|---|---|---|---|---|---|
| Parietal necrosis | 2 weeks | 53 | 0.18 | 2.9E-02 | 0.0E+00 | 0.00E+00 | 0.00E+00 |

ERC10-L = point of departure = 53 ppm

^ Using an RfC = 0.18 ppm; RfC/10 = 0.018 ppm; RfC*10 = 1.8 ppm

Risk estimates were zero (to 30 decimal places) for concentrations near the RfC, but almost 3% at the point of departure. Again, for illustrative purposes, if we assume linearity of risk down to zero, exposures at the ERC10-L would result in 10% risk, and exposures at the RfC (RfC = ERC10-L/300) would result in 0.033% risk.

## 3.5 Application of the Straw Man to COS

### 3.5.1 Estimating a Dose-Response Function and Deriving an RfC with Uncertainty Distributions

We estimated a reference concentration (RfC) for COS exposure based on the 2-week rat data for parietal necrosis using the Straw Man approach. As explained in Section 2.3.3, the Straw Man approach is a Monte Carlo simulation of distributions for the point of departure and uncertainty factors. The output is a distribution of doses corresponding to various risk levels. Following the Straw Man criterion of setting the RfC as the daily dose rate that is expected (with 95% confidence) to produce less than 1/100,000 excess incidence over background of a minimally adverse response in a standard general population of mixed ages and genders, we select the RfC as the 5$^{th}$ percentile value of the Monte Carlo output distribution.

$$\hat{\mathrm{RfC}} = \hat{\mathrm{Dose}}\left(\frac{1}{100{,}000}\right) = \hat{\mathrm{EC}}_{50} \cdot \hat{\mathrm{UF}}_{\mathrm{A}} \cdot \hat{\mathrm{UF}}_{\mathrm{S}} \cdot \hat{\mathrm{D}} \cdot 10^{\sqrt{\left(\log\hat{\mathrm{GSD}}_{\mathrm{PK}}\right)^2 + \left(\log\hat{\mathrm{GSD}}_{\mathrm{PD}}\right)^2} \cdot \mathrm{probit}\left(\frac{1}{100{,}000}\right)}$$

All of the uncertainty distribution terms in the equation are summarized in Table 3-7. First, a log-probit distribution is fit to the Morgan et al. study data ($EC_{50}$). Second, using data from Baird et al., 1996, a lognormal distribution is employed for the subchronic to chronic projection ($UF_s$). Third, following Evans and Baird (1988) a constant value of 3 was employed for database insufficiency (D, lack of developmental studies). Fourth, a central estimate of 1 was chosen to reflect concordance between effects seen in animals compared to humans ($UF_A$). However, the Straw Man uses a lognormal distribution for the animal to human extrapolation, with a GM of 1 and a GSD of 2.6 (following Price et al. 2008). Fifth, Straw Man employs an exogenously derived human interindividual variability term consisting of a pharmacokinetic and a pharmacodynamic component to utilize what is already known about interindividual variability to better characterize the variability in responses ($UF_H$).

**Table 3-7 Geometric means and standard deviations for uncertainty distributions applied in the Straw Man analysis of COS.**

| Term | Uncertainty distribution Type | Geometric Mean | Geometric Standard Deviation | Source |
|---|---|---|---|---|
| $\hat{E}C_{50}$ | Log-Probit | 67.20 | 1.03 | Morgan et al., 2004 |
| $\hat{U}F_S$ | Lognormal | 0.499 | 2.17 | Baird et al., 1996 |
| D | Point | 3 | 0 | Evans and Baird, 1988 |
| $\hat{U}F_A$ | Lognormal | 1.0 | 2.6 | Price et al. 2008 |
| $\hat{U}F_H$ $logGSD_{PK}$ | Lognormal | 0.202 | 0.092^ | Hattis and Lynch, 2007 |
| $\hat{U}F_H$ $logGSD_{PD}$ | Lognormal | 0.246 | 0.161^ | Hattis and Lynch, 2007 |

$\hat{E}C_{50}$: Effective dose associated with a biological effect in 50 percent of the animals (mg/kg/day).
$\hat{U}$Fs: Uncertainty distribution for subchronic to chronic extrapolation.
D: Uncertainty point estimate for database deficiency (can also be distribution)
$\hat{U}F_A$: Uncertainty distribution for interspecies projection.
$\hat{U}F_H$ GSD,PK: Uncertainty distribution for pharmacokinetic variability in humans.
GSD,PD: Uncertainty distribution for pharmacodynamic variability in humans.
^These are the standard deviations of the distributions of the log[log(GSD)]s for the pharmacokinetic and pharmacodynamic variability, based on the general finding that the log(GSD)s for these component variabilities themselves have uncertainties that are approximately lognormally distributed (Hattis and Lynch, 2007).

The output of the Monte Carlo distribution of carbonyl sulfide concentrations (and their likelihood) that produce a 1 in 100,000 increased risk of parietal necrosis is shown in Table 3-8. The concentration necessary to meet the Straw Man risk criterion is about 0.06 ppm or about one third of the RfC of 0.18 ppm estimated by the traditional noncancer risk assessment method. Note that selecting the median or mean concentrations from the output distribution would yield RfC estimates that are 5 to 16 times greater (less conservative) than those estimated using POD derived from the traditional noncancer risk assessment, BMDS, and CatReg methods.

**Table 3-8 Straw Man Distribution of Concentrations Associated with a 1 in 100,000 Increased Risk of Parietal Necrosis^**

| Percentile | Concentration (ppm) | Ratio of Concentration to RfC^^ |
|---|---|---|
| 5th %tile | 0.059 ¥ | 0.33 |
| Median | 0.99 | 5.4 |
| Mean | 2.9 | 16 |
| 95th %tile | 11 | 61 |

Notes:
¥ This is the Straw Man-proposed RfC.
^The Straw Man Criterion is:
The daily dose rate that is expected (with 95% confidence) to produce less than 1/100,000 excess incidence over background of a minimally adverse response in a standard general population of mixed ages and genders.
^^RfC = 0.18 ppm (as calculated in Section 3.2.1.)

***Impact of Straw Man's Interindividual Variability Term on the RfC***

To investigate the impact of interindividual variability, the most complex Straw Man uncertainty distribution, on the risk estimates, we conducted another Monte Carlo analysis where the $GSD_{PK}$ and $GSD_{PD}$ terms were collapsed to their central estimates (Table 3-9). In other words, they are treated like point estimates. Without this important source of uncertainty, the Straw Man RfC is 0.12 ppm, or double (less conservative) the Straw Man estimated RfC when the uncertainty in the estimated interindividual variability is included.

**Table 3-9 Straw Man Distribution of Concentrations Associated with a 1 in 100,000 Increased Risk of Parietal Necrosis (without human interindividual variability uncertainty distribution )^**

| Percentile | Concentration (ppm) | Ratio of Concentration to RfC^^ |
|---|---|---|
| 5th %tile | 0.12¥ | 0.66 |
| Median | 1.2 | 6.7 |
| Mean | 2.7 | 15 |
| 95th %tile | 9.4 | 53 |

Notes:
¥ This is the Straw Man-proposed RfC without the uncertainty distribution for interindividual variability.
^The Straw Man Criterion is:
The daily dose rate that is expected (with 95% confidence) to produce less than 1/100,000 excess incidence over background of a minimally adverse response in a standard general population of mixed ages and genders.
^^RfC = 0.18 ppm (as calculated in Section 3.2.1.)

### 3.5.2 Estimates of COS Risk near the RfC using the Straw Man Approach

By definition, the Straw Man model estimates the dose associated with a 1 in 100,000 (or less, with 95% confidence) risk at the RfC. In the case of parietal necrosis from COS inhalation, the RfC estimated in this way was 0.06 ppm (the 5th percentile of the distribution of concentrations that increased risk by 1 in 100,000; Table 3-8). For comparative purposes, we also estimated the uncertainty in risks that might be expected using the Straw Man calculations for the three-fold larger RfC predicted following the traditional noncancer risk assessment framework (i.e., 0.18 ppm) and a factor of 10 above and below this value (i.e., 0.018 ppm and 1.8 ppm). (Table 3-10)

It can be seen in Figure 3-10 that while the median estimate of risk at the 0.18 ppm level is very small (4.0E-11), the uncertainty is so large that an upper 95th percentile risk is nearly 1 in 1,000, and the mean of the risk distribution is even more— over 2 in 1,000. One result of the risk distributions produced by some Straw Man simulations is extreme skewing to the right produced by the combined action of several sources of uncertainty. As can be seen in Table 3-10, this can cause the mean of the distribution to exceed the 95th percentile. This is not an artifact, but a measure of the great overall uncertainty estimated in the system.

**Table 3-10 Straw Man Risk Distributions Around the Traditional RfC (based on 2-week exposures to COS and necrosis of the parietal cortex)**

| Percentile | RfC/10 = 0.018 ppm | RfC = 0.18 ppm | 10·RfC = 1.8 ppm |
|---|---|---|---|
| 5th %tile | very small | very small | 5.0E-13 |
| Median | very small | 4.0E-11 | 2.4E-04 |
| Mean | 5.4E-05 | 2.2E-03 | 4.6E-02 |
| 95th %tile | 2.4E-08 | 8.0E-04 | 2.9E-01 |

***Impact of Straw Man's Interindividual Variability Term on Risk Estimates Around the RfC***

We also examined the Straw Man risk distributions where the human interindividual uncertainty distribution, represented by the the $GSD_{PK}$ and $GSD_{PD}$ terms, what the authors believed in earlier versions of the Straw Man framework to be the largest source of uncertainty, were collapsed to their central estimates (Table 3-11). The mean risk at the RfC (of 0.18 ppm) was nearly unchanged at 2 in 1,000. However, the estimated 95$^{th}$ percentile risk at the RfC/10 was reduced by several orders of magnitude (from 2.4E-08 to 4.4E-12), while the 5$^{th}$ percentile risk at 10·RfC was nearly one-hundred fold higher (from 5.0E-13 to 3.3E-11). If we consider the percentiles levels to correspond to fractions of the population, in effect, collapsing the human interindividual uncertainty distribution means that the most sensitive individuals (95$^{th}$ percentile) face a lower risk at low doses, while the least sensitive individuals (5$^{th}$ percentile) face a higher risk, although virtually the entire population is estimated to have very small risks at this low exposure level.

**Table 3-11 Straw Man Risk Distributions Around the Traditional RfC (based on two week exposures to carbonyl sulfide and necrosis of the parietal cortex) <u>without</u> human interindividual variability uncertainty distribution**

| Percentile | RfC/10 = 0.018 ppm | RfC = 0.18 ppm | 10·RfC= 1.8 ppm |
|---|---|---|---|
| 5th %tile | very small | very small | 3.3E-11 |
| Median | very small | 3.5E-12 | 9.9E-05 |
| Mean | 2.6E-05 | 1.9E-03 | 4.4E-02 |
| 95$^{th}$ %tile | 4.4E-12 | 1.1E-04 | 2.9E-01 |

## 3.6 Comparison Across Approaches in the Carbonyl Sulfide Case Study Summary

A case study of carbonyl sulfide (which has drawn appreciable scrutiny in recent years, in part because of its increasing use as a grain fumigant to replace methyl bromide, and because it is known to pose neurotoxic risks at extremely high exposure levels) was conducted to explore the potential impact on human health benefits assessment of applying different quantitative methods to derive dose-response functions. In this case study, the estimated dose-response functions fit the dataset summarizing COS-induced parietal necrosis in the brain well. Estimated reference concentrations and risk levels are summarized in Table 3-12. Similar RfCs (~ 0.2 ppm) were estimated using the

traditional noncancer risk assessment method, BMDS, and CatReg, while a somewhat lower or more restrictive value was estimated using the Straw Man (0.06 ppm). Estimates of risk around the RfC were similar using BMDS and CatReg, (essentially zero), while Straw Man-based estimates were substantial (2 in 1000). The estimates of different RfCs and risks between the BMDS, CatReg and Straw Man models suggest model selection could strongly affect benefit-cost analysis. These findings are discussed in greater detail below.

One advantage of applying quantitative methods over the traditional noncancer risk assessment approach is that all data points generated in the critical effect study contribute to the determination of the point of departure. However, in this case study, the traditional, BMDS, and CatReg approaches all identified nearly the same POD (Table 3-12). Achievement of similar PODs using different approaches may be due to high quality study design as well as an unequivocal dose-response curve with a steep slope.

Another advantage of the quantitative methods over the traditional noncancer risk assessment approach is they generate a dose-response function, which can be used to estimate risk for a given exposure (i.e., human equivalent dose). In the COS case study, BMDS and CatReg dose-response functions lead to estimates of zero risk at exposures in the range of estimated RfC values (Table 3-12 and Figure 3-5). This is likely due to the steep dose-response curve and limited variability in animal response data. A finding of no risk at these exposure levels also implies there would be no benefits to a regulation, for example, which reduced concentrations of carbonyl sulfide from 1.8 ppm to 0.18 ppm or lower (i.e., from 10·RfC to the RfC or below the RfC). In contrast, the Straw Man model predicted substantial mean risk of effect (2 in 1,000) when the RfC was 0.18 ppm. These results suggest that a probabilistic approach to analysis, such as the Straw Man model, would also have generated substantial estimates of benefits for reduction of exposures near the RfC. Compared to estimates in CatReg, the BMDS probit model estimated slightly higher risk at low doses and slightly lower risk at high doses.

This greater apparent mean risk at the Straw Man RfC may be due in part to the use of skewed (i.e., lognormal) uncertainty distributions for estimating the subchronic to chronic projection, and human interindividual variability. The compounding of right-skewed lognormal input distributions results in mean estimates that are far higher than the median, thereby giving more weight to distributional tails. It is likely that methodological differences, such as the incorporation of uncertainty factors into risk estimation and the application of exogenous interhuman variability terms, also help drive distributional estimates. The human interindividual variability is expected to have a large effect on the risk output, but its limited impact in the COS case study may be due to the severe endpoint (necrosis of brain tissue) in the critical effect study. For such severe endpoints, human interindividual variability is assumed to be limited (since most respond)[18].

[18] The pharmacodynamic variability term used in the Straw Man model is the product of several terms including functional reserve capacity, which varies by endpoint severity. In this example, the severe endpoint term was 5.42E-07, while the moderate and mild endpoint terms ranged from 0.1 to 0.22, respectively.

Therefore, the factors that determine mean risks seem to be dependent on the interaction of the other uncertainty distributions. On the other hand, collapsing the human interindividual variability term did result in differences at the tails of the risk distributions at low and high doses (a factor of 10 below and above the RfC). Namely, collapsing human interindividual variability to its central estimate led to much smaller estimates of risk for the 95$^{th}$ percentile individual at low exposures, such as at or below the RfC.

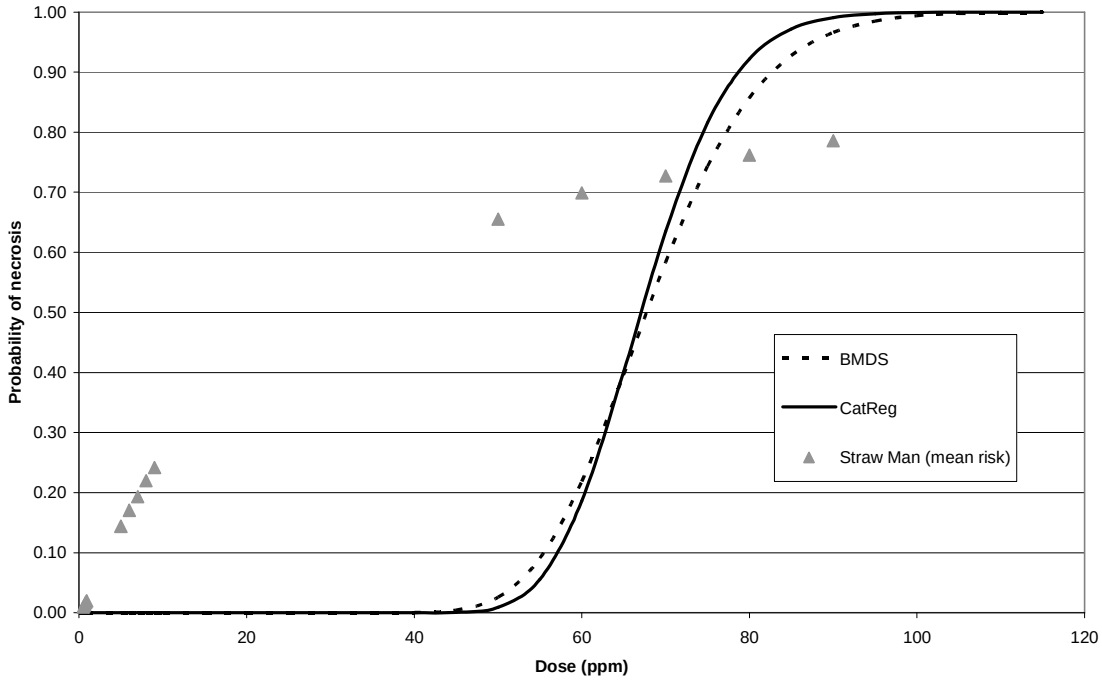
**Table 3-12 Estimated risks of parietal necrosis around an RfC of 0.18 ppm, by model**

| Model | POD (ppm) | RfC (ppm) | Risk at RfC/10 | Risk at RfC | Risk at 10·RfC |
|---|---|---|---|---|---|
| Traditional | 54 | 0.18 | n/a^ | n/a^ | n/a^ |
| BMDS | 49 | 0.16 | 0 | 0 | 0 |
| CatReg | 53 | 0.18 | 0 | 0 | 0 |
| Straw Man | n/a | 0.06¥ | 5.4E-05 | 2.2E-03 | 4.6E-02 |

Notes:
^An estimate (with uncertainty spanning perhaps an order of magnitude) of a continuous inhalation exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious noncancer effects during a lifetime.
¥ While the Straw Man proposes an RfC of 0.06 ppm corresponding to increased risk of parietal necrosis of 1 in 100,000, these risk estimates are the mean values of the Straw Man risk distribution for an exposure level of 0.18 ppm for comparability with the other models.

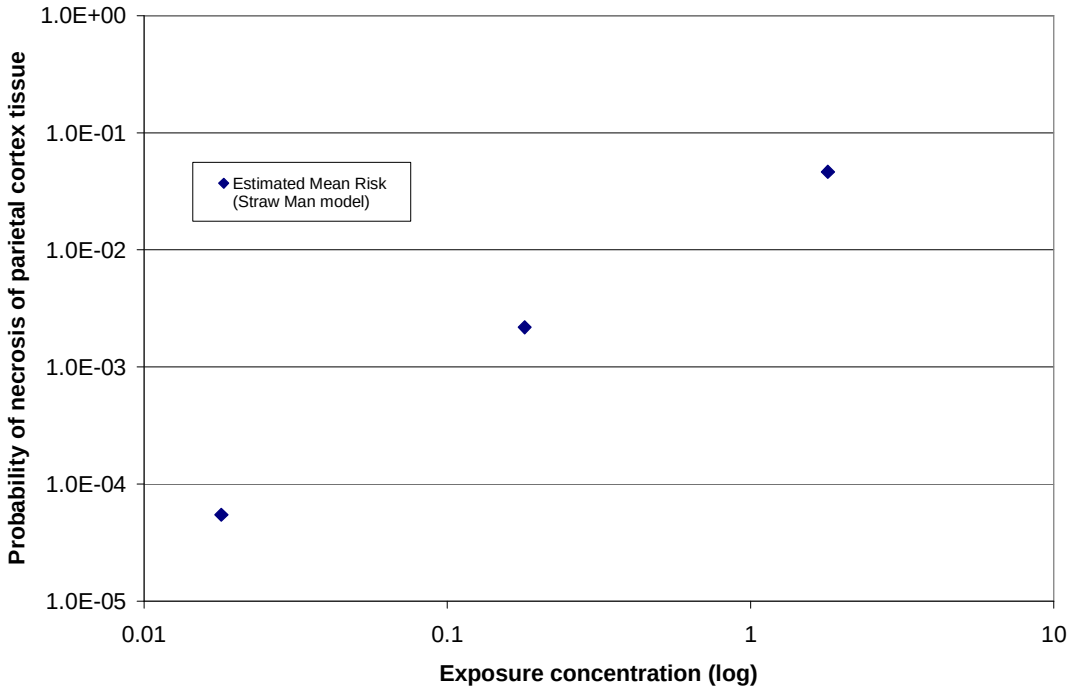
The three quantitative models also estimated different slopes, which lead to different changes in risk for a given change in dose (Figure 3-5). This may be important for benefits analyses, which often involve small shifts in exposure dose. For example, a shift in exposure dose from 80 ppm to 60 ppm leads to an 80 percent decrease in risk in the CatReg model but a 75 percent decrease in the BMDS model. By comparison, the Straw Man model incorporates the uncertainty factors into its dose-response function using uncertainty distributions. The impact of this incorporation is to reduce the slope of the dose-response curve so that the mean response occurs over a dose range at least 3 orders of magnitude (i.e., from about 0.01 to 200 ppm). Using the same example, a shift in exposure from 80 ppm to 60 ppm leads to an 8 percent decrease in risk using the Straw Man model. (It should be noted that potential ambient exposures to COS are all well below 60 ppm.)

**Figure 3- 5 Comparison of BMDS, CatReg, and Straw Man model results for carbonyl sulfide exposure and necrosis of the parietal cortex (Straw Man model includes uncertainty distributions).**




In order to better view the Straw Man estimates of risk at low doses, Figure 3-6 magnifies the Straw Man estimates of risk around the RfC (0.018, 0.18, 1.8 ppm). These risk levels also suggest that there could be differential benefits of a regulation depending on where on the dose-response curve the reduction in exposure occurs. For example, a reduction in lifetime exposure concentration from 1.8 ppm to 0.18 ppm would lead to a change in risk of 5 percent. While an additional reduction in concentration from 0.18 ppm to 0.018 ppm would be associated with a change in risk of 0.2 percent. However, in terms of relative risk these changes are nearly identical, translating to an approximately 95% reduction in risk for each 10-fold reduction in exposure.

**Figure 3- 6 Estimated Straw Man Mean Risk Levels around the RfC (log-log scale)**




On the basis of the COS case study, we found that BMDS and CatReg models provided only slightly different estimates of absolute risk and different rates of change in risk as dose changes (i.e., slopes). The different results between CatReg and BMDS likely occur because of subtle but important differences, which we mention briefly below but do not further quantify. These differences include the mathematical form of the model, the estimation of background incidence, and different approaches to bounding and estimation of the confidence limits. For example, BMDS uses maximum likelihood estimation to generate confidence limits while CatReg uses Wald limits—a potential explanation for the difference seen between the $BMC_L$ and ERC10-L.

The Straw Man model estimated a dose-response curve that was much less steep than the curves estimated by BMDS and CatReg. This stretching of the mean risk at different doses can be attributable to the uncertainty distributions that Straw Man incorporates. The Straw Man results had several important implications for analyses. First was the identification of a different (i.e., 1/3 lower) RfC value than estimated by the other approaches. Second, were quantifiable estimates of risk (and thus benefits) near the RfC. The Straw Man model estimated risk near the RfC occurs in spite of the very steep dose-response curve for the animal data and the use of a severe endpoint for which human interindividual variability is presumed to be limited. This suggests that for the many other chemicals with more gradual dose-response functions and mild to moderate endpoints (which will have greater human interindividual variability), the Straw Man model will be useful for the quantification of benefits near the RfC. On the other hand, the effect of stretching out the risks that correspond to doses—which lead to estimates

of risk near the RfC in this case study, also implies there could be relatively smaller benefits achieved per incremental reduction in exposure, compared to other models.

# 4. Case Study Application of BMDS, CatReg, and Straw Man to 1,2,4,5-Tetrachlorobenzene (TCB)

To help illustrate the difference between the noncancer risk assessment models, we conduct a case study of 1,2,4,5-tetrachlorobenzene (TCB). We first provide background on this compound (hazard assessment), which is expected to act on the kidneys after ingestion. Next, we describe the "traditional" noncancer risk assessment that was conducted by IRIS in 1991. As a point of comparison to the RfD generated in IRIS, we follow the steps suggested in recent proceedings of the EPA Risk Assessment Forum[19]. We also estimate RfDs by determining alternative points of departure using BMDS and CatReg, as well as by following the Straw Man framework. We compare the RfDs, along with the risk around the RfD across the models. Since the critical effect study has such a high rate of response in controls, we apply the Abbott correction to BMDS output to generate risk estimates comparable to Straw Man.

## 4.1 TCB Hazard Assessment

In 1980, the estimated U.S. production of 1,2,4,5-Tetrachlorobenzene (TCB) was $5.4x10^6$ kg (NTP, 1991). Commercial production of TCB in the United States ceased in 1983 (NTP, 1991). Since then it has been used primarily as an intermediate in the manufacture of herbicides, insecticides and defoliants, including 2,4,5−Trichlorophenol and 2,4,5−Trichlorophenoxyacetic acid. It has also been used as a dielectric fluid in electrical transformers and to impart moisture resistance to various substances (NTP, 1991). There are no known natural sources of tetrachlorobenzenes. The chemical consists of a benzene ring with 4 chlorine atoms in the configuration shown in Figure 4-1. 1,2,4,5-TCB is on EPA's list of 31 priority chemicals, which are voluntarily targeted for waste minimization efforts. 1,2,4,5-TCB was included on this list because it is persistent, bioaccumulative, and toxic. However, no Toxic Release Inventory data are available for 7 of these priority chemicals, including TCB.

[19] This document, Harmonization in Interspecies Extrapolation: Use of BW 3/4 as Default Method in Derivation of the Oral RfD (External Review Draft), suggested converting the POD to a HED and adjusting the $UF_A$ to account for pharmacodynamics only.

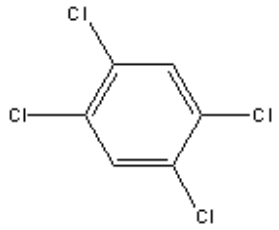



**Figure 4- 1 Chemical structure of 1,2,4,5-Tetrachlorobenzene (TCB)**

Tetrachlorobenzene is expected to exist solely as a vapor in the ambient atmosphere. Vapor-phase TCB is degraded in the atmosphere by reaction with photochemically-produced hydroxyl radicals with an estimated atmospheric half-life of 200 days (NLM, 2003). TCB is expected to have low mobility in soils. Biodegradation of TCB is expected to occur slowly based on a half-life of 29 to 47 days in soil and sediments. TCB is expected to volatilize from water surfaces, but adsorption to sediments may attenuate this process (NLM, 2003). Bioconcentration in aquatic organisms is considered high based on bioconcentration factors (BCF) in the range of 1,600 to 4,800 measured in carp and log BCF values of 3.7-4.1 measured in trout (NLM, 2003).

The general population may be exposed to 1,2,4,5-tetrachlorobenzene via inhalation of ambient air, ingestion of contaminated food and drinking water. TCBs former production and use as an insecticide and current use as an intermediate in the production of herbicides and defoliants may have resulted in its release to the environment through various waste streams (NLM, 2003). TCB is also a degradation byproduct of pentachlorobenzene and hexachlorobenzene (both of which are still used and released to the ambient environment) and therefore may enter the environment as a result of the microbial degradation of these compounds (NLM, 2003).

The Total Diet Study (TDS), sometimes called the market basket study, is a Food and Drug Administration program that determines levels of various contaminants and nutrients in foods. Summary data from 1991 to 2003 indicate trace levels of TCB in one potato sample (n=44) and four salmon samples (n=24). The trace detections (i.e., above the limit of detection but less than the limit of quantification) of TCB in food items ranged from 1E-05 ppm to 4.1E-04 ppm[20]. A recent analysis of persistent, bioaccumulative, and toxic compounds in the United States milk found that levels of chlorobenzene compounds (including TCB) were below the limit of detection (Schaum, 2003).

Animal research indicates TCB is easily absorbed and metabolized slowly relative to its other isomers. It readily accumulates and persists in adipose tissue and in organs with a high fat content (NTP, 1991). The mode of action of TCB appears to be unknown with regard to the observed kidney lesions in rats (Chu et al., 1984). TCB induces the

[20] See summary sheet at, e.g., http://www.cfsan.fda.gov/~acrobat/tds1byps.pdf

hepatic cytochrome p-450 microsomal enzymes in rats. Increased liver microsomal protein and phospholipid concentrations as well as increased serum cholesterol and triglyceride concentrations have been observed in animals dosed with TCB (NTP, 1991). TCB is metabolized to 2,3,5,6-tetrachlorophenol in the rat (Chu et al., 1984). Kohli, et al. (1976a) proposed the formation of the phenol metabolites through corresponding arene oxides. The authors suggested the involvement of an "NIH shift" of the chlorine atom in the formation of the metabolites[21].

#### 4.1.1 TCB Critical Effect Study Identification

The Integrated Risk Information System (IRIS) established a reference dose for TCB in 1991. In IRIS, the principal study upon which the 1,2,4,5-tetrachlorobenzene RfD was based was published by Chu et al. in 1984. This study examined Sprague-Dawley rats that were fed diets containing TCB. In this study, groups of 15 male and 15 female rats were fed diets containing 0, 0.5, 5.0, 50, or 500 ppm TCB for 90 days (approximately 13 weeks). Dose-related increases in the frequency and severity of kidney lesions were observed for male, but not female, rats at doses greater than 5 ppm. However, the severity of the kidney lesions in male rats was only deemed significant at 50 and 500 ppm. (See Table 4-1.)

In the Chu et al. study, the control group had a high rate of response (kidney lesions were found in 10 out of 15, or 67%, of male rats). While Chu et al. do not explain why there is such a high rate of response in the controls, the International Programme on Chemical Safety (IPCS, 1990) explains that, "even with random assignment of individual animals to each group and strict adherence to [good laboratory practice], the incidence of spontaneous neoplastic and other morphologic lesions is often highly variable among control groups." However, the Programme recognizes that high rates of response do not necessarily preclude a study from demonstrating a dose-response relationship. In order to prove that a chemical causes biological change, the Programme suggests that, "the differences between control and treatment groups should show a dose-response relationship and delineate a trend away from the expected norm for the particular species and strain of experimental animal use." The Chu et al. study meets these criteria because it shows a dose-response curve at doses above of 5.0 ppm (TCB in diet) and significant biological and physiological changes at 500 ppm. It is likely these reasons that motivated IRIS to rely on the Chu et al. study to establish NOAEL and LOAEL values for 1,2,4,5-tetrachlorobenzene. Liver lesions were also observed in female rats at feed concentrations of 500 ppm (U.S. EPA, 1991).

In a supporting 28-day study described in IRIS, rats fed diets containing TCB showed dose-related effects for liver and kidney pathology at 3.4 and 32 mg/kg/day (Chu et al., 1983). Males were more susceptible than females for all criteria. Kidney lesions were judged to be mild while liver lesions were reported as mild to moderately severe. Relative liver weight was significantly increased at 32 mg/kg/day. Hepatic microsomal enzymes were induced 2- to 12-fold at 32 mg/kg/day. Adverse effects were not

[21] An NIH shift is a chemical rearrangement where a hydrogen atom on an aromatic ring undergoes an intramolecular migration primarily during a hydroxylation (adding an –OH group) reaction

observed at two lower doses (0.04 and 0.4 mg/kg/day) (U.S. EPA, 1991). In a third supporting study described in IRIS, only higher doses (200 mg/kg/day) were associated with mortality, elevated serum cholesterol, increased organ weights, and liver enzyme induction when administered to pregnant rats for 10 days (U.S. EPA 1991). Non-control doses administered in this third supporting study ranged from 50 mg/kg/day to 200 mg/kg/day, which are all above the highest dose administered in the critical effect study (34 mg/kg/day).

#### 4.1.2 Data Inputs for TCB Case Study

Our application of models was based on quantal observations of the incidence of kidney lesions observed in male rats[22]. Using data from Chu et al. (1984) input files were created. The inputs included variables for concentration in feed (ppm), animal dose (mg/kg/day), the number of animals with kidney lesions, and mean lesion severity per group. The authors provided data on TCB concentration in feed and an animal dose equivalent range of 0.034 to 34 mg/kg/day for male rats based on weight gain and food consumption. To account for differences in metabolism rates across species, these doses were converted to human equivalent doses (HED) using a body weight to the ¾ power normalized by body weight (e.g., $BW^{-1/4}$) scaling formula (Rodricks et al., 2001):

$$D_h = D_a[W_a / W_h]^{1/4} \qquad (15)$$

where:

$D_h$ = human equivalent dose (mg/kg/day)

$D_a$ = animal equivalent dose (mg/kg/day)

$W_h$ = human weight (kg) = 70 kg

$W_a$ = animal weight (kg) = 0.25 kg

It was assumed that the male rats weighed 0.25 kg and the adult human body weight was 70 kg (National Institute for Occupational Safety and Health, 1982). The resulting conversion factor was 0.244 or (0.25/70)^0.25. To facilitate comparisons to the Straw Man approach we modeled the data using a dichotomous model type and a probit model with the log of the dose (i.e., log-probit).

[22] A fuller analysis of the data would have been possible if the full results of the experiment were available (that is, numbers of animals with different degrees of severity of the kidney lesions). Unfortunately this was not possible from the data in the published Chu et al. paper.

**Table 4- 1 Incidence and severity of kidney lesions in male rats exposed to 1,2,4,5-tetrachlorobenzene for 13 weeks (from Chu et al., 1984)**

| | | | Male Rats | |
|---|---|---|---|---|
| **Concentration Group** (ppm) | **Dose Equivalent[a] (mg/kg/day)** | **HED[b] (mg/kg/day)** | Kidney lesions / total animals | Kidney Lesion Severity[d] |
| 0.0 | Control | Control | 10/15[c] | 1.7 |
| 0.5 | 3.40E-02 | 8.31E-03 | 4/15 | 1.3 |
| 5.0 | 3.4E-01 | 8.31E-02 | 14/15 | 2.4 |
| 50.0 | 3.4E+00 | 8.31E-01 | 14/15 | 3.0[e] |
| 500.0 | 3.4+01 | 8.31E+00 | 15/15 | 5.5[e] |

[a] Dose equivalent was determined by Chu et al. based on feed consumption and weight gain.
[b] HED=0.244*Dose. 0.244 is the body weight scaling factor described in section 4.1.2, and equals (animal weight/human weight)^0.25. In this case, the scaling factor=(0.25 kg/70 kg)^0.25= 0.244.
[c] Number of animals with lesions/total examined
[d] Grading range from 1 to 10 based on severity where 1 = normal; 2 = mild interstitial reaction of multiple foci, membranous glomerular changes, hilar dilatation with atrophy of the papilla; 3 to 4 = severe and diffuse interstitial nephritis of both kidneys; 8-10 = carcinoma.
[e] The severity of effects was considered significant only at the 50 and 500 ppm levels because of a high incidence of mild kidney lesions in the controls.

## 4.2 Application of the Traditional Noncancer Risk Assessment Method to TCB

### 4.2.1 Reference Dose (RfD) from IRIS

In 1991, EPA's IRIS database based its noncancer risk assessment of TCB on the Chu et al. (1984) study. As previously described, Sprague Daley rats were fed diets containing 0, 0.5, 5.0, 50, and 500 ppm of 1,2,4,5-tetrachlorobenzene (TCB) for 13 weeks (Chu et al., 1984). The authors reported a dose-related increase in the frequency and severity of kidney lesions in male rats at feed concentrations of 5.0 ppm and higher; however, the severity of effects was statistically significant in pairwise comparisons only at the 50 and 500 ppm feed concentrations because of a high incidence of mild kidney lesions in the controls. (See Table 4-1.)

Based on the principal and supporting studies, IRIS determined the kidney lesions in male rats to be a critical effect and indicated a NOAEL of 5.0 ppm in the diet or 0.34 mg/kg/day as an animal dose. IRIS then applied a composite uncertainty factor of 1,000 to this point of departure (i.e., the NOAEL). This composite uncertainty factor was based on 10 for intraspecies variability (e.g., interindividual or $UF_H$), 10 for interspecies variability (e.g., animal to human or $UF_A$), and 10 for extrapolation of subchronic to chronic effects (e.g., $UF_S$). (No database insufficiency factor was used for this chemical.) The final oral RfD was reported as 3E-4 mg/kg/day (U.S. EPA, 1991).

$$RfD_{Traditional} = \frac{NOAEL_{animal}}{UF_{\text{int}\,erindividual} UF_{animal} UF_{subchronic}} = \frac{0.34\ mg/kg/day}{10x10x10} = 3.4x10^{-4}\ mg/kg/day$$

In 2006, the EPA Risk Assessment Forum (RAF) Workgroup for Harmonization in Interspecies Extrapolation recommended changing the $UF_A$ from a factor of 10 to a factor of 3, to account for the pharmacodynamic portion of the uncertainty, since body weight scaling would account for the pharmacokinetic portion when using human equivalent doses. For the purposes of comparison, we compute the RfD following this protocol. The interindividual variability and subchronic uncertainty factors remain unchanged ($UF_H$ = 10; $UF_s$ = 10), but the POD is the human equivalent dose NOAEL (rather than the animal dose NOAEL) and the animal-to-human uncertainty factor, $UF_A$, is lowered to 3 instead of 10. Note that the RfD estimated following this protocol is approximately 20% lower than the value published in IRIS.

$$RfD_{RAF} = \frac{NOAEL_{HED}}{UF_{\text{int}erindividual} UF_{animal} UF_{subchronic}} = \frac{0.0831\, mg/kg/day}{10x3x10} = 2.77x10^{-4}\ mg/kg/day$$

## 4.3 Application of the BenchMark Dose Software to TCB

#### 4.3.1 Estimating a Dose-Response Function and Deriving an RfD based on the $BMD_L$

Benchmark Dose Software (BMDS) (Version 1.4.1) was used to develop a function to estimate the probability of kidney lesions as a function of exposure dose to 1,2,4,5-tetrachlorobenzene. The software was also used to estimate the benchmark dose (BMD, corresponding to a 10 percent response) and its lower limit ($BMD_L$). Input data were the human equivalent doses (HED) and the prevalence of kidney lesions from Table 4-1.

Several BMDS models were fit to the input data (Table 4- 2). However, all model fits were poor due in part to the high incidence of kidney lesions in the control group. In all cases, the Chi square goodness of fit p-value was less than 0.1, indicating that the fitted model differed significantly from the model that went through all data points. Akaike's Information Criterion (AIC) was used to compare model fits. (Lower AICs indicate a better fit to the data.) The log-logistic model had the lowest AIC (66.7), though it was not substantially different from the log-probit model, which had the second lowest AIC (67.0). As such, to ensure comparability with the Straw Man model, we selected the log-probit model to determine the point of departure and to estimate risks around the RfD.

Using the log-probit model describing kidney lesions in male rats, a 10% increase in response was seen at 9.24E-03 mg/kg/d (BMD). The point of departure (the $BMD_L$) was determined to be 1.2E-03 mg/kg/day. The $BMD_L$s in all the models presented ranged from 1.4E-05 mg/kg/day to 3.6E-02 mg/kg/day.

**Table 4- 2 Selected BMDS Dose-Response Model Results for 1,2,4,5-TCB**

| Model[d] | BMD* (mg/kg/d) | $BMD_L$ (mg/kg/d) | Background | Intercept/ Power[a] | Slope | AIC[b] | Chi Square Goodness of Fit p-value[c] |
|---|---|---|---|---|---|---|---|
| Log-Probit | 9.24E-03 | 1.19E-03 | 0.494 | 1.793 | 0.656 | 67.0 | 0.009 |
| Probit | 6.18E-02 | 3.61E-02 | n/a | 0.24 | 1.71 | 69.3 | 0.0053 |
| Gamma | 5.59E-03 | 1.39E-05 | 0.53 | 0.53 | 1.93 | 69.4 | 0.0036 |
| Logistic | 5.20E-02 | 2.54E-02 | n/a | 0.37 | 3.32 | 69.1 | 0.0058 |
| Log-Logistic | 8.67E-03 | 1.23E-03 | 0.48 | 3.64 | 1.23 | 66.7 | 0.0068 |
| Weibull[d] | 4.49E-03 | 1.14E-04 | 0.51 | 0.62 | 2.93 | 68.8 | 0.0045 |

*BMD = BenchMark Dose for 10% extra risk, 0.95 Confidence Level, Dichotomous Model

[a] The probit, log-probit, logistic, and log-logistic models estimate intercepts; the gamma and Weibull estimate powers.

[b] Lower Akaike Information Criterion (AIC) indicates a better model fit.

[c] p-value tests whether the full model (that goes through all points) differs from the fitted model, with the indicated intercept (or power) and slope. Generally, $p > 0.1$ indicates an acceptable fit.

[d] Quantal-Linear results defaulted to the Weibull model. Quantal Linear is a subset of Weibull, in which the power parameter equals one.

Log-probit model form: P[response] =

Background + (1-Background)·CumNorm(Intercept+Slope·ln(Dose)),

Probit model form: P[response] =

CumNorm(Intercept+Slope·Dose),

where CumNorm(.) is the cumulative normal distribution function.

Gamma model form: P[response] =

Background + (1-Background)·(1/Γ(Power))·(integral from 0 to Dose of $t^{Power-1}$·exp(-t)dt)

Where Γ(t) is the gamma function, or the integral from 0 to ∞ of $t^{-1}$·exp(-t)dt

Logistic model form: P[response] =

1/(1 + exp(-Intercept -Slope·Dose))

Log-logistic model form: P[response] =

Background+(1-Background)/(1 + exp(-Intercept -Slope·ln(Dose)))

Weibull model form: P[response] =

Background + (1-Background)·(1-exp(-Slope·$Dose)^{Power}$)

To obtain a reference dose (RfD) estimate from the BMDS software, the $BMD_L$ (10 percent response) was used as a point of departure and divided by appropriate uncertainty factors. Following the traditional noncancer risk assessment method that was described in IRIS for TCB, the $BMD_L$ value employed is an animal dose (i.e., since we converted animal doses to human equivalent doses for the modeling, we convert back to the animal dose by dividing the $BMD_L$ HED by 0.244, or 1.19E-03 mg/kg/d divided by 0.244, which results in 5.31E-03 mg/kg/d as the animal point of departure). Following the TCB IRIS entry, the uncertainty factors are 10 for human interindividual variability ($UF_H$), 10 for subchronic to chronic extrapolation ($UF_s$), and 10 for interspecies

variability ($UF_A$) for a composite uncertainty factor of 1000. This results in an RfD of 5.3E-06 mg/kg/d.

$$RfD_{BMDS} = \frac{BMDL_{animal-dose}}{UF_{\text{int}erindividual} UF_{animal} UF_{subchronic}} = \frac{0.00531\, mg/kg/day}{10x10x10} = 5.3x10^{-6}\ mg/kg/day$$

For comparative purposes, we also follow the RAF guidelines. In this case, the HED $BMD_L$ is employed (1.19E-03 mg/kg/d) along with a factor of 3 to account for interspecies variability ($UF_A$) because HED accounts for the pharmacokinetic, but not pharmacodynamic, portion of the $UF_A$. The $UF_H$ and $UF_S$ remain unchanged at 10 each. This results in an RfD of 4.0E-06 mg/kg/d. This value is about 25% lower than that determined following the BMDS and the IRIS TCB assessment framework above. Note that the human equivalent dose $BMD_L$ (point of departure) is approximately 8 times lower than the BMD, which is an indication of the protective effect of choosing the 95$^{th}$ percentile lower bound, rather than the BMD as the point of departure.

$$RfD_{BMDS,RAF} = \frac{BMDL_{HED}}{UF_{\text{int}erindividual} UF_{animal} UF_{subchronic}} = \frac{0.00119\, mg/kg/day}{10x3x10} = 4.0x10^{-6}\ mg/kg/day$$

#### 4.3.2 Estimates of Risk around the RfD using BMDS Model

Using the log-probit model fits, we estimated the BMDS predicted risk level at concentrations around the point of departure ($BMD_L$) and reference dose (RfD). For comparability between methods, we use an RfD of 3E-04 mg/kg/day (as estimated by IRIS). Risks are computed at the RfD and an order of magnitude above (3E-03 mg/kg/day) and below (3E-05 mg/kg/day) for consistency with Straw Man output, but they also reflect exposures that might be typical for humans and thus the range at which benefits would be estimated.

##### *4.3.2.1 Abbott Correction*

In this case study the BMDS estimated background response parameter was substantial (e.g., 0.49 for the log-probit model) due to the high incidence of kidney lesions in controls. Therefore, Abbott's formula, which adjusts the incidence of effect (i.e., kidney lesions) in each dose-response group for the incidence occurring in the control group (Finney 1971), was applied. It was important to make such a correction for comparability with Straw Man results, since the Straw Man estimates incremental risk *above* background response. We applied the Abbott correction to the BMDS risk estimates *after* the model parameters were fit to the animal data and adjustments were made for human exposure[23].

[23] The Abbott correction is typically applied *before* BMDS modeling, but this was not possible because the BMDS software could not accommodate negative incidence rates.

Abbott's formula is:

$P_{response} = (P(D)-P(0))/(1-P(0))$ (16)

where:

P(D) = Proportion of animals with kidney lesions at concentration D
P(0) = Proportion of animals with kidney lesions in the control group

Using the log-probit BMDS model, the estimated risk at the IRIS-determined RfD was 2.1E-04. Risk estimates varied substantially between 0.1·RfD and 10·RfD (Table 4- 3). Estimates of risk at the $BMD_L$ (POD) were 4.3E-03. Castorina and Woodruff (2003) recently proposed a method for estimating risk at the RfC/RfD using BMDS output. For sublinear dose-response functions, like log-probit, the authors used the BMDS model to estimate risk at the RfD as we have done. However, for linear dose-response functions, the authors point out that the point of departure ($BMD_L$) is based on a benchmark response (BMR) level of risk such as 10 percent; thus to obtain the level of risk at the RfD (= $BMD_L$/UF), one might divide the BMR by the composite uncertainty factor. In other words, they propose dividing the BMR (in our case, 10%) by the composite uncertainty factor (in our case, 1000 or 300) to estimate risk at the RfD. Performing this calculation would yield risks of 1.0E-04 or 3.3E-04, risk at the RfD, somewhat higher than the estimate based on a probit model. Similarly, the risk at 0.1·RfD would be 1.0E-04 and the risk at 10·RfC would be 1.0E-2. Although the dose-response function for TCB does not appear to be linear, we provide these risk values as points of comparison to the log-probit risks.

**Table 4- 3 Estimated risk of kidney lesions at the point of departure ($BMD_L$) and around the RfD for 13 week male rat exposure to TCB as predicted by BMDS log-probit model (with Abbott's correction) and linear extrapolations**

| Model Type | $BMD_L$ (mg/kg/d) | Risk at the $BMD_L$ | Risk at the 0.1·RfD^ | Risk at the RfD^ | Risk at the 10·RfD^ |
|---|---|---|---|---|---|
| Log-Probit | 1.2E-03 | 4.3E-03 | 2.3E-07 | 2.1E-04 | 2.2E-02 |
| Linear[#] | 5.3E-03 | 1.0E-1 | 1.0E-05[a] | 1.0E-04[a] | 1.0E-03[a] |
| Linear[#] | 1.2E-03 | 1.0E-1 | 3.3E-05[b] | 3.3E-04[b] | 3.3E-03[b] |

^RfD as reported in U.S. EPA IRIS database (i.e., 3E-04 mg/kg/day).
[#] Using method suggested for linear extrapolation by Castorina and Woodruff (2003) for illustrative purposes.
a composite uncertainty factor of 1000
b composite uncertainty factor of 300

## 4.4 Application of CatReg to TCB

### 4.4.1 Estimating a Dose-Response Function and Deriving an RfD based on the Extra Risk Dose (ERD)

CatReg was used to develop a function to estimate the probability of kidney lesions as a function of exposure to TCB. The software was also used to explore the severity data

reported by Chu et al. (1984). Due to limitations in the level of detail of the Chu et al. data, we were unable to obtain meaningful CatReg output.

A log-probit model[24] with a background term was used to generate a dose-response function of the animal data (Chu et al., 2004; See **Appendix D** for data input). However, the background response term (gamma) hit its maximum bound and therefore the model results were considered invalid. Irregular results (negative values) were also observed for the lower bound estimate of the 10 percent extra risk dose, ERD10 (Table 4- 4). Alternative model configurations were examined, including the use of additional information on lesion severity to add another level of effect. However, the gamma term in these models continued to exceed its maximum bound—most likely due to the very high background incidence in the controls and the anomalous decrease in incidence in the low exposure group. It was not possible to construct datasets more conducive to the CatReg analysis because the Chu et al. study reports mean severity levels per dose group and not the individual severity levels per animal. Given the large number of combinations of individual severity levels that could result in the same mean group severity level, the original data for individual animals could not be deduced. Due to the lack of an acceptable model fit, CatReg estimates of risk were not calculated.

**Table 4- 4 CatReg parameter estimates for the probability of kidney lesions as function of exposure concentration 1,2,4,5-Tetrachlorobenzene.**

| Model | ERD10* (mg/kg/day) | ERD10-L‡ (mg/kg/day) | Background γ** | Kidney Lesion Intercept $\alpha_s$ | Concentration Slope Factor $\beta_1$ | AIC | Chi Square Goodness of fit p-value^ |
|---|---|---|---|---|---|---|---|
| log-probit | 0.0032 | -0.0052 | 0.00786 | 1.84 | 0.91 | 68 | 0.7476 |

Model form: $\Pr(Y \geq s|C) = H[\alpha_s + \beta_1 * f_1(C+\gamma)]$

*H* is a probability function taking values between 0 and 1, and is based on the normal distribution.

*Extra risk dose 10 percent (ERD-10) adjusts for background incidence rates of the same effects by applying Abbott's formula, which estimates risk at dose, D, only among the fraction of the population not expected to respond to the secondary (background) causes:

$P_{response}$ = (P(D)-P(0))/(1-P(0)),

where: P(D) = proportion of animals with kidney lesions at dose, D,

and P(0) = proportion of animals with kidney lesions in the control group.

‡ERD10-L = ERD10 Lower Bound.

**Note--γ, or background, hit its maximum bound.

^For CatReg, p-values > 0.05 indicate the model is not a good fit.

[24] A cumulative odds model with a probit link and log of the dose was used to estimate parameters.

## 4.5 Application of the Straw Man to TCB

### 4.5.1 Estimating a Dose-Response Function and Deriving an RfD with Uncertainty Distributions

The Straw Man method was used to estimate a reference dose (RfD) for TCB exposure based on the 13-week rat data for kidney lesions. Table 4- 5 shows the uncertainty distributions that were applied to derive the Straw Man RfD and associated risks. The Straw man proposal suggests the RfD be set at the daily dose rate that is expected (with 95% confidence) to produce less than 1/100,000 excess incidence over background of a minimally adverse response in a standard general population of mixed ages and genders, as is shown the equation below.

$$\hat{RfD} = \hat{Dose}\left(\frac{1}{100{,}000}\right) = \hat{ED}_{50} \cdot \hat{UF}_A \cdot \hat{UF}_S \cdot 10^{\sqrt{\left(\log\hat{GSD}_{PK}\right)^2 + \left(\log\hat{GSD}_{PD}\right)^2} \cdot \text{probit}\left(\frac{1}{100{,}000}\right)}$$

**Table 4- 5 Geometric means and standard deviations for uncertainty distributions applied in Straw Man analysis of TCB.**

| Term | Uncertainty distribution Type | Geometric mean | Geometric Standard Deviation | Source |
|---|---|---|---|---|
| $\hat{ED}_{50}$ | Log-Probit | 0.266 | 1.998 | Estimated from Chu et al., 1984 |
| $\hat{UF}_s$ | Lognormal | 0.499 | 2.17 | Baird et al., 1996 |
| $\hat{UF}_A$ | Lognormal | 0.6 | 2.72 | Price et al. 2008 |
| $\hat{UF}_H$ PK ($\log\hat{GSD}_{PK}$) | Lognormal | 0.202 | 0.092** | Hattis and Lynch, 2007 |
| $\hat{UF}_H$ PD ($\log\hat{GSD}_{PD}$) | Lognormal | 0.330 | 0.161** | Hattis and Lynch, 2007 |

$\hat{ED}_{50}$: Effective dose associated with a biological effect in 50 percent of the animals (mg/kg/day)
$\hat{UF}_s$: Uncertainty distribution for subchronic to chronic extrapolation.
$\hat{UF}_A$: Uncertainty distribution for interspecies projection.
$\hat{UF}_H$ PK: Uncertainty distribution for pharmacokinetic variability in humans.
$\hat{UF}_H$ PD: Uncertainty distribution for pharmacodynamic variability in humans.
**These are the standard deviations of the uncertainty distributions of the log[log(GSD)]s for the pharmacokinetic and pharmacodynamic variability, respectively, based on the general finding that the log(GSD)s for these component variabilities themselves have uncertainties that are approximately lognormally distributed (Hattis and Lynch, 2007).
Note that there was no database insufficiency factor identified in the IRIS TCB assessment, therefore none is included in the Straw Man model.

Table 4- 6 shows doses (mg/kg/day) estimated to have various probabilities of producing 1 in 100,000 risk and the TCB induced kidney lesions observed by Chu et al., 1984, and the other more generic uncertainties built in to the Straw Man analytical model. It is interesting to note that the $ED_{50}$ estimate for TCB has a much higher geometric standard

deviation relative to its geometric mean than the same value estimated in the COS case study. This is likely the result of the high prevalence of lesions in the control group positively affecting the magnitude of the parameter's standard error. The concentration necessary to meet the Straw Man risk criterion—a 95% chance that the risk would be less than 1/100,000--was about 3.5E-05 mg/kg/day or about one tenth of the IRIS-determined RfD value of 3E-04 mg/kg/day. Note that had the median or mean values of the distribution been selected as the tolerable risk management criterion, rather than the 5$^{th}$ percentile, the Straw Man method would yield an RfD estimate over three times higher than traditional noncancer risk assessment method.

**Table 4- 6 Distributions of doses associated with risks meeting the Straw Man Criterion**

| Percentile of the uncertainty distribution for risk | Estimated dose to produce 1 in 100,000 Risk (mg/kg/day) | Ratio of Straw Man Dose for Target Risk divided by RfD^ |
|---|---|---|
| 5th %tile | 3.5E-05* | 0.1 |
| Median | 9.3E-04 | 3.1 |
| Mean | 3.9E-03 | 13.0 |
| 95th %tile | 1.5E-02 | 50.3 |

*The daily dose rate that is expected (with 95% confidence) to produce less than 1/100,000 excess incidence over background of a minimally adverse response in a standard general population of mixed ages and genders.
^ RfD as estimated in IRIS = 3E-04 mg/kg/d

#### 4.5.2 Straw Man Estimates of Risk Around the RfD

The mean risk value at an RfD of 3E-4 mg/kg/day (as estimated by IRIS) was about 2E-3 or 2 in 1000 (Table 4- 7). At this concentration, over a lifetime of continuous daily exposure at the RfD, about an additional 0.2% of the population may experience kidney lesions over those who would develop kidney lesions in the absence of exposure. The magnitude of the mean risk is the result of the highly skewed uncertainty distribution, where the median (50$^{th}$ percentile) risk is estimated to be only 2 in a hundred million, while the 95$^{th}$ percentile is greater than 500,000 in a hundred million.

**Table 4- 7 Straw Man distributions of risk around the RfD (as reported in US EPA IRIS database for 1,2,4,5-tetrachlorobenzene)**

| Percentile | 0.1·RfD | RfD | 10·RfD |
|---|---|---|---|
| 5th %tile | very small | 7.5E-23 | 5.4E-11 |
| Median | 5.2E-16 | 2.0E-08 | 1.5E-03 |
| Mean | 4.3E-05 | 2.0E-03 | 5.2E-02 |
| 95th %tile | 6.1E-06 | 5.4E-03 | 3.1E-01 |

We also compared distributions of risk in Straw Man models that did not include uncertainty distributions for human interindividual variability (i.e., only included the central estimates of the pharmacokinetic and pharmacodynamic variabilities) (Table 4-8). Compared to the model that included uncertainty distributions for human

interindividual variability, mean risk at the traditional RfD was similar: 1.4E-03 compared to 2.0E-03. However, the 95-percentile risk at low dose (i.e., 0.1 RfC) decreased several orders of magnitude (from 6.1E-06 to 7.4E-09). If we consider percentile levels corresponding to fractions of the population, this means that the most sensitive individuals would experience lower risk when the $UF_H$ terms were collapsed. The 5$^{th}$ percentile (least sensitive individuals) risk at 10 RfD increased by 2 orders of magnitude when the $UF_H$ terms were collapsed, indicating that less sensitive individuals would experience a higher risk. However, both estimates were still very small.

**Table 4- 8 Straw Man distributions of risk around the RfD (as reported in US EPA IRIS database for 1,2,4,5-tetrachlorobenzene) without uncertainty distributions for human interindividual variability**

| Percentile | 0.1·RfD | RfD | 10·RfD |
|---|---|---|---|
| 5th %tile | very small | 2.1E-17 | 2.8E-09 |
| Median | 3.7E-17 | 4.3E-09 | 7.5E-04 |
| Mean | 4.5E-06 | 1.4E-03 | 4.8E-02 |
| 95th %tile | 7.4E-09 | 1.0E-03 | 3.1E-01 |

## 4.6 Summary Across Approaches in the 1,2,4,5-Tetrachlorobenzene Case Study

1,2,4,5-tetrachlorobenzene is a persistent, bioaccumulative and toxic compound used primarily as an intermediate in the manufacture of herbicides, insecticides and defoliants, including 2,4,5−Trichlorophenol and 2,4,5−Trichlorophenoxyacetic acid. While 1,2,4,5-TCB is no longer directly applied in agricultural settings, summary data from the Total Diet Study continue to show trace detections of the compound in foods. A case study of TCB was conducted to explore the potential impact on human health benefits assessment of applying different quantitative methods to derive dose-response functions. In the case study of TCB, the quantitative model fits were poor because of high rates of response in controls. Output from the CatReg model was considered invalid. Straw Man and BMDS based estimates of the RfD were similar to each other and about 1/10$^{th}$ of the RfD reported in IRIS (i.e., 3E-04 mg/kg/day). Straw Man based estimates of risk near the RfD were consistently higher than those estimated by BMDS, except at the very highest doses where risk estimates converged. Again, within the Straw Man model, estimates of mean risk near the RfD were not strongly affected by using only the central estimate of the uncertainty distribution for human interindividual variability. Divergent estimates of risks at low doses between the BMDS and Straw Man models suggest model selection could strongly affect benefit-cost analysis. These findings are summarized in greater detail below.

For TCB, the traditional and BMDS methods generated substantially different RfD values (Table 4- 9). These differences are due in part to two methodological differences in the way the RfD was estimated. First, in the traditional noncancer risk assessment approach, determination of the POD (i.e., the NOAEL) was based on a significant increase in both the incidence and the severity of the kidney lesions. Because of the high background incidence in controls and the severity of the lesions did not increase

until the very highest doses, a relatively higher NOAEL was selected. However, for the BMDS, the point of departure, was based only on kidney lesion incidence (i.e., severity was not included) and is calculated based on the lower confidence limit of an extra risk (of 10%)—as opposed to a significant increase in lesion severity. In part because the unusual dose-response curve for kidney lesions from TCB yields a large uncertainty in the $BMD_L$, this means there is a larger than usual difference between a central estimate of the BMD and its lower confidence limit—yielding a relatively lower estimate of the POD and therefore the RfD. A second, smaller, difference results from the RAF guidelines for the human equivalent dose (HED) and subsequent adjustment of the $UF_A$.

**Table 4- 9 Estimated RfD for TCB by approach**

| Model | RfD (mg/kg/day) | Ratio of Model based RfD to Traditional IRIS RfD |
|---|---|---|
| Traditional, IRIS | 3.00E-04 | 1 |
| Traditional, RAF | 2.77E-04 | 0.9 |
| BMDS | 5.3E-06 | 0.02 |
| BMDS, RAF | 4.0E-06 | 0.01 |
| CatReg | n/a | n/a |
| Straw Man | 3.50E-05 | 0.1 |
| Straw Man‡ | 1.10E-04 | 0.4 |

‡with no uncertainty in human interindividual variability

Overall, the model fits were poor, especially for the CatReg model where the estimate of the gamma (background) term hit a boundary and the results were rejected for this analysis. The poor model fits were likely a reflection of the anomalies in the dataset—in particular a very high background incidence of kidney lesions in the controls followed by a much lower incidence of lesions in the first non-zero dose group. Because the BMDS includes background cases in its parameter estimates, the Abbott formula was used to adjust the incidence of effect in each dose-response group for the incidence occurring in the control group. An Abbott correction was not required for the Straw Man output because the results are based only on incremental risk above background response.

Application of quantitative methods allows the derivation of a dose-response function, which can be used to estimate risk for a given exposure, and the associated uncertainty. Using the RfD reported in the U.S. EPA IRIS database we found varying levels of risk depending on the model used in the TCB case study. The Straw Man consistently provided the highest estimates of risk near the NOAEL based RfD, followed by the BMDS. At very low doses (i.e., 0.1·RfD) the Straw Man predicted substantially higher risk than the BMDS. For example, at 0.1·RfD the Straw Man estimate is about 2 orders of magnitude higher than estimates from the BMDS. However, as dose increased to the 10·RfD, the difference in risk estimates decreased between the two models (Table 4- 10, Figure 4- 2).

**Table 4- 10 Estimated risk of kidney lesions at the IRIS-derived RfD (3E-4 mg/kg/day), by model**

| Model | Risk at RfD/10 | Risk at RfD | Risk at 10·RfD |
|---|---|---|---|
| Traditional | n/a* | n/a* | n/a* |
| BMDS (HED) | 2.3E-07 | 2.1E-04 | 2.2E-02 |
| CatReg[#] | n/a | n/a | n/a |
| Straw Man^ | 4.3E-05 | 2.0E-03 | 5.2E-02 |

*An estimate (with uncertainty spanning perhaps an order of magnitude) of a continuous inhalation exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious non-cancer effects during a lifetime
# Risk not calculated due to invalid model fit.
^ Mean risk values estimated by Straw Man

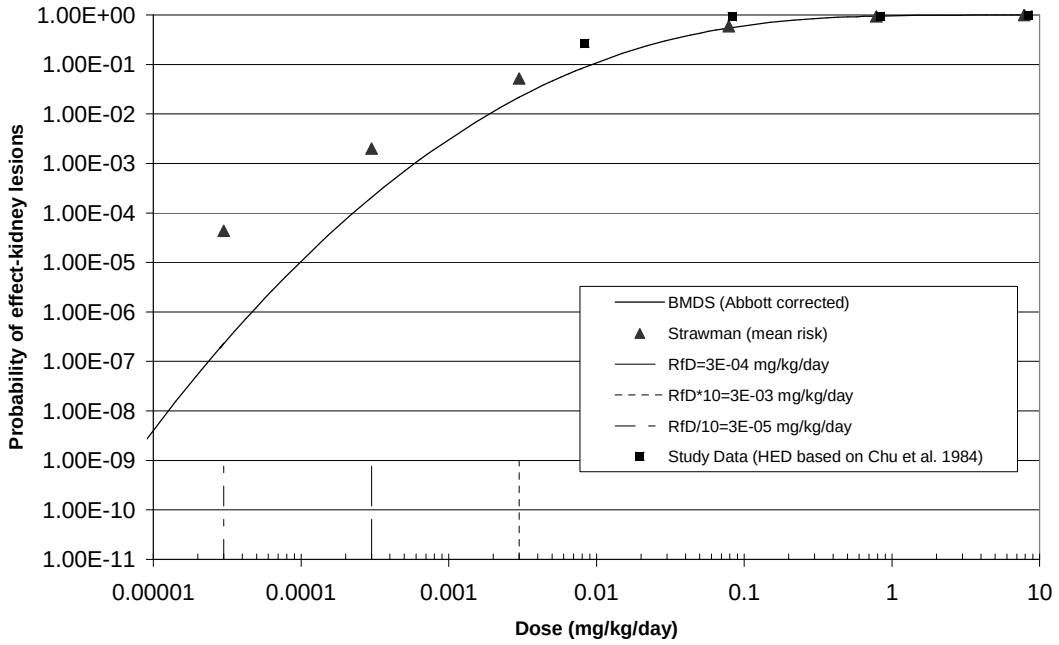



**Figure 4- 2 Comparison of risk estimates for exposure to 1,2,4,5-TCB based on Straw Man and BMDS models using human equivalent dose– note zero dose groups not shown in this figure**

The greater risk at 0.1·RfD predicted by the Straw Man may be due in part to the use of skewed (i.e., lognormal) uncertainty distributions for estimation of the subchronic to chronic projection, interspecies projection, and human interindividual variability. We examined the impact of collapsing human interindividual variability to its central (i.e., mean) estimate and again found little impact on the mean risk value estimated by the Straw Man method. Therefore the factors that determine mean risk seem to be dependent on the interaction of the other uncertainty distributions. On the other hand, collapsing uncertainty in individual variability to its central estimate strongly affected the estimated risk in the tails of the distributions. For example, collapsing human

interindividual variability to its central estimate led to smaller estimates of 95$^{th}$ percentile risk at low dose (i.e., 0.1·RfD).

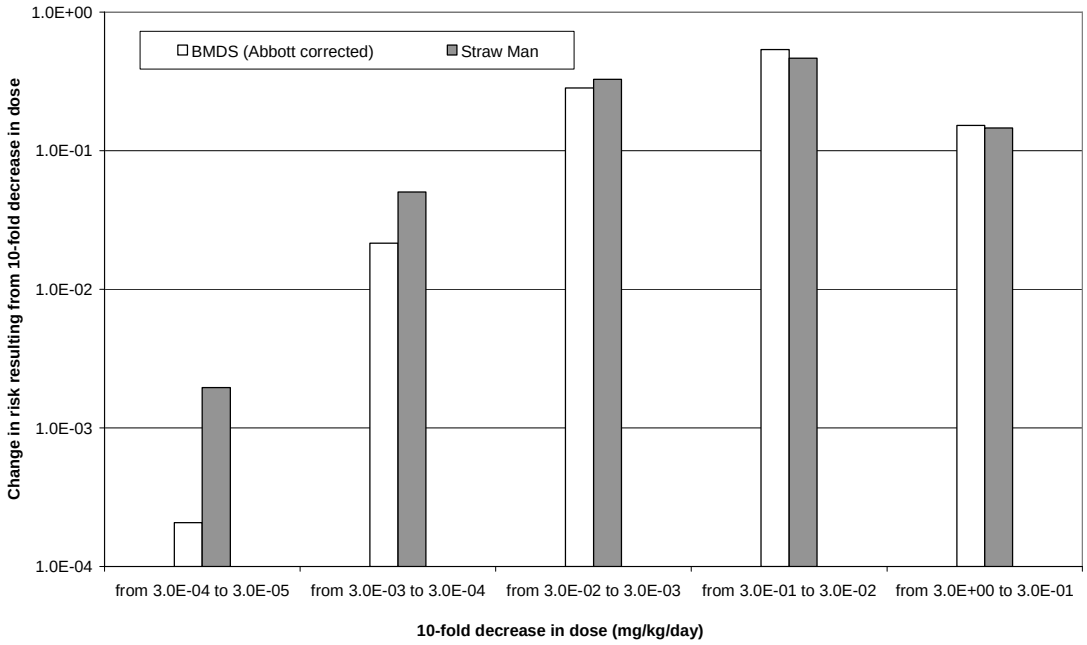



**Figure 4- 3 Change in risk per 10x decrease in dose of 1,2,4,5-TCB for Straw Man and BMDS models. Delta risk represents the change in risk due to a change in dose from a higher dose to the given dose, for example, delta risk(3.00E-05)=Risk(3.00E-4)-Risk(3.00E-5)**

Figure 4- 3 shows the decrease in risk resulting from a ten-fold decrease in dose for the BMDS model (clear bar) and the Straw Man model (grey bar). For example, if the dose were decreased from 3.0E-04 to 3.0E-05 mg/kg/day, the Straw Man model predicts a decrease in risk of approximately 2E-3 and the BMDS model (Abbott-corrected) predicts a decrease in risk of approximately 2E-4. (Note: y-axis is on the log scale).

While the BMDS and Straw Man models appear to generate similar (and substantial) levels of risk at high doses (i.e., 10·RfD) they produce increasingly different levels of risk as dose decreases. Figure 4- 3 suggests that there could be differential benefits of a regulation depending on which model is used, and where on the dose-response curve the reduction in exposure occurs. Not only does Straw Man provide higher levels of risk at low doses, the change in risk (delta risk) is greater as well. This is due, in part to a different response slopes in the low dose regions (Figure 4- 2).

# 5. Discussion

## 5.1 Traditional Noncancer Risk Assessment Method vs. Quantitative methods

Benefits for noncancer health effects are difficult to quantify for two main reasons. First, substantial uncertainties exist about the shape of the dose-response function at typical levels typical of human exposure, the variability of human response to exposures, and the characteristics of any adverse health effect in humans. Second, the traditional noncancer risk assessment method does not attempt to estimate a dose-response function, and therefore does not lend itself to benefits analysis. The inability to quantitatively estimate risk has been overcome to some extent by the availability of several methods (and software packages), including the benchmark dose method (BMDS), categorical regression (e.g., CatReg); and probabilistic methods that explicitly account for some of the uncertainties associated with the use of data from animal experiments and variability in response across humans (e.g., Straw Man).

Through case studies of carbonyl sulfide and tetrachlorobenzene, we examined the ability of three alternative models to the traditional noncancer risk assessment framework (as we have defined it: using a NOAEL or LOAEL as the point of departure) to estimate reference doses/concentrations and to estimate risk levels. Overall, the case studies suggested each quantitative model was an improvement over the traditional noncancer risk assessment method in setting a reference dose or concentration. The alternative models allow for the use of more data points from the animal study than just the NOAEL or LOAEL, and they generate a dose-response function that can be employed in benefits estimation. While the traditional noncancer risk assessment method implies a safe level, there may be a level of risk embedded in it, at least for some individuals. In fact, meta-analyses conducted on toxicological studies have shown a 5% to 25% risk of adverse health effects at the NOAEL (Faustman, 1996). The quantitative models enable the user to specify this risk level (e.g., 10% above background for BMDS and CatReg, or 1 in 100,000 for Straw Man) making it transparent.

The TCB and COS case studies provide interesting results, but it is unclear whether the results would follow the same patterns for additional chemicals. We discuss the utility and validity of the model outputs across the specific dose ranges used in the TCB and COS case studies (Sections 3 & 4), but it is beyond the scope of this work to provide a generalized dose-response theory for when a particular model should be used. However, while it is difficult to outline from first principles all the considerations that must be met to make particular data sets amenable to a certain type of analysis, there are two overarching guidelines. Generally, the use of a model can be justified when (a.) the model reasonably describes the data, and (b.) when the model represents a possible mode of action.

## 5.2 RfD/RfC and Model Fit

For COS, the traditional noncancer risk assessment method, BMDS, and CatReg all generated nearly identical RfC values (~ 0.2 ppm) since the point of departure for each method – NOAEL, $BMC_L$, ERC10-L, respectively – was nearly the same. The high degree of similarity among the RfC's was likely due in part to what appears to be a clear dose-response relationship in the experimental animal data, making the determination of a point of departure straightforward. On the other hand, in the TCB case study the BMDS-derived RfD was lower than the value specified in IRIS following the traditional (e.g., NOAEL/LOAEL) framework. As described earlier, this occurred in part because the traditional approach considered the *severity* of the lesions, which was not considered in the BMDS approach. A review of the literature indicated that on average the BMDS method is no more conservative than the traditional approach (Faustman, 1996). The Straw Man based RfDs were also lower than the traditional RfC and RfD in both case studies. These results are similar to those previously reported by Hattis et al (2002), where "17 of 18 reference doses estimated by the Straw Man method were more conservative (i.e., lower) than the corresponding reference doses estimated by the traditional (NOAEL/LOAEL) method".

Overall, model applicability and fit varied by case study. Where both BMDS and CatReg provided a fit (for COS), the AIC's were comparable, suggesting similar model fit. However, CatReg appeared to have more restrictive requirements on the nature of the data or at the very least, required more data than provided in these case studies. For example, the input data must fit a specific form. There cannot be strongly anomalous patterns in the data as was the case with TCB—where the background term exceeded boundaries set by the software. The data must be categorical and health effects must increase in severity with longer durations and exposures (i.e., there cannot be negative effect estimates as there were for carbonyl sulfide). For parietal necrosis caused by 2 and 12 week exposure to carbonyl sulfide, the CatReg program terminated prematurely because the severity in effects actually decreased slightly with increasing exposure duration. On the other hand, this did provide potentially useful information from a policy point of view. For instance, knowing that effects are severe and irreversible (or do not worsen) after two weeks exposure provides a clear incentive to prevent or reduce initial exposure to the chemical.

The main differences in the uncertainty factors and uncertainty distributions in setting the RfC (COS) or RfD (TCB) had to do with the database deficiency term and the animal to human extrapolation. The COS case study employed an uncertainty factor of 3 for database insufficiency (D, due to the lack of a developmental study). The TCB case study did not employ such a factor. The animal-to-human extrapolation ($UF_A$) was also dealt with differently for the two case studies. For COS inhalation, a 1:1 concordance between animal and humans was assumed (the slower metabolism rate expected in humans is exactly offset by the slower uptake expected in humans relative to the test animals). For TCB ingestion, body weight scaling (¾ power normalized by body weight) was to account for the pharmacokinetic portion of the $UF_A$, while a reduced UFA (3 versus 10) was to account for the pharmacodynamic portion.
.

## 5.3 Benefit-Cost Analysis (Human Health Risk) Implications

To quantify the benefits of a policy option to reduce a chemical's concentration or to quantify the public health burden of a particular chemical, several items would need to be determined: 1. the background incidence of a particular health effect in humans, 2. the increase above background of the health effect attributable to the chemical, 3. chemical levels (and potential reductions), 4. and the unit value (estimated cost) of that health effect. For example, COS exposure results in neuron death by inhibiting a key enzyme involved in the production of cell energy (ATP). Via this mechanism, exposure to COS could contribute to many types of chronic neurological diseases that involve chronic loss of neurons, as neurons do not ordinarily regenerate in adulthood. Such conditions include Parkinson's disease, Alzheimer's and other dementias, or the loss of neural elements that produce chronic age- and noise-induced hearing loss. However, the mechanism by which these diseases might develop in humans via COS exposure has not been established. There are background incidences of Parkinson's disease, for example, and COS exposure may result in more cases, as a function of exposure and duration of exposure. Although toxicological evidence exists suggesting a link with COS exposure and cell death, nothing currently in the literature can extend this link to Parkinson's disease or the other chronic neurological conditions mentioned above.

As a result of significant limitations in extrapolating information from animal studies to humans, the change in the incidence of the health effect in humans and their unit values (of, for example, avoided Alzheimer cases) were not evaluated, but are expected to vary in a linear manner with the model output. Therefore, knowledge of the performance characteristics of these models under certain dose ranges can be highly informative to benefit-cost analysis.

The most notable difference in model performance arose when examining the risk outputs. At low doses (i.e., one tenth of the RfC or RfD), the Straw Man consistently estimated higher mean risk for a given dose than BMDS or CatReg. The Straw Man model also appeared to estimate risk over a wider range of doses. This was particularly apparent in the COS case study where BMDS and CatReg were only able to estimate changes in risk for concentrations between approximately 40 ppm and 100 ppm. Neither BMDS nor CatReg dose-response functions would be able to estimate any useful change in health risk, nor corresponding health benefits, from completely eliminating COS exposures over salt marshes where some of the highest COS levels have been measured (between 0.024 ppm and 0.073 ppm).

The greater risk at low doses/concentrations predicted by the Straw Man may be due in part to the use of skewed (i.e., lognormal) uncertainty distributions for estimation of the subchronic to chronic projection, interspecies projection, and human interindividual variability. We examined the impact of collapsing human interindividual variability to its central (i.e., geometric mean) estimate and for both case studies, but found little impact on the mean risk value estimated by the Straw Man method. In effect, collapsing the human interindividual uncertainty distribution - if we consider the percentiles levels to correspond to fractions of the population - means that the most sensitive individuals

(95$^{th}$ percentile) face a lower risk at low doses, while the least sensitive individuals (5$^{th}$ percentile) face a higher risk, although virtually the entire population is estimated to have very small risks at this low exposure level. A similar effect on the distribution of risk should be expected from the collapse of any other uncertainty distributions, since all input parameter variance is positively correlated with risk distribution variance.

The Straw Man method acknowledges that substantial uncertainties exist for the input data, subchronic extrapolations, database insufficiencies, animal-to-human extrapolations, and human interindividual variability. It quantifies this uncertainty, to the extent possible, incorporating it into a Monte Carlo analysis yielding risk at different exposure levels. Incorporation of these uncertainties and the use of distributions lead to higher risk estimates at low doses and a wider dose-response range in general. While the dose-response function generated from BMDS and CatReg simply reflect a fit to the animal data, the Straw Man output of doses and corresponding distributions of risks reflect both the animal data, as well as the uncertainty factor distributions that attempt to correct for, among other things, animal-to-human differences.

There most divergent estimates of risk across models occur at low doses – the dose range most relevant to EPA's risk assessment goals. The Straw Man model estimates higher risks at low doses than the BMDS or CatReg models. We believe the Straw Man approach provides insights into the uncertainty distributions for risks associated with various doses, because it characterizes each component factor that influences risk. However, each of these uncertainty characterizations ($UF_A$, $UF_S$, D, $UF_H$) is itself uncertain and might be the subject of an extended sensitivity analysis. Even so, the probabilistic results produced by the Straw Man approach may be of interest to the risk assessment and risk management communities given the recently published National Research Council report, which recommends quantitative probabilistic analyses for both cancer and non-cancer effects (NRC 2008).

## 5.4 User considerations for the three software packages

While the BMDS was not developed for benefit-cost analysis in particular, it provides dose-response equations, which are readily applicable using human equivalent doses (HED) as inputs. Of the three models, BMDS is perhaps the most user-friendly. It has a simple interface and provides multiple models to help the user identify the best fit. The output is easy to read, the BMD and $BMD_L$ (or BMC and $BMC_L$) are calculated and a graph of the dose-response curve is provided. In summary, the model performs very well for what it was developed to do (i.e., to estimate a point of departure), with minimal data requirements. If a continuous endpoint were selected for analysis, the BMDS has the particular advantage of being able to quickly model continuous data (like cytochrome c oxidase inhibition). In contrast, the CatReg and Straw Man models both require quantal data inputs (although both CatReg and Straw Man approaches can be used to analyze continuous data by dividing continuous endpoints into several categories and quantifying the numbers of subjects in each category).

The BMDS model has some limitations relative to CatReg and the Straw Man. For instance, it does not have the capacity of CatReg to evaluate and analyze a larger

dataset composed of multiple endpoints and exposure durations. In the COS case study a more complete understanding of the data was achieved with the application of CatReg—and subtle differences in effect associated with duration of exposure or gender were detected that might not have been seen with the use of BMDS. Ultimately, understanding such differences could lead to the selection of different endpoints and results.

CatReg requires a greater level of involvement by the user to create the data input file and to operate the software and make appropriate choices about model configuration. However, the ability to include multiple studies, endpoints, and exposure duration into the models provides the user substantially more ability to understand the dose-response patterns elicited by a chemical. CatReg appears to impose more restrictions on the form and range of values that a model can assume. A simple example of this was termination of the model run when a longer duration of exposure did not result in more severe health effects.

The evaluation of the Straw Man approach is perhaps similar to other complex models—it provides valuable information not available in simpler models but at the cost of ever greater user involvement. Initial data inputs to the Straw Man (to calculate the $ED_{50}$) are essentially identical to the BMDS. Subsequent steps, however, require substantial involvement and judgment on the part of the user to determine the nature of the uncertainty distributions to be applied to the model. The uncertainty distributions are based on empirical observations of other studies (for the subchronic to chronic uncertainty factor) and studies conducted for drug development – a field that has extensive experience converting animal doses and effects to human doses and effects (for the human interindividual variability). There is the potential for measurement error in the database. Furthermore, information does not exist for all chemicals, but for a non-random selection of chemicals.

In addition, the model (a series of Excel worksheets) does not have any formal documentation, although the general process has been described in a publication (Hattis et al., 2002) and an accessible web site (http://www2.clarku.edu/faculty/dhattis). Another distinction in the application of the Straw Man model is that it does not generate a typical dose-response function, as does BMDS or CatReg. This is a reflection of the probabilistic nature of the model. This doesn't present any particular disadvantage but may require additional simulations to generate response plots in dose range of interest.

# 6. References

Axelrad DA, Baetcke K, Dockins C, Griffiths CW, Hill RN, Murphy PA, et al. (2005). Risk assessment for benefits analysis: framework for analysis of a thyroid-disrupting chemical. J Toxicol Environ Health A 68(11-12):837-855.

Baird, S.J.S.; Cohen, J.T.; Graham, J.D.; Shlyakhter, A.I.; Evans, J.S. (1996). Noncancer Risk Assessment: A Probabilistic Alternative to Current Practice. Hum. Ecol. Risk Assess. 1996, 2, 78–99.

Bartholomaeus AR and VS Haritos, (2005). Review of the toxicology of carbonyl sulfide, a new grain fumigant. Food and Chemical Toxicology 43, 1687-1701.

Castorina R and Woodruff TJ (2003). Assessment of Potential Risk Levels Associated with U.S. Environmental Protection Agency Reference Values. Environ Health Perspect, 111(10); 1318-1325.

Chang SJ, Shih TS, Chou TC, Chen CJ, Chang HY, Chen PC, Sung FC (2006). Electrocardiographic abnormality for workers exposed to carbon disulfide at a viscose rayon plant. JOEM, 48(4), 394-399.

Chengelis, C.P., Neal, R.A., (1979). Hepatic carbonyl sulfide metabolism. Biochemical and Biophysical Research Communications 90, 993–999.

Chu I, Villeneuve DC, Valli VE, Secours VE. Toxicity of 1,2,3,4-, 1,2,3,5-, and 1,2,4,5-Tetrachlorobenzene in the Rat: Results of a 90-Day Feeding Study. Drug Chem Toxicol,, 1984; 7: 113-127.

Evans, J.S and Baird, S.J.S. (1998). Accounting for Missing data in Noncancer Risk Assessment. Human and Ecological Risk Assessment, 4; 291–317.

Faustman, E.M. 1996. Review of Noncancer Risk Assessment: Application of the Benchmark Dose Methods. Prepared for the Commission on Risk Assessment and Risk Management. 62 pages. Accessed from http://www.riskworld.com/Nreports/1996/risk_rpt/pdf/faustman.pdf

Finney, DJ. Probit Analysis. Cambridge University Press, Cambridge, 1971.

Hattis, D. and Lynch, M. K. “Empirically Observed Distributions of Pharmacokinetic and Pharmacodynamic Variability in Humans—Implications for the Derivation of Single Point Component Uncertainty Factors Providing Equivalent Protection as Existing RfDs.” In Toxicokinetics in Risk Assessment, J. C. Lipscomb and E. V. Ohanian, eds., Informa Healthcare USA, Inc., 2007, pp. 69-93.

Hattis D, Baird S, Goble R. (2002). A straw man proposal for a quantitative definition of the RfD. Drug Chem Toxicol 25(4):403-436.

International Programme on Chemical Safety (1990). "Principles for the Toxicological Assessment of Pesticide Residues in Food." Accessed at: http://www.inchem.org/documents/ehc/ehc/ehc104.htm in July 2008.

Kholi J, Jones D, Safe S. The Metabolism of Higher Chlorinated Benzene Isomers. Can J Biochem, 1976; 54: 203-208.

Morgan DL, Little PB, Herr DW, Moser VC, Collins B, Herbert R, Johnson GA, Maronpot RR, Harry GJ, Sills RC (2004). Neurotoxicity of carbonyl sulfide in F344 rats following inhalation exposure for up to 12 weeks. Toxicology and Applied Pharmacology, 200; 131-145.

National Institute for Occupational Safety and Health, Registry of Toxic Effects of Chemical Substances (1980 edition, volume 1) Washington, D.C. 20402: U. S. Department of Health and Human Services DHHS (NIOSH) Publication No. 81-116, Government Printing Office, February, 1982.

NLM (2007). National Library of Medicine, Environmental Health and Toxicology, Specialized Information Services: Hazardous Substances Data Bank (HSDB), Carbonyl Sulfide, CASRN 463-58-1.

NLM (2003). National Library of Medicine, Environmental Health and Toxicology, Specialized Information Services: Hazardous Substances Data Bank (HSDB), 1,2,4,5-tetrachlorobenzene, CASRN 95-94-3.

NRC (2008). Committee on Improving Risk Analysis Approaches. Science and Decisions: Advancing Risk Assessment. http://www.nap.edu/catalog/12209.html.

NTP (1991). National Toxicology Program. NTP Report on the Toxicity Studies of 1,2,4,5-Tetrachlorobenzene in F344/N Rats and B6C3F$_1$ Mice (Feed Studies). National Institutes of Health, NTP TOX 7, NIH Publication No. 91-3126.

Price, P.S.; Swartout, J.C.; Schmidt, C.; Keenan, R.E. "Characterizing interspecies uncertainty using data from studies of anti-neoplastic agents in animals and humans" Toxicology and Applied Pharmacology 233 (2008) 64–70

Rodricks JV, Gaylor DW, Turnbull D. "Quantitative Extrapolations in Toxicology". In Principles and Methods of Toxicology, Fourth Edition, Hayes AW, ed. Taylor and Francis, 2001, pp. 372-373.

Sills RC, Morgan DL, Herr DW, Little PB, George NM, Ton TV, Love, NE, Maronpot RR. Johnson GA (2004). Contribution of Magnetic Resonance Microscopy in the 12-Week Neurotoxicity Evaluation of Carbonyl Sulfide in Fischer 344 Rats. Toxicologic Pathology, 32:501–510.

Sills, RC, Harry GJ, Valentine WM, Morgan DL. (2005). Interdisciplinary neurotoxicity inhalation studies: Carbon disulfide and carbonyl sulfide research in F344 rats. Toxicology and Applied Pharmacology, 207; S245-S250.

U.S. Environmental Protection Agency (2006a). 2006. CatReg Software User Manual (R-version). (ORD, ed):U.S.EPA.

U.S. Environmental Protection Agency (2006b). 2005 TRI Public Data Release (PDR), freeze date: December 19, 2006. http://www.epa.gov/tri/

U.S. Environmental Protection Agency (2006c). Harmonization in Interspecies Extrapolation: Use of BW 3/4 as Default Method in Derivation of the Oral RfD (External Review Draft). Washington, D.C., EPA/630/R-06/001

U.S. Environmental Protection Agency (2004). Risk Assessment Principles & Practices. (OSA, ed):U.S.EPA. March 2004 EPA/100/B-04/001

U.S. Environmental Protection Agency (2000). Guidelines for Preparing Economic Analyses. September 2000 EPA 240-R-00-003.

U.S. Environmental Protection Agency (1991). Integrated Risk Information System (IRIS), accessed July 5, 2007. http://www.epa.gov/iris/subst/0107.htm

# Appendix A. Sources and Derivation of Uncertainty Distributions in the Straw Man Approach

### Data for estimating uncertainty from subchronic/chronic projection.

Directly applicable empirical data are available for this uncertainty factor from the work of Baird et al. (1996) for ratios of subchronic/chronic NOAELs. The 61 chemicals for which comparative data are available span a wide range of industrial chemicals, include 50 assays in rats, and also include 11 sets of assays by inhalation (with the remainder via oral dosing). The distribution of the subchronic/chronic ratios is approximately lognormal with a geometric mean of 0.499 and a geometric standard deviation of 2.17.

### Data input for estimating uncertainty from database deficiency

In some of the discussion of the latest neurotoxicity observations of Morgan et al. (2004), the particular vulnerability of some regions of the brain to damage from COS is attributed to relatively high rates of energy consumption. If this is the case, then one might well expect similar or even greater vulnerability when the nervous system is rapidly growing and developing in the perinatal period and when animals are juveniles. Because of this, it is reasonable to expect that the dose rates needed to cause a comparable rate/extent of neural cell death might be even smaller in developing neural structures than is seen for the most energetically active brain areas in young adult animals. We hypothesize that an uncertainty factor of 3 would be assigned for this consideration under the traditional system. To represent the uncertainty resulting from the missing study of developmental neurotoxicity we draw on the work of Evans and Baird (1988) on the effects of the absence of a reproductive/developmental study from an otherwise complete database, as represented by the data bases available for 35 pesticides. In those data for about 3/4 of the cases it was found that a missing reproductive/developmental study would have made no difference in the assessed NOAEL because the chronic toxicity study yielded a lower estimate of the NOAEL. For the remaining 1/4 of the cases, the needed dose multiplier to correspond to the pesticide data was well described by the lowest quarter of a lognormal distribution with approximately the following distribution in Table 3-12:

**Table A- 1 Distribution of Dose Multipliers Required to Estimate Uncertainty Associated with Database Deficiency.**

| %tile | Dose multiplier for database deficiency |
|---|---|
| 1 | 0.062 |
| 5 | 0.201 |
| 10 | 0.353 |
| 15 | 0.544 |
| 20 | 0.770 |
| 25+ | 1 |

In other words, a dose reduction factor of 3 (multiplier of dose of 0.333) would correspond to approximately the 9$^{th}$ percentile of the pesticide data—in the remaining 91% of cases, a smaller reduction in dose would be indicated to compensate for the missing data.

**Data input for uncertainty in interspecies projection (rats to humans)**

For the initial implementation of the Straw Man methodology for COS we have elected to assume that the central estimate of neurotoxic sensitivity in humans is the same as that derived for rats. That is log($ED_{50}$ human) = log($ED_{50}$ animal) on a ppm basis (i.e., a 1:1 concordance in effect). There is, however appreciable uncertainty in this, which is modeled based on a prior analysis of the Price et al. (unpublished, 2002) data for cases where only information from rats is available (Hattis et al., 2002). The log(GSD), from Hattis et al., 2002, for interspecies projection uncertainty of the $ED_{50}$ was 0.416 (corresponding to a GSD of about 2.6 in antilog terms). The 5% - 95% confidence limits about the projection of the human $ED_{50}$ are therefore about 0.21 – 4.8 times the central estimate.

**Data input for Human Interindividual Variability in Sensitivity**

The variability in response expected for different individuals is treated by combining estimates of interindividual variability in pharmacokinetic and pharmacodynamic parameters from a database of human susceptibility to a variety of biological responses. Interindividual variability in pharmacokinetics (absorption, distribution, metabolism, and excretion) is represented as a central estimate log(GSD) of 0.202 with a lognormal uncertainty represented as a log[log(GSD)] of 0.092.

Three components make up the pharmacodynamic variability: active site availability, functional parameter change, and functional reserve capacity. In previous work the variability in functional reserve capacity was found to be smaller for more severe effect endpoints (Hattis and Lynch 2007). The effect observed in the rats by Morgan et al. (2004) is best characterized as a “severe/irreversible” adverse effect. In humans, this would be expected to have the least interindividual variability in pharmacodynamic susceptibility. Summing the variances for each of the pharmacodynamic steps yields:

Overall central estimate of pharmacodynamic log(GSD)
= [(log(GSD) for active site availability)$^2$ + (log(GSD for functional parameter change)$^2$ + log(GSD) for functional reserve capacity) $^2$]$^{0.5}$
= [(0.0917)$^2$ + (0.229)$^2$ + (0)$^2$]$^{0.5}$ = 0.246

Similar calculations for moderate and mild log(GSD)s for overall pharmacodynamic interindividual variability result in larger central estimates of log(GSD)s of 0.267 and 0.330, respectively. In all cases the uncertainty in the estimates of the pharmacodynamic interindividual variability is taken as a log[log(GSD)] of 0.161 (Hattis and Lynch 2007).

# Appendix B. Analysis of Cytochrome C Oxidase Inhibition

In Appendix B, we derived a different RfC based on concentration related decreases in cytochrome c oxidase activity in the posterior colliculus and the parietal cortex areas of the brain of male and female rats exposed to COS for 24, 52, and 86 days (Morgan et al., 2004). The inhibition of cytochrome c oxidase in various parts of the brain—indicate appreciable (and statistically significant) effects at lower exposure levels (down to 200 ppm) than the brain morphology/histology observations.

## B.1 Quantalizing Data Inputs

Using data from Morgan et al. (2004), input data files were created. Response data for different exposure durations (i.e., 24, 52, and 86 days) within each dose category were consolidated. Male and female data were also combined. A standard normal distribution was used to identify the first percentile value of cytochrome c oxidase activity in the controls. This value also serves as the demarcation between incidence of effect and no effect. A first percentile estimate was chosen as the threshold because in a sample of 30 animals per dose group it would represent a rare event in untreated animals, though other percentiles could be used. For each exposure category, inhibition of cytochrome c oxidase activity was defined as a distribution and the percentage of that distribution that was less than the first percentile of the control values was used to calculate the incidence of effect.

**Table B-1 Input Animal Dose-Response Data for Benchmark Dose Software and CatReg**

| Dose* (ppm) | N | Mean response (umol/min/mg protein) | Standard Deviation (umol/min/mg protein) | Percent below the 1st percentile in controls | Incidence of response |
|---|---|---|---|---|---|
| 0 | 30 | 1777 | 245 | 1% | 0 |
| 36 | 30 | 1504 | 222 | 9% | 3 |
| 53 | 30 | 1259 | 156 | 37% | 11 |
| 71 | 30 | 1236 | 197 | 44% | 13 |

*Dose adjusted to a 24-hour 7 day/week continuous exposure.
(Source: from Morgan et al., 2004 with changes).

## B.2 Estimating an RfC Using the Traditional Framework

Based on a review of the cytochrome c oxidase experiments by Morgan et al (2004) there is no NOAEL associated with the data. Instead, lowest observed adverse effect level (LOAEL) for cytochrome c oxidase inhibition would be 200 ppm. Animals in the study conducted by Morgan et al. (2004) were not exposed to levels of COS around-the-clock, but for 6

hours/day for 5 days/week. Using Haber's rule[25], the LOAEL of 200 ppm (reported by the researchers) was converted to a continuous (24 hours/day for 7 days/week) exposure by noting that the animals were exposed for 30 hours out of a possible 168 hours per week, or approximately 18% of the time. As such, the LOAEL corresponding to a continuous exposure is 18% of 200 ppm, or 36 ppm.

We estimate the composite uncertainty factor (the product of the relevant uncertainty factors) to be 3000— the product of 10 for human interindividual variability, 10 for subchronic/chronic exposure uncertainty, 10 for using a LOAEL in place of a NOAEL and 3 for uncertainty relating to the absence of potentially salient studies from database (i.e. in this case a developmental neurotoxicity study, which we believe is indicated because of the apparent central neurotoxic action of this toxicant). No adjustment was made for interspecies extrapolation because we assumed a 1:1 concordance of effect between the observed adverse events in the experimental rat and the parallel signs and pathologies in the exposed human.

Therefore, the RfC, as estimated using the traditional noncancer risk assessment method is 36 ppm (LOAEL, the point of departure) divided by the composite uncertainty factor of 3000, or 0.012 ppm. This RfC value is 1/15$^{th}$ the value of the RfC estimated (i.e., 0.18 ppm) based on quantal observations of unequivocally adverse morphological changes in rat brains in Section 3.

$$RfC_{Traditional,Cyt-c} = \frac{LOAEL_{continuous}}{UF_{NOAEL->LOAEL} UF_{\text{int}erindividual} UF_{subchronic} UF_{animal} UF_{database}} = \frac{36\,ppm}{10x10x10x1x3} = 0.012\,ppm$$

## *B.3 BMDS*

The log-probit BMDS model fit was acceptable (chi square p-value > 0.1). Using this method, the estimated point of departure ($BMC_L$) was 18 ppm. Using the same composite uncertainty factor derived earlier (3000), this translates to an RfC of 0.06 ppm.

[25] Haber's rule states that equal values of concentration*time are expected to produce equal biological effects. This would be expected, for example, if the agent causes effects that depend on a reaction between the toxicant and a cellular macromolecule that is not ordinarily reversed on the time scale of the exposures. This assumption can either under- or over-state the biologically effective dose and resulting adverse effects that would take place during a continuous exposure delivering the same product of toxicant concentration and time, depending on the details of pharmacokinetic handling and the pharmacodynamic processes leading to cell death or impaired function.

**Table B- 2 Benchmark Dose Model Results (log probit model) for Carbonyl Sulfide exposure cytochrome c oxidase inhibition**

| Dose Categories (ppm) | BMC* (ppm) | $BMC_L$ (ppm) | Back ground | Intercept** $\alpha$ | Slope*** $\beta$ | Chi Square Goodness of fit p-value^ | AIC |
|---|---|---|---|---|---|---|---|
| 0, 36, 54, 71 | 33 | 18 | 0 | -6.94 | 1.6 | 0.6496 | 105 |

Model form: P[Response] = $\Phi\{\alpha + \beta \log(\text{Dose})\}$, since BMDS set background to zero
ppm = parts per million
The natural logarithm of the dose is used in the log-probit model.
*BMC corresponds to a BMR of 10%
**The 95% confidence interval for the intercept was -11.3 to -2.6.
***The 95% confidence interval for the slope was 0.5 to 2.7.
^ p-value tests whether the full model (that goes through all points) differs from the fitted model, with the indicated intercept and slope. Generally, p > 0.1 indicates an acceptable fit.

$$RfC_{BMDS,Cyt-c} = \frac{BMC_L}{UF_{NOAEL->LOAEL} UF_{interindividual} UF_{subchronic} UF_{animal} UF_{database}} = \frac{18\,ppm}{10x10x10x1x3} = 0.006\,ppm$$

## *B.4 CatReg*

The log-probit CatReg model fit was acceptable (chi square p-value > 0.05). Using this method, the estimated point of departure ($BMC_L$) was nearly identical to the one derived using BMDS - 19 ppm versus 18 ppm. Using the same composite uncertainty factor derived earlier (3000), this translates to an RfC of 0.063 ppm.

**Table B- 3 CatReg parameter estimates for the probability of inhibition of cytochrome c oxidase as a function of exposure to Carbonyl Sulfide.**

| Model | ERC10 (ppm) | ERC10-L | Background | Intercept ($\alpha$) | Concentration Slope ($\beta_1$) | AIC | Chi Square Goodness of fit p-value^ |
|---|---|---|---|---|---|---|---|
| log-probit | 33 | 19 | 0 | -6.94 | 3.71 | 105 | 0.81 |

Model form: $\Pr(Y \geq s|C,T) = H[\alpha_s + \beta_1 * f_1(C)]$

*H* is a probability function taking values between 0 and 1, and is based on the Gaussian distribution.
^ Tests whether fitted model is significantly different from full model.

$$RfC_{CatReg,Cyt-c} = \frac{ERC10 - L}{UF_{NOAEL->LOAEL} UF_{interindividual} UF_{subchronic} UF_{animal} UF_{database}} = \frac{19\,ppm}{10x10x10x1x3} = 0.0063\,ppm$$

## *B.5 Straw Man*

The RfC estimated from the Straw Man fit to the data is 0.06 ppm, slightly lower than that estimated using BMDS and CatReg as the points of departure.

**Table B- 4 Distributions of doses associated with risks meeting the Straw Man Criterion***

| Percentile | Estimated Dose (ppm) to produce 1 in 100,000 Risk | Ratio of Estimated Dose to RfC^ calculated using other methods |
|---|---|---|
| 5th %tile | 0.078 | 0.8 |
| Median | 1.2 | 2 |
| Mean | 3.3 | 2 |
| 95th %tile | 13 | 2 |

*The daily dose rate that is expected (with 95% confidence) to produce less than 1/100,000 excess incidence over background of a minimally adverse response in a standard general population of mixed ages and genders, or the daily dose rate that is expected (with 95% confidence) to produce less than a 1/1000 excess incidence over background of a minimally adverse response in a definable sensitive subpopulation.
^ RfC = 0.012 ppm (as calculated using the traditional method, in Section B.1)

## *B.6 RfC and Risk Estimates by Model*

The RfC's that were based on the points of departure estimated from the BMDS and CatReg models were the lowest, followed by the Straw Man estimated RfC, with the traditional method resulting in the largest RfC.

The Straw Man model estimated the highest risks at all doses.

**Table B- 5 Estimated risk of cytochrome c oxidase inhibition at the Traditional RfC (or 0.012 ppm).**

| Model | RfC (ppm) | Risk at RfC/10 | Risk at RfC | Risk at 10·RfC |
|---|---|---|---|---|
| Traditional | 0.012 | n/a | n/a | n/a |
| BMDS | 0.006 | 2.1E-70 | 6.2E-45 | 2.5E-25 |
| CatReg | 0.0063 | 5.4E-71 | 3.1E-45 | 2.0E-25 |
| Straw Man* | 0.08 | 1.9E-06 | 2.1E-05 | 1.1E-03 |

* Mean risk values presented

# Appendix C. Input Dataset for CatReg Analysis of Carbonyl Sulfide

| Exp. | Ref.Id | Group | Species | Sex | Target | ppm | mg/m3 | Weeks | SevLo | Nsub | Incid |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2004131 | 1 | RT | B | Colliculus | 0 | 0 | 2 | 0 | 18 | 18 |
| 1 | 2004131 | 2 | RT | B | Colliculus | 54 | 132.7 | 2 | 0 | 10 | 10 |
| 1 | 2004131 | 3 | RT | B | Colliculus | 71 | 174.4 | 2 | 0 | 16 | 11 |
| 1 | 2004131 | 3 | RT | B | Colliculus | 71 | 174.4 | 2 | 1 | 16 | 5 |
| 1 | 2004131 | 4 | RT | B | Colliculus | 89 | 218.7 | 2 | 0 | 13 | 2 |
| 1 | 2004131 | 4 | RT | B | Colliculus | 89 | 218.7 | 2 | 1 | 13 | 11 |
| 1 | 2004131 | 1 | RT | F | Colliculus | 0 | 0 | 2 | 0 | 8 | 8 |
| 1 | 2004131 | 2 | RT | F | Colliculus | 54 | 132.7 | 2 | 0 | 0 | 0 |
| 1 | 2004131 | 3 | RT | F | Colliculus | 71 | 174.4 | 2 | 0 | 9 | 6 |
| 1 | 2004131 | 3 | RT | F | Colliculus | 71 | 174.4 | 2 | 1 | 9 | 3 |
| 1 | 2004131 | 4 | RT | F | Colliculus | 89 | 218.7 | 2 | 0 | 10 | 2 |
| 1 | 2004131 | 4 | RT | F | Colliculus | 89 | 218.7 | 2 | 1 | 10 | 8 |
| 1 | 2004131 | 1 | RT | M | Colliculus | 0 | 0 | 2 | 0 | 10 | 10 |
| 1 | 2004131 | 2 | RT | M | Colliculus | 54 | 132.7 | 2 | 0 | 10 | 10 |
| 1 | 2004131 | 3 | RT | M | Colliculus | 71 | 174.4 | 2 | 0 | 7 | 5 |
| 1 | 2004131 | 3 | RT | M | Colliculus | 71 | 174.4 | 2 | 1 | 7 | 2 |
| 1 | 2004131 | 4 | RT | M | Colliculus | 89 | 218.7 | 2 | 1 | 3 | 3 |
| 1 | 2004131 | 1 | RT | B | Parietal | 0 | 0 | 2 | 0 | 20 | 20 |
| 1 | 2004131 | 2 | RT | B | Parietal | 54 | 132.7 | 2 | 0 | 20 | 19 |
| 1 | 2004131 | 2 | RT | B | Parietal | 54 | 132.7 | 2 | 1 | 20 | 1 |
| 1 | 2004131 | 3 | RT | B | Parietal | 71 | 174.4 | 2 | 0 | 20 | 7 |
| 1 | 2004131 | 3 | RT | B | Parietal | 71 | 174.4 | 2 | 1 | 20 | 13 |
| 1 | 2004131 | 4 | RT | B | Parietal | 89 | 218.7 | 2 | 1 | 16 | 16 |
| 1 | 2004131 | 1 | RT | F | Parietal | 0 | 0 | 2 | 0 | 10 | 10 |
| 1 | 2004131 | 2 | RT | F | Parietal | 54 | 132.7 | 2 | 0 | 10 | 9 |
| 1 | 2004131 | 2 | RT | F | Parietal | 54 | 132.7 | 2 | 1 | 10 | 1 |
| 1 | 2004131 | 3 | RT | F | Parietal | 71 | 174.4 | 2 | 0 | 10 | 2 |
| 1 | 2004131 | 3 | RT | F | Parietal | 71 | 174.4 | 2 | 1 | 10 | 8 |
| 1 | 2004131 | 4 | RT | F | Parietal | 89 | 218.7 | 2 | 1 | 10 | 10 |
| 1 | 2004131 | 1 | RT | M | Parietal | 0 | 0 | 2 | 0 | 10 | 10 |
| 1 | 2004131 | 2 | RT | M | Parietal | 54 | 132.7 | 2 | 0 | 10 | 10 |
| 1 | 2004131 | 3 | RT | M | Parietal | 71 | 174.4 | 2 | 0 | 10 | 5 |

| Exp. | Ref.Id | Group | Species | Sex | Target | ppm | mg/m3 | Weeks | SevLo | Nsub | Incid |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2004131 | 3 | RT | M | Parietal | 71 | 174.4 | 2 | 1 | 10 | 5 |
| 1 | 2004131 | 4 | RT | M | Parietal | 89 | 218.7 | 2 | 1 | 6 | 6 |
| 2 | 2004131 | 1 | RT | B | Colliculus | 0 | 0 | 12 | 0 | 18 | 18 |
| 2 | 2004131 | 2 | RT | B | Colliculus | 54 | 132.7 | 12 | 0 | 18 | 18 |
| 2 | 2004131 | 3 | RT | B | Colliculus | 71 | 174.4 | 12 | 0 | 18 | 6 |
| 2 | 2004131 | 3 | RT | B | Colliculus | 71 | 174.4 | 12 | 1 | 18 | 12 |
| 2 | 2004131 | 1 | RT | F | Colliculus | 0 | 0 | 12 | 0 | 9 | 9 |
| 2 | 2004131 | 2 | RT | F | Colliculus | 54 | 132.7 | 12 | 0 | 9 | 9 |
| 2 | 2004131 | 3 | RT | F | Colliculus | 71 | 174.4 | 12 | 0 | 9 | 4 |
| 2 | 2004131 | 3 | RT | F | Colliculus | 71 | 174.4 | 12 | 1 | 9 | 5 |
| 2 | 2004131 | 1 | RT | M | Colliculus | 0 | 0 | 12 | 0 | 9 | 9 |
| 2 | 2004131 | 2 | RT | M | Colliculus | 54 | 132.7 | 12 | 0 | 9 | 9 |
| 2 | 2004131 | 3 | RT | M | Colliculus | 71 | 174.4 | 12 | 0 | 9 | 2 |
| 2 | 2004131 | 3 | RT | M | Colliculus | 71 | 174.4 | 12 | 1 | 9 | 7 |
| 2 | 2004131 | 1 | RT | B | Parietal | 0 | 0 | 12 | 0 | 20 | 20 |
| 2 | 2004131 | 2 | RT | B | Parietal | 54 | 132.7 | 12 | 0 | 20 | 20 |
| 2 | 2004131 | 3 | RT | B | Parietal | 71 | 174.4 | 12 | 0 | 20 | 11 |
| 2 | 2004131 | 3 | RT | B | Parietal | 71 | 174.4 | 12 | 1 | 20 | 9 |
| 2 | 2004131 | 1 | RT | F | Parietal | 0 | 0 | 12 | 0 | 10 | 10 |
| 2 | 2004131 | 2 | RT | F | Parietal | 54 | 132.7 | 12 | 0 | 10 | 10 |
| 2 | 2004131 | 3 | RT | F | Parietal | 71 | 174.4 | 12 | 0 | 10 | 6 |
| 2 | 2004131 | 3 | RT | F | Parietal | 71 | 174.4 | 12 | 1 | 10 | 4 |
| 2 | 2004131 | 1 | RT | M | Parietal | 0 | 0 | 12 | 0 | 10 | 10 |
| 2 | 2004131 | 2 | RT | M | Parietal | 54 | 132.7 | 12 | 0 | 10 | 10 |
| 2 | 2004131 | 3 | RT | M | Parietal | 71 | 218.7 | 12 | 0 | 10 | 5 |
| 2 | 2004131 | 3 | RT | M | Parietal | 71 | 218.7 | 12 | 1 | 10 | 5 |

*Species*RT = rat
*Sex*B = Both male and female combined, M = male, F = female
*SevLo*0 = no effect, 1 = necrosis
*Nsub*Number of subjects
*Incid*Incidence of SevLo effect

# Appendix D. Input Dataset for CatReg Analysis of 1,2,4,5-tetrachlorobenzene

| Exp | Group | Species | Target | Sex | Mg/kg/day | Hours | Nsub | Incid | SevLo |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | RT | K | M | 0 | 90 | 15 | 5 | 0 |
| 1 | 1 | RT | K | M | 0 | 90 | 15 | 10 | 1 |
| 1 | 2 | RT | K | M | 0.007861 | 90 | 15 | 11 | 0 |
| 1 | 2 | RT | K | M | 0.007861 | 90 | 15 | 4 | 1 |
| 1 | 3 | RT | K | M | 0.078607 | 90 | 15 | 1 | 0 |
| 1 | 3 | RT | K | M | 0.078607 | 90 | 15 | 14 | 1 |
| 1 | 4 | RT | K | M | 0.786071 | 90 | 15 | 1 | 0 |
| 1 | 4 | RT | K | M | 0.786071 | 90 | 15 | 14 | 1 |
| 1 | 5 | RT | K | M | 7.860712 | 90 | 15 | 0 | 0 |
| 1 | 5 | RT | K | M | 7.860712 | 90 | 15 | 15 | 1 |