CHARACTERIZATION OF DATA UNCERTAINTY AND VARIABILITY IN IRIS ASSESSMENTS
PRE-PILOT VS PILOT/POST-PILOT

Contract No. 68-C-99-238
Task Order No. 2

Prepared for:

Karen Hogan
National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C.

Prepared by:

Versar, Inc.
6850 Versar Center
Springfield, VA 22151

July 28, 2000

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

This report presents the results of a scientific review of uncertainty and variability analysis and documentation in health assessments contained in the Integrated Risk Information System (IRIS). The review was conducted by six individuals with expertise in various areas of toxicology, human health risk assessment, and related fields. The objective was to perform an in-depth study of uncertainty and variability documentation in IRIS assessments, based on sixteen chemicals randomly selected from the 537 assessments currently in the IRIS database. Specifically, the review addressed uncertainty documentation in sixteen assessments that spanning the years 1988 to 2000. The first group of eight were selected from the assessments generated by EPA during the time period 1987 through 1994, also referred to as "pre-Pilot" assessments. The second group of eight was selected from assessments released from 1995 to present, also referred to as "Pilot/post-Pilot assessments." Each IRIS assessment (the summary sheets and the Toxicological Review) was reviewed to determine how uncertainties in the data bases, modeling extrapolation, and other factors were considered in the development of toxicity values. Comparisons between the pre-Pilot and Pilot/post-Pilot groups were made as part of this review. Such a comparison facilitates a better understanding of the evolution in the procedures used to develop IRIS assessments and the extent to which uncertainty and variability have been incorporated into IRIS assessments.

The remainder of Section 1 presents an introduction to this topic, a description of the purpose and focus of the study, and background information to provide historical perspective on how uncertainty and variability have been addressed in the IRIS program. Section 2 describes the methodology for conducting this review. Section 3 presents the evaluation of the extent to which the sixteen assessments document uncertainty and variability. A broader comparison of pre-Pilot versus Pilot/post-Pilot assessments in general is provided in Section 4. Section 5 presents discussion, conclusions, and recommendations from this evaluation.

1.1 FOCUS AND PURPOSE

The Integrated Risk Information System (IRIS) is an online database containing 537 toxicity assessments generated over a period of nearly 20 years by the U.S. EPA. Recently, in H.R. 106-379, the U.S. Congress directed U.S. EPA to evaluate the extent to which uncertainty and variability in IRIS assessments have been considered and documented in deriving toxicity endpoints for chemical toxicants. Accordingly, U.S. EPA carried out a preliminary screening of the full IRIS assessment data-set to facilitate selection of a smaller, representative sample of sixteen IRIS assessments for an in-depth evaluation, to be carried out by six independent experts in the field. This report summarizes the results of that analysis. It focuses on the extent to which uncertainty and variability in available data were explored, documented, and

reflected in the IRIS assessments at different points in time. The primary objective was not to critique the numeric values of either the uncertainty factors or the quantitative toxicity endpoints themselves.

IRIS contains consensus EPA estimates of "safe" subthreshold dose values for a human population (within one order of magnitude), calculated by dividing a NOAEL or LOAEL value associated with an adverse toxic effect defined as the "critical" effect by an uncertainty factor. Uncertainty factors are applied to compensate for such uncertainties as: 1) lack of or scanty human epidemiological data; 2) inter-human population variability; 3) extrapolation from chronic duration laboratory animal toxicity studies to humans, where it is unclear whether humans are more or less sensitive than the animal species tested; 4) extrapolation from subchronic to chronic toxicity; 5) extrapolation from a LOAEL to NOAEL; or 6) some inadequacy or deficiency in the existing data-base (e.g., only one study in only one species) or in a key study. Uncertainty factors typically range between 1 and 10, but the default value of 10 is believed to be conservative.

Relatively recently, the science behind establishing uncertainty factors has progressed to the point that risk assessors speak of "data-derived" uncertainty factors to derive non-cancer toxicity values. The idea is to fine-tune the uncertainty values to incorporate chemical-specific data. Several approaches have been offered. One example would be the use of toxicokinetic and toxicodynamic information in deriving an intra- or inter-species extrapolation factor. *Toxicokinetic* information takes into account absorption, distribution, metabolism and excretion of a toxic compound. Variations in all of these sub-factors may occur both within and between species. *Toxicodynamic* variations refer to inherent sensitivities in species or individuals "influencing the target organ's response to a specified dose" (Dourson et al., 1996). Others have developed approaches in which a distribution for each area of uncertainty is established using toxicity data on groups of chemicals for which RfDs have been developed.

The process and techniques used to characterize uncertainty in IRIS assessments have evolved over time. Therefore, U.S. EPA also tasked the expert team to compare how uncertainty was considered between older "pre-Pilot" assessments (pre-1995) and those generated between 1995 and the present (Pilot/post-Pilot assessments). Overall findings are discussed in Sections 3 and 4.

1.2 HISTORICAL EVOLUTION OF UNCERTAINTY/VARIABILITY EVALUATIONS IN IRIS ASSESSMENTS

Health and environmental organizations world-wide utilize the concept of a "safe" dose in the dose-response evaluation of the noncancer toxicity of chemical substances. For many years, the concept of an "acceptable daily intake (ADI)" was used in pharmacology, toxicology, and risk assessment as a chemical-

specific daily dose unlikely to induce adverse health effects in humans, with an appropriate margin of safety. The concept of "reference dose" was originally introduced in 1983-1984 for internal EPA use to promote consistency among program offices in the assessment of chemicals with noncancer toxicity and to avoid the use of prejudicial terms such as "safety" and "acceptable." The assumptions used to develop RfDs were similar to those for ADIs, and involved judgments on principal study and critical effect, application of uncertainty (or safety) factors to provide a margin of safety, and the concept of a threshold (i.e., a dose level below which noncancer toxicity is not likely to occur). A somewhat different approach was taken in the derivation of virtually safe human doses (VSD) for cancer-causing chemical agents. Because the development of cancer may involve genetic or DNA damage, it has historically been assumed that there is no threshold for carcinogenic effects and that every exposure to a carcinogen has some probability, however small, of inducing cancer. For this reason, an inherently conservative no-threshold mathematical model has been used to estimate the VSD. Therefore, the need to apply numerous uncertainty factors to cancer incidence data in order to provide an adequate margin of safety for the protection of human health was minimized.

Information on how EPA utilized available scientific information on chemical hazards to assess dose-response and characterize health risks first became available to the general public in 1987 as the Integrated Risk Information System (IRIS) data base. In 1994, EPA instituted a policy of externally peer reviewing all new or revised IRIS summaries, in addition to conducting a consensus internal review. In 1997, the IRIS data base became available to the general public over the Internet, accompanied by electronically accessible review documents (Toxicological Reviews) detailing the scientific underpinnings of the IRIS summaries. IRIS chemical assessments are now utilized by government, industry, academia and nonprofit organizations nationally and world-wide to assess and manage human health risks.

In the early years of the IRIS data base, variability or uncertainty in chemical-specific information and evaluations were not separately assessed. In the early assessments, default uncertainty factors were applied to toxicity data to address a number of data-based or biologically-based uncertainties, according to standard operating *ad hoc* procedures (e.g., Barnes and Dourson, 1988) and early noncancer risk assessment guidelines (e.g., U.S. EPA, 1986). The default value for these uncertainty factors was always 10, and any factors used were multiplied together. Although standard practices theoretically allowed for departure from 10, in practice this was rarely done. Composite uncertainty factors were generally believed to be protective of human health; however, they not only compounded conservatism in estimating human risks but also often resulted in the use of very large uncertainty factors.

In the past decade, the science behind uncertainty in risk assessment has progressed somewhat beyond the use of standard default values for numerous uncertainties. Increased knowledge of the biology

and toxicology of chemical compounds, development of such tools as PB/PK models, in conjunction with the availability of more detailed data, has resulted in more frequent, scientifically justified departures from default uncertainty positions (Dourson et al., 1996). In the early to mid-1990's, EPA began to modify the way uncertainty factors are calculated and derived both by combining certain uncertainty concepts into a single factor and by moving towards a scientific framework for reducing default uncertainty factor values by using biological data.

The utilization of biological or toxicologic data to get beyond use of default uncertainty values in human non-cancer risk assessment is referred to as using "data-derived" uncertainty factors. To date, few EPA IRIS chemical assessments employ data-derived uncertainty factors. However, it is anticipated that this approach will become more widespread as the scientific framework evolves for incorporating chemical-specific data into uncertainty analysis. By integrating all available information to characterize the weight-of-evidence for human toxicity and cancer, scientists and regulators will improve the way in which uncertainty is evaluated in experimental data, reduce uncertainty, and increase confidence in the assessment.

2.0 METHODOLOGY FOR CONDUCTING IRIS REVIEW

This section describes the methodology used to conduct this review of uncertainty and variability in IRIS assessments. The procedures used were designed to statistically select a small number of IRIS assessments to be analyzed in depth by six individuals with expertise in toxicology and human health risk assessment. These six expert reviewers are presented in Table 2-1 and detailed information on the reviewer selection process is described later in this section. A random statistical sample was taken comprising sixteen of the 537 IRIS assessments available on-line as of January 31, 2000. These sixteen chemicals are summarized in Table 2-2. Also presented in this section are the charge to the reviewers, which established the scope of the review, and the materials provided to the reviewers which included the IRIS summary sheets, Toxicological Reviews (when available), and available supporting documents for the sixteen IRIS chemicals.

2.1 STATISTICAL SAMPLING OF IRIS ASSESSMENTS

This study of uncertainty and variability in IRIS assessments was carried out on a subset of sixteen of the 537 chemicals contained in the on-line database. The first group of eight assessments was selected from the assessments generated by EPA from the time period 1987 through 1994, also referred to as "pre-Pilot" assessments. The second group of eight was selected from assessments written from 1995 to present, also referred to as "Pilot/post-Pilot assessments." The statistical sampling procedures consisted of a two-step process. First a preliminary stratified sampling and assessment screen was conducted by EPA/NCEA to select 67 of the 537 assessments. Then Versar conducted a random sampling of sixteen final chemicals which became the basis for the in-depth evaluation. While these sixteen chemicals were selected in a statistically-based manner, it should be noted that conclusions drawn on these sixteen selected chemicals may not necessarily represent the entire IRIS database.

Table 2-1. Reviewers of IRIS Uncertainty and Variability Documentation

Anthony Cox, Ph.D.	Ph.D. Risk Analysis
Brent L. Finley, Ph.D., DABT	Ph.D. Toxicology/Pharmacology
Patricia M. McGinnis, Ph.D., DABT	Ph.D. Toxicology
Russell E. Keenan, Ph.D.	Ph.D. Environmental Biology
Bonnie R. Stern, Ph.D., M.P.H.*	Ph.D. Animal Behavior & Ecology, and Reproductive Physiology & Endocrinology
Curtis Travis, Ph.D.	Ph.D. Applied Mathematics

* Lead Reviewer

Table 2-2. IRIS Assessments Selected for Uncertainty and Variability Review

Pre-Pilot IRIS Assessments
4-Methylphenol
Danitol
Dibromo-3-chloropropane, 1,2-
Dichlorodiphenyltrichloroethane, p,p'-
Hexachlorobenzene
Manganese
Prochloraz
Toluene diisocyanate mixture, 2,4-/2,6-
Pilot/Post-Pilot IRIS Assessments
Acetonitrile
Benzene
Beryllium and compounds
Chlordane
Chromium III, insoluble salts
Ethylene glycol monobutyl ether
Methyl methacrylate
Naphthalene

2.1.1 Preliminary Assessment and Screen

EPA/NCEA conducted a random sampling and preliminary screening on 67 of the 537 IRIS assessments on-line as of January 31, 2000. This screening was used to document the extent of uncertainty and variability documentation in IRIS assessments and to develop the statistical sample from which the study sample would be selected. Specifically, from this initial sample of 67, a smaller subset of sixteen chemicals would be randomly selected (see below) to facilitate an in-depth evaluation of IRIS assessments. EPA's simple random sample comprised 10 percent of the pre-Pilot IRIS assessments (52/522), plus all of the fifteen Pilot and post-Pilot IRIS assessments (for a total of 67). These were stratified into three categories, those with "none/minimal," "some/moderate," or "extensive" presentation and discussion of uncertainty and variability (Table 2-3).

Of the 52 pre-Pilot IRIS assessments screened by EPA/NCEA, 3/52 (6 percent) had extensive, 16/52 (31 percent) some or moderate, and 31/52 (64 percent) none or minimal presentation or discussion of variability and uncertainty. In sharp contrast, almost all (14/15, or 93 percent) Pilot/post-Pilot assessments showed extensive treatment of variability and uncertainty.

Table 2-3. Extent of Presentation and Discussion of Variability and Uncertainty

Category	Variability	Uncertainty
None/ Minimal	Any studies relevant to the conclusions are listed, only qualitative dose-responses indicated; no discussion.	Uncertainty factors listed, and overall confidence stated; no discussion.
Some/ Moderate	Conditions for Minimal met, plus adverse effect levels provided; some discussion.	Uncertainty factors listed; some discussion of uncertainty and confidence in the assessment.
Extensive	Conditions for Some/ Moderate met, plus measures of variability or discussion of variability of the results.	Discussion of the strengths and weaknesses of the available studies, with an assessment of the level of confidence in the body of evidence.

Source: Provided by EPA/NCEA.

2.1.2 Random Selection of Final Sixteen IRIS Assessments

The second step in the statistical sampling process was to randomly select sixteen IRIS assessments, eight pre-pilot and eight Pilot/post-Pilot, in a manner that would result in a sample representative of the IRIS database. Specifically, eight pre-Pilot and eight Pilot/post-Pilot assessments were randomly selected from those 67 that EPA/NCEA screened, with the goal of obtaining four each from the some/moderate and extensive categories within each of the pre-Pilot and Pilot/post-Pilot sets. However, there were fewer than four assessments in two of the targeted subgroups. Specifically, only three assessments fell in the extensive subset of the pre-Pilot assessments sampled, so an additional some/moderate assessment was evaluated among the pre-Pilot assessments, resulting in a total of five some/moderate and three extensive pre-Pilot assessments. Similarly, among the Pilot/post-Pilot assessments, only one assessment fell in the some/moderate category, so seven extensive assessments were evaluated to complete the sample of eight Pilot/post-Pilot assessments.

The process used consisted of a random number generator, review of the results, and in one case, a random substitution (see below). The chemicals selected were examined to ensure that they covered ranges of scientific complexities, toxic endpoints of concern, and IRIS entry dates. This sample of sixteen IRIS assessments was the basis for the in-depth review which was intended to: (1) to characterize the extent to which EPA considered uncertainty and variability in the data when developing the IRIS assessments and (2) to compare how uncertainty was considered between older assessments (pre-1995) and those generated between 1995 and the present.

2.1.2.1 *Pre-Pilot Chemicals*

A random selection procedure was performed on the 52 pre-Pilot assessments selected by EPA/NCEA using the statistical software package STATISTICA. This process had three basic steps. The first was to simply to assign a numerical value to each of the 52 assessments. Then we determined the shape of the distribution to ensure that each assessment had an equal probability of being selected. Finally, the last step was to generate random numbers, select the assessments, and review them to ensure that the eight selected assessments covered a broad range of dates of entry into the database, toxic endpoints, type of chemical, and level of complexity. This review of the eight pre-Pilot assessments confirmed that we had met these criteria. If one or more of the criterion had not been met, then we would have repeated/resampled until all criteria were adequately represented.

The eight pre-Pilot assessments selected included all three having extensive documentation and five of the sixteen with some/moderate documentation. This selection procedure achieved EPA's goal for assessments that exhibit a range of scientific complexities, toxic endpoints, and entry dates. The eight pre-Pilot assessments selected cover the years from 1988 to 1995, with six of the eight years represented. With respect to chemical types, four of the chemicals are pesticides, one is a metal and three are other types of organic chemicals. Finally, these assessments cover a broad range of toxic endpoints, including neurologic effects, liver effects, and cancer. Three RfCs, five RfDs, and five cancer assessments are included for this group of chemicals. The specific pre-Pilot assessments selected are summarized in Table 2-4.

2.1.2.2 *Pilot/Post-Pilot Chemicals*

A total of fifteen Pilot/post-Pilot assessments were available, from which eight were randomly chosen for detailed review (Table 2-5). All but one (14 of 15) of these were determined by EPA/NCEA to represent extensive treatment of variability and uncertainty, with only one having some/moderate documentation. The Pilot/post-Pilot assessment with some/moderate documentation was “automatically selected.” Random sampling of seven chemicals was performed from the remaining list of fourteen possible candidates having extensive documentation. This selection yielded eight assessments that exhibit a range of scientific complexities, toxic endpoints, and entry dates. The eight Pilot/post-Pilot assessments selected covered the years from 1998 to 2000, with each year represented. With respect to chemical types, two are pesticides, two are metals, one is a solvent, and three are other types of organic chemicals. These randomly-selected assessments cover a broad range of toxic endpoints, including leukemia, lung cancer, and respiratory effects. Six RfCs, four RfDs, and eight cancer assessments are included for this group of chemicals.

Table 2-4. Final Eight Pre-Pilot IRIS Assessments

Chemical	Overall Documentation	Last Significant Revision	Chemical Type	Key Health Endpoints
Dibromo-3-chloropropane, 1,2-	Extensive	10/01/1991	Pesticide	RfC: testicular effects
Toluene diisocyanate mixture, 2,4-/2,6-	Extensive	09/01/1995	Other	RfC: chronic lung-function decline
Manganese	Extensive	11/01/1995	Metal	RfD: CNS effects RfC: impaired neurobehavioral function Cancer: D, no human data, animal data inadeq.
Dichlorodiphenyltrichloroethane, p,p'-	Some/Moderate	08/02/1988	Pesticide	RfD: liver lesions Cancer: B2, liver tumors, benign and malignant
Prochloraz	Some/Moderate	01/01/1989	Pesticide	RfD: increased SAP and liver wt, liver histopath Cancer: C, liver tumors
Methylphenol, 4-	Some/Moderate	09/01/1990	Other	Cancer: C, skin papillomas
Hexachlorobenzene	Some/Moderate	03/01/1991	Pesticide	RfD: liver effects Cancer: B2, liver, thyroid, kidney tumors
Danitol	Some/Moderate	10/01/1994	Pesticide	RfD: tremors

Table 2-5. Final Eight Pilot/Post-Pilot IRIS Assessments

Chemical	Overall Documentation	Last Significant Revision	Chemical Type	Key Health Endpoints
Chlordane	Extensive	02/07/1998	Pesticide	RfD: hepatic necrosis RfC: hepatic effects Cancer: B2/L, hepatocarcinomas
Beryllium and compounds	Extensive	04/03/1998	Metal	RfD: small intestinal lesions RfC: beryllium sensitivity, progression to CBD Cancer: B1/L, lung cancer
Methyl methacrylate	Extensive	03/02/1998	Other	RfD: none (at any dose tested) RfC: olfactory epithelium degeneration Cancer: E/NL,
Naphthalene	Extensive	09/17/1998	Pesticide;	RfD: decreased BW RfC: respiratory lesions Cancer: C/CBD, respiratory tract tumors
Acetonitrile	Extensive	03/03/1999	Other	RfC: mortality Cancer: D/CBD, (see description)
Ethylene glycol monobutyl ether	Extensive	12/30/1999	Other	RfD: MCV changes RfC: red blood cell count changes Cancer: C/CBD, pheochromocytoma
Benzene	Extensive	01/19/2000	Solvent	Cancer: A/K, leukemia
Chromium III, insoluble salts	Some/Moderate	09/03/1998	Metal	RfD: none (at any dose tested) RfC: inadequate data Cancer: D/CBD, (see description)

For the list of Pilot/post-Pilot chemicals, one random substitution was made. One of the originally-selected chemicals, methylene diphenyl diisocyanate (MDI) was found to have been subjected to extensive review by EPA's Science Advisory Board (SAB) in the past two years as part of a review of the RfC process. Therefore, it was felt that in-depth review of this chemical would not be as beneficial as another randomly selected chemical from the Pilot/post-Pilot list. Another random sampling was performed from the remaining chemicals and methyl methacrylate was selected. Addition of this chemical to the list was found to adequately meet the above criteria for a range of chemical types, endpoints, and date of entry. The final list of eight Pilot/post-Pilot assessments selected are summarized in Table 2-7.

2.2 SELECTING REVIEWERS

Six nationally-recognized experts in toxicology, human health risk assessment, and related disciplines were selected to perform this review of IRIS uncertainty and variability (Table 2-1). These individuals were selected using a rigorous process that included consideration of more than twenty candidates. Extensive review of potential reviewers' credentials (and possible conflicts of interest) was conducted to produce this final review panel. Examples of core criteria included subject matter expertise, publication in the field of study, discipline, peer review experience, and objectivity (i.e., absence of conflict of interest). In this case, we identified individuals who have performed human health risk assessments, who are familiar with EPA's methodologies for establishing toxicity values, and who are expert in evaluating and documenting data variability and risk uncertainty. Several of the individuals have published articles on this topic within the last few years. Other candidates were considered whose research interests include uncertainty in risk assessments. These individuals have particular experience with human health risk and uncertainty/variability analysis and related chemicals and are recognized as leaders in their fields. Additionally, several of the individuals have particular interest/expertise in children's exposures and risks. Therefore, their collective experience covers assessing human health effects of many types of chemicals and many different endpoints (cancer and noncancer). We also sought to get a diverse panel of experts with somewhat different perspectives; some with more biologically-oriented backgrounds and others that would provide more of a mathematical/statistical viewpoint on the uncertainty and variability evaluations presented in the IRIS assessments.

In addition to certifying their scientific expertise, written certification was obtained to ensure that the six reviewers were free of bias or conflicts of interest (see below). One of the reviewers was designated the lead scientist. This individual had additional responsibilities for managing the process, promoting consistency in the reviews, answering questions from the other reviewers, and contributing to this report.

2.3 ASSIGNMENT OF IRIS ASSESSMENTS AND CHARGE TO REVIEWERS

Each reviewer was tasked to produce an independent analysis of each of the eight IRIS assessment packages to which he/she was assigned. The sixteen IRIS assessments were assigned to the six reviewers in a manner that ensured that each chemical was evaluated by three individuals (each reviewer examined eight chemicals). Table 2-6 presents a summary of the assignment of chemicals to reviewers. In general, certain chemicals were assigned to reviewers with particular expertise in that or similar chemicals. In one instance, chemicals were selected in a manner to ensure that no potential existed for conflict of interest.

Table 2-6. Matrix of IRIS Assessment Assignments

Cox	Finley	Keenan	McGinnis	Stern	Travis
1,2-dibromo-3-chloropropane	Toluene diisocyanate mixture	Chromium III	p,p'-Dichloro diphenyltrichloroethane	Methyl methacrylate	4-Methylphenol
Hexachloro benzene	Danitol	Chlordane	Prochloraz	Beryllium	Naphthalene
Acetonitrile	Ethylene glycol monobutyl ether	Benzene	Manganese	1,2-dibromo-3-chloropropane	Toluene diisocyanate mixture
Manganese	p,p'-Dichloro diphenyltrichloroethane	Prochloraz	4-Methylphenol	Hexachloro benzene	Danitol
Chlordane	Methyl methacrylate	Beryllium	Benzene	Acetonitrile	Ethylene glycol monobutyl ether
Benzene	Chromium III	1,2-dibromo-3-chloropropane	Toluene diisocyanate mixture	Manganese	p,p'-Dichloro diphenyltrichloro ethane
Prochloraz	4-Methylphenol	Hexachloro benzene	Danitol	Chlordane	Methyl methacrylate
Beryllium	Naphthalene	Acetonitrile	Ethylene glycol monobutyl ether	Napththalene	Chromium III

A detailed charge to the reviewers was prepared to focus the evaluation on select information in each IRIS package. This charge/checklist was intended to be used as a tool to focus on those issues most central to this effort and to assist with preparing written comments that could be incorporated into this report. The questions under number 1 below were the major issues that EPA wished to be addressed. The remaining questions were a "checklist" to assist the reviewers in evaluating information in each IRIS assessment in a methodical manner. The charge/instructions are summarized below:

1. EPA wishes to have your comments on the following questions:

   A. Did EPA characterize to an appropriate extent the uncertainty and variability in data used to develop these IRIS health assessments?

   B. How does this compare between pre-Pilot and Pilot/post-Pilot assessments?

   C. Did EPA appropriately address:

      (i) strengths and weaknesses of the scientific evidence from available studies, and sources of variability in the data used in the assessment?
      (ii) uncertainties in the underlying data, and uncertainties in the qualitative and quantitative judgements given in the assessment?

   D. Are there other relevant observations or comments that you would like to raise?

2. Identify all risk factors addressed by the IRIS assessment under review: e.g. Q-star or Slope Factor, RfD, RfC, etc.

3. List what relevant background data for each risk factor were available, either on EPA's website, or as bibliographic information, which required review.

   Were all necessary documents reviewed? If not, what was missing?

4. Identify the basis for each risk factor: (e.g. NOAEL, NOEL, LOAEL, LOEL, BMD, etc.)

5. Identify:

   A. Was the risk factor based on human data (describe) or on animal data? How many subjects in the critical study(ies)?
   B. Critical effect
   C. Route of exposure that yielded the critical effect
   D. Mechanism of action for the critical effect observed, if known
   E. For human data: was a sensitive sub-population included?
   F. For animal data: was the species/strain known to be genetically sensitive? To have any genetic peculiarity with regard to the toxicity of the compound?

6. Identify:

   A. Uncertainty Factor & Basis, breaking down the UF into its component parts, as necessary. Were the following considered?
      - inter-species variability
      - inter-human variability
      - extrapolation from less-than-chronic to chronic toxicity
      - extrapolation from LOAEL to NOAEL
      - data-base insufficiencies (e.g. too few studies, limited types of studies)

7. Was the UF data-derived? Was there sufficient information to make a quantitative assessment of uncertainty? If not, how was available information used to derive UF?

8. Identify the MF (modifying factor) used, if any, and its basis.

9. In your judgement, is the critical effect identified relevant to humans? Is the route of exposure relevant to humans?

10. What is the overall confidence rating for the data used to derive the overall slope factor or RfD, or RfC? Do you concur?

2.4 MATERIALS DISTRIBUTED FOR THE REVIEW

The materials distributed to the reviewers consisted of the documents summarized in Table 2-7. For each chemical, a package was assembled and distributed to the assigned reviewers consisting of the IRIS summary sheets and source/supporting documentation. These documents generally consisted of Toxicological Reviews (available for the Pilot/post-Pilot chemicals), health assessment documents, drinking water criteria documents, and related documents that presented the data bases from which EPA prepared the IRIS assessments. The reviewers were instructed to consider the IRIS summary sheets and the Toxicological Reviews collectively as the available documentation (sometimes referred to as the IRIS assessment or IRIS database).

Table 2-7. Documents Sent to Reviewers for IRIS Review

Acetonitrile
- S IRIS Summary Sheets
- S Toxicological Review of Acetonitrile

Benzene
- S IRIS Summary Sheets
- _ Carcinogenic Effects of Benzene: An Update
- S Extrapolation of the Benzene Inhalation Unit Risk Estimate to the Oral Route of Exposure
- S Interim Quantitative Cancer Unit Risk Estimates Due to Inhalation of Benzene

Beryllium and Compounds
- S IRIS Summary Sheets
- S Toxicological Review of Beryllium and Compounds
- S Health Assessment Document for Beryllium

Chlordane (Technical)
- S IRIS Summary Sheets
- S Toxicological Review of Chlordane
- S Carcinogenicity Assessment of Chlordane and Heptachlor/Heptachlor Epoxide

Chromium (III), Insoluble Salts
- S IRIS Summary Sheets
- S Toxicological Review of Chromium (III)

Danitol
- S IRIS Summary Sheets
- S Danitol - Review of a Developmental Toxicity Study in Rats - 7/21/92
- S Review of Chronic Toxicity Study of Fenpropathrin Technical in Dogs - 9/27/95
- S Fenpropathrin: Correction of a DER for a Rat Chronic/Oncogenicity Study - 5/4/93
- S 3 Month Feeding - Rat - 1975
- S 3 Month Feeding - Rat - 1976
- S 3 Month Feeding -Dog - 1980
- S 3-Generation Reproduction - Rat 1979b
- S Developmental Toxicity - Rabbit 1985a
- S 3-Generation Reproduction - Rat 1986
- S 2-Year Feeding - Mouse 1985b
- S 2-Year Feeding - Rat 1986
- S Statistical Evaluation of Summary Rat Oncogenicity Data for Fenpropathrin - 1985
- S Danitol (fenpropathrin) - Data Review and Registration - 9/18/88

1,2,-Dibromo-3-chloropropane (DBCP)
- S IRIS Summary Sheets
- S Drinking Water Criteria Document for DBCP

p,p'-Dichlorodiphenyltrichloroethane (DDT)

- IRIS Summary Sheets
- The Carcinogen Assessment Group's Evaluation of the Carcinogenicity of Dicofol (Keltane), DDT DDE, and DDD (TDE)

Ethylene glycol monobutyl ether (EGME)

- IRIS Summary Sheets
- Toxicological Review of Ethylene Glycol Monobutyl Ether (EGME)

Hexachlorobenzene

- IRIS Summary Sheets
- Drinking Water Criteria Document for Hexachlorobenzene
- Health Assessment Document for Chlorinated Benzenes

Manganese

- IRIS Summary Sheets
- Drinking Water Quality Document for Manganese
- Health Assessment Document for Manganese

Methyl methacrylate

- IRIS Summary Sheets
- Toxicological Review of Methyl Methacrylate

4-Methylphenol

- IRIS Summary Sheets
- Health and Environmental Effects Document for 4-Methylphenol
- Health and Environmental Effects Profile for Cresols
- The Carcinogen Assessment Group's Preliminary Risk Assessment on Cresols Type I - Air Program

Naphthalene

- IRIS Summary Sheets
- Toxicological Review of Napthalene

Prochloraz

- IRIS Summary Sheets
- Memorandum dated April 29, 1987; Prochloraz, Mouse Study-Quantitative Risk Assessment
- Memorandum dated April 29, 1988; Toxicology Branch Peer Review Committee Draft Document on Prochloraz
- Memorandum dated Jan 27, 1988; Prochloraz-Rat Study, Qualitative Risk Assessment
- U.S. EPA - OPP/HED Tox Oneliners; dated 20-Apr-2000

2,4-/2,6-Toluene diisocyanate mixture (TDI)

- IRIS Summary Sheets
- Health and Environmental Effects Document for Toluene Diisocyanate (TDI)

3.0 EVALUATION OF THE EXTENT TO WHICH ASSESSMENTS DOCUMENT THE RANGE OF UNCERTAINTY AND VARIABILITY OF THE DATA

Each of the sixteen IRIS assessments selected was assigned to three experts to review independently. Reviewers were asked to comment on the following three major questions taken directly from EPA's charge. These questions were to be answered in the context of the practices and information available at the time. Reviewers were asked to assess whether:

- EPA characterized to an appropriate extent the uncertainty and variability in data used to develop these IRIS health assessments?

- EPA appropriately addressed uncertainties in the underlying data, and uncertainties in the qualitative judgments given in the assessment?

- EPA appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time, and sources of variability in the data used in the assessment?

A summary of reviewers' findings with regard to these three issue-areas is presented below in Table 3-1 (complete text of each reviewer's comments is attached in Appendix A). This table summarizes the reviewers' overall opinions on each chemical with respect to the whether "EPA characterized to an appropriate extent the uncertainty and variability" in the assessment. These "ratings" of positive, mixed, and negative for each reviewer's evaluation provide a summary of each chemical. The table also shows the range of evaluations from the three reviewers assigned to each IRIS assessment.

Table 3-1. Summary of IRIS Assessments Reviewed

Pre-Pilot IRIS Assessments Reviewed		Pilot/Post-Pilot IRIS Assessments Reviewed	
Chemical	**Reviewer Evaluation (Positive/Mixed/Negative)**	**Chemical**	**Reviewer Evaluation (Positive/Mixed/Negative)**
4-Methylphenol	0 - 0 - 3	Acetonitrile	1 - 0 - 2
Danitol	2 - 0 - 1	Benzene	1 - 0 - 2
DBCP	1 - 2 - 0	Beryllium	1 - 0 - 2
DDT	1 - 1 - 1	Chlordane	0 - 1 - 2
HCB	0 - 0 - 3	Chromium III	0 - 1 - 2
Manganese	2 - 0 - 1	Ethylene Glycol Monobutyl Ether	2 - 1 - 0
Prochloraz	0 - 1 - 2	Methyl Methacrylate	3 - 0 - 0
Toluene Diisocyanate	2 - 0 - 1	Naphthalene	1 - 2 - 0

A comparison of these summary ratings against the categories that EPA assigned (extensive, or some/moderate) in the preliminary screening can be made to examine agreement between the preliminary screening and the in-depth review. Table 3-2 presents a summary of this comparison. For example, of the eight pre-Pilot chemicals, EPA found that three had extensive documentation. These three chemicals (and two others) received positive ratings after the in-depth review. Specifically, two of the three chemicals were rated "positively" by two of the three reviewers. Other pre-Pilot assessments did not receive any positive ratings. These results from the extensive and some/moderate chemicals from the pre-Pilot group show some degree of agreement with the preliminary screening. There is also some agreement between the preliminary screen and the review for the Pilot/post-Pilot chemicals. Seven of the eight Pilot/post-Pilot chemicals were found by EPA to have extensive levels of documentation. The reviewers were in general agreement with this finding, since six of the seven chemicals with extensive documentation received a positive rating from at least one reviewer (and two of the chemicals received either two or three positive ratings). The one chemical that EPA felt only had some/moderate documentation was trivalent chromium. This chemical was one of only two Pilot/post-Pilot chemicals to receive no positive ratings.

Table 3-2. Comparison Between EPA's Preliminary Screening and the "Ratings" Provided From the In-Depth Review of Sixteen IRIS Chemicals

IRIS Assessment	EPA's Evaluation of Extent of Documentation	In-depth Rating (Positive/mixed/negative)
Pre-Pilot Chemicals		
Dibromo-3-chloropropane, 1,2-	Extensive	1-2-0
Toluene diisocyanate mixture, 2,4-/2,6-	Extensive	2-0-1
Manganese	Extensive	2-0-1
Dichlorodiphenyltrichloroethane, p,p'-	Some/Moderate	1-1-1
Prochloraz	Some/Moderate	0-1-2
Methylphenol, 4-	Some/Moderate	0-0-3
Hexachlorobenzene	Some/Moderate	0-0-3
Danitol	Some/Moderate	2-0-1
Pilot/post-Pilot Chemicals		
Chlordane	Extensive	0-1-2
Beryllium and compounds	Extensive	1-0-2
Methyl methacrylate	Extensive	3-0-0
Naphthalene	Extensive	1-2-0
Acetonitrile	Extensive	1-0-2
Ethylene glycol monobutyl ether	Extensive	2-1-0
Benzene	Extensive	1-0-2
Chromium III	Some/Moderate	0-1-2

The remainder of this section provides short summaries of reviewers' responses to the three questions listed above for each IRIS assessment reviewed. Pre-Pilot chemicals are listed first, followed by Pilot/post-Pilot chemicals. Table 3-3 summarizes uncertainty and modifying factors presented in each of the sixteen assessments for non-cancer endpoints. This table illustrates several points about the sample of IRIS assessments reviewed. First, it may be seen that the uncertainty factors for the pre-Pilot assessments are often smaller than those calculated for the Pilot/post-Pilot assessments. Second, with a single exception, uncertainty factors for the pre-Pilot assessments employed default values, while Pilot/post-Pilot IRIS assessments often employed factors other than 10.

3.1 4-METHYLPHENOL (PRE-PILOT - 1990)

EPA only partially characterized the uncertainty and variability in data used to develop this IRIS health assessment. Reviewers pointed out that the IRIS assessment contained no discussion of uncertainty because the non-cancer toxicity value (i.e. oral RfD) had been withdrawn. No confidence rating was assigned or discussed for the oral cancer slope factor.

One reviewer pointed out that two years after the IRIS assessment was published, an EPA document classified 4-methylphenol as a Group D carcinogen, indicating inadequate data for categorization as a human carcinogen.

EPA did not appropriately address uncertainties in the underlying data. Two reviewers stated that the IRIS assessment contained no discussion of 4-methylphenol's mechanism of action, and no discussion of the relevance of the critical effect to humans, both of which bear on the uncertainty of the oral slope factor.

EPA did not appropriately address strengths and weaknesses in the scientific evidence from studies available at the time. Uncertainties with regard to the completeness of the data-base used to support the oral slope factor were incompletely described. Examples included: i) reliance on case reports, in some cases involving single individuals exposed to multiple chemicals; ii) skin papillomas being found only in animal studies designed to evaluate the promoting activity of 4-methylphenol. IRIS does not report that no malignancies were found at any dose; iii) the genotoxicity data base supporting the oral slope factor seemed to cite only positive data from studies with *mixtures* of methylphenol isomers, but did not cite negative data (obtained with 4-methylphenol alone) referenced in supporting documents (and therefore available at the time). One reviewer noted that all published, peer-reviewed genotoxicity studies were negative; only unpublished studies with mixtures of methylphenol isomers yielded positive results.

Table 3-3. Uncertainty and Modifying Factors Presented in 16 IRIS Assessments

	Pre-Pilot Assessments											
	4-Methylphenol	**Danitol**	**DBCP**	**DDT**		**HCB**		**Manganese**		**Prochloraz**		**Toluene Diisocyanate**
	CSF	RfD	RfC	RfD	CSF	RfD	CSF	RfD	RfC	RfD	CSF	RfC
UF	—	100	1,000	100	—	100	—	1	1,000	100	—	30
MF	—	1	1	1	—	1	—	3	1	1	—	—
Total	—	100	1,000	100	—	100	—	3	1,000	100	—	30

	Pilot/Post-Pilot Assessments														
	Acetontonitrile	**Benzene**	**Beryllium**			**Chlordane**			**Chromium III**	**Ethylene Glycol Monobutyl Ether**		**Methyl Methacrylate**		**Naphthalene**	
	RfC	CSF	RfD	RfC	CSF	RfD	RfC	CSF	RfD	RfD	RFC	RfD	RfC	RfD	RfC
UF	100	—	300	10	—	300	300	—	100	10	30	100	10	3,000	3,000
MF	10	—	1	1	—	1	1	—	10	1	1	1	1	1	1
Total	1,000	—	300	10	—	300	300	—	1,000	10	30	—	—	3,000	3,000

3.2 DANITOL (PRE-PILOT - 1994)

Reviewers did not agree as to whether EPA appropriately characterized the uncertainty and variability in data used to develop this IRIS health assessment. One of the reviewers felt these were properly characterized with respect to the oral RfD (which was the only toxicity endpoint defined). Another reviewer found that although characterization was lacking, this was in accord with the standards of the time (i.e., the RfD was verified in 1993). The third reviewer was dissatisfied with EPA's characterization, stressing that the assessment lacked narrative text or interpretation describing either the rationale for the uncertainty factors chosen or variability across the aggregated data-set.

EPA only partially addressed uncertainties in the underlying data. Standard default values to account for inter- and intra-species differences were used in accord with the practice at the time. The IRIS assessment contained no discussion of the mechanism of action, variations in species susceptibility, and no discussion of the relevance of the critical effect to humans (or the relevance of the dosing regimen to likely human exposure pathways), all of which bear on the uncertainty of the oral RfD.

EPA only partially addressed strengths and weaknesses in the scientific evidence from studies available at the time. This chemical is a pesticide, and much of the supporting data is not available in the open literature. One reviewer suggested that the documentation for these files differs from others on IRIS and explain why. Another reviewer pointed out that the summary makes no mention of human exposure data or toxic effects. The third reviewer pointed out that although the IRIS assessment cites 11 studies in support of the RfD (in addition to the critical study), it did not provide a satisfactory discussion of the uncertainty associated with these data. For example, EPA provided no justification as to why a 1-year dog-feeding study (designated the principal study) was superior to the 2-year mouse and rat feeding studies. Another major source of uncertainty in the dog-feeding study was that the animals were fed *ad libitum* and actual doses given were therefore unknown. It was unclear whether this uncertainty applied to the rat and mouse feeding studies, and if not, why these lifetime dosing studies were considered to be less reliable than the shorter dog-feeding study.

3.3 DIBROMOCHLOROPROPANE (DBCP - PRE-PILOT - 1991)

EPA adequately characterized the uncertainty and variability in data used to develop this IRIS health assessment. The reviewers felt that although all appropriate uncertainty factors were considered in the uncertainty section, and a very brief synopsis of the rationale for use of each was provided, EPA did not adequately support choices made for the uncertainty factors used.

EPA appropriately addressed uncertainties in the underlying data. The critical study was well conducted and uncertainty and variability in the critical study were well characterized. The uncertainty factor was, in part, data-derived, in that a dosimetric adjustment was used to convert the inhalation animal NOAEL to an inhalation human NOAEL, and the interspecies extrapolation uncertainty factor was reduced to 3.

EPA did not appropriately address strengths and weaknesses in the scientific evidence from studies available at the time. First, the IRIS assessment may be biased towards consideration of testicular toxicity data, for which there was extensive review of supplemental documentation. This results in additional uncertainty with regard to the data base. The IRIS file understandably focused on male reproductive toxicity in humans and animals, and on standard animal bioassay data (i.e., chronic and subchronic studies). However, a MEDLINE search of scientific literature available at the time the IRIS assessment was drafted, revealed that other relevant toxicity data existed but were not included in the supplemental documentation, including studies on: female reproductive toxicity, pharmacokinetics and metabolism, carcinogenicity (e.g., a 1982 NTP bioassay was available at the time), mutagenicity/ genotoxicity, and liver/kidney toxicity. As a result, uncertainties in the underlying data concerning non-testicular target organs of toxicity are not well addressed.

3.4 DICHLORODIPHENYLTRICHLOROETHANE (DDT - PRE-PILOT - 1987 & 1991)

EPA partially characterized the uncertainty and variability in data used to develop this IRIS health assessment. All reviewers stated that little information on uncertainty or variability was presented. One reviewer notes that each component of the aggregate uncertainty factor should be discussed, even if the individual factor is only assigned a value of "1." For example, uncertainty as to whether the database was sufficient was not mentioned - it is therefore unclear whether this potential source of uncertainty was considered. This was, however, in accordance with the standards of the time, as the database uncertainty factor was not routinely considered by EPA before 1993.

EPA only partially addressed uncertainties in the underlying data. The IRIS assessment contained no discussion of the mechanism of action, potential for existence of sensitive subpopulations, relevance of the critical effect to humans, or intra-species difference in metabolism, all of which bear on the uncertainty associated with the oral RfD, the cancer slope factor, and the inhalation unit risk values. One reviewer notes that there was no discussion of the uncertainty associated with using oral exposure data to develop an inhalation slope factor. The cancer slope factor referenced in the assessment is identified by EPA as deriving from the geometric mean of ten slope factors (over a 13 fold range) derived from six mouse and rat studies. EPA makes it plain that there is little variability in the potency data, even across

different species and strains. However, uncertainty and limitations in the database for cancer assessment are not well laid out.

EPA did not appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. Two of the reviewers felt that assessment was deficient in that no discussion of the strengths and weaknesses in the scientific evidence (or the qualitative or quantitative judgments made relative to uncertainties or variability within it) was presented. For example, there was no discussion or interpretation as to the merit or validity of the key studies, one of which was performed in 1950. Reproductive studies are presented as having conflicting results, however, EPA does not interpret these findings or address uncertainty associated with potential relevance to humans. One of the reviewers felt that the supporting document for the cancer assessment was superior to the IRIS summary in weighing the evidence and pointing out limitations in the studies, but felt that specific types of uncertainty were not discussed, in accord with the practice at the time.

3.5 HEXACHLOROBENZENE (PRE-PILOT - 1988)

EPA only partially characterized the uncertainty and variability in data used to develop this IRIS health assessment. Only default uncertainty factors are referenced; support for these values is lacking. However, this was in accord with the standards of the time.

EPA did not appropriately address uncertainties in the underlying data. Reviewers stated that EPA should have considered: (i) the uncertainty of extrapolating unit risks from rats to humans; (ii) the uncertainty of extrapolating from sub-chronic to chronic study; and (iii) the uncertain relevance of animal toxicity data to the human porphyria effects observed.

EPA did not appropriately address strengths and weaknesses in the scientific evidence from studies available at the time. Important information is not discussed in this IRIS assessment, however some of this information became available only after the assessment was completed. The main omissions are as follows: (i) discussion of the known role of hexachlorobenzene as a liver tumor promotor in rats and mice (e.g., Shirai, 1978); (ii) discussion of the mechanism of toxicity in general; (iii) discussion of the role of enzyme induction and interspecies differences; (iv) discussion of mechanistic data (e.g., Stewart, 1989) applicable to cancer risk; (v) discussion of epidemiological data (e.g., Currier, 1980; Sala, 1999); (vi) discussion of the probability that hexachlorobenzene is (or is not) a human carcinogen, and to factor in that information to the quantitative assessment. Genotoxicity/mutagenicity data to support the tumor results and dose-response information are not presented.

3.6 MANGANESE (PRE-PILOT - 1995)

Reviewers did not agree as to whether EPA appropriately characterized the uncertainty and variability in data used to develop this IRIS health assessment. Opinion was divided among the reviewers as to whether EPA appropriately characterized uncertainty and variability; two felt that EPA characterized uncertainty fairly well, while the dissenter felt that a more rigorous approach with regard to mathematics and statistical treatment would have been useful. However, such an approach was not in accordance with standard practice at the time the assessment was developed by EPA. [One reviewer noted that the carcinogenicity assessment was performed in 1987, and placed in IRIS in 1988. The RfD and RfC were developed and loaded into IRIS in 1995 and 1993, respectively.]

EPA appropriately addressed uncertainties in the underlying data. Two reviewers stated that EPA characterized uncertainty and variability in the data well when setting the oral RfD. The reviewer also stated that "the quantitative assessment discusses and takes into account both the essentiality and toxicity of manganese, the lack of a reliable animal model for manganese toxicity, individual variability with regard to toxicity and essentiality requirements, pharmacokinetic information, [possible] mechanisms of neurotoxicity, and the limitations of deriving an RfD for an essential trace element."

EPA appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. One reviewer stated that "uncertainties concerning pharmacokinetics, individual variability, and sensitive subgroups such as infants and persons with pre-existing liver disease are thoroughly discussed."

3.7 PROCHLORAZ (PRE-PILOT - 1989)

Reviewers did not agree as to whether EPA characterized appropriately the uncertainty and variability in data used to develop this IRIS health assessment. Default uncertainty values were applied to yield the oral RfD; uncertainty associated with the RfD was not characterized comprehensively. One reviewer felt that this was consistent with the standards of the time. Other reviewers noted that this IRIS assessment accounted only for inter- and intra-species differences. Uncertainty associated with database insufficiencies was not mentioned (however, the practice of considering database uncertainty post-dates this assessment) and very little information was provided regarding the confidence in the underlying data for the RfD. With regard to the carcinogenicity assessment, the reviewers did not agree as to whether EPA adequately captured uncertainty associated with the oral slope factor. One reviewer felt strongly that because the slope factor was extrapolated by default from a linear no-threshold dose-response curve, use of

such an upper-bound estimate from a linear model precluded a proper evaluation of uncertainty. However, use of such a model was in accordance with EPA's 1986 *Guidelines for Carcinogen Risk Assessment*.

EPA did not appropriately addressed uncertainties in the underlying data. The IRIS assessment did not discuss prochloraz' mechanism of action, and there was no discussion of the relevance of the critical effect (benign and malignant mouse liver tumors) to humans, both of which bear on the uncertainty of the oral slope factor.

EPA only partially addressed strengths and weaknesses in the scientific evidence from studies available at the time. In general, with regard to the RfD, the reviewers felt that the IRIS assessment lacked a discussion of the strengths and weaknesses of the underlying database. With regard to the cancer assessment, the evaluation could have gone further and laid out how the other information in the database factored into the weight of evidence, e.g., the lack of positive tumorigenic response in rats, no indication of tumors in a 2-year dog study, negative genotoxicity test results, and structure:activity relationships to known carcinogens (e.g. 2,4,5-T, Silvex, 2,4,6-trichlorophenol). One reviewer felt that EPA should consider contemporary evidence that prochloraz might act purely as a non-genotoxic promotor, and whether evidence exists that it causes liver damage in humans at low doses. Another reviewer pointed out that confidence in the cancer assessment was increased, knowing that it had been reviewed by the OPP Peer Review Group and the FIFRA SAP. This reviewer suggested that EPA could explain why documentation for pesticides differs from other chemicals on IRIS.

3.8 TOLUENE DIISOCYANATE (TDI - PRE-PILOT - 1995)

EPA characterized fairly well uncertainty and variability in data used to develop this IRIS health assessment. One reviewer stated that discussion of uncertainties identified was inadequate and too brief, while another agreed but pointed out that this was in accord with the standards of the time. The third reviewer felt that uncertainty and variability were thoroughly discussed. The latter stated: "There are several epidemiological studies of worker exposure to TDI and the summary does a good job of sorting through them and identifying the uncertainties and variabilities in the overall data set. [This is] particularly [so] with respect to co-exposure to other chemicals and smoking effects."

EPA only partially addressed uncertainties in the underlying data. A number of uncertainties were not discussed. These included: (i) whether peak exposure or chronic exposure is the proper exposure variable to consider, (ii) potential effects of co-exposure to other chemicals and increased degradation of lung functions, (iii) the fact that the LOAEL is based on a measurable decrease in lung function, but the most sensitive end point may be hypersensitivity to low, nonirritating concentrations of TDI, (iv) the fact

that hypersensitivity has been shown to occur in humans following acute exposures and, thus, that the LOAEL may not be protective against human hypersensitivity, (v) the mechanism of action for both decreased lung function and hypersensitivity, and (vi) existence of sensitive subpopulations.

There is no discussion as to whether the calculated RfC will be protective against the development of sensitization, the critical effect known to occur in humans. The assessment uses an additional uncertainty factor of 3 to account both for subchronic to chronic extrapolation and the lack of developmental toxicity data in a second species. The document does not explain why a five-year occupational study is considered subchronic, nor does the document explain why an additional uncertainty factor is needed because of the lack of developmental studies in a second species.

EPA only partially addressed strengths and weaknesses in the scientific evidence from studies available at the time. Potentially supporting studies which were not evaluated included many additional studies of occupational exposure studies and one chronic animal inhalation study. There was a multigenerational reproductive/developmental study. No use is made of the additional human studies.

3.9 ACETONITRILE (PILOT/POST-PILOT - 1999)

EPA did not appropriately characterize the uncertainty and variability in data used to develop this IRIS health assessment. Two of three reviewers felt that EPA did not adequately support choices made for the uncertainty factors used. For example, one reviewer felt that the MF of 10 was too conservative, given: (i) the lack of clarity as to whether exposure via inhalation was relevant to the development of the forestomach lesions seen after oral exposure to acetonitrile in animal studies, and (ii) the questionable relevance of mouse forestomach lesions to human toxicity. The justification for this uncertainty was not discussed in the IRIS assessment. Another MF value could have been considered for severity of effect, and to protect against more subtle, unreported symptoms which may precede death (the critical effect). The MF of 10 was, therefore, inappropriate for forestomach toxicity, and should have been either 1 or 3.

EPA did not appropriately address uncertainties in the underlying data. The IRIS assessment contained no discussion of the physicochemistry, pharmacokinetics, or mechanism of action of acetonitrile, nor was the relevance of the critical effect (mortality) to humans discussed, all of which bear on the uncertainty associated with the inhalation RfC. Uncertainties associated with using a frank effect such as mortality as a critical effect should have been more fully discussed. It would have been useful to discuss what caused mortality at the higher doses and whether less severe effects might have been observed at

lower doses if the existence of other, less toxic effects (e.g., neurotoxicity or cardiotoxicity) had been explored. Uncertainties associated with the limited hematologic data were not well presented.

EPA did not appropriately address strengths and weaknesses in the scientific evidence from studies available at the time. The critical effect was mortality (death of a single female mouse), and although increases in forestomach hyperplasia were seen, they were probably associated with grooming of contaminated animal fur and/or mucociliary clearance with ingestion. The relevance of this effect for humans is questionable. Also, conflicting findings of mortality in the subchronic rodent study versus lack of mortality in the chronic bioassay with the same species could have been better discussed. Concerns about developmental toxicity data are unfounded as the critical effect in the principal study protects against all adverse effects occurring at higher doses and developmental studies indicate that adverse effects are only observed at very high doses and usually in conjunction with maternal toxicity.

3.10 BENZENE (PILOT/POST-PILOT - 2000)

Reviewers disagreed as to whether EPA appropriately characterized uncertainty and variability in data used to develop this IRIS health assessment. Two of three reviewers felt that EPA did not adequately characterize uncertainty. The approving reviewer stated that the quantitative risk assessment was difficult to follow and understand, even with the assistance of the support documentation.

Reviewers disagreed as to whether EPA appropriately addressed uncertainties in the underlying data. The approving reviewer felt that in the absence of a well-accepted mechanism for benzene-associated leukemia in humans, EPA adequately laid out in narrative form a number of important uncertainties particularly well, including: (i) the exposure measurements for the principal study (Rinsky et al., 1981, 1987); (ii) reconstruction or estimation of exposures for the early years of this cohort for dose-response assessment; (iii) mode of leukemogenesis and hematoxicity; (iv) metabolism of benzene; (v) shape of the dose response model; (vi) choice of extrapolation model; and (vii) existence of potentially sensitive subpopulations. The unit risk expressed as a range quantitatively captures uncertainty in the exposure-response relationship as well as in the choice of model (linear model).

One of the two dissenting reviewers felt that: "EPA did not acknowledge serious uncertainty in the scientific community as to whether benzene can cause leukemia at concentrations less than 10 ppm (i.e., q_1* may be zero at low concentrations). Substantial epidemiological and mechanistic evidence of a leukemic response threshold is not discussed. The quantitative analysis ignores uncertainties and variabilities (e.g., uncertainties about mechanism of action, model form, model parameters, inter-individual variability, exposure uncertainties, extrapolation of uncertainties across routes and regimens). Thus, the

assessment presents an artificially narrow uncertainty range that belies the true uncertainties. "Uncertainty" is not mentioned as a topic. "Variability" is discussed only in the context of obtaining an oral risk factor from the inhalation [risk factor]."

"Uncertainties in the following were either not included or ignored:

- uncertainties about model form and exposure data were identified as crucial in the qualitative discussion but ignored in the quantitative analysis.
- uncertainty in exposure estimates was not included in statistical models.
- uncertainty in the oral absorption factor was ignored by assuming it was 0.5
- the fact that more years of follow-up data weakened the associations claimed by Rinsky is not discussed.
- uncertainty related to sub-chronic to chronic extrapolation was ignored. This is a major source of uncertainty. For example, effects seen above 10 ppm are not seen below 10 ppm in mice (Farris, 1997). Evidence of threshold effects in the Pliofilm (Schnatter 1996; Wong, 1995) and other (Wong and Raabe, 1995) worker populations was not discussed.
- the fact that the leukemias in the Pliofilm cohort occurred primarily at a plant for which no exposure data were available was not addressed."

"Variability in the following were not included or ignored:

- inter-human variability was ignored. The known inter-individual variability in benzene metabolism based on CYP2E1 was not addressed in the model.
- variability in data found across plant locations was not addressed."

Reviewers disagreed as to whether EPA appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. One of the dissenting reviewers stated that: "Relevant animal and epidemiological studies suggesting that the linear model is not appropriate for benzene were not discussed (e.g. Farris, 1997; Schnatter, 1996; Wong, 1995; Wong & Raabe, 1995). Human, animal, and *in vivo* studies showing a non-linear mechanism... were ignored.... The very likely mechanism of action is aneuploidy induced by binding of reactive metabolites (p-benzoquinone) to thiol/sulfhydryl groups on tubulin in CD34+ CFU-GM stem cells. This disrupts spindle formation and segregation of chromosomes during mitosis - an inherently non-linear mechanism. Dose-dependent aneuploidy has been found in benzene-exposed workers in vivo (Zhang et al., 1998). The above mechanism has been demonstrated in detail *in vitro* (Pfeiffer and Metzler, 1996). The IRIS document does not cite these or other relevant literature (including key papers by Irons and Stillman). It [employs] a linear

model as a default, on the grounds that the non-linear aneuploidy mechanism just described "has not been shown conclusively." It also conjectures that multiple mechanisms may be involved, without citing specifics, and concludes that not enough is known to justify deviating from a low-dose linear model."

3.11 BERYLLIUM (PILOT/POST-PILOT - 1997)

Reviewers disagreed as to whether EPA appropriately characterized the uncertainty and variability in data used to develop this IRIS health assessment. Two of three reviewers felt that uncertainty for each of the three toxicity endpoints was handled very unevenly. One reviewer felt that the toxicity and uncertainty assessments for the inhalation RfC were of higher quality than those documented for the oral RfD or the oral slope factor.

With regard to the carcinogenicity assessment, one reviewer felt that the uncertainty and variability were very poorly characterized, both qualitatively and quantitatively.

EPA did not appropriately address uncertainties in the underlying data. One reviewer stated that because the toxicokinetics and toxicity of metal compounds differ markedly by route of exposure, the use of inhalation data to raise concerns about oral exposure is of limited relevance. The IRIS assessment contained no discussion of beryllium toxicokinetics, or beryllium speciation and toxicity, nor was the relevance of the critical non-cancer effect (gastrointestinal inflammation) to humans discussed, all of which bear on the uncertainty associated with the toxicity endpoints derived.

EPA did not appropriately address strengths and weaknesses in the scientific evidence from studies available at the time. Reviewers felt the strengths and weaknesses of the scientific evidence and uncertainties in qualitative and quantitative judgements were poorly addressed. The data-base was lacking in studies regarding effects of beryllium on the immune system, which is the key target organ for the development of the sensitization response progressing to chronic beryllium disease which in humans is the critical effect after inhalation exposure.

Furthermore, with regard to the oral non-cancer data, one reviewer stated the following: "The IRIS file reports in error that adverse effects were seen at the two highest dose levels. Statistical significance was not reported but it is unlikely that the 1/10 incidence of gastrointestinal lesions reported in the second highest dose group tested is significant. Therefore, there were no significant adverse findings in this dose group. The only animals to demonstrate significant gastrointestinal toxicity were in the 500 ppm group, the group which was terminated after 33 weeks of exposure because of "overt signs of toxicity." This toxicity was not well described but appeared to involve systemic blood infection resulting from perforation of the

gastrointestinal tract and consequent bacterial invasion. Because toxic [effects] were severe enough to warrant termination of exposure after 33 weeks, the Maximum Tolerated Dose was exceeded. Therefore, the critical effect, observed only in this group, is not appropriate. Further, the study cannot be considered to be of chronic duration because the only dose group to show significant incidence of the critical effect was terminated at week 33. If the 500 ppm-dose group were to be excluded from the study, then the study NOAEL would be 50 ppm, the highest dose group to undergo long-term exposure (172 weeks)."

Also, it was felt that each of the studies in the underlying carcinogenicity data base, human and animal, had many deficiencies and these were not well summarized. One reviewer stated: "Shortcomings in the cancer epidemiology studies which may have led to an overestimate of excess lung cancer risk associated with beryllium exposure were not clearly described. There were no exposure measurements in any of the principal epidemiology studies and a convoluted method for using estimates by NIOSH of the upper and lower bounds of airborne exposure concentrations (spanning one order of magnitude) for quantitative assessment of unit cancer risk was incompletely presented. It is unclear how NIOSH arrived at these estimates, whether they can be justifiably applied to the principal study data, and how excess relative risks were converted to excess cancer incidences to derive the inhalation unit risk. "

3.12 CHLORDANE (PILOT/POST-PILOT - 1999)

EPA did not appropriately characterize the uncertainty and variability in data used to develop this IRIS health assessment. All reviewers were largely in agreement that uncertainty and variability were not adequately characterized and quantified in this IRIS assessment.

EPA did not appropriately address uncertainties in the underlying data. Key data uncertainties such as: (i) available mechanistic knowledge, (ii) evidence of sub-linear dose-response, (iii) evidence that rodent endpoints (liver responses) are not relevant for humans, were recognized in the qualitative discussions, but were not addressed in the quantitative assessment and uncertainty factors, (iv) uncertainties in within-species variability, and (v) data base deficiencies which include not only the lack of a multi-generation reproductive study but also the lack of a second developmental toxicity study in rabbits in accordance with current test guidelines.

With regard to the carcinogenicity assessment, both uncertainty and variability are poorly characterized. Discussion of the evidence for sublinearity or nonlinearity of the dose-response of chlordane carcinogenicity in the low-dose region of the dose-response curve is restricted to a single sentence in the Confidence Section. [The results of the cell proliferation study, the numerous mutagenicity/ genotoxicity studies, the evidence supporting the sublinearity/nonlinearity at low doses, and the cytotoxicity

accompanying the high-dose tumorigenic findings (which are supportive of a cytotoxicity-mediated carcinogenic response) are not appropriately presented or integrated into the discussion.] Only the linear multi-stage model was used. Models for non-genotoxic and known tumor promotor effects were not used. Evidence for sub-linearity at low doses and the possibility of a zero slope at the origin were not separately evaluated. The route-to-route extrapolation for derivation of an inhalation cancer slope factor was not fully justified and the use of a default inhalation absorption rate of 100% was likely to be an overestimate due to the semi-volatile nature of chlordane.

EPA did not appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. The reproductive study by Narotsky and Kavlock (1995) is not cited; this study found a significant increase in the percent loss of pups per litter at both doses tested. Relevant data for hematotoxic effects (a toxic effect seen in humans) were not obtained. This uncertainty was mentioned but ignored in the quantitative assessment. Very little data was presented on pharmacokinetics: specifically, absorption, distribution, metabolism, and excretion. Metabolic pathway information was not presented. These data were available at the time of the IRIS assessment.

3.13 CHROMIUM III (PILOT/POST-PILOT - 1998)

EPA did not appropriately characterize uncertainty and variability in data used to develop this IRIS health assessment. Reviewers were in general agreement that EPA did not adequately characterize uncertainty and variability in this IRIS assessment. Examples of deficiencies identified included: (i) uncertainty associated with extrapolating from subchronic to chronic and LOAEL to NOAEL values, (ii) no justification for the 10-fold factors applied to interspecies and intrahuman variability, and (iii) why the "database deficiency" was addressed using a modifying factor and not an uncertainty factor.

EPA did not appropriately address uncertainties in the underlying data. No critical toxic effects were identified in the principal study, however those sought were limited to body weight loss and gross histological impacts. One reviewer felt that EPA should have noted that, although there was no observed critical effect in the principal study, other studies might have produced a NOAEL at even higher doses, or alternatively, other studies looking for more sensitive endpoints (e.g. kidney toxicity or neurotoxicity) might have led to a lower RfD being defined. The IRIS assessment contained no discussion of the toxicokinetics of chromium III.

EPA only partially appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. One reviewer pointed out that the O'Flaherty physiologically-based model for chromium disposition (available at the time of the update) might have been used to reduce the

factor of 10 uncertainty for interspecies extrapolation, and noted that two other post-Pilot assessments reviewed had reduced the interspecies UF from 10 to 3 based on the existence of a pharmocokinetic model. Although there were references to human exposure data being available, outcomes and uncertainties associated with their use were not discussed further.

3.14 ETHYLENE GLYCOL MONOBUTYL ETHER (EGME - PILOT/POST PILOT - 1999)

Reviewers largely (but not completely) agreed that EPA appropriately characterized uncertainty and variability in data used to develop this IRIS health assessment. One reviewer felt that EPA more than adequately characterized uncertainty and variability in this IRIS assessment.

EPA partially addressed uncertainties in the underlying data. One reviewer approved use of "the maximum concentration of the metabolite BAA in blood ... as the dose metric. This was an innovative step forward in using biological data to reduce uncertainty." The same reviewer approved use of a PB/PK model as "another excellent example of the use of data to quantitatively reduce uncertainty, and at the same time, provide a mechanistically more realistic approach to estimating the RfD." However, uncertainties associated with the use of the model needed to be addressed in more detail.

Another reviewer felt that EPA failed to justify the relevance of the critical effect observed to humans. This reviewer stated: "[The only human data discussed are case reports in which no hemolytic effects were observed. Since the EPA's analysis is based on the assumption that hemolytic effects will occur in humans at sufficient doses of EGME, this uncertainty needs to be discussed. Similarly, the summary does not discuss whether BAA formation in the blood, which is used as the target metric for the PB/PK model, is thought to be a precursor to or even related to toxic effects in humans. Again, this is a critical underpinning of EPA's analysis, and the uncertainties associated with this assumption are not mentioned."

EPA appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. The uncertainties and variabilities in the animal data supporting the reference doses and concentrations were well-described, particularly with respect to inter-species (rat vs. mouse) and inter-gender sensitivities and the possible effects of age.

3.15 METHYL METHACRYLATE (MMA - PILOT/POST-PILOT - 1998)

EPA appropriately characterized uncertainty and variability in data used to develop this IRIS health assessment. All three reviewers felt that EPA characterized uncertainty and variability in the data

used to derive the toxicity endpoints appropriately. Uncertainties were "addressed clearly, succinctly, in a focused, integrated manner." The uncertainty factors applied were partially data-derived.

EPA appropriately addressed uncertainties in the underlying data. One reviewer pointed out that uncertainty associated with extrapolating from subchronic to chronic studies, and also for extrapolating from a LOAEL to NOAEL, were not discussed. [Although it was possible that] EPA [had actually] considered them and chose to assign them a value of 1, the reviewer felt this should have been mentioned for completeness' sake. Reviewers suggested that the assessment should have included a section explaining the mechanism of action and MMA metabolism. Also, the assessment should have discussed interspecies extrapolation in terms of pharmacokinetics and pharmacodynamics.

All in all, EPA appropriately addressed strengths and weaknesses in the scientific evidence from studies available at the time. The critical study was conducted in 1964, and may not meet current GLP criteria. The IRIS assessment would be improved if it discussed: what aspects of the study would fail to meet GLP, and how does this influence the conclusions regarding the uncertainty of the study and the RfD overall.

The portal-of-entry target organ for MMA toxicity is identified as the rodent forestomach. One reviewer pointed out that: "The relevance of forestomach toxicity is unclear because no mention of a forestomach effect in either the IRIS assessment or the Toxicological Review was located. Therefore, it is unclear why this rodent target organ was discussed in the Uncertainty section."

3.16 NAPHTHALENE (PILOT/POST-PILOT - 1998)

EPA appropriately characterized uncertainty and variability in data used to develop this IRIS health assessment. Reviewers generally expressed that overall EPA did a good job of characterizing uncertainty. However, all felt more text was required to adequately support the uncertainty factors used.

EPA incompletely addressed uncertainties in the underlying data. The uncertainty factors for the RfD and the RfC are both 3,000. The reviewers felt these high values require justification because EPA did not provide enough explanation, and the reviewers were left with a number of questions. For example, the interspecies variability, intra-human variability, and subchronic to chronic extrapolation factors were all set at 10, when no toxic effects were seen in rodent studies, and no sensitive subpopulation of humans had been defined relative to the critical effect (nasal toxicity).

Another reviewer felt that: “evidence for the relevance of the critical effect to humans was lacking because no data suggest that decreased body weight occurs in humans exposed to naphthalene. [EPA only states] that it is not possible to determine whether the RfD is protective of hemolytic anemia in humans.”

One reviewer pointed out that: “Additional text on the uncertainties associated with calculating the HEC for the RfC would be appropriate. In particular, the uncertainty in the assumption regarding the mouse:human blood/gas partition coefficients should be addressed. Because these coefficients are “not available,” a default ratio of 1 was used. It is unclear to the reader whether any level of confidence should be assigned to this value.”

EPA mostly addressed strengths and weaknesses in the scientific evidence from studies available at the time. One reviewer stated: “Supporting data for the principal study are well presented. However, a review of other data available at the time turned up gaps, including reproductive and developmental toxicity studies and pharmacokinetic data. Numerous species and tissue differences between rodents and primates and the implications for interspecies extrapolation were not well-characterized. Human studies demonstrating hemolytic effects following human exposure to naphthalene, as well as animal and human studies showing cataract formation, were not adequately summarized in the IRIS assessment [summary sheets] and this is an omission. These studies are well reviewed in the Toxicological Review.”

4.0 PRE-PILOT IRIS ASSESSMENTS CONTRASTED WITH PILOT/POST-PILOT IRIS ASSESSMENTS

In general, reviewers found that the pre-Pilot IRIS assessments provided limited information on uncertainty and variability, although this level of documentation appeared to be consistent with the practices at the time. Default values for inter- and intra- species extrapolation were consistently used without discussing either uncertainty, in general, or presenting more than the briefest accompanying rationale or scientific justification for the default values used. In general, pre-Pilot assessments that were reviewed did not use an uncertainty factor to account for data base insufficiencies, and did not discuss relevance of a critical effect identified in an animal study to human exposure or toxicity. Pre-Pilot assessments often did not utilize existing human data to interpret the relevance of toxic effects in animals (e.g., TDI, HCB) to humans, even when human data would seem to support the consideration of other toxic endpoints. Pre-Pilot assessments often presented either: (1) very brief summaries of supporting data/studies; (2) no supporting study information; or (3) in the case of pesticides, a two-line paragraph indicating the type of supporting study, the species, and the LOAEL/NOAEL. Another characteristic of pre-Pilot IRIS assessments that were reviewed is that, in general, no data are presented on pharmacokinetics, metabolism, mode of action, strain-species sensitivities. Route-to-route extrapolation, for both cancer and noncancer effects, was routinely conducted without any apparent scientific justification.

While most of the pre-Pilot assessments were prepared in a manner that was consistent with Agency guidelines and practice at the time, some pre-Pilot IRIS assessments were of higher quality than others. It is worth noting that reviewers often commented that the discussion of uncertainty was superior when the overall assessment was "well done." A summary of some of the characteristics of a well done assessment are presented in Table 4-1. Examples of particularly fine assessments, when judged according to contemporary practices at the time included DBCP and manganese, for both of which human toxicity data were available. The DBCP assessment was noticeably more comprehensive than other pre-Pilot assessments. It presented evidence of a high level of concordance between human and animal toxicity studies for testicular effects. Uncertainty and variability in the

Table 4-1. Characteristics of a "Well Done" IRIS Assessment

- Clear discussion of rationale for uncertainty factors
- Extensive detail and discussion of database, identification of critical effects, principal study, doses, etc.
- Quantitative discussion of sources of uncertainty
- Synthesis of relevant information on dose, response, and modeling to assist with assessment of uncertainty
- Discussion of mechanisms of action, metabolism, PBPK modeling
- Discussion of susceptible subpopulations

principal study were well characterized and the use of a default value of 10 for extrapolation from subchronic to chronic duration was justified in the documentation. The 1995 manganese assessment was also much more comprehensive and thorough than most pre-Pilot assessments. Since manganese is an essential trace element, human data were the focus of the assessment. Furthermore, as with all essential trace elements, there was a concern for adverse health effects resulting from manganese deficiency as well as toxicity, and thus a comprehensive review and evaluation of the literature was performed.

Pilot/post-Pilot IRIS assessments typically presented more information but also varied in quality. Some of the chemicals considered in this study were originally classified as pre-Pilot assessments, then underwent minor modifications in later years and were reclassified as Pilot/post-Pilot; these assessments often retained many of the characteristics of pre-Pilot assessments, and contain little information on, or discussion of, uncertainty and variability. Other Pilot/post-Pilot assessments were distinctly more comprehensive than pre-Pilot assessments and included more description and better discussion of data gaps and end points such as reproductive/developmental or neurological effects, as well as physicochemical information relevant to pharmacokinetics and toxicity and more complete synopses of conclusions for each supporting study. Nevertheless, default values and default approaches to characterize uncertainty and variability were generally used, often with little explanation or scientific justification.

Application of the RfC methodology (U.S. EPA, 1994) in the Pilot/post-Pilot assessments reduced the uncertainty associated with the extrapolation of findings in animal studies to humans since interspecies adjustments were based on relative lung parameters and dose intake. In the Pilot/post-Pilot assessments reviewed, the default uncertainty factor for interspecies extrapolation for inhaled chemicals was decreased from 10 to 3 (one-half log 10). However, the rationale for this reduction was only briefly discussed in the uncertainty section of the IRIS noncancer risk assessments reviewed.

One reason for the improvement in the Pilot/post-Pilot assessments has been the availability of the Toxicological Review documents that accompany the summary sheets on the IRIS website. Overall, the best Pilot/post-Pilot assessments contained a more comprehensive discussion of the mechanism of action, the relevance of the critical effect to humans, or the impact of pharmacokinetic and/or metabolic information on inter-species variability. These latter data were appropriately applied to modify or fine-tune the uncertainty factors used (e.g., EGBE, MMA). In general, Pilot/post-Pilot IRIS assessments were more detailed, provided more chemical-specific information, and rarely extrapolated from route-to-route in either noncancer or cancer assessments. The animal-to-human dosimetric adjustment for inhalation chemical exposures is used in the Pilot/post-Pilot assessments to reduce the inter-species uncertainty factor from 10 to 3. [Incidentally, this method has become the “default” method for noncancer inhalation exposures.] A few Pilot/post-Pilot assessments have started to incorporate newer approaches and studies, and discuss

pharmacokinetics, metabolism, mode-of-action, human relevance of the critical effect, and incorporate these data into the uncertainty analyses. One innovation is the application of the benchmark dose (BMD) modeling approach for non-cancer endpoints. However, Pilot/post-Pilot assessments were more likely to present both the BMD and a NOAEL/LOAEL to calculate the RfD (the two values were usually similar). Uncertainty factors appeared to be applied similarly for a BMD/BMC as compared with a NOAEL/LOAEL approach. [When setting an RfD or RfC, Agency practice is generally to use the lowest LOAEL or the highest NOAEL or NOEL identified in the database of toxicity studies. Thus, the RfD/RfC should be protective against toxic effects observed at higher doses in other studies, and only uncertainty factors for inter- and intra-species variability were generally needed.]

The subset of IRIS assessments which relate to pesticides constitutes a special grouping amongst those assessments that were reviewed. Many of the toxicology studies available for commercially marketed pesticides are proprietary or "confidential business information" and are not available in the peer-reviewed scientific literature (not that this diminishes the value of the information). Typically, these IRIS assessments tended to present very short summaries of the supporting and principal studies, which constitute the only information about the studies available to the readers. Default uncertainty factor values were generally identified without further discussion and data variability was generally not mentioned. However, this subgroup also has the following characteristics: (1) toxicology studies were usually performed according to Agency mandated test guidelines at the time; (2) the studies were subject to (often extensive) review by EPA Office of Pesticide Program scientists, internal review panels, and the OPP Scientific Advisory Panel, and (3) often, the data base would be extensive and comprehensive, because of registration requirements mandated by the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA).

Another special grouping within the IRIS summaries reviewed is the essential trace elements (ETE). The reviewers evaluated two IRIS assessments that fell into this grouping: chromium III and manganese. Both dietary deficiency and overexposure to ETE may produce adverse health effects; therefore, application of the usual default uncertainty factors without critical evaluation, discussion, and integration of the data is not feasible. Compounded conservatism tends to overestimate risk and be overly protective of public health; for ETEs, this approach is more likely to harm public health by lowering the reference dose to a level where it is below the minimum daily human requirement. Consequently, the ETE literature is in general more comprehensively reviewed, and uncertainties and data variability are addressed. Because reliable animal models for these ETEs were not usually available, human data were used for risk assessment thus eliminating uncertainties associated with interspecies extrapolation. Individual variability with regard to toxicity and essentiality requirements, data variability, pharmacokinetic information, putative mechanisms of toxicity, and the limitations in deriving an RfD for an ETE are uncertainties which

tend to be characterized and discussed. A comprehensive justification for the use of specific uncertainty factors is thus far more likely to appear in an ETE IRIS assessment than in a chemical IRIS assessment.

IRIS cancer assessments reviewed for this study, whether pre-Pilot or Pilot/post-Pilot, seem to have been performed in accordance with EPA's 1986 *Guidelines for Carcinogen Risk Assessment*. This often meant that a highly conservative, non-threshold model was used to generate an upper bound estimate of low-dose human risk (after applying, as necessary, a scaling factor to the animal data to adjust for animal-to-human differences). Uncertainties in carcinogenicity assessments were generally not discussed. This approach meant that inter-human variability was not adjusted for, because using an upper bound estimate of low-dose risk is very conservative and is assumed to be protective of sensitive subgroups. Similarly, a carcinogenicity study is, by definition a lifetime study; therefore, no adjustment is generally necessary for study duration. Still, Pilot/post-Pilot cancer assessments tend to contain a more comprehensive presentation of the weight-of-evidence for animal or human carcinogenicity, and some uncertainties -- notably those concerned with genotoxic versus nongenotoxic mode of action -- were described. Structural similarities to other known carcinogens or noncarcinogens were sometimes presented to reduce uncertainty about a proposed cancer category designation.

5.0 DISCUSSION AND CONCLUSIONS

There is no question that EPA's years of labor in providing biologically-based, consensus IRIS toxicity values to the scientific community has been of inestimable value, at the very least because the process has been instrumental in clarifying issues and suggesting research needs in the developing field of risk assessment. IRIS is indeed a useful tool for public health and risk assessment. EPA's intention has been to serve the risk assessment needs of public health and environmental scientists at large, not just those having profound familiarity with the primary toxicological literature for IRIS chemicals. It is clear that the content and quality of IRIS assessments has inevitably expanded over time, in step with the development of the risk assessment field in general, which is still evolving. The addition of supporting documentation online in the form of the *Toxicology Reviews*, and its availability to the general public via the Internet since 1997, has tended to increase user expectations for IRIS as well.

1. Documentation of Uncertainty in IRIS Assessments: The level of documentation varied considerably in the sixteen IRIS assessments reviewed. Also, the comments from the reviewers covered a wide spectrum of responses. There are several reasons why. First, the very meaning of the terms "uncertainty" and "variability" in the detection and evaluation of a chemical hazard is not at all easy to define. As Bailar and Bailer (1999) have pointed out, the meaning of these concepts, and more importantly, the way they are quantified, varies tremendously depending on the background of the risk assessor. Bailar and Bailer defined uncertainty as "a lack of precise knowledge about the state of nature," and described the differing views scientists hailing from different disciplines may have of these terms and how to deal with them. The validity of their thesis was confirmed in this study, where experts in biostatistics, biology/toxicology, and risk assessment often manifested very different reactions to the same dataset. Those reviewers with a more mathematical bent towards data analysis tended to be dissatisfied with IRIS' uncertainty analyses.

A second reason is that the Agency does not appear to have a solid set of established consensus guidelines on either what kinds of uncertainty should be addressed or how uncertainty should be characterized. Procedures for departing from, reducing default values, and using data-derived uncertainty factors are still not available. Also, current procedures for uncertainty analysis are not amenable to incorporating mode of action or other data-derived information for reducing uncertainty and dealing with data variability. They were not designed for these inputs. [This is true even though the concept of uncertainty is discussed, in general, in several recent EPA documents (e.g., U.S. EPA, 1997a,b; U.S. EPA 1992).] Therefore, it was inevitable that as the reviewers evaluated each pre-Pilot and Pilot/post-Pilot IRIS assessment, the absence of an established standard to apply (before or since 1995), would result in

lack of harmony in their findings. Lastly, it is the very nature of uncertainty, that it is always there, and a rigorously "accurate" value for it will be somewhat elusive.

The practical effect of the above is that EPA has, at times, responded to these challenges by resorting to policy decisions, rather than by handling uncertainty in a quantitative manner. These policy decisions have been consistent with EPA's mission to protect public health and the environment and to set conservative limits on human exposures to carcinogens and other toxic agents in the face of poorly characterized variability and uncertainties. However, there have been occasions when EPA's decisions on how to handle data uncertainties and variability have been advertised as science-based when, in reality, it would be more appropriate to describe them as grounded in policy decisions.

2. *Uneven Extent of Characterization of Uncertainty*: In general, the reviewers found that uncertainty and variability in the data have been unevenly documented in IRIS assessments. The level of scientific analysis and characterization of variability and uncertainty varied between chemicals as well as within the IRIS assessment (e.g. RfD, RfC, slope factor) for the same chemical. Sometimes the existence of uncertainty or variability is acknowledged but not further characterized. The cause seems to be variations in the depth and breadth of scientific information, as well as lack of a standardized approach. Although EPA decision makers appear to recognize the usefulness of more complete uncertainty analyses, they have made only limited progress in replacing *ad hoc* procedures based on a few simple but sweeping assumptions with procedures based on the range of risk values consistent with data-derived information about biologic mechanisms of carcinogenic or other toxic effects, chemical disposition in the body, actual human exposures, and other factors influencing the range of biologically relevant risk values.

3. *Use of New Techniques Still Not Widespread*: The reviewers recognized that although the clear trend in risk science is a movement towards use of pharmacokinetic and biological data to fine-tune uncertainty and variability analyses, "data-derived" techniques are still under development. There is also a movement towards use of innovative mathematical approaches in risk and uncertainty assessment, but expertise in their use is not widespread and consensus on how and when to employ them is needed. Clearly, it is important that an understanding of the potential effects (pro and con) of using sophisticated mathematical techniques is needed. One of the more mathematically-inclined reviewers in this study strongly supported the replacement of all of IRIS individual point estimates of uncertainty with generic, composite uncertainty distributions, and referenced several published techniques currently in the literature. The reviewer even demonstrated use of such an approach with respect to chromium III, citing Swartout et al. (1998). Such ideas have been considered by the EPA's FIFRA Scientific Advisory Panel relatively recently (U.S. EPA, 1999). However, other reviewers expressed the opinion that application of such techniques to derivation of IRIS toxicity values is still experimental, and needs to be careful to respect an

essential harmony between the biological basis of toxicity and the mathematics. One reviewer felt that application of these techniques may not be transparent or useful to the intended IRIS audience, the general health professional.

4. *Improved Discussion of Cancer Assessment Uncertainties Needed*: With regard to cancer risk assessment, more detail is given to characterizing the data and developing a weight-of-evidence argument for cancer classification in Pilot/post-Pilot assessments. However, the reviewers noted that discussion of such topics as the human relevance of the tumorigenic effect, the pharmacokinetics and dynamics of the compound in biological systems and species differences, and the mode of action is limited. Modeling uncertainties are generally not addressed, and the same linearized model is used to extrapolate from high animal doses to the low doses typical of human exposure. It is anticipated that once the 1999 proposed *Guidelines for Carcinogen Risk Assessment* are finalized, a comprehensive narrative documenting and characterizing uncertainties will be available for IRIS summaries. However, narratives were lacking in the IRIS assessments reviewed.

5. *Default Usage of No-threshold Models for Derivation of Cancer Endpoints*: All of the cancer assessments reviewed in this study employed a no-threshold model to derive cancer toxicity endpoints. Several of the reviewers objected vehemently to use of such a default no-threshold approach and offered other modeling options for quantifying cancer toxicity endpoints. In particular, reviewers objected to application of non-threshold models to the derivation of cancer risk values for apparent non-genotoxic or promoting chemicals. It is to be hoped that increasing use of the EPA's 1996 proposed *Guidelines for Carcinogen Risk Assessment* should result in application of other models and methods and update of older IRIS assessments.

6. *Factoring in Human Data*: Still generally lacking is discussion explaining why: (i) humans are considered to be more sensitive than rodents, when (in some cases) data may exist to indicate that, for the selected critical effect, this may not be true; (ii) adjustment for less-than-lifetime to lifetime duration is needed when (for some chemicals) pharmacokinetic and physicochemical data may exist to indicate that bioaccumulation and tissue retention are unlikely to occur; and (iii) a particular animal health effect is being used to estimate human risk when human data demonstrate that the critical human health concerns are entirely different. Although discussion might not alter the choice of uncertainty factors, it would present a reasoned and transparent justification for the selection. Recently an analysis of many of these issues has become available (Dourson et al., 2000). These researchers affirm that "use of human data is a public health protective policy that should be encouraged."

7. *Updating Older IRIS Assessments*: All of the reviewers pointed out the continuing, well-recognized, need to periodically revisit the older IRIS assessments. The reviewers acknowledged, however, that some systematic policy would be needed to prioritize such an effort to provide for the best use of resources. EPA may wish to develop criteria for such prioritization of effort.

In some cases, the discussion of uncertainty in IRIS Toxicology Reviews is far more comprehensive than that appearing in the IRIS summary. It has been argued that readers interested in uncertainty have the opportunity to examine the support documents, which EPA has made available along with the IRIS summary on the IRIS website for more recent assessments. Several reviewers felt that because the IRIS summaries are geared toward the general health scientist, not the specialized toxicologist, they should be complete and "stand-alone" resources. A thorough, but succinct, discussion of uncertainty in the context of the IRIS summary, which integrates a range of scientific data with the range of associated uncertainties not only provides transparency to the reader but also increases confidence in the assessment.

6.0 REFERENCES

Bailar JC 3rd, Bailer AJ. (1999). Risk assessment--the mother of all uncertainties. Disciplinary perspectives on uncertainty in risk assessment. Ann N Y Acad Sci 1999;895:273-85.

Barnes, DG, Dourson, ML. (1988). Reference dose (RfD): Description and use in health risk assessment. Reg. Toxicol. Pharmacol. 8:471-486.

Daniel, Wayne W. (1978). *Biostatistics: A Foundation for Analysis in the Health Sciences.* John Wiley & Sons: New York..

Dourson, ML, SP Felter and D Robinson. (1996). Evolution of science-based uncertainty factors in noncancer risk assessment. Reg. Tox. Pharmacol., 24: 108-120.

Dourson, ML, Anderson, ME, Erdreich, L, MacGregor J. (2000) Using human data to protect public health. Submitted to *Toxicological Sciences* for publication. http://tera.org/news.

National Research Council (1994). *Science and Judgment in Risk Assessment*. National Academy Press: Washington, D.C.

Swartout JC; Price PS; Dourson ML; Carlson-Lynch HL; Keenan RE. (1998) A probabilistic framework for the reference dose (probabilistic RfD). *Risk Anal* 1998 Jun;18(3):271-82

U.S. Environmental Protection Agency (1986) Guidelines for carcinogen risk assessment. Federal Register 51(185):33992-34003.

U.S. Environmental Protection Agency (1992). Guidelines for exposure assessment, Notice, 1992. Federal Register 57(104):22888-22938.

U.S. Environmental Protection Agency (1996) Proposed guidelines for carcinogen risk assessment, Notice, 1996. Federal Register 61(79):17960-18011.

U.S. Environmental Protection Agency (1997a). *Policy for Use of Probabilistic Analysis in Risk Assessment*. National Center for Environmental Assessment: Washington, DC.

U.S. Environmental Protection Agency (1997b). Exposure Factors Handbook. Office of Research and Development: Washington, D.C.

U.S. Environmental Protection Agency (1999). Report of the FIFRA Scientific Advisory Panel Meeting, February 23-24, 1999, Arlington, Virginia. SAP Report No. 99-02, March 25, 1999.

U.S. Environmental Protection Agency (1999). Proposed guidelines for carcinogen risk assessment. Washington, D.C.