

EPA Summary Report

**Characterization of Data Variability and Uncertainty:
Health Effects Assessments in the Integrated Risk Information
System (IRIS)**

In response to Congress, HR 106-379

National Center for Environmental Assessment
Office of Research and Development
US Environmental Protection Agency
Washington, DC

DISCLAIMER

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

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
1. INTRODUCTION AND PURPOSE	2
2. BACKGROUND	3
2.1. Hazard and Dose-Response Assessment	4
2.2. IRIS Program and Data Base	6
2.3. Uncertainty and Variability	8
3. EVALUATION APPROACH	10
3.1. Protocol Development	10
3.2. Screening Evaluation	10
3.3. In-depth evaluation	11
4. SUMMARY OF RESULTS	13
4.1. Screening Evaluation	13
4.2. In-Depth Evaluation	13
5. DISCUSSION	16
6. CONCLUSIONS	18
7. REFERENCES	20

ATTACHMENTS

- EPA Screening Evaluation Report: Presentation and Discussion of Uncertainty and Variability in IRIS Assessments
- Versar Report: Characterization of Data Uncertainty and Variability in IRIS Assessments, Pre-Pilot vs. Pilot/post-Pilot
- Appendix A of Versar Report: Individual reports of experts assembled by Versar, Inc.

EXECUTIVE SUMMARY

In response to a Congressional directive contained in HR 106-379 regarding EPA's appropriations for FY2000, EPA has undertaken an evaluation of the characterization of data variability and uncertainty in its Integrated Risk Information System (IRIS) health effects information database. Through consultation with EPA's Science Advisory Board, EPA developed and implemented a systematic plan to select a representative sample of chemical assessments in IRIS to be evaluated in-depth by an independent panel of experts for the extent to which EPA has documented uncertainty and variability. EPA conducted a screening evaluation on 10% percent of the IRIS summaries of chemical assessments completed during the period of 1988-1994 (52 of 522 pre-Pilot assessments) and all 15 Pilot/post-Pilot IRIS summaries and Toxicological Reviews (completed after 1995) for overall documentation of data variability and uncertainty. An EPA contractor then selected 16 assessments (IRIS summaries and support documents) for in-depth examination from the screening sample (8 of 52 pre-Pilot and 8 of 15 Pilot/post-Pilot). The contractor selected six independent experts (outside EPA) in the field of human health risk assessment, who performed this in-depth review.

In general, the outside experts concluded that the characterization of data variability and uncertainty varied across the assessments they reviewed. While the documentation of data variability and uncertainty has generally improved since the IRIS Pilot's introduction of Toxicological Reviews to substantiate IRIS summaries, the reviewers found that the quality of the characterization of data variability and uncertainty varied among the Pilot/post-Pilot assessments. The reviewers also suggested ways to describe uncertainty and variability, and a number of scientific improvements, especially the need to update older assessments with more recent scientific data and risk assessment methods.

This study supports EPA's goal to make the scientific bases for risk assessment conclusions more transparent. EPA will continue to look into ways to improve the characterization and documentation of data variability and uncertainty in future IRIS assessments.

Note: This report reflects the review and comments of the Environmental Health Committee of EPA's Science Advisory Board, as discussed publically August 30, 2000, and documented in their final report to the EPA Administrator, dated September 26, 2000.

(See <http://www.epa.gov/science1/drrep.htm>).

1. INTRODUCTION AND PURPOSE

The Integrated Risk Information System (IRIS) data base contains EPA's consensus scientific positions on potential adverse human health effects that may result from chronic exposure to specific chemical substances in the environment. As of January 31, 2000, the IRIS data base contained 537 chemical-specific assessments. IRIS is widely used by regulator programs and risk assessors at all levels of government and by the public. First publically available in 1988, these assessments provide the summary results of EPA deliberations culminating in consensus hazard and dose-response conclusions for cancer and noncancer health effects. Since 1995 (when the "IRIS Pilot" program was undertaken), EPA has taken several steps to ensure that the best available scientific information is included in chemical assessments made available on IRIS, including improvements in documentation of scientific decisions, and external peer reviews of all subsequent assessments.

Regarding IRIS, Congress issued the following directive, which was contained in the October 1999 report from Congress (HR 106-379) regarding EPA's appropriations for FY2000:

"The conferees are concerned about the accuracy of information contained in the Integrated Risk Information system [IRIS] data base which contains health effects information on more than 500 chemicals. The conferees direct the Agency to consult with the Science Advisory Board (SAB) on the design of a study that will a) examine a representative sample of IRIS health assessments completed before the IRIS Pilot Project, as well as a representative sample of assessments completed under the project and b) assess the extent to which these assessments document the range of uncertainty and variability of the data. The results of that study will be reviewed by the SAB and a copy of the study and the SAB's report on the study sent to the Congress within one year of enactment of this Act."

In response to the Congressional directive, EPA has undertaken an evaluation of the characterization of data variability and uncertainty in IRIS assessments. This report addresses Congress's directive. Section 2 of the report provides background information about EPA's approaches to health hazard and dose-response assessments, and describes the IRIS program and the kinds of health information available in IRIS. It also discusses the sources of scientific uncertainties and variability related to the risk assessment process, and defines these terms in the context of the purpose of this EPA study, i.e., characterization of data variability and uncertainty of chemical assessments in IRIS. Section 3 describes the study protocol, and the summary findings of the study are provided in section 4. Details of the study protocol and results can be found in the three attachments. Discussion of study results, study conclusions, and references are provided in sections 5, 6 and 7, respectively.

2. BACKGROUND

Risk assessment is the process EPA uses to identify and characterize environmentally-related human health problems. As defined by the National Academy of Sciences (NAS, 1983), risk assessment entails the evaluation of all pertinent scientific information to describe the likelihood, nature, and extent of harm to human health as a result of exposure to environmental contaminants. EPA has used the basic NAS paradigm as a foundation for its published risk assessment guidance, and as an organizing system for many individual environmental chemical assessments. There are four components to every complete risk assessment - hazard assessment, dose-response assessment, exposure assessment, and risk characterization. *Hazard assessment* describes qualitatively the likelihood that an environmental agent can produce adverse health effects under certain environmental exposure conditions. *Dose-response assessment* quantitatively estimates the relationship between the magnitude of exposure and the degree and/or probability of occurrence of a particular health effect. *Exposure assessment* determines the extent of human exposure. *Risk characterization* integrates the findings of the first three components to describe the nature and magnitude of health risk associated with environmental exposure to a chemical substance or a mixture of substances.

There are many uncertainties associated with environmental risk assessments due to the complexity of the exposure-dose-effect relationship, and the lack of, or incomplete, knowledge and information about the physical, chemical, and biological processes within and between human exposure to an environmental substance(s) and health effects. Major sources of uncertainty include the use a wide range of data from many different disciplines (e.g., epidemiology, toxicology, biology, chemistry, statistics), the use of many different predictive models and methods in lieu of actual measured data, the use of many scientific assumptions and science policy choices, i.e., scientific positions assumed in lieu of scientific data, in order to bridge the information and knowledge gaps in the environmental risk assessment process. These diverse elements, along with varying interpretations of the scientific information, can result in divergent results in the risk assessment process, an outcome that leads to risk assessment controversies. Thus, EPA risk assessment guidelines stress the importance of identifying uncertainties and variability and presenting them as part of risk characterization.

Over the years, EPA has conducted health hazard and dose-response assessments for many environmental chemical contaminants. The summary findings and outcomes of these assessments which represent scientific consensus positions across the Agency are made available in the IRIS data base. Information on IRIS can be used with an exposure assessment for a specific exposure scenario to perform a complete risk assessment. The following sections

provide an overview of EPA's historical and current approaches to health hazard and dose-response assessments, describe EPA's IRIS program and the kinds of information available in IRIS, and define variability and uncertainty in the context of hazard and dose-response assessments and available information in IRIS.

2.1. Hazard and Dose-Response Assessment

In general, chemicals often affect more than one organ or system of the body (e.g., liver, kidney, nervous system) and can produce a variety of health endpoints (e.g., cancer, respiratory allergies, infertility), depending on the conditions of exposure such as the amount, frequency, duration, and route of exposure (i.e., ingestion, inhalation, dermal contact). For most environmental chemicals, available health effects information is generally limited to high exposures in studies of humans (e.g., occupational studies of workers) or laboratory animals. Thus, evaluation of potential health effects associated with low levels of exposure generally encountered in the environment involves inferences based on the understanding of the mechanisms of chemical-induced toxicities. *Mechanism of action* is defined as the complete sequence of biological events that must occur to produce an adverse effect. In cases where only partial information is available, the term *mode of action* is used to describe only major (but not all) biological events which are judged to be sufficient to inform about the shape of the dose-response curve beyond the range of observation.

For effects that involve the alteration of genetic material (e.g., most cancers, heritable mutations), there are theoretical reasons to believe that such a mode of action would not show a threshold, or dose below which there are no effects. On the other hand, a threshold is widely accepted for most other health effects, based on considerations of compensatory homeostasis and adaptive mechanisms. The threshold concept presumes that a range of exposures from zero to some finite value can be tolerated by an individual without adverse effects. Accordingly, different approaches have traditionally been used to evaluate the potential carcinogenic effects and health effects other than cancer, referred to as "noncancer" effects.

Carcinogenic Effects Cancer hazard assessment involves a qualitative weight-of-evidence evaluation of potential human carcinogenicity. This evaluation is a synthesis of all pertinent information in addressing the question of "*How likely an agent is to be a human carcinogen.*" The EPA's 1986 Guidelines for Carcinogen Risk Assessment (USEPA, 1986) provide a classification system for the characterization of the overall weight-of-evidence for potential human carcinogenicity based on human evidence, animal evidence, and other supportive data. The EPA's 1996 Proposed Guidelines for Carcinogen Risk Assessment (USEPA, 1996a) and the subsequent revised external review draft (USEPA, 1999), emphasize the need for characterizing cancer hazard in addition to hazard identification. Accordingly, the

question to be addressed in hazard characterization is expanded to “*How likely an agent is to be a human carcinogen, and under what exposure conditions a cancer hazard may be expressed.*” In addition, the revised guidelines stress the importance of considering the mode(s) of action information of the agent for making an inference about potential cancer hazard beyond the range of observation. To express the weight-of-evidence for potential human carcinogenicity, the EPA’s proposed revised guidelines emphasize using a hazard narrative in place of the classification system. However, in order to provide some measure of consistency, standard *hazard descriptors* are used as part of the hazard narrative to express the conclusion regarding the weight-of-evidence for potential human carcinogenicity.

Dose-response assessment for carcinogenic effects usually involves the use of a linear extrapolation model(s) to estimate an upper bound on cancer risks at a given low level of exposure. The linear low dose extrapolation approach is considered appropriate for cases where there is insufficient understanding of the mode of action, or when available data indicate a linear dose-response curve at low doses, but there are not enough data that would allow the development of biologically-based dose-response models. This risk estimate is known as cancer unit risk for inhalation exposure and slope factor for oral exposure. It is recognized that such an estimate may not give a realistic prediction of risk and the true value of risk may be as low as zero. However, the use of such models puts a ceiling on what the risk might be. When there is sufficient evidence for a nonlinear mode of action, but not enough data to construct a biologically-based model for the relationship, EPA’s proposed revised cancer guidelines (USEPA, 1996a) call for the use of a margin of exposure analysis as a default procedure. A margin of exposure analysis compares the point of departure (i.e., the lower 95% confidence limit of the dose or exposure associated with 10% risk of cancer or precursor effects) with the dose associated with the environmental exposure(s) of interest, and determines whether or not the exposure margins are adequate. Both default approaches may be used for a specific cancer assessment, if it is mediated by multiple modes of action which may include linear and nonlinear modes of action.

Noncancer Effects The Agency has published several guidelines for assessing specific noncancer health endpoints including developmental toxicity, reproductive toxicity, and neurotoxicity, (USEPA, 1991, 1996b, 1998, respectively). Like the cancer guidelines, these guidelines set forth principles and procedures to guide EPA scientists in the interpretation of epidemiologic, toxicologic and mechanistic studies to make inferences about the potential hazard of these specific health endpoints. Following a review and evaluation of the spectrum of potential health effects associated with the chemical of interest (i.e., hazard identification), a dose-response assessment is then performed on the “critical effect” (i.e., the adverse effects or its known precursor which occurs at the lowest dose) to derive a chronic reference dose (RfD) or

reference concentration (RfC) for oral and inhalation exposure, respectively. The RfD/RfC is defined as “*an estimate (with uncertainty spanning perhaps an order of magnitude) of a continuous oral/inhalation exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious noncancer effects during a lifetime*” (Barnes and Dourson, 1988; USEPA, 1994a). The RfD/RfC approach assumes that if exposure can be limited so that a critical effect does not occur, then no other noncancer effects will occur. Thus, this approach fulfills the regulatory needs for various EPA’s regulatory programs for defining an exposure level(s) below which there is negligible risk of adverse noncancer health effects.

2.2. IRIS Program and Data Base

The IRIS data base was created in 1986 as a mechanism for developing consistent intra-Agency consensus positions on potential health effects of chemical substances. EPA Program Offices and Regions were regulating some of the same substances, and determined that in many cases the Agency needed to use consistent scientific judgments on potential health effects in risk-based decisions. Chemical assessments prepared by Program and Regional Offices were peer reviewed by three intra-agency workgroups (i.e., RfD, RfC, and Carcinogen Risk Assessment Verification Endeavor, or CRAVE, workgroups) comprising health scientists across the Agency. Summary results of these consensus assessments were collected and made available on IRIS. Combined with site-specific or national exposure information, the summary health information in IRIS could then be used by risk assessors and other staff to evaluate potential public health risks from environmental contaminants. Summary information in IRIS consists of three components: derivation of oral chronic RfD and inhalation chronic RfC, for noncancer critical effects, cancer classification (and cancer hazard narrative for the more recent assessments) and quantitative cancer risk estimates.

IRIS summaries were originally written for an internal EPA audience. For this reason, IRIS information has focused on the documentation of toxicity values (i.e., RfD, RfC, cancer unit risk and slope factor) and cancer classification. The bases for these numerical values and evaluative outcomes are provided in an abbreviated and succinct manner. Details for the scientific rationale can be found in supporting documents, and references for these assessment documents, and key studies are provided in the bibliography sections. Moreover, it was not considered necessary to articulate every default assumption used in individual chemical assessments as these assumptions have been explicitly discussed and supported in the Agency’s published risk assessment guidance. It is also important to note that the three components of IRIS information (RfD, RfC, and cancer evaluation) were added to the database at different times, depending on the regulatory needs, without an explanation of why other endpoints were

not assessed.

As external interest in the information on IRIS grew, EPA made the IRIS data base publically available in 1988 via the National Library of Medicine's TOXNET system. In 1995, EPA undertook the IRIS Pilot Program to evaluate and implement a number of improvements in the documentation of summary information in IRIS and in the scientific peer review process. Individual chemical hazard and dose-response assessments for cancer and noncancer health effects are now provided in a single supporting document known as the IRIS "Toxicological Review" (or an equivalent support document). This procedure was subsequently adopted in response to the need for a more integrated health assessment as harmonized dose-response approaches become available for cancer and noncancer effects. In addition, there has been an increased demand for more transparency in the default assumptions and methods used in these chemical assessments, in response to the Agency policy on risk characterization (USEPA, 1995), as well as for developing and documenting the scientific bases for moving away from default methods (e.g., use of chemical-specific data to replace default values of uncertainty factors). In order to make the scientific quality of the assessments more uniform, an external peer review process was included in the Pilot program into the preparation of each chemical assessment, in response to EPA's Peer Review Policy (USEPA, 1994b). Since 1997, IRIS summaries and accompanying support documents, including a summary and response to external peer review comments, have been publically available in full text on the IRIS web site at <http://www.epa.gov/iris>. The Internet site is now EPA's primary repository for IRIS. Together they comprise the "IRIS assessment" for a given chemical substance.

The information currently on IRIS represents the state-of-the-science and state-of-the-practice in risk assessment as it existed when each assessment was prepared, often 10 or more years ago. When EPA reassesses older IRIS entries, an opportunity exists to update the science and apply more current methodologies. EPA uses an annual priority-driven approach to determine which chemical substances are most in need of assessment or reassessment.

The Office of Research and Development, National Center for Environmental Assessment (NCEA) coordinates the Agency-wide IRIS priority-setting process as part of its broader role of managing the IRIS program. The criteria that drive EPA's priorities are usually Program Offices' and Regions' statutory, regulatory, and programmatic needs. Availability of new scientific information to perform reassessments is also a strong criterion. The determination of the annual IRIS agenda is further modified by the availability of EPA scientific staff with appropriate expertise and other resources in various IRIS-sponsoring Offices to develop and manage individual assessments. NCEA's IRIS Staff therefore works with other parts of the Agency to refine the compilation of priority needs with consideration of available resources to accomplish the work. The resulting annual IRIS agenda, published in the Federal Register each

winter, therefore reflects both the Agency's priority chemicals for assessment or reassessment and internal commitments to lead the work.

Much work will be needed over the coming years in order to update even the highest priority substances. In an effort to improve the pace of the assessment process and leverage resources, EPA is currently evaluating ways to work cooperatively with external parties on assessment development. Five cooperative efforts are currently in progress, three with private organizations and two with other federal agencies. Others are under consideration. Under a cooperative arrangement, an external party may submit an assessment for EPA's consideration in developing an EPA IRIS document; however, EPA's consensus position must be documented separately. EPA is continuing to look for opportunities to improve the IRIS process and the pace of data base update.

2.3. Uncertainty and Variability

Because the Congressional language was to address "*uncertainty and variability of the data*," this report uses an expansive definition of the term "variability." As used in this report, "variability" encompasses any aspect of the risk assessment process that can have varying results, including the potential interpretations of the available data, the availability of different data sets collected under different experimental protocols, and the availability of different models and methods. Several of these would be considered as sources of uncertainty under the definitions of variability and uncertainty used by the NRC (1994) and EPA (1992, 1997). These stricter definitions use "variability" to refer to differences attributable to diversity in biological sensitivity or exposure parameters; these differences can be better understood, but not reduced by further research. "Uncertainty" refers to lack of knowledge about specific factors, parameters, or models, and generally can be reduced through further study. This section summarizes key uncertainties and data variability generally encountered in hazard and dose-response evaluations for cancer and noncancer effects.

Hazard Assessment For most chemical substances for which there are insufficient data in humans, a major uncertainty in the evaluation of potential health effects to humans is the reliance on animal studies of high exposure to predict human response at lower exposure, particularly in the absence of an understanding of how an agent causes the observed toxicologic effects in the animals, and in the face of the varying results frequently obtained with different animal species under different exposure conditions. Even when there are human data, there is uncertainty about average response at lower exposures and there is variability in individual response around this average. Therefore, EPA has adopted a number of scientific assumptions as science policy choices in the face of data and knowledge gaps.

Major assumptions used in hazard assessment (unless there are data to the contrary)

include the following: (a) effects observed in one human population are predictive of other human populations; (b) in the absence of human data, effects seen in laboratory animals are assumed to be relevant to humans, and humans may respond similarly (although not identically) to the most sensitive animal species; and (c) effects seen at high exposure are relevant for evaluation of potential effects at low exposure. These scientific assumptions or science policies have also been articulated further in EPA's peer-reviewed risk assessment guidance documents, as discussed above.

Reference Values for Noncancer Effects To derive a RfD/RfC for a noncancer critical effect, the common practice is to apply standard "uncertainty factors" (UFs) to the no-observed adverse effect level (NOAEL), lowest-observed adverse effect level (LOAEL) or benchmark dose/concentration (BMCL_x)¹ (US EPA, 1995c). These UFs are used to account for the extrapolation uncertainties (e.g., inter-individual variation, interspecies differences, duration of exposure) and database adequacy. A modifying factor (MF) is also used as a judgment factor to account for the confidence in the critical study (or studies) used in the derivation of the RfD/RfC. Replacements for default UFs are used when chemical-specific data are available to modify these standard values. This is known as the "data-derived" approach. Moreover, the use of pharmacokinetic or dosimetry models can obviate the need for an UF to account for differences in toxicokinetics across species.

A number of related factors can lead to significant uncertainty of the RfD/RfC. Among these is the selection of different observed effects as a critical effect, which may vary within and across available studies. Also significant is the choice of different data sets for the identification of the NOAEL, LOAEL, or bench mark dose analysis, the use of different values for the various UFs, and additional judgments which impact the MF.

Cancer Risk Estimates Cancer dose-response assessment generally involves many scientific judgments regarding the selection of different data sets (benign and malignant tumors or their precursor responses) for extrapolation, the choice of low dose extrapolation approach based on the interpretation and assessment of the mode of action for the selected tumorigenic response(s), the choice of extrapolation models, methods to account for differences in dose across species, and the selection of the point of departure for low dose extrapolation. Given that many judgments need to be made in the many steps of the assessment process in the face of data variability, along with the use of different science policy choices and default procedures and methods to bridge data and knowledge gaps, it is generally recognized that uncertainty exists in

¹BMCL_x is defined as the lower 95% confidence limit of the dose that will result in a level of "x"% response (e.g., BMCL₁₀ is the lower 95% confidence limit of a dose for a 10% increase in a particular response).

cancer risk estimates.

3. EVALUATION APPROACH

The following sections describe the overall approach for this evaluative study and the study protocols for the screening step and the in-depth evaluation of the documentation of data variability and uncertainty of available health information in IRIS. Details of the study protocols can be found in the attachments (EPA Screening Evaluation Report, and Versar In-Depth Report).

3.1. Protocol Development

Following the Congressional directive, EPA consulted with the Executive Committee of EPA's Science Advisory Board (SAB) about a proposed approach to this study. The agreed-upon approach involved assembling a team of independent, qualified individuals, external to EPA, to evaluate a representative set of IRIS assessments for the extent of documentation of variability and uncertainty. The use of external experts would avoid internal bias and the appearance that the IRIS program was "reviewing itself." The assessments would be reviewed simultaneously by multiple evaluators, in order to obtain a range of opinions from experts with a variety of relevant backgrounds. In order to address Congress's point concerning pre-Pilot and Pilot assessments, half of the sample would be from the set of pre-Pilot assessments (completed before 1995) and half from the later assessments.

The SAB supported EPA's overall approach, and recommended a number of enhancements. First, they recommended a tiered approach to selecting a representative sample of assessments, in which a sample of at least 10% of the available assessments would first be screened for their treatment of variability and uncertainty. This screening was to consider broad categories of documentation, and be verified by an independent reviewer. A smaller set of assessments would be chosen from the screening sample for in-depth review.

The SAB also encouraged examining as large a set of assessments in-depth as possible. They felt that three reviews per assessments would provide a sufficient range of opinions, given an adequate range of subject area expertise among the evaluators. This decision made it possible to target a sample of 16 assessments, to be reviewed by a total of six independent evaluators.

3.2. Screening Evaluation

An EPA scientist (IRIS Program Staff) carried out the screening evaluation, which is detailed in the attached EPA report. As recommended by the SAB, a 10% sample of pre-Pilot IRIS assessments (52 of 522) was identified. These, and the 15 Pilot/post-Pilot IRIS assessments

completed by January 31, 2000, a total of 67 assessments, were classified into three broad categories of overall documentation: none/minimal, some/moderate, or extensive (see Table 2, attached EPA Screening Report). The purpose of the preliminary screening was to survey broadly the extent of documentation of uncertainty and variability of health effects information in IRIS, in order to facilitate an in-depth evaluation of a smaller, but representative set of chemical assessments in IRIS. Due to the large volume of pre-Pilot assessment materials (52 sets of an IRIS summary plus supporting EPA Source Document(s)), only the IRIS Summaries were examined. For the later IRIS assessments, the IRIS summary and the Toxicological Review were examined. Consequently, this screening addressed only the overall approach to providing information concerning variability and uncertainty in the on-line assessments, not the completeness of the summarized information, nor the cited scientific literature available at the time of each assessment.

The first category, “None/Minimal,” describes assessments which presented conclusions, with overall uncertainty and confidence statements, but no incidence rates or other quantitative health effect levels for the available studies (such as, percent weight loss), nor any rationale for the confidence statements. Assessments with “Some or Moderate” documentation contained quantitative effect levels and some discussion of variability of effects, including variability across dose groups. In addition, these assessments contained some discussion of the reasons for overall confidence in the assessment. Assessments with “Extensive” documentation contained quantitative information (such as confidence intervals), some comparison of results across related studies, discussion of sources of uncertainty, comparison of uncertainties across available studies, and rationales for confidence in the available studies and conclusions drawn in the assessment. A listing of the categorized assessments was provided to the contractor to facilitate choosing the random sample for in-depth evaluation of the treatment of variability and uncertainty.

As recommended by the SAB Executive Committee, a second reviewer (an EPA health scientist without routine involvement in preparing or reviewing IRIS assessments) repeated the above evaluative step, without any knowledge of the results of the first round of review. The details of this second evaluation are also provided in the attached EPA Screening Report.

3.3. In-depth evaluation

The in-depth evaluation then focused on 16 IRIS assessments, half (8) from the pre-Pilot assessments and the other half from the Pilot/post-Pilot assessments. Within these two subsets, the assessments were randomly selected from the “some/moderate” and “extensive” documentation categories as evenly as possible. The assessments in the “none/minimal” category were not included in this part of the evaluation; it was not clear whether it would be a good use

of the experts' effort to review these assessments, as they likely contained limited characterization of uncertainty and variability, at least based on the summary information. EPA's contractor (Versar, Inc.) selected the sample of 16 assessments for in-depth evaluation. The materials for in-depth review of the pre-Pilot assessments included the IRIS summaries and the supporting EPA Source Document(s) identified in each summary. For the Pilot/post-Pilot assessments, the materials were the IRIS summary and Toxicological Review. The selection process and assessments chosen are provided in the attached Versar report.

EPA's contractor assembled and coordinated a set of six independent experts to carry out the review. These experts were selected on the basis of their in-depth knowledge of EPA's human health risk assessment methodologies, familiarity with IRIS, knowledge of current practices for evaluating and documenting uncertainty and variability in data used in health assessments, and expertise in how these factors relate to sensitive subpopulations including children. They represented a range of professional affiliations and of health science backgrounds among cancer and noncancer toxic endpoints. The experts evaluated the documentation of uncertainty and variability in assessments on the basis of the data available at the time each assessment was conducted, focusing on the presentation of available data and variability in that data, discussion of confidence and uncertainty, including any uncertainty factors applied. The evaluators self-certified that they had not been involved in the development or peer review of the assessments under review for the study, and that they could perform independently, free of conflict of interest. Each evaluator was assigned 8 assessments to review, generally evenly divided between pre-Pilot and Pilot/post-Pilot assessments. Each chemical assessment was independently reviewed by three evaluators. The evaluators and their assigned assessments are listed in Table 2-6 of the attached Versar report.

The evaluators were asked to answer the following questions:

- Considering the data available at the time each assessment was performed, and the EPA guidelines and methodologies operative at the time of the assessment, did EPA characterize to an appropriate extent the uncertainty and variability in data used to develop these IRIS health assessments? How does this compare between pre-Pilot and Pilot/post-Pilot assessments?
- Did EPA appropriately address the strengths and weaknesses of the scientific evidence from available studies, and sources of variability in the data used in each assessment?
- Did EPA appropriately address the uncertainties in the underlying data, and uncertainties in the qualitative and quantitative judgments given in each assessment?

The evaluators were also encouraged to raise other relevant observations or comments.

4. SUMMARY OF RESULTS

The summary findings of the screening and in-depth evaluations are provided below. Details of review results can be found in the attached EPA report (screening evaluation) and Versar report (overall summary of in-depth review and Appendix A, containing individual reviewers' findings).

4.1. Screening Evaluation

The results of the screening evaluation of the 52 pre-Pilot IRIS summaries by the first EPA reviewer were that: 3/52 had extensive, 16/52 some or moderate, and 33/52 none or minimal presentation or discussion of variability and uncertainty. Nearly all of the Pilot/post-Pilot assessments (14/15) showed extensive documentation of variability and uncertainty in the IRIS summary and Toxicological Review. It should be noted that a proper comparison between the two groups of assessments (pre-Pilot versus Pilot/post-Pilot) cannot be made as it requires an evaluation of a comparable set of assessment documentation (the source documents for the pre-Pilot assessments were not evaluated in the screening phase). The independent verification of the screening evaluation by a second EPA reviewer produced similar results (see attached EPA Screening Report, Table 5), with a Spearman rank correlation coefficient of 0.82. For 15 assessments, the ratings for the reviewers differed by one category.

Given the valuable input from the verification step, it is reasonable to consider the results of the two rankings together. Among the 52 pre-Pilot summaries, then, approximately two-thirds (63-79%) contained none to minimal documentation of variability and uncertainty information. Almost all (93-100%) of the assessments carried out after 1995 demonstrated extensive documentation of variability and uncertainty information.

4.2. In-Depth Evaluation

The report of the in-depth evaluation (attached Versar Report) summarizes the collective findings and conclusions of the six evaluators in responding to EPA's questions. The evaluators' individual reports are provided in Appendix A of the Versar report. The primary conclusions to each question are summarized below.

Considering the data available at the time each assessment was performed, and the EPA guidelines and methodologies operative at the time of the assessment, did EPA characterize to an appropriate extent the uncertainty and variability in data used to develop these IRIS health assessments? How does this compare between pre-Pilot and Pilot/post-Pilot assessments? As described above, six independent evaluators examined, in-depth, a sample of 16 IRIS assessments which had been found to have either a "some/moderate" or "extensive" degree of

documentation of variability and uncertainty in the screening evaluation. Each chemical assessment consisting of an IRIS summary and any supporting document(s) was reviewed by three independent evaluators. There was a range of opinions concerning the adequacy of documentation of data variability and uncertainty for the individual assessments among the reviewers. This range extended from two assessments (pre-Pilot assessments from 1988 and 1990) considered by all 3 reviewers to have been inadequately characterized, to one assessment (post-Pilot assessment from 1998) unanimously considered to demonstrate thoroughly adequate documentation. The evaluations for each of the other 13 assessments were not unanimous but were still informative (see Versar report, Table 3-2). These evaluations are discussed further below.

The evaluators generally concluded that the pre-Pilot IRIS summaries provided limited information on uncertainty and variability, although this was consistent with the practice at the time. Further, a number of evaluators felt that pre-Pilot assessments often did not utilize existing human data to interpret the relevance of toxic effects in animals to humans, even when the human data seemed to support the consideration of other toxic endpoints. Some noted that route-to-route extrapolation, for both cancer and noncancer effects, was routinely carried out without any apparent scientific justification. Despite these shortcomings, evaluators did point out that two (1,2-dibromo-3-chloropropane and manganese) of the eight pre-Pilot summaries were especially well characterized regarding uncertainty and variability (see Versar report, Section 4), when judged according to practices standard at the time.

The evaluators noted that the Pilot/post-Pilot IRIS summaries typically presented more information than the pre-Pilot summaries, but at the same time varied in quality. More specifically, they concluded that some Pilot/post-Pilot summaries contained little discussion of variability and uncertainty, while others were distinctly more comprehensive than pre-Pilot assessments. The more comprehensive assessments included more description and better discussion of data gaps and endpoints 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. 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 or metabolic information on interspecies variability. Two of these better assessments (ethylene glycol monobutyl ether and methyl methacrylate) were highlighted for using this additional information to adjust uncertainty factors away from the default values.

The evaluators appreciated the availability of the Toxicological Review documents that accompany the IRIS summaries on the IRIS website.

Did EPA appropriately address the strengths and weaknesses of the scientific evidence

from available studies, and sources of variability in the data used in the assessment? The evaluators concluded that the strengths and weaknesses of the scientific evidence from available studies were not thoroughly addressed in the earlier IRIS assessments, relative to the later assessments. It was found that only one of the eight pre-Pilot assessments appropriately addressed all of the substantive studies available at the time of the assessments. On the other hand, the evaluators considered six of the eight Pilot/post-Pilot assessments appropriately addressed the strengths and weaknesses of the substantive studies available at the time of the assessments (see Versar report, section 3).

Did EPA appropriately address the uncertainties in the underlying data, and uncertainties in the qualitative and quantitative judgments given in the assessment? In addition to verifying whether the standard uncertainty factors of the time were applied appropriately to develop the provided RfD/RfC, the evaluators determined whether additional issues contributing to variability and uncertainty had been considered, such as mechanism of action, variations in species susceptibility, potential for existence of sensitive subpopulations, relevance of the dosing regimen to likely human exposure pathways, and relevance of the critical effect to humans. The evaluators found that these latter issues tended not to be addressed in the pre-Pilot summaries, with the exception of two (1,2-dibromochloropropane and manganese).

The evaluators raised similar concerns about the Pilot/post-Pilot summaries with respect to these issues. Except for one assessment (methyl methacrylate) for which there was full agreement that uncertainties of the assessment had been adequately addressed, there was a range of opinions for the other seven Pilot/post-Pilot IRIS summaries. That is, there was usually at least one evaluator who was dissatisfied with these summaries, on the basis of the lack of coverage of these more advanced scientific issues.

Reviewers' Recommendations In addition to responding to the three questions above, there were some general themes in the evaluators' individual recommendations for improving IRIS assessments. First, the reviewers recommended development of a standardized approach to handling variability and uncertainty in IRIS assessments. It was also recommended that data quality issues should be clarified in IRIS assessments. Specifically, toxicological experiments carried out before the advent of Good Laboratory Practices (GLPs) should be earmarked as such, since there could be more uncertainty attached to data carried out before this standardization was implemented. Also, data from unpublished or non-peer-reviewed sources could carry similar uncertainties. The evaluators also emphasized that there did not appear to be enough consideration of the relevance of specific findings in animals to humans, both in choice of critical effects and exposure conditions. They also felt that the presumption that humans are more sensitive to environmental toxicants required more justification and discussion in most assessments.

In their individual reports (Appendix A of the Versar report), the evaluators made specific recommendations for improving those assessments they reviewed. These recommendations generally addressed inclusion of more recent scientific information (such as, mode of action or discussion of concordance of animal and human health endpoints) and pointed out instances where these data might support the use of more recently developed risk assessment methods (e.g., benchmark dose, quantitative uncertainty analysis).

5. DISCUSSION

The characterization of the extent of documentation of variability and uncertainty in chemical assessments in IRIS was accomplished using a tiered strategy, first by screening for the degree of this documentation in broad terms in a random sample, then in-depth in a smaller, targeted subsample. The representativeness of the in-depth evaluations for characterizing the rest of the database, first for the pre-Pilot IRIS assessments, then for the later IRIS assessments is discussed below.

The screening evaluation of 10% of the pre-Pilot IRIS data base provided a baseline for characterizing the IRIS database. Recall that about two-thirds (63-79%) of the sample of pre-Pilot IRIS summaries were found to have none to minimal documentation of variability and uncertainty (see section 4.1 above). Given the subjective nature of this evaluation, the additional review and consensus-building necessary to narrow this estimate did not appear warranted. Thus, it was concluded that approximately one-third (21-37%) of the pre-Pilot IRIS summaries demonstrated at least some documentation of the variability and uncertainties in deriving the toxicity values provided.

There was reasonable concordance for the pre-Pilot assessments between the screening evaluation and the in-depth review, given the different purposes of the two steps of the overall evaluation. In particular, two assessments (hexachlorobenzene and Prochloraz) were considered by the evaluators in their in-depth review to have inadequate documentation (see Versar report, section 3). These assessments were also judged to have minimal rather than moderate documentation in the independent verification stage of the screening evaluation (EPA Screening Report, Appendix B). At the other end of the scale, the two assessments highlighted as the most thoroughly documented of the pre-Pilot in-depth sample (1,2-dibromochloropropane and manganese) were also considered to be extensively documented in the screening evaluation.

One apparent outlier involved an assessment determined in the screening evaluation to have moderate documentation, yet was considered unanimously by the in-depth evaluators to have inadequate documentation of uncertainties (4-methylphenol; see Versar report, section 3.1). While the degree of discussion in the summary was more detailed than was otherwise typical at

the time (1990), the evaluators concluded that important aspects of uncertainty had been overlooked, e.g., incomplete use of data available at the time, and uncritical use of data from structural analogues that were not clearly relevant.

The correspondence of the screening evaluation and the in-depth evaluation for the Pilot/post-Pilot assessments was also complementary. It was found in the screening evaluation that the IRIS summary and Toxicological Review for the Pilot/post-Pilot assessments generally contained extensive documentation of variability and uncertainty. In the in-depth evaluation, the reviewers further examined the completeness of the discussions provided. While they concluded that the quality of the discussions varied, it was also not always clear whether these remarks were addressed to the IRIS summary alone, the Toxicological Review alone, or to both.

In conclusion, the statistical sampling approach taken in choosing the assessments to review allows some generalization of the results of the screening evaluation and the in-depth evaluation to the rest of the IRIS data base. That is, based on a 10% sample, approximately two-thirds of the pre-Pilot IRIS summaries can be expected to contain minimal discussion of the variability and uncertainty inherent in the available toxicity values. The remaining third of the pre-Pilot IRIS summaries can be expected to contain at least moderate documentation of variability and uncertainty. Among assessments with at least moderate documentation of variability and uncertainty, in their in-depth review, the evaluators found that coverage of relevant uncertainty and variability issues was uneven across the assessments they reviewed, with two of the eight assessments noticeably more comprehensive than the other pre-Pilot assessments. Among the Pilot/post-Pilot assessments, all but one demonstrated extensive documentation of variability and uncertainty, partly through the ready availability of the accompanying Toxicological Reviews. The evaluators' in-depth reviews of eight of these assessments noted a range in quality of the discussion of relevant uncertainties in these assessments as well. One Pilot/post-Pilot assessment was highlighted as being more comprehensive than all of the other assessments examined in-depth.

The independent evaluators also made several recommendations for improving IRIS assessments, including the need for updating assessments. EPA recognizes that many assessments in the IRIS database have not been updated and therefore either may not reflect the latest scientific findings or current risk assessment methods. With respect to current risk assessment methods, EPA has been applying the revised cancer guidelines in all assessments underway since they were proposed, as noted in Section 2 above, but acknowledges that some unevenness in documentation exists while the Agency gains experience in applying them. Concerning "data-derived" uncertainty factors, it should be noted that EPA-published risk assessment guidelines support the use of relevant data to replace these defaults. Limitations in developing data-derived factors are mostly due to the unavailability of useful data to justify

departure from defaults. EPA is developing guidance for risk assessors in the application of the “data-derived” approach to facilitate the maximum use of scientific data in replacing default UFs. Moreover, EPA acknowledges that discussion of many of these underlying uncertainties in IRIS assessments can be improved.

One of the more recent risk assessment methods encouraged by several reviewers was quantitative uncertainty analysis. The goal of a quantitative uncertainty analysis is to clarify the overall degree of variability and uncertainty and the confidence that can be placed in the analysis and its findings, through a systematic approach to account for relationships among the inputs or assumptions (in the case of risk assessment for IRIS, all of the data choices and uncertainty decisions discussed above) which contribute to a risk decision (in this case a toxicity value). Quantitative choices must be made for each input, even for qualitative decisions. A number of recent documents have emphasized the importance of adequately characterizing variability and uncertainty in risk assessments and discuss quantitative uncertainty analysis in more detail (US EPA, 1992, 1995, 1997a,b; National Academy of Sciences, 1994). EPA’s current *Policy for Use of Probabilistic Analysis in Risk Assessment* (1997b) provides that, “For human health risk assessments, the application of Monte Carlo and other probabilistic techniques has been limited to exposure assessments in the majority of cases. The current policy, Conditions for Acceptance and associated guiding principles are not intended to apply to dose response evaluations for human health risk assessment until this application of probabilistic analysis has been studied further.”

Since it is the function of IRIS to implement Agency-approved published methodologies and Agency-wide policies, implementation of newer risk assessment methods in IRIS waits for an Agency-level mandate. In the meantime, EPA agrees that a thorough description of the available data and its related uncertainties can provide the IRIS user with a level of confidence in a particular assessment, and can lay the groundwork for later uncertainty analysis, should it be considered practical.

6. CONCLUSIONS

The results of the screening evaluation indicated that about a third of IRIS summaries for pre-Pilot chemical assessments had at least some documentation of data variability and uncertainty, while a large majority of Pilot/post-Pilot assessments (consisting of both IRIS summaries and Toxicological Reviews) had extensive documentation. While the documentation in assessments has improved overall since the IRIS Pilot’s introduction of Toxicological Reviews to substantiate IRIS summaries, the results of the in-depth evaluation indicate that the quality of the characterization of data variability and uncertainty varies among the Pilot/post-Pilot

assessments.

This study supports EPA's commitment to provide more transparent scientific bases for risk assessment conclusions. EPA will continue to look into ways to improve documentation of variability and uncertainty issues in future Toxicological Reviews, and recapitulate this information in IRIS summaries.

7. REFERENCES

Barnes, DG, and Dourson, M. (1988) Reference dose (RfD) description and use in health risk assessments. *Regulatory Toxicology and Pharmacology*. 8: 4471-4486.

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

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

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

U.S. Environmental Protection Agency (1991) Guidelines for Developmental Toxicity Risk Assessment, dated December 5, 1991. *Fed. Reg.* 56, No.234: 63798-63826.

U.S. Environmental Protection Agency (1992) Guidelines for Exposure Assessment *Federal Register* 57: 22888-22938. EPA/600Z-92/001.

US Environmental Protection Agency (1994a) *Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry*. Office of Health and Environmental Assessment, National center for Environmental Assessment, RTP, NC. EPA/600/8-90-066F.

US Environmental Protection Agency (1994b) Peer Review and Peer Involvement at the US EPA. Memorandum of the Administrator, Carol M. Browner, June 7.

US Environmental Protection Agency (1995) EPA Guidance on Risk Characterization. Memorandum of the Administrator, Carol M. Browner, March 21.

U.S. EPA. (1995c) Use of the Benchmark Dose Approach in Health Risk Assessment. Office of Research and Development. EPA/630/R-94/007, February 1995.

U.S. Environmental Protection Agency (1996a. Proposed Guidelines for Carcinogen Risk Assessment, Notice, 1996. *Federal Register* 61(79):17960-18011.

U.S. Environmental Protection Agency (1996b) U.S. EPA, 1996b. Guidelines for Reproductive Toxicity Risk Assessment. October 31. *Federal Register* 61 (212): 56274-56322.

US Environmental Protection Agency (1997a) Guiding Principles for Monte Carlo Analysis. Office of Research and Development. EPA/630/R-97/001.

US Environmental Protection Agency (1997b) Policy for Use of Probabilistic Analysis in Risk Assessment. Memorandum of the Deputy Administrator, Fred Hansen, May 15.

U.S. Environmental Protection Agency (1998) Guidelines for Neurotoxicity Risk Assessment, dated October 4, 1995. *Federal Register* 63 (93): 26926-26954.

U.S. Environmental Protection Agency (1999) Proposed Guidelines for Carcinogen Risk Assessment. External Review Draft.