

PB90-183534

Workshop Report on EPA (Environmental Protection Agency)  
Guidelines for Carcinogen Risk Assessment: Use of Human Evidence  
Held in Washington, DC. on June 26-27, 1989

Eastern Research Group, Inc., Arlington, MA

Prepared for:

Environmental Protection Agency, Washington, DC

Sep 89

2023708871

U.S. Department of Commerce  
National Technical Information Service



BIBLIOGRAPHIC INFORMATION

PB90-183534

Report Nos: none

Title: Workshop Report on EPA (Environmental Protection Agency) Guidelines for Carcinogen Risk Assessment: Use of Human Evidence. Held in Washington, DC. on June 26-27, 1989.

Date: Sep 89

Performing Organization: Eastern Research Group, Inc., Arlington, MA.

Performing Organization Report Nos: EPA/625/3-90/017

Sponsoring Organization: \*Environmental Protection Agency, Washington, DC. Office of Health and Environmental Assessment.

Contract Nos: EPA-68-02-4404

Supplementary Notes: Also available from Supt. of Docs. Sponsored by Environmental Protection Agency, Washington, DC. Office of Health and Environmental Assessment.

NTIS Field/Group Codes: 57Y, 68G

Price: PC A05/MF A01

Availability: Available from the National Technical Information Service, Springfield, VA. 22161

Number of Pages: 98p

Keywords: \*Meetings, Epidemiology, Studies, Exposure, Methodology, Procedures, Humans, \*Carcinogenicity tests, \*Risk assessment, \*Health planning guidelines, Dose-response relationships.

Abstract: The U.S. Environmental Protection Agency (EPA) issued guidelines for assessing human risk from exposure to environmental carcinogens. The guidelines set forth principles and procedures to guide EPA scientists in the conduct of Agency risk assessments, to promote high scientific quality and Agency-wide consistency, and to inform Agency decision-makers and the public about these scientific procedures. In publishing the guidance, EPA emphasized that one purpose of the guidelines was to 'encourage research and analysis that will lead to new risk assessment methods and data,' which in turn would be used to revise and improve the guidelines. Thus, the guidelines were developed and published with the understanding that risk assessment is an evolving scientific undertaking and that continued study would lead to changes.

2023708872

PA 90-183534

**EPA**  
**Workshop Report on**  
**EPA Guidelines for**  
**Carcinogen Risk**  
**Assessment Use of**  
**Human Evidence**

**RISK ASSESSMENT FORUM**

REPRODUCED BY  
U.S. DEPARTMENT OF COMMERCE  
NATIONAL TECHNICAL INFORMATION SERVICE

2023708873

EPA/625/3-90/017

September 1989

## **Workshop Report on EPA Guidelines for Carcinogen Risk Assessment: Use of Human Evidence**

Assembled by:

Eastern Research Group, Inc.  
6 Whittemore Street  
Arlington, MA 02174  
EPA Contract No. 68-02-4404

for the

Risk Assessment Forum  
Technical Panel on Carcinogen Guidelines

U.S. Environmental Protection Agency  
Washington, DC 20460

2023708874

NOTICE

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This workshop was organized by Eastern Research Group, Inc., Arlington, Massachusetts, for the EPA Risk Assessment Forum. ERG also assembled and produced this workshop report. Sections from individual contributors were edited somewhat for clarity, but contributors were not asked to follow a single format. Relevant portions were reviewed by each workshop chairperson and speaker. Their time and contributions are gratefully acknowledged. The views presented are those of each contributor, not the U.S. Environmental Protection Agency.

## CONTENTS

|  | PAGE |
|--|------|
| INTRODUCTION . . . . .   | 1    |
| MEETING AGENDA . . . . .   | 5    |
| COLLECTED WORKSHOP MATERIALS . . . . .   | 7    |
| • Study Design and Interpretation . . . . .  | 9    |
| • Chair Summary . . . . .  | 15   |
| • EPA Classification System for Categorizing Weight of Evidence<br>for Carcinogenicity from Human Studies . . . . .      | 23   |
| • Chair Summary . . . . .  | 27   |
| • Dose-Response Assessment . . . . .   | 39   |
| • Chair Summary . . . . .  | 42   |
| APPENDICES   |      |
| Appendix A EPA Risk Assessment Forum Technical Panel on<br>Carcinogen Guidelines and Associates . . . . .                | 47   |
| Appendix B List of Participants . . . . .  | 51   |
| Appendix C List of Observers . . . . .   | 57   |
| Appendix D Introductory Plenary Session Comments<br>(Drs. Philip Enterline, Raymond R. Neutra, and Gerald Ott) . . . . . | 61   |
| Appendix E 1986 Guidelines for Carcinogen Risk Assessment . . . . .  | 81   |

**WORKSHOP REPORT ON EPA GUIDELINES FOR CARCINOGEN RISK ASSESSMENT:  
USE OF HUMAN EVIDENCE**

**June 26-27, 1989**

**Washington, DC**

**INTRODUCTION**

**1: Guidelines Development Program**

On September 24, 1986, the U.S. Environmental Protection Agency (EPA) issued guidelines for assessing human risk from exposure to environmental carcinogens (51 Federal Register 33992-34003). The guidelines set forth principles and procedures to guide EPA scientists in the conduct of Agency risk assessments, to promote high scientific quality and Agency-wide consistency, and to inform Agency decision-makers and the public about these scientific procedures. In publishing this guidance, EPA emphasized that one purpose of the guidelines was to "encourage research and analysis that will lead to new risk assessment methods and data," which in turn would be used to revise and improve the guidelines. Thus, the guidelines were developed and published with the understanding that risk assessment is an evolving scientific undertaking and that continued study would lead to changes.

As expected, new information and thinking in several areas of carcinogen risk assessment, as well as accumulated experience in using the guidelines, has led to an EPA review to assess the need for revisions in the guidelines. On August 26, 1988, EPA asked the public to provide information to assist this review (53 Federal Register 52656-52658). In addition, EPA conducted two workshops to collect further information. The first workshop for analysis and review of these issues was held in Virginia Beach, Virginia, on January 11-13, 1989 (53 Federal Register 49919-20). That workshop brought together experts in various areas of carcinogen risk assessment to study and comment on the use of animal evidence in considering both qualitative issues in classifying potential carcinogens and quantitative issues in dose-response and

extrapolation. The report from this workshop was made available to the public on April 24, 1989 (54 Federal Register 16403).

On June 16, 1989, the Agency announced that a workshop for the study and review of the use of human evidence in risk assessment would be held in Washington, D.C., on June 26 and 27, 1989 (54 Federal Register 25619). This report is a compilation of the discussions and presentations from that meeting. As with the Virginia Beach meeting, the Agency's intention was not to achieve consensus on or resolution of all issues. It was hoped instead that these workshops would provide a scientific forum for objective discussion and analysis.

These workshops are part of a three-stage process for reviewing and, as appropriate, revising EPA's cancer risk assessment guidelines. The first stage began with several information-gathering activities to identify and define scientific issues relating to the guidelines. For example, EPA scientists and program offices were invited to comment on their experiences with the 1986 cancer guidelines. Also, the August 1988 Federal Register notice asked for public comment on the use of these guidelines. Other information was obtained in meetings with individual scientists who regularly use the guidelines. Information from the workshops and these other sources will be used to decide when and how the guidelines should be revised.

In the second stage of the guidelines review process, EPA is analyzing the information described above to make decisions about changing the guidelines, to determine the nature of any such changes and, if appropriate, to develop a formal proposal for peer review and public comment. EPA's analysis of the information collected so far suggests several possible outcomes, ranging from no changes at this time to substantial changes for certain aspects of the guidelines.

In the third stage of this Agency review, any proposed changes would be submitted to scientific experts for preliminary peer review, and then to the general public, other federal agencies, and EPA's Science Advisory Board for



comment. All of these comments would be evaluated in developing final guidance.

## 2. Human Evidence Workshop

On June 26 and 27, 1989, epidemiologists and others met in Washington, D.C. to study and comment on the scientific foundation for possible changes in the human evidence sections of the 1986 carcinogen guidelines. In general, although these guidelines emphasize that reliable human evidence takes precedence over animal data, guidance on the use of human evidence is considerably less detailed than that for animal data. Thus, workshop discussions focused on the possibilities of expanding and clarifying the guidelines by adding new language for (1) study design and interpretation, (2) quantification of human data, and (3) weight-of-evidence analyses for human data.

The workshop participants met both in plenary sessions and in separate work groups to consider "strawman" language for potential inclusion in revised guidelines for carcinogen risk assessment. They also addressed related questions posed by the EPA Technical Panel. The work group on study design issues was chaired by Dr. Marilyn Fingerhut, Chief, Industrywide Studies Branch, at the National Institute for Occupational Safety and Health (NIOSH). Dr. Philip Enterline, Emeritus Professor of Biostatistics at the University of Pittsburgh School of Public Health, chaired the work group on dose-response issues. The work group on weight of evidence classification issues was chaired by Dr. Raymond Neutra, Chief of the Epidemiologic Studies Section of the California Department of Health Services. Dr. Enterline was the overall chair for the workshop. The strawman language and questions were developed by a subcommittee of the EPA Technical Panel. These documents were intended to initiate and guide work group discussions rather than to formally propose specific language or policy. Members of the EPA Technical Panel also participated in each work group. Other EPA scientific staff and the public attended the workshop as observers.

As a scientific forum for objective discussion and analysis among the invited panelists, the workshop was designed to assist EPA epidemiologists and scientists in developing the scientific foundation for proposed guidance on the use of human evidence in risk assessment. Broader policy issues will become important later in the process when the public is invited to review any proposed changes in the guidelines.

2023708880

**EPA CANCER GUIDELINES REVIEW**

**WORKSHOP ON HUMAN EVIDENCE**

**June 26-27, 1989**

**AGENDA AND WORKGROUP ASSIGNMENTS**

**Chairman, Dr. Philip Enterline**

**Monday, June 26**

| <u>Time</u>     | <u>Topic</u>                       | <u>Principals</u>          |
|-----------------|------------------------------------|----------------------------|
| 7:30 a.m.       | Registration Check-in              | All                        |
| 8:30 a.m.       | Welcome                            | Dr. Patton                 |
| 8:35 a.m.       | Opening Comments                   | Dr. Enterline              |
| 8:50 a.m.       | Public Interest Views              | Dr. Neutra and Panelists   |
| 9:20 a.m.       | Private Sector Views               | Dr. Ott and Panelists      |
| 9:50 a.m.       | Administrative Announcements       | Ms. Schalk                 |
| 10-10:20 a.m.   | COFFEE BREAK                       |                            |
| 10:20 a.m.      | Observer Comments                  |                            |
| 11:30 a.m.      | Charge to Work Groups              | Dr. Farland                |
| 12:00-1:15 p.m. | LUNCH                              |                            |
| 1:15 p.m.       | Workgroups                         |                            |
|                 | A: Study Design and Interpretation | Dr. Fingerhut, Chair       |
|                 | B: Weight of Evidence              | Dr. Neutra, Chair          |
|                 | C: Dose Response                   | Dr. Enterline, Chair       |
| 3:15-3:30 p.m.  | COFFEE BREAK                       |                            |
| 3:30 p.m.       | Workgroups                         |                            |
|                 | A & B                              | Dr. Neutra & Dr. Fingerhut |
|                 | C                                  | Dr. Enterline              |
| 5:30 p.m.       | Adjourn                            |                            |
| 5:30-7:00 p.m.  | Cash Bar Reception                 |                            |

Tuesday, June 27

| <u>Time</u>      | <u>Topic</u>                     | <u>Principals</u>   |
|------------------|----------------------------------|---|
| 8:00 a.m.        | Workgroup Reports and Discussion | Workgroup Chairs<br>(Drs. Enterline, Fingerhut & Neutra); Panelists |
| 10:15-10:30 a.m. | BREAK                            |   |
| 10:30 a.m.       | Observer Comments and Discussion |   |
| 11:30 a.m.       | Workgroup Recommendations        | Drs. Fingerhut, Neutra, Enterline                                   |
| 12:15 p.m.       | Wrap-up                          | Dr. Enterline   |
| 12:30 p.m.       | ADJOURNMENT                      |   |

**WORKGROUP ASSIGNMENTS**

Workgroup A: Study Design and Interpretation

Chair: Dr. Fingerhut

Members: Drs. Matanoski, Hulka, Buffler, Cantor, Friedlander, D. Hill, Halperin, Hogan, Koppikar

Workgroup B: Weight of Evidence

Chair: Dr. Neutra

Members: Drs. Cole, Ott, Blair, Falk, Infante, M. Chu, Bayliss, Blondell, Margosches

Workgroup C: Dose-Response

Chair: Dr. Enterline

Members: Drs. Crump, Checkoway, Gibb, Raabe, Smith, Krewski, Chen, Nelson, K. Chu, Farland, Scott

EPA Technical Panel: Drs. Farland, R. Hill, Patton, M. Chu, Rhomberg, Wiltse, Rees, Gibb, Bayliss, Blondell, Chen, D. Hill, Hogan, Margosches, Nelson, Scott

Risk Assessment Forum Staff: Drs. Patton, Rees

**COLLECTED WORKSHOP MATERIALS**

- **Study Design and Interpretation**
  - Strawman Language and Related Questions**
  - Chair Summary of Work Group Session**
- **EPA Classification System for Categorizing Weight of Evidence for Carcinogenicity from Human Studies**
  - Strawman Language and Related Questions**
  - Chair Summary of Work Group Session**
- **Dose-Response Assessment**
  - Strawman Language and Related Questions**
  - Chair Summary of Work Group Session**

## STUDY DESIGN AND INTERPRETATION

### Strawman Language and Related Questions

#### Introduction and Study Types<sup>1</sup>

Epidemiologic studies provide unique information about the response of humans who have been exposed to suspect carcinogens. These studies allow the possible evaluation of the consequences of an environmental exposure in the precise manner in which it occurs and will continue to occur in human populations (OSTP, 1985). There are various types of studies or study designs that are well-described and defined in various textbooks and other documents (e.g., Breslow et al., 1980, 1987; Kelsey et al., 1986; Lilienfeld et al., 1979; Mausner et al., 1985; Rothman, 1986). The more common types are described below.

A variety of study designs are considered to be hypothesis-generating. In general, these studies utilize already existing collections of data (e.g., vital statistics, census data), but only produce indirect associations, because they are based on broadly defined group or population characteristics. Studies depending on case reports typically are also considered hypothesis-generating, because the relatively limited numbers of cases and the absence of comparison groups generally do not permit causal inferences. Generally cross-sectional studies are hypothesis-generating. Sometimes a population may be well enough followed or restricted that bias is unlikely to arise from migration, mortality or similar removal from the observed group; a cross-sectional or prevalence study may then offer some sort of risk estimate.

Epidemiologic studies designed to test a specific hypothesis, such as case-control and cohort studies, are more useful in assessing risks to exposed humans. These studies examine the characteristics of individuals within a

---

<sup>1</sup>Editor's Note: Unless otherwise noted, paragraph numbers refer to the Human Studies section of the 1986 Guidelines for Carcinogen Risk Assessment in Appendix E of this document.

population. Case-control studies can provide reasonable estimates of population-based risk when controls are properly chosen, while cohort designs have the best capability to provide accurate estimates of population-based risk. Under certain circumstances, case-report studies may support causal associations, and prevalence studies may provide population-based risks.

Issues:

- a. We have noted that use of the "descriptive/analytical" characterization of the array of study designs may provoke classification disagreements, detracting from the desired focus on which designs have what utility for use in risk assessment. For that reason, two sentences previously in the Guidelines have been omitted. Does the Panel believe they should be restored or supplied in some other fashion, or does the current text provide sufficient discussion?
- b. Should PMR studies or clusters be specifically addressed? Where do they fit in?
- c. Should the guidelines point out that studies designed specifically to test hypotheses also can generate other hypotheses, but the distinction between these types of information should be maintained.

Adequacy (to replace current paragraph 2)

Criteria for the adequacy of epidemiologic studies for risk assessment purposes include, but are not limited to, factors which depend on the study design and conduct:

1. The proper selection and characterization of study and comparison cases or groups.
2. The adequacy of response rates and methodology for handling missing data.
3. Clear and appropriate methodology for data collection and analysis.
4. The proper identification and characterization of confounding factors and bias.

5. The appropriate consideration of latency effects.
6. The valid ascertainment of the causes of morbidity and death.
7. Complete and clear documentation of results.

For studies claiming to show no evidence of human carcinogenicity associated with an exposure, the statistical power to detect an appropriate outcome should be included in the assessment, if it can be calculated. It should be noted that sufficient statistical power alone does not determine the adequacy of a study.

Although not unique to human studies, it is important to reiterate that sufficient and thorough evaluation of suspect carcinogens requires that evidence be available in a form, quality, and quantity suitable for assessment. In some cases, the availability of and access to raw data may be important. Guidelines for reporting epidemiological research results have been previously published (IRLG, 1981; others?)

Issues:

- a. Application of various criteria are dependent on study type; most listed here usually apply to case-control and cohort studies. Is this a problem?
- b. Should we address rationales for combining sites and tumor types in this section, as done in the animal section?
- c. Given the discussion in the weight-of-evidence section, do we need much more here on what constitutes an adequate study?
- d. Any suggestions for appropriate citations?
- e. Is there too much emphasis on statistical power?

Criteria for Causality (paragraphs 3 and 4 deleted; new text)

Epidemiologic data are often used to infer causal relationships. Many forms of cancer are stated as causally related to exposure to agents for which there is no direct biological evidence, most notably, cigarette smoking and



lung cancer. As insufficient knowledge about the biological basis for disease in humans makes it difficult to classify exposure to an agent as causal, epidemiologists and biologists have provided a set of criteria that define a set of relationships about data. A causal interpretation is enhanced for studies that meet the following criteria. None of these criteria actually proves causality; actual proof is rarely attainable when dealing with environmental carcinogens. The absence of any one or even several of these criteria does not prevent a causal interpretation; none of these criteria should be considered either necessary or sufficient in itself.

Criteria for causality are:

1. Consistency: Several independent studies of the same exposure in different populations, all demonstrating an association which persists despite differing circumstances, usually constitute strong evidence for a causal interpretation (assuming the same bias or confounding is not also duplicated across studies). This criterion also applies if the association occurs consistently for different subgroups in the same study.

Issue: Diverse responses from similar populations (races or species) lend weight to human conclusions but seem to detract from animal ones. How should we address this inconsistency?

2. Strength (magnitude) of association: The greater the estimate of the risk of cancer due to exposure to the agent, the more credible will be a causal interpretation. It is less likely that nonrandom error (e.g., bias or some confounding variable) or chance can explain the association because these factors themselves have to be highly associated with the disease. A weak association might be more readily explained by the presence of chance or bias.

Issues:

- a. Should we provide a guideline value, such as a relative risk of 5.0, since magnitude of association can also depend on, for instance, the variety and range of magnitudes of exposure present and the rarity of the cancer?

- b. Should the example of smoking-alcohol-esophageal cancer be considered here?

3. Temporal relationship: The disease occurs within a reasonable enough time frame after the initial exposure to account for the health effect. Cancer requires a latent period during which transformation of neoplasia into malignancy occurs and a period of time passes before discovery. While latency periods vary, existence of the period is acknowledged. Since the time of transformation is seldom known, however, the initial period of exposure to the agent is the accepted starting point in most epidemiologic studies.

4. Dose-response or biologic gradient: An increase in the measure of effect is correlated positively with an increase in the exposure or estimated dose. A strong dose-response relationship across several categories of exposure can be considered to be evidence for causality if confounding effects are unlikely to be correlated with exposure levels. The absence of a dose-response gradient, however, may mean only that the maximum effect had already occurred at the lowest dose or perhaps all gradients of exposure were too low to produce a measurable effect. The absence of a dose-response relationship should not be construed as evidence of a lack of a causal relationship.

5. Specificity of the association: If a single, clearly-defined exposure is associated with an excess risk of one or more site-specific cancers, while other sites show no association, it increases the likelihood of a causal interpretation. Different agents, however, may be responsible for more than one site-specific cancer. Replication of the specific association(s) in different population groups (cf. consistency) would then be needed to provide strong support for a causal interpretation.

Issue: Shall we retain this last sentence? A comment has been made that specific locations (i.e., microenvironments) influence expression of cancer.

In some cases, conclusions regarding an association may be based on a mixture of chemicals rather than the specific chemical in question. In these cases, judgment on the causal relationship to the specific chemical will

depend on such other information as the pharmacokinetics of the chemical or other biologic or epidemiologic data.

Issue: Shall we include: In some instances, it may be concluded only the mixture can be held culpable, e.g., in the process to produce benzyl chloride.

6. Biological plausibility: The association makes sense in terms of what is known about the biologic mechanisms of the disease or other epidemiologic knowledge. It is not inconsistent with biological knowledge about how the exposure under study could produce the cancer.

7. Collateral evidence: A cause-and-effect interpretation is consistent with what is known about the natural history and biology of the disease. A proposed association that conflicted with existing knowledge would have to be examined with particular care.

#### References

Breslow, N.E. and Day, N.E. Statistical Methods in Cancer Research, Vol. 1. The Analysis of Case-Control Data. 1980.

Breslow, N.E. and Day, N.E. Statistical Methods in Cancer Research, Vol. 2. The Design and Analysis of Cohort Studies. 1987.

Interagency Regulatory Liaison Group (IRLG). Guidelines for documentation of epidemiologic studies. American Journal of Epidemiology. 114:609-613. 1981.

Kelsey, J.L., Thompson, W.D., and Evans, A.S. Methods in Observational Epidemiology. 1986.

Lilienfeld, A.M. and Lilienfeld, D. Foundations of Epidemiology, 2nd ed. 1979.

Mausner, J.S. and Kramer, S. Epidemiology, 2nd ed. 1985.

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

Rothman, K.J. Modern Epidemiology. 1986.

Chair Summary of Work Group Session on  
Study Design and Interpretation  
Chair: Dr. Marilyn Fingerhut

Introduction

The Study Design Work Group focused on three questions contained in the strawman language suggested by the EPA staff as a preliminary revision of Section II.B.7: 1) What types of epidemiologic studies are acceptable to the EPA for risk assessment? 2) What characteristics are desirable in a study to be used for risk assessment? and 3) What criteria strengthen the view that an epidemiologic association may reflect a causal relationship?

The Study Design Work Group and the Weight-of-Evidence Work Group met together to consider the first question and to agree upon the types of studies to be discussed in each Work Group. The members of the Study Design Group discussed Questions 2 and 3.

The sections below briefly describe the discussions pertaining to the three questions, and identify the recommendations and suggestions made to EPA. The sections contain revised strawman language for Section II.B.7, which reflects the ideas suggested by the Study Design Work Group. This chair's summary presents the work group's views within the context of the strawman document.

Question 1: What types of human studies are acceptable to  
the EPA for purposes of risk assessment?

The members of both the Study Design and Weight-of-Evidence groups discussed this question in some detail and concluded that all valid epidemiologic studies can contribute information to an EPA risk assessment. Consequently, the panelists rejected suggestions by a few members to weight certain study types more heavily than others. There was general agreement that the various types of epidemiologic studies, properly conducted, could be

useful. These include cohort, case-control, cross-sectional, proportional mortality (incidence) ratios, clusters, clinical trials, and correlational studies. Each type has strengths and limitations. It was agreed that "case reports" do not constitute studies, but that some of these should be reviewed by EPA during a risk assessment effort, because series of case reports have provided key information about human risk for several chemicals. Vinyl chloride was one example cited.

Both groups strongly recommended that EPA needs additional experienced epidemiologists to evaluate epidemiologic data and to assist in risk assessments, because professionally sophisticated judgments are required when evaluating the studies.

The Study Design Group reviewed the EPA strawman language suggested as a replacement for the current paragraph 1 of Section II.B.7 Human Studies. The group generally agreed that the proposed revision not be used. The group suggested a brief replacement paragraph:

Introduction and Study Types (to replace the current paragraph 1)

Epidemiologic studies with various study designs can provide unique information about the response of humans who have been exposed to suspect carcinogens. Each study must be evaluated for its individual strengths and limitations. Conclusions about causal associations usually also include consideration of the entire body of literature, including toxicology and biologic mechanisms.

The Study Design Work Group suggested that guidelines be written for use by experienced epidemiologists. Therefore, the following responses were given to the questions posed by the EPA on page 2 of the strawman text (p. 9 of this document). There is no need to distinguish studies as analytical vs. descriptive, or hypothesis-generating vs. hypothesis-testing, or complete vs. incomplete (as suggested by one member of the group) because experienced

epidemiologists, who are aware of strengths and limitations of the various study designs, will judge studies by their inherent validity and applicability to the particular risk assessment. For this reason, there is no need to specifically address proportional mortality ratios (PMRs) or clusters, or to address the distinction between hypothesis-generating and hypothesis-testing in the guidelines.

The Study Design Work Group recognized that it may be desirable to provide in the guidelines an overview of epidemiologic principles and study types. The information would be useful for professionals trained in other disciplines. The information could also explain to the public how the EPA uses human studies in risk assessment. The group suggested that this overview of epidemiology might be placed in an appendix.

The group suggested that EPA continue to provide epidemiologic training to nonepidemiologists in the Agency who are involved with the risk assessment activities. However, the key judgments on epidemiologic studies should be made by experienced epidemiologists. Upon learning that EPA has very few epidemiologists on staff, the group recommended expanding this expertise in the Agency. A few of the members of the combined Study Design and Weight-of-Evidence Groups suggested that EPA might wish to consider using, for risk assessment, the approach of the International Agency for Research on Cancer (IARC) in which the entire evaluation of human data is conducted by expert epidemiologists, and is thus free from political interference. Other participants observed that political concerns may influence any such group. They suggested that the regular use of EPA staff provides an objective approach to risk assessment. There was some discussion but no agreement in the groups about this point.

Question 2: What characteristics are desirable in a study  
used for risk assessment?

The EPA had provided strawman language to replace paragraph 2 of Section II.B.7 Human Studies, which focused on the question of "criteria for adequacy of epidemiologic studies for risk assessment purposes." Because the members of the group generally agreed with the view that all valid epidemiologic studies may contribute information to a risk assessment, discussion by the Study Design Work Group led to substitution of a different question: "What characteristics are desirable in a study used for risk assessment?" Since each type of study has particular characteristics, strengths, and limitations, the group suggested revising of the EPA strawman language for paragraph 2 to describe characteristics desirable (rather than required) for risk assessment. Several new characteristics were added to those identified by the EPA version.

The following suggested revision is a restatement of ideas from the Group and should not be considered a polished or finished revision.

Adequacy (to replace current paragraph 2)

Criteria for the adequacy of epidemiologic studies are well recognized. Considerations made for risk assessment should recognize the characteristics, strengths, and limitations of the various epidemiologic study designs. Characteristics which are desirable in the epidemiologic studies are listed here.

1. Relevance
  - The study deals with the exposure-response relationship central to the risk assessment.
2. Adequate Exposure Assessment
  - Study subjects have exposure.
  - Analysis deals with time-related measures as far as study type permits, e.g., duration, intensity, age at first exposure, etc.

3. Proper Selection and Characterization of Study and Comparison groups
  - Selection and characterization are carefully described.
  - Source population is appropriate.
  - Results are generalizable to populations to be protected by the risk assessment.
4. Identification of a Priori Hypotheses
5. Adequate Sample Size
6. Adequate Response Rates and Methodology for Handling Missing Data
7. Clear and Appropriate Methodology for Data Collection and Analysis
8. Proper Identification and Characterization of Confounding and Bias
9. Appropriate Consideration of Latency Effects
10. Valid Ascertainment of Causes of Morbidity and Death
11. Complete and Clear Documentation of Results

The panelists recommended that EPA continue to actively seek available unpublished studies, if an unpublished report (or the documentation for a published report) might contribute to the risk assessment process.

Question 3: What criteria strengthen the view that an epidemiologic association may reflect a causal relationship?

Strawman language had been provided by EPA to substitute for paragraphs 3 and 4 of Section II.B.7 Human Studies. The Study Design Work Group suggested that the EPA staff consider rewriting this text to express an historical



approach, indicating that Koch's postulates were modified by Bradford Hill for use in environmental studies, and that his criteria have been modified by EPA for considerations relevant to risk assessment.

The panelists were in general agreement on most points that are contained in the suggested text below. Some members suggested deleting "specificity." They viewed it as misleading or incorrect, based upon the view that most agents are observed to cause several effects. However, all agreed that as expressed below, it is a useful criterion when it is present.

The panelists agreed that only one criterion (temporal relationship) was essential for causality. The presence of other criteria may increase the credibility of a causal association, but their absence does not prevent a causal interpretation. The panelists viewed all but specificity and coherence as applicable to an individual study.

The panelists' ideas for a suggested revision follow:

Criteria for Causality (paragraphs 3 and 4 deleted; new text).

Epidemiologic data are often used to infer causal relationships. A causal interpretation is enhanced for studies to the extent that they meet the criteria described below. None of these actually establishes causality; actual proof is rarely attainable when dealing with environmental carcinogens. The absence of any one or even several of the others does not prevent a causal interpretation. Only the first criterion (temporal relationship) is essential to a causal relationship: with that exception, none of the criteria should be considered as either necessary or sufficient in itself. The first six criteria apply to an individual study. The last criterion (coherence) applies to a consideration of all evidence in the entire body of knowledge.

1. Temporal relationship: This is the single absolute requirement, which itself does not prove causality, but which must be present if causality

is to be considered. The disease occurs within a biologically reasonable time frame after the initial exposure to account for the specific health effect. Cancers require certain latency periods. While latency periods vary, existence of the period is acknowledged. The initial period of exposure to the agent is the accepted starting point in most epidemiologic studies.

2. Consistency: When compared to several independent studies of a similar exposure in different populations, the study in question demonstrates a similar association which persists despite differing circumstances. This usually constitutes strong evidence for a causal interpretation (assuming that the same bias or confounding is not also duplicated across studies). This criterion also applies if the association occurs consistently for different subgroups in the same study.
3. Strength (magnitude) of association: The greater the estimate of risk and the more precise (narrow confidence limits), the more credible the causal association.
4. Dose-response or biologic gradient: An increase in the measure of effect is correlated positively with an increase in the exposure or estimated dose. A strong dose-response relationship across several categories of exposure, latency, and duration is supportive although not conclusive for causality, assuming confounding effects are unlikely to be correlated with exposure levels. The absence of a dose-response gradient, however, may be explained in many ways. For example, it may mean only that the maximum effect had already occurred at the lowest dose, or perhaps all gradients of exposure were too low to produce a measurable effect. If present, this characteristic should be weighted heavily in considering causality. However, the absence of a dose-response relationship should not be construed by itself as evidence of a lack of a causal relationship.

5. Specificity of the association: In the study in question, if a single exposure is associated with an excess risk of one or more cancers also found in other studies, it increases the likelihood of a causal interpretation. Most known agents, however, are responsible for more than one site-specific cancer. Therefore, if this characteristic is present, it is useful. However, its absence is uninformative.
6. Biological plausibility: The association makes sense in terms of biological knowledge. Information from toxicology, pharmacokinetics, genotoxicity, and in vitro studies should be considered.
7. Coherence: This characteristic is used to evaluate the entire body of knowledge about the chemical in question. Coherence exists when a cause-and-effect interpretation is in logical agreement with what is known about the natural history and biology of the disease. A proposed association that conflicted with existing knowledge would have to be examined with particular care.

In a joint session of the Study Design and Weight-of-Evidence Groups at the end of the meeting, some panelists noted the desirability of having epidemiologic data available for use in risk assessment at the time that animal studies are completed by the National Toxicology Program (NTP). A suggestion was made by some that EPA consider undertaking an effort to assess the feasibility of conducting a human epidemiologic study at the same time the Agency recommends that NTP undertake an animal study. There was only limited discussion of this point. Some panelists objected to it, mainly because of logistic difficulties.

EPA CLASSIFICATION SYSTEM FOR CATEGORIZING WEIGHT OF EVIDENCE FOR  
CARCINOGENICITY FROM HUMAN STUDIES

Strawman Language and Related Questions

Assessment of Weight of Evidence for Carcinogenicity from Studies in Humans

There are a variety of sources of human data. When the totality of human evidence is considered, the conditions under which the information has been collected are of importance in defining the limits of its inference. These limits are particularly critical for studies where no positive results have been seen, although they contribute to conclusions in all circumstances.

In the evaluation of carcinogenicity based on epidemiologic studies it is necessary to consider the roles of extraneous factors such as bias and other nonrandom error and chance (random error) and how they might affect evaluation and estimates of an agent's effects. Some extraneous factors of concern are selection bias, information bias, and confounding. Five classifications of human evidence are established in this section. The following discussion includes some interpretation and illustration of their use.

1. The category of sufficient implies the existence of a causal relationship between the exposure in question and an elevation of cancer risk. Most if not all of the criteria for causality as defined in Section II. B. 7. should be met. Most agents or mixtures falling into this category would require at least one methodologically sound epidemiologic study meeting most of the criteria for causality and whose results cannot be explained by chance, bias, or confounding.

Issues:

- a. If one such study is available, would others be needed as confirmatory?
- b. Is it necessary to specify what are "most criteria?"

Sometimes a case series will present data that drive a causal conclusion.

Issues:

- a. Are supporting studies needed? Language that might serve is: One or more supporting epidemiologic studies that also demonstrate a relationship between the exposure and cancer should be available. The latter studies need not be definitive by themselves although the stronger they are in terms of their validity the more credible will be a "sufficient" categorization of the epidemiologic data.
- b. Sometimes studied populations will differ only in cumulative dose or in dose rate. Should a conclusion of carcinogenicity be limited to circumstances of exposure?

Corollary: Shall we include discussion of the evaluation of a body of studies where some show effects at one site and some show them at another or where some are of different ethnic or geographic groups or where there are age and sex differences?

- c. The Agency receives studies, some by statute, from a variety of sources, that may not have appeared in the open or peer-reviewed literature. Should comment be made regarding our intent to use such studies?

2. The category of limited implies that a causal interpretation is more credible than nonrandom error, although it cannot be entirely ruled out as an explanation for the statistically significant positive association found in at least one or more epidemiologic studies. Such studies would typically include a vigorous effort by the author or be carefully reviewed by the Agency to explain why nonrandom error (confounding, information bias, etc.) is unlikely to account for the association.

Also included in the limited category are agents for which the evidence consists of some number of independent studies exhibiting statistically significant positive associations between the exposure and the same site-

specific cancer but for which nonrandom error could not be ruled out entirely as the explanation for the association in each study. This category may also include substances for which a series of epidemiologic studies (some number of which must be considered valid) exhibit apparent but not significant positive associations for the same site-specific cancer without any series of valid studies in which there is apparent lack of association to counter the observed association.

Issues:

- a. How many studies would support each conclusion?
- b. Is it necessary for responses in a series of studies to be specific to site in order to fall into the limited category?

3. The category of inadequate implies that the data, although perhaps suggestive, do not meet the criteria for a limited categorization of the evidence. This would include studies that demonstrate statistically significant positive associations that could be explained by the presence of nonrandom error and which are not specific with respect to site. Also included in this category are studies deemed of insufficient quality or statistical power, and where there is no confidence in any particular interpretation. For example, results may be consistent with a chance effect, or exposure may not clearly be tied to the agent in question. Alternately, a report of a study may render it incapable of being evaluated owing to insufficient documentation.

Issue: Is it proper to modify or downgrade a category by such language as "Studies showing no positive results can be used to lower the classification from limited to inadequate . . . " ?

Possible Choices to Complete This Statement Might Be:

"only if the exact same conditions (including sensitivity) have been replicated in a statistically significantly positive study of the kind described under this category. The results would thus be contradictory and the net effect of the latter study would be to negate the findings of the former."

or

"if they are at least as likely to detect an effect as an already-completed study providing limited evidence."

Corollary: How explicit should we be?

4. The category of no data indicates no data are available directly regarding humans.

5. The category of evidence of not being a carcinogen in humans is reserved for circumstances in which the body of evidence indicates that no association exists between the suspected agent and an increased cancer risk. It should be recognized that alterations in the conditions under which a study is done may lead to statistically significant risk estimates where they did not exist before. Studies of uncertain quality with no positive results should not be used to reduce the weight of evidence.

Issues:

- a. We are reluctant to use the word "negative" because it has come to mean a variety of things including (1) a study with no cases of cancer, (2) a study judged statistically to have no excess cases of cancer, and (3) a study that leads readers to believe there is no need for concern about carcinogenicity. We do not wish to perpetuate the misuse of the term "negative" when referring to certain epidemiologic studies. Have we adequately described the circumstances under which we would conclude that the body of epidemiologic evidence suggests an agent is not a carcinogen?
- b. Do we need this category? Will it ever be used?

Chair Summary of Work Group Session on Classification System  
for Categorizing Weight of Evidence for Carcinogenicity  
from Human Studies

Chair: Dr. Raymond R. Neutra

I. INTRODUCTORY DISCUSSIONS

Weighing evidence refers to the act of reviewing and summarizing human evidence ranging from case studies to randomized trials. Evidence which suggests positive, null, or even protective carcinogenic effects is considered while taking note of the quality of each piece of information. The group seemed to advocate a procedure that considers all study results regardless of direction, rather than only positive studies, in other words, a "weight of evidence approach" rather than a "strength of evidence approach," respectively.

One workshop participant pointed out that those who review human evidence assign some informal prior probability to the hypothesis that the substance under investigation causes cancer in humans. Without advocating formal Bayesian statistical procedures, it should be noted that this "prior probability" is influenced by the nature of the substance, information on its metabolism, and behavior in short-term tests. The group decided that results of animal bioassays or subchronic tests should not influence judgments on the prior probability or in the interpretation of the human studies, since a separate process, in EPA deals with the weight of animal evidence. In a subsequent process, the two streams of evidence will be combined by scientists to give a final "posterior" characterization of the evidence.

There was a discussion of nomenclature. It was agreed that the adjective "negative" should be avoided, as it is ambiguous. It has been used to mean "bad," "protective effect," "no effect," or "absent." For the purposes of the work group it was agreed that "null" would be used for a study that had a relative risk close to 1.0 with confidence limits which included 1.0, and that "inverse association" is the appropriate terminology for a study that showed a



relative risk less than 1.0 and confidence limits which did not include 1.0. A positive study is one with a relative risk greater than 1.0 with confidence limits which do not include one.

The group recognized that the evaluation of a body of evidence depends on, but is different from, the process of evaluating individual studies. The latter process was discussed by the Study Design and Evaluation work group, which listed a series of study characteristics and criteria for likely causality that should be considered in characterizing a single study. In the process of evaluating the evidence from a single study, one of the following categories might be designated: clear, some, equivocal, or no evidence of human carcinogenicity. The study design group felt that it was not possible or desirable to have a rigid algorithm for making this categorical determination on a particular study. For the purposes of discussion in the weight-of-evidence work group, these terms were used even though definitions were not developed by that group. Similarly the weight-of-evidence group was not in favor of a rigid algorithm for combining evidence among individual studies. Instead, it was suggested that a review of all the studies by a qualified group could lead to an ordinal classification. It was agreed that assigning a ratio scale numerical score of evidentiary sufficiency would not be helpful since the Agency would need to categorize the score anyway for action purposes. An artificial numerical score might well complicate rather than simplify the regulatory process.

The general consensus of the workshop was to avoid too narrowly defined guidelines, e.g., the use of specific numerical standards in defining "tight confidence limits."

It was further noted that overly specific guidelines had a number of drawbacks. First, they could never capture all conceivable contingencies and thus would be a Procrustean bed. Second, a cookbook could be inappropriately used. This is a particular problem in an agency such as EPA where the prevalence of epidemiologists is below 1/1000 (15 total in this organization of 16,000 employees). The group encouraged EPA to increase the number of

epidemiologists on their staff and to have continuing education for epidemiologists and others to foster interdisciplinary work.

One participant urged that an IARC-like process using external experts should be employed for weighing evidence. A number of drawbacks were pointed out.

There was discussion about the methods that EPA should employ in summarizing bodies of evidence. For example, should there be an appendix presenting techniques for meta analysis and its graphical presentation? No consensus emerged because of concern that these techniques could be misused.

## II. PROPOSED MODIFICATION OF THE STRAWMAN CATEGORIES FOR WEIGHED EVIDENCE

The strawman language suggested the categories of Sufficient, Limited, Inadequate, No Human Data, Evidence of Not Being a Carcinogen in Humans.

The work group suggested the categories: Sufficient Evidence for Human Carcinogenicity, Limited Evidence for Human Carcinogenicity, Inconclusive Evidence for Human Carcinogenicity, No Human Data. There was no consensus and considerable argumentation about two additional possible categories: Human Evidence Not Suggestive of Carcinogenicity and Sufficient Evidence for Lack of Human Carcinogenicity. The work group's understanding of the categories and its responses to specific strawman issues raised by EPA staff in each respective section are dealt with below.

### III. SPECIFIC COMMENTS ON EACH EVIDENTIARY CATEGORY

#### 1. Sufficient Evidence for Human Carcinogenicity

##### Issues A and B. Required Number and Quality of Studies

In some circumstances where the information "prior probability" was high and the study was particularly strong, the work group felt that a single study with clear evidence could provide sufficient evidence for a substance. In most cases, there would probably be more than one study with clear evidence. The group did not want to provide a cookbook to define the criteria for sufficiency.

##### Issues A, B and C. Need for Supportive Information and Peer Review

The work group felt that only in the rarest circumstances would a case series provide sufficient evidence for human carcinogenicity, e.g., vinyl chloride. They were reluctant, however, to provide a rigid algorithm which always requires supporting studies.

The work group did not wish to limit hazard identification to the dose scenario covered in the epidemiological study. That is dealt with during dose-response assessment. It also advised that a series of studies which show an increased risk of cancer at a particular site should be given more weight than a series showing an increased risk at various sites. In the latter case, the mechanism of causation should enter into the weighing process. Also, hazard identification should not be limited to the particular races, sex, or age group covered by the epidemiological studies.

Most work group members thought that unpublished studies should be considered in weighing evidence. Two strong caveats were voiced. There should be deadlines for submission to prevent a continual stream of last-minute submissions which delay the regulatory process indefinitely. There

should be regulatory peer review and perhaps a requirement that any journal acceptance or rejection correspondence be submitted to the Agency.

## 2. Limited Evidence For Human Carcinogenicity

### Issues A and B. Number of Required Studies and Sites Specificity

One or more studies providing "some evidence" even if there are some "null" studies, will qualify for this classification. Alternatively, a series of positive equivocal evidence studies in the absence of any null studies would qualify as well. The work group did not have suggestions for an algorithm to deal with these issues.

## 3. Inconclusive Evidence For Human Carcinogenicity

The work group preferred the term "inconclusive" to the term "inadequate" because the latter implies poor quality evidence when in fact the studies may be of good quality but contradictory.

The evidence may gain this characterization under three contingencies: 1) the evidence is a mixture of equivocal and null studies; 2) the evidence is a mixture of imprecise null studies which do not add up to a precise null study; or 3) The evidence does not meet the criteria for the other categories.

### Issue A. Modifying and Downgrading Categories

The work group did not discuss exact wording to cover the situation in which a positive study is followed by a null study so that a substance would fall from the limited category into the inconclusive category.

#### 4. No Human Data

The category of No Data indicates no data are available directly regarding humans. The subcommittee had no comments on this self-evident category.

#### 5. No Evidence for Human Carcinogenicity (Human Evidence Not Suggestive of Carcinogenicity)

There was considerable discussion about the concept of this and the following category and of the names which should properly apply to them. It should be kept in mind that this epidemiological categorization was to be based on human studies and interpreted in the light of short-term studies and mechanistic insights, but not subchronic or chronic cancer animal bioassays. The ultimate classification would weight the two streams of evidence.

The sensitivity of this and the following category has to do with the weight required for human studies to overcome sufficient animal evidence. The weight of evidence for a substance would fall into this category rather than the "inconclusive" category if all the human studies had been null studies yet animal risk assessment would have predicted null studies at the human dose delivered to the population size exposed, and there were possible mechanisms of action which would predict a nonthreshold dose-response curve. In this case, a series of good null studies are simply not good enough to definitively cancel out sufficient animal evidence. Yet the evidence of a series of null studies with individually tight confidence intervals or tight intervals when taken together somehow warrants more than an "inconclusive" label. The weight of evidence regarding EDB is an example of this situation. The substance is genotoxic and animal risk extrapolations would have predicted that the worker studies carried out could not have detected the fairly small relative risks expected from the doses received.

There was some discussion about what it meant to be "taken together." Subjecting a series of small studies to a Mantel-Haenszel procedure was suggested. There were some technical objections, but it was agreed that consensus might be found for some analogous procedure.

#### 6. Sufficient Evidence for Noncarcinogenicity in Humans

There was no consensus on this classification. Many members felt that a series of strong null studies with tight confidence limits would qualify if coupled with a widely accepted mechanistic understanding that suggested that the agent should not cause cancer at doses to which humans could be accidentally exposed at work or in the environment. This kind of mechanistic and epidemiological evidence could cancel out a series of positive animal bioassays for the purpose of hazard identification.

A few others in the workshop pointed out that null studies could be used to determine if humans were substantially less sensitive than animals in the dose-response stage of risk assessment. In the hazard identification stage, however, only a much more stringent criterion was appropriate. The confidence limits around the null value needed to be so tight that they excluded the possibility of added risk of public health and regulatory concern. One proposal supported by two of the workshop participants was that "risk of regulatory concern" be quantified.

The majority who disagreed with this proposal seemed to have two views. First, that it was unwise to tie the Agency's hand with a number which might change and which was an issue of risk management. Second, it was perceived that no study would be able to rule out all risks of potential regulatory concern. Such power is not practically achievable, and even if it were, one would not trust epidemiology's ability to control confounding sufficiently to accurately assess relative risks so close to the null. This stringent requirement might also create disincentives for government and corporate sponsors who fund epidemiological studies in the hope that they would give the

candidate chemical a "clean regulatory bill of health" in the hazard identification phase. Everyone recognized that it is not possible to prove an absolute zero risk.

The advocates of the more stringent definition, e.g., to require exclusion of all risks of regulatory concern, responded that this was exactly the point. Sponsors should give up the vain hope that null epidemiological evidence could be used in the hazard identification process to get their substances "off the list"<sup>1</sup> when there is sufficient evidence from animal studies.

Although epidemiology can rarely get a substance off the hazard identification list once an animal study has put it there, a series of good quality null studies would put the substance in the Human Evidence Not Suggestive of Carcinogenicity category. If the human response was considerably lower than that predicted from animal studies, this may lead to higher tolerated industrial emissions of the substance, which in turn, may, in some cases, have important economic implications. Thus incentives exist for carrying out epidemiological investigations even if these studies can rarely be used to justify delisting a substance.

Saccharin was cited as an example of a substance that should not get a "clean bill of health." A series of strong null human studies was still easily compatible with the animal predictions of 800 extra cases per year in the United States. Although this is equivalent to a relative risk of 1.01, small by epidemiological standards, one is hard pressed to exonerate a substance with a study which does not have the power to exclude the very number predicted by animal risk assessment. It is for this reason that some members in the subcommittee demanded a null study with the power to exclude the low added lifetime risks of interest to regulatory agencies. The wording used by IARC for a similar category was proposed with the addition of a sentence dealing with the need for power to exclude risks of regulatory

---

<sup>1</sup>Editor's note. Concept of a list arose during the work group discussion, and was not previously referred to the strawman language.

interest. This line of argument was unfamiliar and even irritating to many of the epidemiologists present.

The proponents for the category, Sufficient Evidence for Noncarcinogenicity in Humans, responded that saccharin was a good candidate for the category since there was experimental and mechanistic evidence that bladder cancer in rats should only occur at high doses and that downward extrapolation of risk to dietary levels in humans was not warranted. There was a question as to whether there was scientific consensus on this.

Although most of the discussants did not question the use of widely accepted mechanistic arguments separating man from animal, there were a few concerns that there could always be other carcinogenic mechanisms that were shared between humans and rodents that might still operate. This is something which needs to be examined carefully.

A consensus did seem to emerge against the more permissive strawman language which suggested that a series of unopposed null studies constituted Sufficient Evidence for Noncarcinogenicity in Humans.

The work group agreed with EPA staff about not using the word "negative" because of the many different interpretations that can be given to this word.

The discussion in the strawman document about how to interpret null studies was not adequate and prompted the arguments outlined above. The work group did not come to a consensus about categories that dealt with null studies.



#### IV. MISCELLANEOUS OBSERVATIONS

##### 1. Control of Smoking

When a series of studies show an effect but smoking has not been controlled for, this should not automatically disqualify the studies for consideration. It should not always be assumed that controlling for smoking would weaken an observed chemical effect. Highly exposed individuals may smoke less.

##### 2. Multiple Exposures

There was a discussion of the problem of concomitant exposure to other chemicals in a series of studies. One should determine if all of the studies were characterized by exposure to the same set of chemicals. If not, some of the other chemicals could be removed as confounders. If so, the participants recommended following the IARC policy of implicating the process as a whole.

##### 3. Carcinogenic Metabolites

Sometimes a substance produces by metabolism another substance that has achieved some degree of evidentiary sufficiency even though the parent compound has not been studied. If the target organ of main exposure is the same, the workshop members agreed that the parent compound should be classified similarly to the metabolite. As circumstances deviate from this paradigm, more judgment will be needed in the classification process.

##### 4. Proper Use of Power Calculations

The work group came to a consensus that it made no sense to calculate the power of the study after the fact. Instead, one should inspect the confidence limits to see if they include the expected effect.

## 5. Epidemiologic Research Needs

A number of participants suggested that EPA fund an epidemiological study every time a request was made to the National Toxicology Program (NTP) for an animal study. Others felt that this would be wasteful of scarce epidemiological resources and that one should wait for positive animal results before embarking on an epidemiological study. Still others felt that we could be missing human carcinogens that happened to have negative animal results. There seemed to be some consensus that a search be initiated for an exposed cohort and an exposure assessment be carried out, in parallel with each NTP study.

## DOSE-RESPONSE ASSESSMENT

### Strawman Language and Related Questions<sup>1</sup>

1. Selection of data: As indicated in Section II.D., guidance needs to be given by the individuals doing the qualitative assessment (epidemiologists, toxicologists, pathologists, pharmacologists, etc.) to those doing the quantitative assessment as to the appropriate data to be used in the dose-response assessment. This is determined by the quality of the data, its relevance to the likely human modes of exposure, and other technical details.

A. Human studies. Estimates based on adequate human epidemiologic data are preferred over estimates based on animal data. Intraindividual differences, including age- and sex-related differences, should be considered where possible. If adequate exposure data exist in a well-designed and well-conducted epidemiologic study that has shown no positive results for any relevant endpoints, it may be possible to obtain an upper-bound estimate of risk from that study. Animal-based estimates, if available, also should be presented when such upper bound estimates are calculated. More carefully executed dose-response assessments benefit from the availability of data that permit the ages of exposure and onset of disease and the level of exposure and duration to that exposure to be incorporated in the assessment.

#### Issues:

- a. Should an upper-bound risk estimate be made from a nonpositive human study if it is the only risk estimate that can be made?

---

<sup>1</sup>The first two paragraphs of the strawman language provided are intended to be parallel to the first two paragraphs of III.A.1. of the current guidelines (see Appendix E). The second three paragraphs of the strawman language are intended to be parallel to the first three paragraphs of III.A.2. of the current guidelines.

- b. Should we describe what is minimal and what is preferred data? A discussion of "preferred data" could become so extensive that description in the guidelines would be cumbersome, such data may never be obtainable, and the discussion in the guidelines would probably never be exhaustive. Alternately, would it be possible to specify levels of preferred data?
- c. In the absence of dose-rate information, is the use of cumulative dose an appropriate default position? Should that be specified in the guidelines?
- d. Should the guidelines be made to reflect possible differences in dose-response between children and adults because of differences in tissue growth, metabolism, food and fluid intake, etc.?
- e. Should the use of person-years of observation be counted from the beginning or the end of exposure for dose-response assessment? Should this be discussed in the guidelines?

2. Choice of mathematical extrapolation model: Since risks at low-exposure levels cannot be measured directly either by animal experiments or by epidemiologic studies, a number of mathematical models have been developed to extrapolate from high to low dose. Models should make optimal use of biologic data where possible. Different extrapolation . . . (The language here would be the same as that in the current guidelines.) . . . A rationale will be included to justify the use of the chosen model.

A. Human data. Dose-response assessments with human data should consider absolute as well as relative risk models when the data are available. Where possible, results from both models should be presented. If selecting one model over another, the rationale should be described. In the absence of information to the contrary, a dose-response model that is linear at low doses will be employed with human data. A point estimate from the model may be used to estimate risk at doses below the observable range.

B. Animal data. For animal data, the linearized multistage procedure will be employed in the absence of information to the contrary. The linearized multistage model is a curve-fitting procedure. It does not model what is believed to be a multistage process of tumor development. It is appropriate as a default procedure, however, in that it is linear at low

doses. Where appropriate, the results of different extrapolation models may be presented for comparison with the linearized multistage procedure. When longitudinal data . . . (Continue discussion in current guidelines.)

Issues:

- a. Point estimates from models of human data have been used in the Agency in the past for dose response assessment. This differs from risk estimates made from animal data in that the estimates from animal data are statistical upper bounds. The rationale for using a point estimate with human data is that (1) there is no cross-species extrapolation with the human data (2) exposures to humans in the epidemiologic data sets used for modeling (usually occupational studies) are much closer than the doses used in animal studies to the environmental exposures of concern and (3) the point estimate, though not a statistical upper bound, provides an upper bound in the sense that the response at lower doses is likely to be less than that predicted by a model with low-dose linearity. Should statistical upper-bound dose-response estimates be used with human data for consistency with the dose-response estimates from animal data?
- b. The linearized multistage model is recognized as a curve-fitting procedure. It does not model stages of cancer. Is it appropriate to recommend the linearized multistage procedure in absence of information to the contrary for the dose-response assessment of animal data? Would it be more appropriate to simply recommend a model that is linear at low doses in the absence of information to the contrary?
- c. Are there examples of agents which have supralinear dose response for humans at low doses? If so, a model with low dose linearity may not be protective of public health.

Chair Summary of Work Group Session on  
Dose-Response Assessment

Chair: Dr. Philip Enterline

The Dose-Response Work Group discussed two major questions posed by EPA in the strawman language. These were:

- (1) How should the most appropriate data be selected for use in dose-response assessment?
- (2) How should the most appropriate extrapolation model be selected for estimating risks from human data sets and should these models be consistent with those used for animal data?

The sections below apply to Section III.A.1 and 2 of the current guidelines (see Appendix E). Some suggestions are given for changes to the strawman language.

1. Selection of Data

While estimates based on adequate human epidemiologic data are preferred over estimates based on animal data, many issues need to be considered so that these estimates will be scientifically sound. The following paragraphs address some of EPA's critical issues in choosing data for dose-response assessment.

Issue A. Estimating Risk from Nonpositive Studies

The group felt that when no positive evidence (either animal or human) is available, an upper bound risk estimate should not be made from a nonpositive human study. In the presence of a good positive animal study, however, it was felt that a human study could be used and that, under appropriate conditions, the upper bound from the human study could be used rather than the upper bound based on animal studies.

#### Issue B. Acceptable Quality of Data

With regard to the kind of epidemiologic data needed, the committee felt that the new strawman language proposed, which appears on page 1 (p. 39 of this document), is adequate: "more carefully executed dose-response assessments benefit from the availability of data that permit the ages of exposure and onset of disease and the level of exposure and duration of exposure to be incorporated in the assessment."

#### Issue C. Default Position for Dose-Rate Information

The committee felt that cumulative dose is an appropriate default position. It is assumed that the wording that now appears in the first full paragraph on page 91 of this document (Appendix E), applies to epidemiologic information. Clearly, the use of daily average or cumulative dose is not always ideal and dose rate information by time would be desirable. Extrapolation from occupational studies that deal with only part of a lifetime-to-lifetime risk may not always be appropriate. The committee felt that using lifetime daily averages will probably not grossly understate risk but in some cases might cause an overstatement.

#### Issue D. Adjustments for Children as Compared to Adults

It was felt that the assumption that there is no difference between children and adults is a default position. After taking dosimetry into account, other factors such as remaining lifetime, tissue growth, metabolism, food and fluid intake, etc., should be taken into consideration wherever possible.

#### Issue E. Options for Counting Person Years of Observation

The committee couldn't comment directly on the issue of whether person years of observation should be counted from the beginning or the end of exposure for dose response assessment. The committee did feel, however, that

latency should be considered in calculating dose for the purpose of examining dose-response relationships. This might take the form of lagging (5 years, 10 years, etc.) or of weighting dose by a time-to-tumor distribution with little weight given to times distant from some average value.

The workshop group suggested the following changes in the strawman language:

Section IIIA. Dose-Response Assessment, Paragraph 1. Selection of Data. On page 1 of the strawman text (pp. 39 and 89 of this document) delete "by the individuals doing the qualitative assessment (epidemiologists, toxicologists, pathologists, pharmacologists, etc.)."

Section IIIA. Dose-Response Assessment, Paragraph 2. Selection of Data: Human studies: On page 1 of the strawman text (pp. 39 and 89 of this document) add to the first line the word "positive" so that the first sentence reads, "Estimates based on adequate positive epidemiologic data are preferred over estimates based on animal data."

## 2. Choice of Mathematical Extrapolation Model

While mathematical models must be relied upon, since risks at low exposure levels cannot be measured directly, a range of choices exists regarding the type of model and its assumptions. The following paragraphs provide guidance based on the work group discussions for the critical issues identified by EPA in the strawman document.

### Issue A. Use of Statistical Upper Bound Dose-Response Estimates

The committee felt that when dose-response estimates are made from positive human data, statistical upper bounds should be used so as to be consistent with dose-response estimates from animal data. While it is true that there is no cross-species extrapolation with the human data and that exposures are closer to those actually experienced by humans in risk



assessment calculations, it was felt that this might be offset by the fact that the general population to which risk assessments apply may be more heterogeneous in terms of susceptibility than the data sets (often based on occupationally exposed groups) from which risk is estimated. Moreover, the committee was not certain that the true dose-response relationship for humans was always concave upward, and thus linear extrapolation may not always provide a margin of safety. In addition to upper bound estimates, the committee felt that point estimates as presently calculated by the EPA should be shown.

#### Issue B. Use of the Linearized Multistage Model

The committee discussed the appropriateness of the linearized multistage model as compared with simple linear models. It was felt that the question of appropriate models was a subject that might be better dealt with at a workshop where this was the main focus and in a context where other models could be presented and discussed.

#### Issue C. Modeling Nonlinear Dose Response

The committee felt that where dose is environmentally determined, there are agents (e.g., radiation, arsenic) where response is concave downward. Under these conditions an assumption of low-dose linearity would not be protective of the public. This was considered in the decision to recommend the calculation of statistical upper bounds from human data.

The workshop group suggested the following change in the strawman language:

Section IIIA. Dose-Response Assessment. Choice of Mathematical Extrapolation Model: Human data. On page 2 of the strawman text (p. 40 of this document) delete the following sentence: "a point estimate for the model may be used to estimate risk if dose is below the observable range."

APPENDIX A

EPA RISK ASSESSMENT FORUM TECHNICAL PANEL  
AND SUBCOMMITTEE ON EPIDEMIOLOGY

WORKSHOP ON CARCINOGEN RISK ASSESSMENT  
EPA RISK ASSESSMENT FORUM TECHNICAL PANEL AND ASSOCIATES

Richard Hill, William Farland, Co-Chairs

Margaret Chu  
Lorenz Rhomberg  
Jeanette Wiltse

Dorothy Patton, Chair, Risk Assessment Forum  
Cooper Rees, Science Coordinator, Risk Assessment Forum

Subcommittee on Epidemiology

David Bayliss  
Jerry Blondell  
Chao Chen  
Herman Gibb, Subcommittee Chair  
Doreen Hill  
Karen Hogan  
Aparna Koppikar  
Elizabeth Margosches  
Neal Nelson  
Cheryl Siegel-Scott

WORKSHOP PARTICIPANTS:

Study Design and Interpretation

Patricia Buffler  
Kenneth Cantor  
Marilyn Fingerhut, Chair  
Barry Friedlander  
William Halperin  
Doreen Hill  
Karen Hogan  
Barbara Hulka  
Renata Kimbrough  
Aparna Koppikar  
Genevieve Matanoski

Weight of Evidence

David Bayliss  
aaron Blair  
Jerry Blondell  
Margaret Chu  
Philip Cole  
Henry Falk  
Elizabeth Margosches  
Raymond Neutra, Chair  
Gerald Ott

**Dose Response**

Harvey Checkoway  
Chao Chen  
Kenneth Chu  
Kenneth Crump  
Philip Enterline, Chair  
William Farland  
Herman Gibb  
Daniel Krewski  
Neal Nelson  
Gerhard Raabe  
Lorenz Rhomberg  
Cherly Siegel Scott  
Allan Smith

**CONTRACTOR ASSOCIATES:**

Kate Schalk, Conference Services, Eastern Research Group  
Trisha Hasch, Conference Services, Eastern Research Group  
Elaine Krueger, Environmental Health Research, Eastern Research Group  
Norbert Page, Scientific Consultant, Eastern Research Group

**APPENDIX B**  
**LIST OF PARTICIPANTS**

U.S. Environmental Protection Agency

Cancer Risk Assessment Guidelines  
Human Evidence Workshop

June 26-27, 1989  
Washington, DC

FINAL LIST OF ATTENDEES

Mr. David Bayliss  
Office of Health and Environmental  
Assessment (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-5726

Dr. Aaron Blair  
National Cancer Institute  
Executive Plaza North, Room 418  
6130 Executive Blvd.  
Rockville, MD 20892  
(301) 496-9093

Mr. Jerry Blondell  
Hazard Evaluation Division (TS-769C)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 557-2564

Dr. Patricia Buffler  
University of Texas  
Health Science Center at Houston  
School of Public Health  
Epidemiology Research Unit  
P.O. Box 20186  
Houston, TX 77225  
(713) 792-7458

Dr. Kenneth Cantor  
National Cancer Institute  
Environmental Studies Section  
6130 Executive Blvd.  
Rockville, MD 20892  
(301) 496-1691

Dr. Harvey Checkoway  
Department of Environmental Health  
SC 34  
University of Washington  
Seattle, WA 98195  
(206) 543-4383

Dr. Chao Chen  
Office of Health and  
Environmental Assessment (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-5719

Dr. Kenneth Chu  
National Cancer Institute  
9000 Rockville Pike  
Bethesda, MD 20892  
301-496-8544

Dr. Margaret Chu  
Office of Health and Environmental  
Assessment (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-7335

Dr. Philip Cole  
School of Public Health  
University of Alabama  
203 TH  
UAB Station  
Birmingham, AL 35294  
(205) 934-6707

Dr. Kenneth Crump  
Clement Associates  
1201 Gaines Street  
Ruston, LA 71270  
(318) 255-4800

Dr. Philip Enterline  
University of Pittsburgh  
School of Public Health  
Room A410  
130 DeSoto Street  
Pittsburgh, PA 15261  
(412) 624-1559  
(412) 624-3032

Dr. Henry Falk  
Center for Disease Control  
EHHC/CEHC  
Mailstop F-28  
1600 Clifton Road, N.E.  
Atlanta, GA 30333  
(404) 488-4772

Dr. William Farland  
Office of Health and Environmental  
Assessment (RD-689)  
Office of Research and Development  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-7315

Dr. Marilyn Fingerhut  
National Institute of  
Occupational Health and Safety  
4676 Columbia Parkway (R-13)  
Cincinnati, OH 45226  
(513) 841-4203

Dr. Barry Friedlander  
Monsanto Company  
800 North Lindbergh -A3NA  
St. Louis, MO 63167  
(314) 694-1000

Dr. Herman Gibb  
Human Health Assessment Group  
(RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-5720

Dr. Bill Halperin  
51 Jackson Street  
Newton Center, MA 02159  
(617) 732-1260

Dr. Doreen Hill  
Analysis and Support Division  
(ANR-461)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 475-9640

Ms. Karen Hogan  
Exposure Evaluation Division  
(TS-798)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-3895

Dr. Barbara Hulka  
Department of Epidemiology  
Rosenau Hall  
CB 7400  
University of North Carolina,  
Chapel Hill, NC 27514  
(919) 966-5734

Dr. Peter Infante  
Health Standards Programs  
N3718  
OSHA/DOL  
200 Constitution Avenue, N.W.  
Washington, DC 20210  
(301) 523-7111

Dr. Renata Kimbrough  
Associate Administrator for Regional  
Operations (A101)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-4727

Dr. Aparna Koppikar  
Office of Health and  
Environmental Assessment (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 475-6765

Dr. Daniel Krewski  
Health and Welfare  
Environmental Health Center  
Room 117  
Ottawa, Ontario  
CANADA K1A 0L2  
(613) 954-0164

Dr. Elizabeth Margosches  
Exposure Evaluation Division  
(TS-798)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-3511

Dr. Genevieve Matanoski  
Johns Hopkins School of Hygiene  
and Public Health  
615 North Wolfe Street  
Baltimore, MD 21205  
(301) 955-8183  
(301) 955-3483 (main office)

Dr. Neal Nelson  
Analysis and Support Division  
(ANR-461)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 475-9640

Dr. Raymond Neutra  
California Department  
of Health Services  
2151 Berkeley Way  
Berkeley, CA 94704  
(415) 540-2669

Dr. Gerald Ott  
Arthur D. Little  
25 Acorn Park  
Cambridge, MA 02140  
(617) 864-5770 (ext. 3136)

Dr. Dorothy Patton  
Risk Assessment Forum (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 475-6743

Dr. Gerhard Raabe  
Mobil Corporation  
150 E. 142nd Street  
New York, NY 10017  
212-883-5368

Dr. David Cooper Rees  
Risk Assessment Forum (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 475-6743

Dr. Lorenz Rhomberg  
Office of Health and Environmental  
Protection (RD-689)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-5723

Ms. Cheryl Siegel Scott  
Exposure Evaluation Division  
(TS-798)  
U.S. Environmental Protection Agency  
401 M Street, S.W.  
Washington, DC 20460  
(202) 382-3511



Dr. Allan Smith  
University of California  
at Berkeley  
315 Warren Hall  
Berkeley, CA 94720  
(415) 642-1517 (office)  
(415) 843-1736 Health Risk  
Associates

**APPENDIX C**  
**LIST OF OBSERVERS**

U.S. Environmental Protection Agency

Cancer Risk Assessment Guidelines  
Human Evidence Workshop

June 26-27, 1989  
Washington, DC

LIST OF OBSERVERS

Steven Bayard  
U.S. EPA (RD-689)  
401 M Street, S.W.  
Washington, DC 20460  
202-382-5722

Judith Bellin  
383 'C' Street, S.W.  
Washington, DC 20024  
202-479-0664

Greg Beumel  
Combustion Engineering  
C.E. Environmental  
1400 16th Street, N.W., Suite 720  
Washington, DC 20036  
202-797-6407

Karen Creedon  
Chemical Manufacturers Assoc.  
2501 M Street, N.W.  
Washington, DC 20037  
202-881-1384

Maggie Dean  
Georgia-Pacific Corp.  
1875 Eye Street, Suite 775  
Washington, DC 20006  
202-659-3600

R.J. Dutton  
Risk Science Institute  
1126 16th Street N.W.  
Washington, DC 20036  
202-659-3306

Joel Fisher  
International Joint Commission  
2001 S. Street, N.W., Room 208  
Washington, DC 20440  
202-673-6222

Robert Gouph  
Toxic Material News  
951 Pershing  
Silver Spring, MD 20310  
301-587-6300

Stanley Gross  
U.S. EPA (H7509C)  
401 M Street, S.W.  
Washington, DC 20460  
202-557-4382

Cheryl Hogue  
Chemical Regulation Reporter  
1231 25th Street, N.W.  
Washington, DC 20037  
202-452-4584

Allan Katz  
Technical Assessment Systems  
1000 Potomac Street, N.W.  
Washington, DC 20007  
202-337-2625

Bob Ku  
Syntax Corporation  
3401 Hillview Avenue  
Palo Alto, CA 94301  
415-852-1981

Susan LeFevre  
Grocery Manufacturers of America  
1010 Wisconsin Avenue, N.W.  
Washington, DC 20007  
202-337-9400

Lisa Lefferts  
Center for Science in Public  
Interest  
1501 16th Street, N.W.  
Washington, DC 20036  
202-332-9110

George Lin  
Xerox Corporation  
Building 843-16S  
800 Salt Road  
Webster, NY 14580  
716-422-2081

Bertram Litt  
Litt Associates  
3612 Veasey Street, N.W.  
Washington, DC 20008  
202-686-0191

Donna Martin  
Putnam Environmental Services  
2525 Meridian Parkway  
P.O. Box 12763  
Research Triangle Park, NC 27709  
919-361-4657

Ray McAllister  
Madison Building, Suite 900  
1155 15th Street, N.W.  
Washington, DC 20005  
202-296-1585

Robert E. McGaughy  
U.S. EPA (RD-689)  
401 M Street, S.W.  
Washington, DC 20460  
202-382-5898

Mark Morrel  
Front Royal Group  
7900 W. Park Drive, Suite A 300  
McLean, VA 22102  
703-893-0900

Nancy Nickell  
Right-to-Know News  
1725 K Street, N.W., Suite 200  
Washington, DC 20006  
202-872-1766

Jacqueline Prater  
Beveridge & Diamond  
1350 Eye Street N.W. Suite 700  
Washington, DC 20005  
202-789-6113

Charles Ris  
U.S. EPA (RD 689)  
401 M Street, S.W.  
Washington, DC 22205  
202-382-5898

Robert Schnatter  
Exxon Biomedical Sciences  
Mettlers Road CN-2350  
East Millstone, NJ 08875  
201-873-6016

Melanie Scott  
Business Publishers, Inc.  
951 Pershing Drive  
Silver Spring, MD 20910  
301-587-6300

Sherry Selevan  
U.S. EPA (RD 689)  
401 M Street, S.W.  
Washington, DC 20460  
202-382-2604

Tomiko Shimada  
Shin Nippon Biomedical Laboratory  
P.O. Box 856  
Frederick, MD 21701  
301-662-1023

Betsy Shirley  
Styrene Information & Research  
Center  
1275 K Street, N.W.  
Washington, DC 20005  
202-371-5314

Arthur Stock  
Shea & Gardener  
1800 Massachusetts Ave., N.W.  
Washington, DC 20036  
202-828-2147

Jane Teta  
Union Carbide Corporation  
Health, Safety and Environmental  
Affairs  
39 Old Ridgebury Road  
Danbury, CT 06817-0001  
203-794-5884

Sandra Tirey  
Chemical Manufacturers Association  
2501 M Street, N.W.  
Washington, DC 20037  
202-887-1274

Keith Vanderveen  
U.S. EPA (TS 798)  
401 M Street, S.W.  
Washington, DC 20460  
202-382-6383

Frank Vincent  
James River Corp.  
P.O. Box 899  
Neenah, WI 54976  
414-729-8152

**APPENDIX D**

**INTRODUCTORY PLENARY SESSION**

- Opening Comments, Dr. Philip Enterline
- Public Interest Views, Dr. Raymond Neutra
- Private Sector Views, Dr. Gerald Ott

OPENING REMARKS: EPA CANCER GUIDELINES REVIEW WORKSHOP ON HUMAN EVIDENCE

Philip Enterline, Ph.D.  
Professor Emeritus of Biostatistics  
University of Pittsburgh  
School of Public Health

I was pleased to learn of EPA's decision to expand and clarify its guidelines for the use of human evidence in quantitative risk assessment. As perhaps many of you are aware, of the fairly large number of risk assessments that have thus far been made, only a handful are based upon human evidence. Most are based on extrapolations from animal experimental data.

In principal, there is no difference between epidemiologic evidence and experimental evidence. The problem lies in the design and analysis of these studies. When I first became interested in epidemiology, it was not considered to be a hard science and many of the "best" scientists were quite undecided as to how much faith to put on epidemiologic observations. One of the doubters was Bradford Hill, a then well-known British medical statistician, who suggested that perhaps epidemiology could be useful if in designing these studies "the experimental approach was kept firmly in mind." I think that is truly the key to good epidemiologic investigations. Somehow we must conduct epidemiologic studies so as to approach the conditions of an experiment as closely as possible.

We have made much progress here with a great boost from advances in statistical methodology and computers. Perhaps the major problem with epidemiologic studies as a tool in quantitative risk assessment is a lack of firm environmental data, although as Allan Smith has pointed out, it is difficult to imagine how the environmental data could be more in error than animal to human extrapolations. I also feel that producers of epidemiologic data need more guidance from consumers as to what kind of data is needed. Most of us are primarily concerned with answering the question, "Is there a disease excess?" rather than "What is the potency of the agent?"

It is my feeling that all well-designed epidemiologic studies involving defined exposures provide some information that can be useful in risk assessment. This is true even if positive findings are not statistically significant. For a number of years I taught a course in Introductory Biostatistics and in that course we covered measurements and tests of significance, with the latter being particularly difficult for many of my students. As part of my final examination, I sometimes ask the following question, "Suppose your grandmother has a cancer and your parents, wanting to take full advantage of your place in the medical science field, ask you to see what kind of treatment is currently in vogue. You search the literature and find two treatments that are being viewed favorably. Results from a large recent clinical trial show treatment A to give a 60 percent five-year survival and treatment B to give a 75 percent five-year survival. Numbers of subjects studied were about the same in each of the two treatment groups and the difference in survival rates is not statistically significant. Which treatment would you select for your grandmother?"

Perhaps not surprisingly most of my students conclude that since the difference in treatment was not statistically significant, there was no difference. If pressed they would simply toss a coin to decide on a treatment for their grandmother. There are, however, a few students who would notice that one of the treatments actually gave better results than the other.

Epidemiologic studies are not different from clinical trials. All contain some information. Some studies are more positive or some more negative than others and this fact alone may be important. A relative risk of 1.2, even if not statistically significant, may mean more in a particular setting than a relative risk of .9. Perhaps the former might be called nonpositive and the latter called negative. EPA clearly recognizes the usefulness of such [nonpositive] epidemiologic data when they evaluate animal data. A very typical situation is one in which there is a positive animal study and a nonpositive or negative human study. While EPA might dismiss the human study because of a belief that there is no such thing as negative epidemiology, they do use the upper confidence interval of the human data to set an upper limit



of risk calculated from the animal study. I think that is a fair way to view human evidence, since the confidence interval is both a function of the power of the study, that is, the sample size, and of the actual results of the study. Incidentally, I don't think it is proper to calculate power after the study has been completed, since it ignores what was found in the study and the situation is clearly different than it was before the study was ever undertaken.

Some people seem to feel that the doses of toxic agents received by humans are too small to cause disease excesses large enough to be detectable by epidemiologic studies. In fact, my students often comment to me that it must have been great in the good old days when there were so many things to discover. I would point out here, however, that for the most part, in the "good old days," there was little to guide us in terms of what to look for or where to look. I recall when the first U.S. epidemiologic study of cigarette smoking and cancer was reported in the early 1950s, there was a great deal of debate as to whether this could be in fact true. Why did it take us so long to find such a grand relationship? Even Bill Hueper of NIH, who was probably our greatest prophet as to environmental causes of cancer, had missed this relationship, attributing only cancer of the tongue and cheek to the use of tobacco. Of course, it was the human evidence that led to what is now perhaps our greatest effort in the field of preventive medicine - the anti-smoking campaign.

I feel that there are many discoveries yet to be made from epidemiologic studies. Some of these involve simply a careful review of existing literature, while others will require some new investigations guided perhaps by observations made in animal experiments as well as the new field of structural activity research (SAR). In studies of working populations, we really need to take a hard look at studies that show large numbers of statistically significant deficits in various diseases. Can these all be attributed to worker selection or is it possible that something in the design or execution of these types of studies is systematically causing understated risks?

In closing, let me assure you based on a couple of years experience as a member of EPA's Science Advisory Board, that EPA is one federal agency that listens to its consultants. Your work during the next day and a half could have an important impact on the quality of EPA's risk assessment activity in the future.

EPIDEMIOLOGICAL RISK ASSESSMENT:  
SOME OBSERVATIONS FROM THE PUBLIC SECTOR

Raymond Richard Neutra, M.D., Dr.P.H.  
Chief - Epidemiological Studies Section  
California Department of Health Services

I. Should We Be Regulating Substance by Substance?

It should be noted that there are 60,000 chemicals in commercial use and that our regulatory scheme has been to regulate them one by one after years of scientific debate. This is analogous to regulating fecal pathogens one by one instead of simply separating people from chemicals the way we have separated them from feces. This is worth pondering before we plunge into the difficulties which this general approach presents us.

II. Epidemiologically Non-Detectable Risks May Be of Societal Concern.

Figure 1 gives a schematized example of diseases according to their baseline, lifetime, cumulative rates, and the relative risks conveyed by the hypothetical carcinogens in each case. Common cancers, whose baseline risks are multiplied many times by a carcinogen, are easy to detect and are of societal concern. Rare cancers affected by carcinogens that convey small relative risks are neither important nor detectable. But what of moderately rare cancers exposed to agents that convey relative risks less than two? They may convey lifetime risks greater than one in a million or one in a hundred thousand, and not be detectable with epidemiology. Saccharin is an example of a widely used agent for whom the animal risk assessment suggested an added burden of 800 bladder cancers a year. Yet this was a relative risk of only 1.01! Even the enormous case control studies that were done could not rule out this added risk. Epidemiologists said this study showed there was no risk of public health concern, but eight hundred cases a year (if identifiable) would attract public and legal attention (compare it to the number of Guillaine-Barre (GB) cases in swine flu vaccine, or the number of rabies cases a year). The public is not calmed by the fact that only a small percentage of

*Rate Ratio Conveyed by Toxicant*

|                     | Small  | Large                               |
|---------------------|--|-------------------------------------|
| $\leq 10^{-6}$      | Not Important<br>Socially<br>Not<br>Detectable | Socially<br>Important<br>Detectable |
| $10^{-5} - 10^{-4}$ | Socially<br>Important<br>Not<br>Detectable     | Socially<br>Important<br>Detectable |
| $\geq 10^{-3}$      | Socially<br>Important<br>Detectable            | Socially<br>Important<br>Detectable |

*Lifetime Risk in Unexposed*

Most EPA environmental risk in this cell

Figure 1. Possible Environmental Risk Scenarios.

all GB cases were attributable to the vaccine nor would they be calmed by a similar claim for diet drinks. The null Hoover study was useful in ruling out some of the outlier risk assessments, and in reassuring us that humans are not dramatically more sensitive than animals. Ultimately, mechanistic evidence may lay the issue to rest. The point here is that just because an epidemiologist can't see it doesn't mean it is unimportant. Hence, epidemiology will rarely, if ever - by itself, be able to give a clean bill of health during the Hazard Identification phase of risk assessment.

### III. Keep the Four Kinds of Evaluation Conceptually Separate.

One evaluates individual studies for Hazard Identification (Clear Evidence, Some Evidence, Equivocal Evidence, No Evidence).

One weighs a body of evidence for Hazard Identification (Sufficient, Limited, Inconclusive, Evidence Not Suggestive of Carcinogenicity, Sufficient Evidence for Noncarcinogenicity).

One evaluates individual studies for their usefulness in Dose-Response Assessment (Very Useful, Somewhat Useful, Not Useful)

One combines useful studies for the purposes of dose-response assessment. (This has no nomenclature since a summary number comes out of the combined dose-response assessment.)

The strawman document was sometimes unclear about the distinction between these various activities.

IV. Systematically Anticipate What to Do When Dose-Response Assessment Is Based on Conflicting Human and Animal Data

California health department staff have found unforeseen scenarios in which animal and human dose-response assessment may or may not agree. Figure 2 shows a simplified flow diagram spelling out the possible combinations. This diagram could serve as a guide to generate "scenario queries" to the participants, e.g., "How would you handle this one?"

What do you do if you have "sufficient animal evidence" but human evidence is "inconclusive" because the only human studies gave null results? According to this flow diagram one would choose the "best" null study to see if the upper confidence level risk was lower than the risk predicted from animals. If so, (as was the case with cadmium) the human upper confidence level risk would be used.

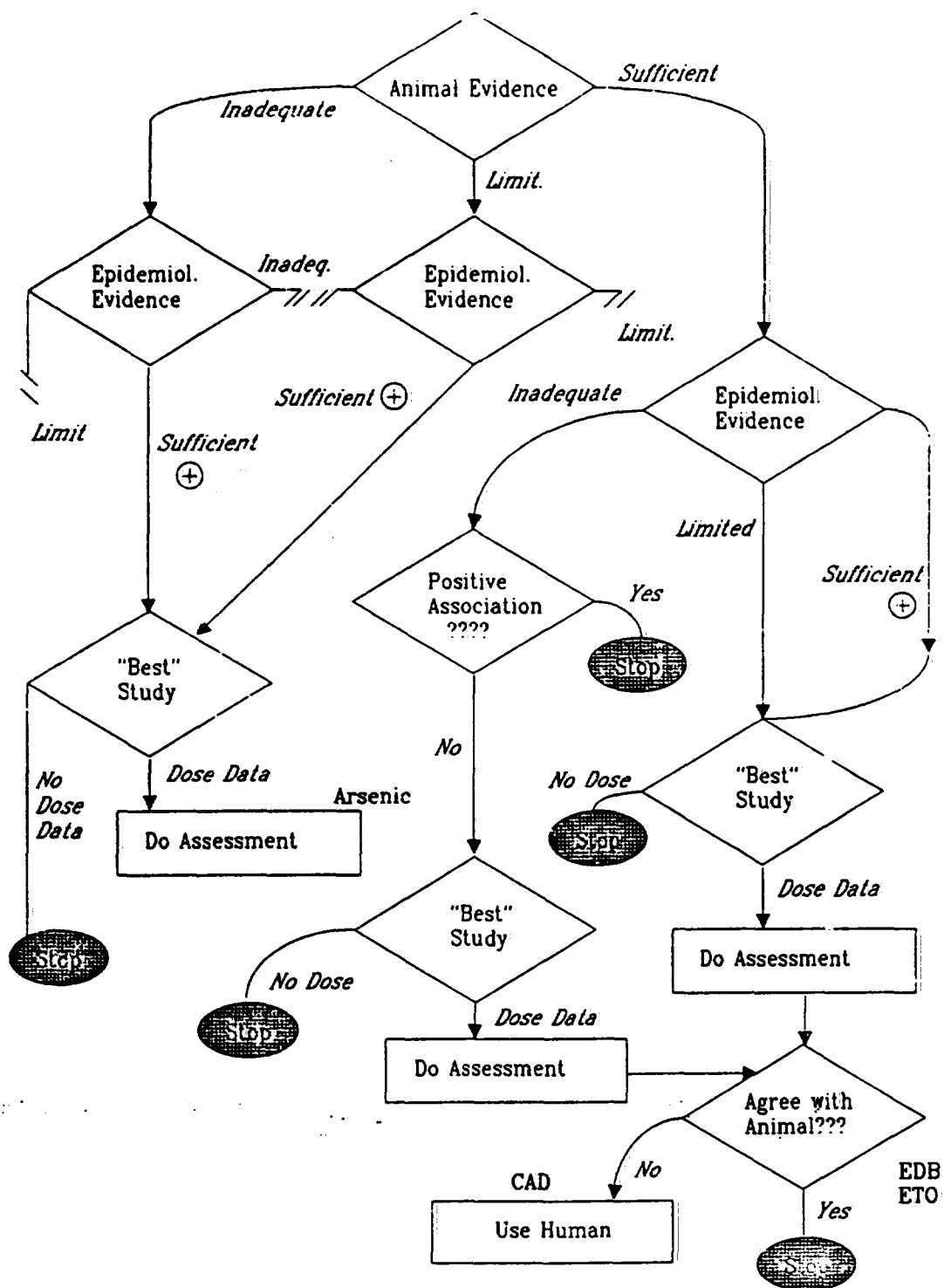


Figure 2. Flow Diagram of Possible Combinations of Associations between Human and Animal Data.

## USE OF HUMAN EVIDENCE IN RISK ASSESSMENT - A PRIVATE SECTOR PERSPECTIVE

Gerald Ott, Ph.D.  
Senior Consultant, Epidemiology  
Arthur D. Little, Inc.

### Background

The views which I present today undoubtedly reflect my experience as an occupational and environmental epidemiologist working in the private sector; however, they are my own views and not necessarily those of any specific organization.

Our standard of living in the United States has been achieved through the individual and collective efforts of people to convert resources, some renewable and others not, into useful products. This production of goods and services almost inevitably leads to the generation of waste byproducts that may be released to the environment. Wastes are any materials that are deemed to be of no discernable value and to have no utility to individuals, institutions, or society in general. Because of the costs of recovery, the unintended release of valued products to the environment may also render these products classifiable as waste.

In an increasingly congested world, there is ample reason to be concerned about the release of hazardous materials to the work and general environments. Human health and environmental quality have been adversely affected by hazardous wastes in the past when both the production of goods was at a lower level and people were not residing in such close proximity to one another. To minimize adverse impacts on human health and the environment, waste control problems must be recognized and addressed using all of the scientific knowledge and appropriate resources available to us.

The impact of wastes may be controlled through:

- Decreased production of goods and services.



- Waste minimization (e.g., continuous production in enclosed systems versus batch production in open systems).
- Recycling (extracting greater value from materials otherwise viewed as useless).
- Confinement in on-site and off-site disposal areas.
- Incineration (with subsequent dispersion and/or confinement of residual materials).
- Intentional dilution or dispersion in various environmental media.

None of these approaches to waste management is free of risks. With each approach, there may be impacted populations that do not share proportionately in the costs and benefits of the enterprises producing the products.

In Table 1, various waste control approaches are listed together with the populations that may be impacted. These include employee populations, community populations, ecologic populations, and global populations. By ecologic populations, I mean the interacting biologic species that exist within an impacted ecosystem. Clearly, there are tradeoffs in control strategies that could differentially impact the various populations. For example, venting used to reduce the likelihood of employee exposure may subject the community population to greater exposure opportunities.

The United States Congress and various governmental agencies have recognized the need to assure that employees and communities are informed of the potential risks and are afforded an opportunity to participate in risk management decisions related to hazardous wastes. This recognition has been reflected in recent employee and community right-to-know laws and regulations and in regulations related to the siting of hazardous waste facilities. Attendant with right-to-know is an obligation to inform people of both the potential health effects of substances to which they may be exposed and the risks projected to result from those exposures. Quantitative risk assessment has become an important tool for informing persons about the projected risks associated with hazard control decisions.

### Epidemiology and Quantitative Risk Assessment

The epidemiologic approach to assessing health risks of environmental factors relies on inductive reasoning, that is, reasoning from a particular set of facts to general principles. Consequently, epidemiology is data-driven. The epidemiologic approach requires (1) a characterization of both exposures and health outcomes in the selected human population of interest, and (2) analyses of the relationships between exposures and the health outcomes in that population. There are, of course, both strengths and weaknesses in the epidemiologic approach. Among the strengths is its direct relevance to the subject at hand, namely, determining the effects of exposures on human health.

Major limitations are:

- The size of the population available for study may be too small to allow detection of low level but important health risks.
- The observation period may not have been sufficiently long; for chronic health effects to have occurred.
- The study design may not address all alternative explanations for the observed health findings.
- The occurrence of real adverse health effects can only be demonstrated after the fact.

This latter limitation suggests that the toxicity endpoints evaluated should emphasize early indicators of reversible adverse effects. An additional, frequently cited, limitation is the lack of quantitative exposure assessments in support of epidemiologic studies. However, with more extensive use of modeling techniques to estimate exposures and with increasingly precise measurement procedures, it may be possible to minimize the practical importance of this limitation in both occupational and environmental settings.

Quantitative risk assessment has emerged as the major scientific tool for deductively determining the likelihood that harm will come to people as a consequence of predictable exposure to hazardous substances. The four steps

of the quantitative risk assessment process are hazard identification, toxicity or dose-response assessment, exposure assessment, and risk characterization. Since quantitative risk assessment utilizes external toxicity information to define a hazard profile for the environmental agents of concern, it is appropriate to include the results of epidemiologic research as well as animal bioassays and other toxicity tests in identifying specific hazards (e.g., establishing the carcinogenicity of a particular substance), in describing exposures and identifying sensitive populations, and in assessing the dose-response relationship.

There are several notable strengths of the quantitative risk assessment approach. First, a quantitative risk assessment establishes what effects could take place in the absence of intervention measures. Thus, there may be opportunities to initiate corrective actions before injury to health has occurred. Secondly, the quantitative risk assessment approach is "highly risk sensitive". This stems from the use of models to predict risks that are far below the risk levels that could be detected in a health study of the subject population. Through the use of across species and low dose extrapolations, acceptable exposure concentrations can be calculated which would yield virtually safe doses provided the assumptions of the risk assessment are valid.

For the remainder of this presentation, I would like to discuss the role of epidemiologic evidence in several specific aspects of the quantitative risk assessment process. These are (1) the selection of relevant epidemiologic studies to be included in the evaluation of risks, (2) the methods by which evidence for or against a particular effect is combined across studies, and (3) the use of epidemiologic evidence as a final check of the dose-response assessment.

### The Selection of Relevant Epidemiologic Studies

Evaluating the consequences of exposure to an agent requires a critical review of the available toxicologic and epidemiologic data for that substance and other interrelated substances. The purpose of the review is to determine appropriate toxicity endpoints and, in particular, to determine the evidence for and against carcinogenicity. In evaluating the available epidemiologic data, two important decisions need to be made. The first decision is whether or not a particular study is relevant (admissible) to the evaluation process. The second decision relates to how the evidence is to be combined across the relevant epidemiologic and toxicologic studies to assess the overall evidence for human carcinogenicity.

In addressing the first decision, it is necessary to identify and characterize each candidate study on the basis of both relevance and methodological strengths. Studies under consideration may range from case reports to cohort studies which specifically examine the exposure of interest. To be admissible, each study should address a relevant biologic outcome, there should be a reasonable basis for ascribing exposure to the agent of interest, and the research should be methodologically sound within the context of its intended purpose.

An assessment of the internal evidence for or against a causal relationship should not be part of the admissibility criteria. In other words, studies should not be selected based on their outcomes or conclusions. From a methodologic viewpoint, the guidelines for evaluating a study should be consistent with "good laboratory practices" and with guidelines developed by the National Academy of Sciences and other professional organizations for judging the quality of epidemiologic research.

Based on these guidelines and relevancy criteria, studies can be classified as (1) not relevant, (2) relevant but methodologically unsound, or (3) admissible by virtue of both relevancy and soundness of methodology. The decision that a study has utilized sound methodology strengthens the basis of

its admissibility. However, studies that have marginal or even important methodologic deficiencies should not be excluded at this point except where other clearly superior studies are available.

While it is essential that peer review takes place, the available studies should not be restricted to those appearing only in the peer-reviewed literature. This is important for two reasons. First, highly relevant and admissible studies may otherwise be excluded from consideration while awaiting publication. This would seem a harsh penalty to exact because of time constraints. Secondly, there are indications that publication bias may result in a shift of the published literature toward positive findings, thus making it difficult to combine evidence across studies without aggregating the bias component. The effort to review critically those few relevant studies not already published would appear to be effort well spent.

#### Combining Evidence Across Studies

A variety of approaches have been proposed to assist the risk assessor in combining evidence across studies. These include expert or judgment-based approaches, categorical accept-reject analysis, classical statistical approaches, and meta-analysis. These approaches to decision-making are conceptually similar in assigning weights to each component study and combining the weighted evidence in some fashion to arrive at an overall judgment regarding causality. They may differ considerably in the methods for determining how much weight to assign to each study and how explicit to be in assigning weights. The system used by the International Agency for Research on Cancer in evaluating the weight of evidence for human carcinogenicity is well known and relies primarily on expert judgment to combine evidence across studies.

Methods for aggregating biological evidence may differ fundamentally from those used to combine statistical evidence. For example, the statistical power to detect a difference can be increased by combining results across

comparably conducted parallel studies and yield a different conclusion than any of the studies viewed in isolation. It is certainly plausible that two studies, neither of which provides statistical evidence of an effect alone, could demonstrate statistical significance when combined. Biologic support for a hypothesis may come from dissimilar studies that demonstrate a connection between different aspects of a particular disease process. Thus, a study demonstrating that methylene chloride is metabolized to carbon monoxide in the liver and a separate study on the cardiovascular effects of carbon monoxide may convincingly link methylene chloride exposure to adverse cardiovascular effects.

While statistical attributes of a study may be judged on the basis of that study alone, biological attributes include elements that are frequently external to the study. This is evidenced by the common organizational structure of most research reports. When reporting findings, researchers typically provide a summary of their own study results followed by an interpretation of their results in the light of existing studies and accepted biologic knowledge.

This suggests that evidence for a causal relationship is primarily aggregated on a biologic plane. The biologic evidence may be combined either in serial or parallel fashion, whereas statistical evidence is only aggregated in parallel. Various explanations may be invoked to describe a viable argument for causation, with the most direct argument consistent with biologic knowledge and observation and requiring the fewest assumptions being accorded the highest status. Statistical evidence secures our confidence in the validity of component statements within the various arguments but does not make the case for causality in and of itself. Broad statements that arise from ecologic or correlational studies generally provide weak arguments since they typically require many assumptions about component statements in the argument.

Evaluating the strengths of various causal arguments may represent an alternative approach to performing weight-of-evidence determinations for human

carcinogenicity. In other words, competing arguments would be put forward and then evaluated to determine the most plausible arguments. The plausible arguments themselves would then be subjected to a weight-of-evidence determination. One potential advantage of this approach is that future research may be more readily directed to address the weakest links in existing causality arguments.

#### A Final Check of the Dose-Response Assessment

In addition to their other contributions, epidemiologic studies may also serve as an overall validation check on the final proposed dose-response model for a given substance. The field of observation potentially open to epidemiologic assessment includes chemical process employees in direct contact with the substance of interest, employees whose assignments involve intermittent direct contact with the substance, employees assigned to the same production site, but whose assignments result in indirect exposure to the substance, members of communities surrounding the production sites and perhaps customers who purchase the substance for subsequent use or who purchase products contaminated by the substance. Quantitative exposure estimates would be required to carry out this exercise; however, continued improvements in exposure modeling suggest that it is feasible to develop appropriate exposure estimates.

Presumably, the most sensitive validation test would be one which compares the total projected excess of cases throughout the period of observation and across all exposed populations with the observed excess of cases developed through epidemiologic followup of the impacted populations. An additional consistency test might be one which examines only persons in the upper portions of the exposure distribution with sufficient latency to allow for manifestation of a carcinogenic response. From the standpoint of determining that a dose-response model overestimates the risks, this exercise is only viable for substances that were in commercial use prior to the mid-1950s. For evaluating the possibility that risks have been underestimated by the model, the approach has merit under a broader range of circumstances.

TABLE I. HAZARDOUS WASTE CONTROL APPROACHES AND IMPACTED POPULATIONS

| CONTROL OF HAZARDOUS WASTES                | IMPACTED POPULATIONS   |
|--|--|
| <hr/>                                      |  |
| <b>ON-SITE CONTROLS</b>                    |  |
| Waste Minimization                         | <u>Employee Populations</u>  |
| Recycling of Wastes                        | Informed/Uninformed<br>Voluntary/Involuntary   |
| Effective on-site Confinement              |  |
| <b>OFF-SITE CONTROLS</b>                   |  |
| Decreased Production of Goods and Services | <u>Employee Populations</u><br><u>Community Populations</u><br><u>Ecologic Populations</u> |
| Effective off-site Confinement             | <u>Global Populations</u>  |
| Ineffective Confinement                    | Informed/Uninformed<br>Voluntary/Involuntary   |
| Recycling of Wastes (off-site)             |  |
| Intentional Dilution                       |  |
| Intentional Dispersion                     |  |



**APPENDIX E**  
**1986 GUIDELINES FOR CARCINOGEN RISK ASSESSMENT**

# **GUIDELINES FOR CARCINOGEN RISK ASSESSMENT**

**SUMMARY:** On September 24, 1986, the U.S. Environmental Protection Agency issued the following five guidelines for assessing the health risks of environmental pollutants.

Guidelines for Carcinogen Risk Assessment

Guidelines for Estimating Exposures

Guidelines for Mutagenicity Risk Assessment

Guidelines for the Health Assessment of Suspect Developmental Toxicants

Guidelines for the Health Risk Assessment of Chemical Mixtures

This section contains the Guidelines for Carcinogen Risk Assessment.

The Guidelines for Carcinogen Risk Assessment (hereafter "Guidelines") are intended to guide Agency evaluation of suspect carcinogens in line with the policies and procedures established in the statutes administered by the EPA. These Guidelines were developed as part of an interoffice guidelines development program under the auspices of the Office of Health and Environmental Assessment (OHEA) in the Agency's Office of Research and Development. They reflect Agency consideration of public and Science Advisory Board (SAB) comments on the Proposed Guidelines for Carcinogen Risk Assessment published November 23, 1984 (49 FR 46294).

This publication completes the first round of risk assessment guidelines development. These Guidelines will be revised, and new guidelines will be developed, as appropriate.

## **FOR FURTHER INFORMATION CONTACT:**

Dr. Robert E. McGaughy  
Carcinogen Assessment Group  
Office of Health and Environmental Assessment  
(RD-689)  
401 M Street, S.W.  
Washington, DC 20460  
202-382-5898

**SUPPLEMENTARY INFORMATION:** In 1983, the National Academy of Sciences (NAS) published its book entitled *Risk Assessment in the Federal Government: Managing the Process*. In that book, the NAS recommended that Federal regulatory agencies establish "inference guidelines" to ensure

consistency and technical quality in risk assessments and to ensure that the risk assessment process was maintained as a scientific effort separate from risk management. A task force within EPA accepted that recommendation and requested that Agency scientists begin to develop such guidelines.

## **General**

The guidelines are products of a two-year Agencywide effort, which has included many scientists from the larger scientific community. These guidelines set forth principles and procedures to guide EPA scientists in the conduct of Agency risk assessments, and to inform Agency decision makers and the public about these procedures. In particular, the guidelines emphasize that risk assessments will be conducted on a case-by-case basis, giving full consideration to all relevant scientific information. This case-by-case approach means that Agency experts review the scientific information on each agent and use the most scientifically appropriate interpretation to assess risk. The guidelines also stress that this information will be fully presented in Agency risk assessment documents, and that Agency scientists will identify the strengths and weaknesses of each assessment by describing uncertainties, assumptions, and limitations, as well as the scientific basis and rationale for each assessment.

Finally, the guidelines are formulated in part to bridge gaps in risk assessment methodology and data. By identifying these gaps and the importance of the missing information to the risk assessment process, EPA wishes to encourage research and analysis that will lead to new risk assessment methods and data.

## **Guidelines for Carcinogen Risk Assessment**

Work on the Guidelines for Carcinogen Risk Assessment began in January 1984. Draft guidelines were developed by Agency work groups composed of expert scientists from throughout the Agency. The drafts were peer-reviewed by expert scientists in the field of carcinogenesis from universities, environmental groups, industry, labor, and other governmental agencies. They were then proposed for public comment in the *FEDERAL REGISTER* (49 FR 46294). On November 9, 1984, the Administrator directed that Agency offices use the proposed guidelines in performing risk assessments until final guidelines become available.

After the close of the public comment period, Agency staff prepared summaries of the comments, analyses of the major issues presented by the commentors, and proposed changes in the language of the guidelines to deal with the issues raised. These analyses were presented to review panels of the SAB on March 4 and April 22-23, 1985, and to the Executive Committee of the SAB on April 25-26, 1985. The SAB meetings were announced in the *FEDERAL REGISTER* as follows: February 12, 1985 (50 FR 5811) and April 4, 1985 (50 FR 13420 and 13421).

In a letter to the Administrator dated June 19, 1985, the Executive Committee generally concurred on all five of the guidelines, but recommended certain revisions, and requested that any revised guidelines be submitted to the appropriate SAB review panel chairman for review and concurrence on behalf of the Executive Committee. As described in the responses to comments (see Part B: Response to the Public and Science Advisory Board Comments), each guidelines document was revised, where appropriate, consistent with the SAB recommendations, and revised draft guidelines were submitted to the panel chairmen. Revised draft Guidelines for Carcinogen Risk Assessment were concurred on in a letter dated February 7, 1986. Copies of the letters are available at the Public Information Reference Unit, EPA Headquarters Library, as indicated elsewhere in this section.

Following this Preamble are two parts: Part A contains the Guidelines and Part B, the Response to the Public and Science Advisory Board Comments (a summary of the major public comments, SAB comments, and Agency responses to those comments).

The Agency is continuing to study the risk assessment issues raised in the guidelines and will revise these Guidelines in line with new information as appropriate.

References, supporting documents, and comments received on the proposed guidelines, as well as copies of the final guidelines, are available for inspection and copying at the Public Information Reference Unit (202-382-5926), EPA Headquarters Library, 401 M Street, S.W., Washington, DC, between the hours of 8:00 a.m. and 4:30 p.m.

I certify that these Guidelines are not major rules as defined by Executive Order 12291, because they are nonbinding policy statements and have no direct effect on the regulated community. Therefore, they will have no effect on costs or prices, and they will

[51 FR 33993]

have no other significant adverse effects on the economy. These Guidelines were reviewed by the Office of

Management and Budget under Executive Order 12291.

August 22, 1986

Lee M. Thorpe,

Administrator

## CONTENTS

### Part A: Guidelines for Carcinogen Risk Assessment

#### I. Introduction

#### II. Hazard Identification

- A. Overview
- B. Elements of Hazard Identification
  - 1. Physical-Chemical Properties and Routes and Patterns of Exposure
  - 2. Structure-Activity Relationships
  - 3. Metabolic and Pharmacokinetic Properties
  - 4. Toxicologic Effects
  - 5. Short-Term Tests
  - 6. Long-Term Animal Studies
  - 7. Human Studies
- C. Weight of Evidence
- D. Guidance for Dose-Response Assessment
- E. Summary and Conclusion

#### III. Dose-Response Assessment, Exposure Assessment, and Risk Characterization

- A. Dose-Response Assessment
  - 1. Selection of Data
  - 2. Choice of Mathematical Extrapolation Model
  - 3. Equivalent Exposure Units Among Species
- B. Exposure Assessment
- C. Risk Characterization
  - 1. Options for Numerical Risk Estimates
  - 2. Concurrent Exposure
  - 3. Summary of Risk Characterization

#### IV. EPA Classification System for Categorizing Weight of Evidence for Carcinogenicity from Human and Animal Studies (Adapted from IARC)

- A. Assessment of Weight of Evidence for Carcinogenicity from Studies in Humans
- B. Assessment of Weight of Evidence for Carcinogenicity from Studies in Experimental Animals
- C. Categorization of Overall Weight of Evidence for Human Carcinogenicity

#### V. References

### Part B: Response to Public and Science Advisory Board Comments

#### I. Introduction

#### II. Office of Science and Technology Policy Report on Chemical Carcinogens

#### III. Inference Guidelines

#### IV. Evaluation of Benign Tumors

#### V. Transplacental and Multigenerational Animal Bioassays

#### VI. Maximum Tolerated Dose

#### VII. Mouse Liver Tumors

#### VIII. Weight-of-Evidence Categories

#### XI. Quantitative Estimates of Risk

## Part A: Guidelines for Carcinogen Risk Assessment

### I. Introduction

This is the first revision of the 1976 Interim Procedures and Guidelines for Health Risk Assessments of Suspected Carcinogens (U.S. EPA, 1976; Albert et al., 1977). The impetus for this revision is the need to incorporate into these Guidelines the concepts and approaches to carcinogen risk assessment that have been developed during the last ten years. The purpose of these Guidelines is to promote quality and consistency of carcinogen risk assessments within the EPA and to inform those outside the EPA about its approach to carcinogen risk assessment. These Guidelines emphasize the broad but essential aspects of risk assessment that are needed by experts in the various disciplines required (e.g., toxicology, pathology, pharmacology, and statistics) for carcinogen risk assessment. Guidance is given in general terms since the science of carcinogenesis is in a state of rapid advancement, and overly specific approaches may rapidly become obsolete.

These Guidelines describe the general framework to be followed in developing an analysis of carcinogenic risk and some salient principles to be used in evaluating the quality of data and in formulating judgments concerning the nature and magnitude of the cancer hazard from suspect carcinogens. It is the intent of these Guidelines to permit sufficient flexibility to accommodate new knowledge and new assessment methods as they emerge. It is also recognized that there is a need for new methodology that has not been addressed in this document in a number of areas, e.g., the characterization of uncertainty. As this knowledge and assessment methodology are developed, these Guidelines will be revised whenever appropriate.

A summary of the current state of knowledge in the field of carcinogenesis and a statement of broad scientific principles of carcinogen risk assessment, which was developed by the Office of Science and Technology Policy (OSTP, 1985), forms an important basis for these Guidelines; the format of these Guidelines is similar to that proposed by the National Research Council (NRC) of the National Academy of Sciences in a book entitled *Risk Assessment in the Federal Government: Managing the Process* (NRC, 1983).

These Guidelines are to be used within the policy framework already provided by applicable EPA statutes and do not alter such policies. These Guidelines provide general directions for analyzing and organizing available data. They do not imply that one kind of data or another is prerequisite for regulatory action to control, prohibit, or allow the use of a carcinogen.

Regulatory decision making involves two components: risk assessment and risk management. Risk assessment defines the adverse health consequences of exposure to toxic agents. The risk assessments will be carried out independently from considerations of the consequences of regulatory action. Risk management combines the risk assessment with the directives of regulatory legislation, together with socioeconomic, technical, political, and other considerations, to reach a decision as to whether or how much to control future exposure to the suspected toxic agents.

Risk assessment includes one or more of the following components: hazard identification, dose-response assessment, exposure assessment, and risk characterization (NRC, 1983).

Hazard identification is a qualitative risk assessment, dealing with the process of determining whether exposure to an agent has the potential to increase the incidence of cancer. For purposes of these Guidelines, both malignant and benign tumors are used in the evaluation of the carcinogenic hazard. The hazard identification component qualitatively answers the question of how likely an agent is to be a human carcinogen.

Traditionally, quantitative risk assessment has been used as an inclusive term to describe all or parts of dose-response assessment, exposure assessment, and risk characterization. Quantitative risk assessment can be a useful general term in some circumstances, but the more explicit terminology developed by the NRC (1983) is usually preferred. The dose-response assessment defines the relationship between the dose of an agent and the probability of induction of a carcinogenic effect. This component usually entails an extrapolation from the generally high doses administered to experimental animals or exposures noted in epidemiologic studies to the exposure levels expected from human contact with the agent in the environment; it also includes considerations of the validity of these extrapolations.

The exposure assessment identifies populations exposed to the agent, describes their composition and size, and presents the types, magnitudes, frequencies, and durations of exposure to the agent.

[51 FR 33994]

In risk characterization, the results of the exposure assessment and the dose-response assessment are combined to estimate quantitatively the carcinogenic risk. As part of risk characterization, a summary of the strengths and weaknesses in the hazard identification, dose-response assessment, exposure assessment, and the public health risk estimates are presented. Major assumptions, scientific judgments, and, to the extent possible, estimates of the uncertainties embodied in the assessment are also presented, distinguishing clearly between fact, assumption, and science policy.

The National Research Council (NRC, 1983) pointed out that there are many questions encountered in the risk assessment process that are unanswerable given current scientific knowledge. To bridge the uncertainty that exists in these areas where there is no scientific consensus, inferences must be made to ensure that progress continues in the assessment process. The OSTP (1985) reaffirmed this position, and generally left to the regulatory agencies the job of articulating these inferences. Accordingly, the Guidelines incorporate judgmental positions (science policies) based on evaluation of the presently available information and on the regulatory mission of the Agency. The Guidelines are consistent with the principles developed by the OSTP (1985), although in many instances are necessarily more specific.

## II. Hazard Identification

### A. Overview

The qualitative assessment or hazard identification part of risk assessment contains a review of the relevant biological and chemical information bearing on whether or not an agent may pose a carcinogenic hazard. Since chemical agents seldom occur in a pure state and are often transformed in the body, the review should include available information on contaminants, degradation products, and metabolites.

Studies are evaluated according to sound biological and statistical considerations and procedures. These have been described in several publications (Interagency Regulatory Liaison Group, 1979; OSTP, 1985; Peto et al., 1980; Mantel, 1980; Mantel and Haenszel, 1959; Interdisciplinary Panel on Carcinogenicity, 1984; National Center for Toxicological Research, 1981; National Toxicology Program, 1984; U.S. EPA, 1983a, 1983b, 1983c; Haseman, 1984). Results and conclusions concerning the agent, derived from different types of information, whether indicating positive or negative responses, are melded together into a weight-of-evidence determination. The strength of the evidence supporting a potential human carcinogenicity judgment is developed in a weight-of-evidence stratification scheme.

### B. Elements of Hazard Identification

Hazard identification should include a review of the following information to the extent that it is available.

1. *Physical-Chemical Properties and Routes and Patterns of Exposure.* Parameters relevant to carcinogenesis, including physical state, physical-chemical properties, and exposure pathways in the environment should be described where possible.

2. *Structure-Activity Relationships.* This section should summarize relevant structure-activity

correlations that support or argue against the prediction of potential carcinogenicity.

3. *Metabolic and Pharmacokinetic Properties.* This section should summarize relevant metabolic information. Information such as whether the agent is direct-acting or requires conversion to a reactive carcinogenic (e.g., an electrophilic) species, metabolic pathways for such conversions, macromolecular interactions, and fate (e.g., transport, storage, and excretion), as well as species differences, should be discussed and critically evaluated. Pharmacokinetic properties determine the biologically effective dose and may be relevant to hazard identification and other components of risk assessment.

4. *Toxicologic Effects.* Toxicologic effects other than carcinogenicity (e.g., suppression of the immune system, endocrine disturbances, organ damage) that are relevant to the evaluation of carcinogenicity should be summarized. Interactions with other chemicals or agents and with lifestyle factors should be discussed. Prechronic and chronic toxicity evaluations, as well as other test results, may yield information on target organ effects, pathophysiological reactions, and preneoplastic lesions that bear on the evaluation of carcinogenicity. Dose-response and time-to-response analyses of these reactions may also be helpful.

5. *Short-Term Tests.* Tests for point mutations, numerical and structural chromosome aberrations, DNA damage/repair, and *in vitro* transformation provide supportive evidence of carcinogenicity and may give information on potential carcinogenic mechanisms. A range of tests from each of the above end points helps to characterize an agent's response spectrum.

Short-term *in vivo* and *in vitro* tests that can give indication of initiation and promotion activity may also provide supportive evidence for carcinogenicity. Lack of positive results in short-term tests for genetic toxicity does not provide a basis for discounting positive results in long-term animal studies.

6. *Long-Term Animal Studies.* Criteria for the technical adequacy of animal carcinogenicity studies have been published (e.g., U.S. Food and Drug Administration, 1982; Interagency Regulatory Liaison Group, 1979; National Toxicology Program, 1984; OSTP, 1985; U.S. EPA, 1983a, 1983b, 1983c; Feron et al., 1980; Mantel, 1980) and should be used to judge the acceptability of individual studies. Transplacental and multigenerational carcinogenesis studies, in addition to more conventional long-term animal studies, can yield useful information about the carcinogenicity of agents.

It is recognized that chemicals that induce benign tumors frequently also induce malignant

tumors, and that benign tumors often progress to malignant tumors (Interdisciplinary Panel on Carcinogenicity, 1984). The incidence of benign and malignant tumors will be combined when scientifically defensible (OSTP, 1985; Principle 8). For example, the Agency will, in general, consider the combination of benign and malignant tumors to be scientifically defensible unless the benign tumors are not considered to have the potential to progress to the associated malignancies of the same histogenic origin. If an increased incidence of benign tumors is observed in the absence of malignant tumors, in most cases the evidence will be considered as limited evidence of carcinogenicity.

The weight of evidence that an agent is potentially carcinogenic for humans increases (1) with the increase in number of tissue sites affected by the agent; (2) with the increase in number of animal species, strains, sexes, and number of experiments and doses showing a carcinogenic response; (3) with the occurrence of clear-cut dose-response relationships as well as a high level of statistical significance of the increased tumor incidence in treated compared to control groups; (4) when there is a dose-related shortening of the time-to-tumor occurrence or time to death with tumor; and (5) when there is a dose-related increase in the proportion of tumors that are malignant.

Long-term animal studies at or near the maximum tolerated dose level (MTD) are used to ensure an adequate power for the detection of carcinogenic

[51 FR 33995]

activity (NTP, 1984; IARC, 1982). Negative long-term animal studies at exposure levels above the MTD may not be acceptable if animal survival is so impaired that the sensitivity of the study is significantly reduced below that of a conventional chronic animal study at the MTD. The OSTP (1985; Principle 4) has stated that,

The carcinogenic effects of agents may be influenced by non-physiological responses (such as extensive organ damage, radical disruption of hormonal function, saturation of metabolic pathways, formation of stones in the urinary tract, saturation of DNA repair with a functional loss of the system) induced in the model systems. Testing regimes inducing these responses should be evaluated for their relevance to the human response to an agent and evidence from such a study, whether positive or negative, must be carefully reviewed.

Positive studies at levels above the MTD should be carefully reviewed to ensure that the responses are not due to factors which do not operate at exposure levels below the MTD. Evidence indicating that high exposures alter tumor responses by indirect mechanisms that may be unrelated to effects at lower exposures should be dealt with on an individual basis. As noted by the OSTP (1985), "Normal metabolic activation of carcinogens may possibly also be altered and carcinogenic potential reduced as a consequence [of high-dose testing]."

Carcinogenic responses under conditions of the experiment should be reviewed carefully as they relate to the relevance of the evidence to human carcinogenic risks (e.g., the occurrence of bladder tumors in the presence of bladder stones and implantation site sarcomas). Interpretation of animal studies is aided by the review of target organ toxicity and other effects (e.g., changes in the immune and endocrine systems) that may be noted in prechronic or other toxicological studies. Time and dose-related changes in the incidence of preneoplastic lesions may also be helpful in interpreting animal studies.

Agents that are positive in long-term animal experiments and also show evidence of promoting or cocarcinogenic activity in specialized tests should be considered as complete carcinogens unless there is evidence to the contrary because it is, at present, difficult to determine whether an agent is only a promoting or cocarcinogenic agent. Agents that show positive results in special tests for initiation, promotion, or cocarcinogenicity and no indication of tumor response in well-conducted and well-designed long-term animal studies should be dealt with on an individual basis.

To evaluate carcinogenicity, the primary comparison is tumor response in dosed animals as compared with that in contemporary matched control animals. Historical control data are often valuable, however, and could be used along with concurrent control data in the evaluation of carcinogenic responses (Haseeman et al., 1984). For the evaluation of rare tumors, even small tumor responses may be significant compared to historical data. The review of tumor data at sites with high spontaneous background requires special consideration (OSTP, 1985; Principle 9). For instance, a response that is significant with respect to the experimental control group may become questionable if the historical control data indicate that the experimental control group had an unusually low background incidence (NTP, 1984).

For a number of reasons, there are widely diverging scientific views (OSTP, 1985; Ward et al., 1979a, b; Tomatis, 1977; Nutrition Foundation, 1983) about the validity of mouse liver tumors as an indication of potential carcinogenicity in humans when such tumors occur in strains with high spontaneous background incidence and when they constitute the only tumor response to an agent. These Guidelines take the position that when the only tumor response is in the mouse liver and when other conditions for a classification of "sufficient" evidence in animal studies are met (e.g., replicate studies, malignancy; see section IV), the data should be considered as "sufficient" evidence of carcinogenicity. It is understood that this classification could be changed on a case-by-case basis to "limited," if warranted, when factors such as the following, are observed: an increased incidence

of tumors only in the highest dose group and/or only at the end of the study; no substantial dose-related increase in the proportion of tumors that are malignant; the occurrence of tumors that are predominantly benign; no dose-related shortening of the time to the appearance of tumors; negative or inconclusive results from a spectrum of short-term tests for mutagenic activity; the occurrence of excess tumors only in a single sex.

Data from all long-term animal studies are to be considered in the evaluation of carcinogenicity. A positive carcinogenic response in one species/strain/sex is not generally negated by negative results in other species/strain/sex. Replicate negative studies that are essentially identical in all other respects to a positive study may indicate that the positive results are spurious.

Evidence for carcinogenic action should be based on the observation of statistically significant tumor responses in specific organs or tissues. Appropriate statistical analysis should be performed on data from long-term studies to help determine whether the effects are treatment-related or possibly due to chance. These should at least include a statistical test for trend, including appropriate correction for differences in survival. The weight to be given to the level of statistical significance (the p-value) and to other available pieces of information is a matter of overall scientific judgment. A statistically significant excess of tumors of all types in the aggregate, in the absence of a statistically significant increase of any individual tumor type, should be regarded as minimal evidence of carcinogenic action unless there are persuasive reasons to the contrary.

**7. Human Studies.** Epidemiologic studies provide unique information about the response of humans who have been exposed to suspect carcinogens. Descriptive epidemiologic studies are useful in generating hypotheses and providing supporting data, but can rarely be used to make a causal inference. Analytical epidemiologic studies of the case-control or cohort variety, on the other hand, are especially useful in assessing risks to exposed humans.

Criteria for the adequacy of epidemiologic studies are well recognized. They include factors such as the proper selection and characterization of exposed and control groups, the adequacy of duration and quality of follow-up, the proper identification and characterization of confounding factors and bias, the appropriate consideration of latency effects, the valid ascertainment of the causes of morbidity and death, and the ability to detect specific effects. Where it can be calculated, the statistical power to detect an appropriate outcome should be included in the assessment.

The strength of the epidemiologic evidence for carcinogenicity depends, among other things, on the

type of analysis and on the magnitude and specificity of the response. The weight of evidence increases rapidly with the number of adequate studies that show comparable results on populations exposed to the same agent under different conditions.

It should be recognized that epidemiologic studies are inherently capable of detecting only comparatively large increases in the relative risk of

[51 FR 33996]

cancer. Negative results from such studies cannot prove the absence of carcinogenic action, however, negative results from a well-designed and well-conducted epidemiologic study that contains usable exposure data can serve to define upper limits of risk; these are useful if animal evidence indicates that the agent is potentially carcinogenic in humans.

#### C. Weight of Evidence

Evidence of possible carcinogenicity in humans comes primarily from two sources: long term animal tests and epidemiologic investigations. Results from these studies are supplemented with available information from short-term tests, pharmacokinetic studies, comparative metabolism studies, structure-activity relationships, and other relevant toxicologic studies. The question of how likely an agent is to be a human carcinogen should be answered in the framework of a weight-of-evidence judgment. Judgments about the weight of evidence involve considerations of the quality and adequacy of the data and the kinds and consistency of responses induced by a suspect carcinogen. There are three major steps to characterizing the weight of evidence for carcinogenicity in humans: (1) characterization of the evidence from human studies and from animal studies individually, (2) combination of the characterizations of these two types of data into an indication of the overall weight of evidence for human carcinogenicity, and (3) evaluation of all supporting information to determine if the overall weight of evidence should be modified.

EPA has developed a system for stratifying the weight of evidence (see section IV). This classification is not meant to be applied rigidly or mechanically. At various points in the above discussion, EPA has emphasized the need for an overall, balanced judgment of the totality of the available evidence. Particularly for well-studied substances, the scientific data base will have a complexity that cannot be captured by any classification scheme. Therefore, the hazard identification section should include a narrative summary of the strengths and weaknesses of the evidence as well as its categorization in the EPA scheme.

The EPA classification system is, in general, an adaptation of the International Agency for Research on Cancer (IARC, 1982) approach for classifying the

weight of evidence for human data and animal data. The EPA classification system for the characterization of the overall weight of evidence for carcinogenicity (animal, human, and other supportive data) includes: Group A -- Carcinogenic to Humans; Group B -- Probably Carcinogenic to Humans; Group C -- Possibly Carcinogenic to Humans; Group D -- Not Classifiable as to Human Carcinogenicity; and Group E -- Evidence of Non-Carcinogenicity for Humans.

The following modifications of the IARC approach have been made for classifying human and animal studies.

For human studies:

(1) The observation of a statistically significant association between an agent and life-threatening benign tumors in humans is included in the evaluation of risks to humans.

(2) A "no data available" classification is added.

(3) A "no evidence of carcinogenicity" classification is added. This classification indicates that no association was found between exposure and increased risk of cancer in well conducted, well-designed, independent analytical epidemiologic studies.

For animal studies:

(1) An increased incidence of combined benign and malignant tumors will be considered to provide sufficient evidence of carcinogenicity if the other criteria defining the "sufficient" classification of evidence are met (e.g., replicate studies, malignancy; see section IV). Benign and malignant tumors will be combined when scientifically defensible.

(2) An increased incidence of benign tumors alone generally constitutes "limited" evidence of carcinogenicity.

(3) An increased incidence of neoplasms that occur with high spontaneous background incidence (e.g., mouse liver tumors and rat pituitary tumors in certain strains) generally constitutes "sufficient" evidence of carcinogenicity, but may be changed to "limited" when warranted by the specific information available on the agent.

(4) A "no data available" classification has been added.

(5) A "no evidence of carcinogenicity" classification is also added. This operational classification would include substances for which there is no increased incidence of neoplasms in at least two well-designed and well-conducted animal studies of adequate power and dose in different species.

#### D. Guidance for Dose-Response Assessment

The qualitative evidence for carcinogenesis should be discussed for purposes of guiding the dose-response assessment. The guidance should be given in terms of the appropriateness and limitations of specific studies as well as pharmacokinetic considerations that should be factored into the dose-

response assessment. The appropriate method of extrapolation should be factored in when the experimental route of exposure differs from that occurring in humans.

Agents that are judged to be in the EPA weight-of-evidence stratification Groups A and B would be regarded as suitable for quantitative risk assessments. Agents that are judged to be in Group C will generally be regarded as suitable for quantitative risk assessment, but judgments in this regard may be made on a case-by-case basis. Agents that are judged to be in Groups D and E would not have quantitative risk assessments.

#### E. Summary and Conclusion

The summary should present all of the key findings in all of the sections of the qualitative assessment and the interpretive rationale that forms the basis for the conclusion. Assumptions, uncertainties in the evidence, and other factors that may affect the relevance of the evidence to humans should be discussed. The conclusion should present both the weight-of-evidence ranking and a description that brings out the more subtle aspects of the evidence that may not be evident from the ranking alone.

#### III. Dose-Response Assessment, Exposure Assessment, and Risk Characterization

After data concerning the carcinogenic properties of a substance have been collected, evaluated, and categorized, it is frequently desirable to estimate the likely range of excess cancer risk associated with given levels and conditions of human exposure. The first step of the analysis needed to make such estimations is the development of the likely relationship between dose and response (cancer incidence) in the region of human exposure. This information on dose-response relationships is coupled with information on the nature and magnitude of human exposure to yield an estimate of human risk. The risk-characterization step also includes an interpretation of these estimates in light of the biological, statistical, and exposure assumptions and uncertainties that have arisen throughout the process of assessing risk.

The elements of dose-response assessment are described in section III.A. Guidance on human exposure assessment is provided in another EPA

[51 FR 33997]

document (U.S. EPA, 1986); however, section III.B. of these Guidelines includes a brief description of the specific type of exposure information that is useful for carcinogen risk assessment. Finally, in section III.C. on risk characterization, there is a description of the manner in which risk estimates should be presented so as to be most informative.

It should be emphasized that calculation of quantitative estimates of cancer risk does not



require that an agent be carcinogenic in humans. The likelihood that an agent is a human carcinogen is a function of the weight of evidence, as this has been described in the hazard identification section of these Guidelines. It is nevertheless important to present quantitative estimates, appropriately qualified and interpreted, in those circumstances in which there is a reasonable possibility, based on human and animal data, that the agent is carcinogenic in humans.

It should be emphasized in every quantitative risk estimation that the results are uncertain. Uncertainties due to experimental and epidemiologic variability as well as uncertainty in the exposure assessment can be important. There are major uncertainties in extrapolating both from animals to humans and from high to low doses. There are important species differences in uptake, metabolism, and organ distribution of carcinogens, as well as species and strain differences in target-site susceptibility. Human populations are variable with respect to genetic constitution, diet, occupational and home environment, activity patterns, and other cultural factors. Risk estimates should be presented together with the associated hazard assessment (section III.C.3.) to ensure that there is an appreciation of the weight of evidence for carcinogenicity that underlies the quantitative risk estimates.

#### A. Dose-Response Assessment

1. *Selection of Data.* As indicated in section II.D., guidance needs to be given by the individuals doing the qualitative assessment (toxicologists, pathologists, pharmacologists, etc.) to those doing the quantitative assessment as to the appropriate data to be used in the dose-response assessment. This is determined by the quality of the data, its relevance to human modes of exposure, and other technical details.

If available, estimates based on adequate human epidemiologic data are preferred over estimates based on animal data. If adequate exposure data exist in a well-designed and well-conducted negative epidemiologic study, it may be possible to obtain an upper-bound estimate of risk from that study. Animal-based estimates, if available, also should be presented.

In the absence of appropriate human studies, data from a species that responds most like humans should be used, if information to this effect exists. Where, for a given agent, several studies are available, which may involve different animal species, strains, and sexes at several doses and by different routes of exposure, the following approach to selecting the data sets is used: (1) The tumor incidence data are separated according to organ site and tumor type. (2) All biologically and statistically acceptable data sets are presented. (3) The range of the risk estimates is presented with due regard to

biological relevance (particularly in the case of animal studies) and appropriateness of route of exposure. (4) Because it is possible that human sensitivity is as high as the most sensitive responding animal species, in the absence of evidence to the contrary, the biologically acceptable data set from long-term animal studies showing the greatest sensitivity should generally be given the greatest emphasis, again with due regard to biological and statistical considerations.

When the exposure route in the species from which the dose-response information is obtained differs from the route occurring in environmental exposures, the considerations used in making the route-to-route extrapolation must be carefully described. All assumptions should be presented along with a discussion of the uncertainties in the extrapolation. Whatever procedure is adopted in a given case, it must be consistent with the existing metabolic and pharmacokinetic information on the chemical (e.g., absorption efficiency via the gut and lung, target organ doses, and changes in placental transport throughout gestation for transplacental carcinogens).

Where two or more significantly elevated tumor sites or types are observed in the same study, extrapolations may be conducted on selected sites or types. These selections will be made on biological grounds. To obtain a total estimate of carcinogenic risk, animals with one or more tumor sites or types showing significantly elevated tumor incidence should be pooled and used for extrapolation. The pooled estimates will generally be used in preference to risk estimates based on single sites or types. Quantitative risk extrapolations will generally not be done on the basis of totals that include tumor sites without statistically significant elevations.

Benign tumors should generally be combined with malignant tumors for risk estimates unless the benign tumors are not considered to have the potential to progress to the associated malignancies of the same histogenic origin. The contribution of the benign tumors, however, to the total risk should be indicated.

2. *Choice of Mathematical Extrapolation Model.* Since risks at low exposure levels cannot be measured directly either by animal experiments or by epidemiologic studies, a number of mathematical models have been developed to extrapolate from high to low dose. Different extrapolation models, however, may fit the observed data reasonably well but may lead to large differences in the projected risk at low doses.

As was pointed out by OSTP (1985; Principle 26),

No single mathematical procedure is recognized as the most appropriate for low-dose extrapolation in carcinogenesis. When relevant biological evidence on mechanism of action exists (e.g., pharmacokinetics, target organ dose), the models or procedures

employed should be consistent with the evidence. When data and information are limited, however, and when much uncertainty exists regarding the mechanism of carcinogenic action, models or procedures which incorporate low dose linearity are preferred when compatible with the limited information.

At present, mechanisms of the carcinogenesis process are largely unknown and data are generally limited. If a carcinogenic agent acts by accelerating the same carcinogenic process that leads to the background occurrence of cancer, the added effect of the carcinogen at low doses is expected to be virtually linear (Crump et al., 1976).

The Agency will review each assessment as to the evidence on carcinogenesis mechanisms and other biological or statistical evidence that indicates the suitability of a particular extrapolation model. Goodness-of-fit to the experimental observations is not an effective means of discriminating among models (OSTP, 1985). A rationale will be included to justify the use of the chosen model. In the absence of adequate information to the contrary, the linearized multistage procedure will be employed. Where appropriate, the results of using various extrapolation models may be useful for comparison with the linearized multistage procedure. When longitudinal data on tumor development are available, time-to-tumor models may be used.

It should be emphasized that the linearized multistage procedure leads to

[51 FR 33998]

a plausible upper limit to the risk that is consistent with some proposed mechanisms of carcinogenesis. Such an estimate, however, does not necessarily give a realistic prediction of the risk. The true value of the risk is unknown, and may be as low as zero. The range of risks, defined by the upper limit given by the chosen model and the lower limit which may be as low as zero, should be explicitly stated. An established procedure does not yet exist for making "most likely" or "best" estimates of risk within the range of uncertainty defined by the upper and lower limit estimates. If data and procedures become available, the Agency will also provide "most likely" or "best" estimates of risk. This will be most feasible when human data are available and when exposures are in the dose range of the data.

In certain cases, the linearized multistage procedure cannot be used with the observed data as, for example, when the data are nonmonotonic or flatten out at high doses. In these cases, it may be necessary to make adjustments to achieve low-dose linearity.

When pharmacokinetic or metabolism data are available, or when other substantial evidence on the mechanistic aspects of the carcinogenesis process exists, a low-dose extrapolation model other than the linearized multistage procedure might be considered more appropriate on biological grounds. When a different model is chosen, the risk

assessment should clearly discuss the nature and weight of evidence that led to the choice. Considerable uncertainty will remain concerning response at low doses; therefore, in most cases an upper-limit risk estimate using the linearized multistage procedure should also be presented.

**3. Equivalent Exposure Units Among Species.** Low-dose risk estimates derived from laboratory animal data extrapolated to humans are complicated by a variety of factors that differ among species and potentially affect the response to carcinogens. Included among these factors are differences between humans and experimental test animals with respect to life span, body size, genetic variability, population homogeneity, existence of concurrent disease, pharmacokinetic effects such as metabolism and excretion patterns, and the exposure regimen.

The usual approach for making interspecies comparisons has been to use standardized scaling factors. Commonly employed standardized dosage scales include mg per kg body weight per day, ppm in the diet or water, mg per m<sup>2</sup> body surface area per day, and mg per kg body weight per lifetime. In the absence of comparative toxicological, physiological, metabolic, and pharmacokinetic data for a given suspect carcinogen, the Agency takes the position that the extrapolation on the basis of surface area is considered to be appropriate because certain pharmacological effects commonly scale according to surface area (Dedrick, 1973; Freireich et al., 1966; Pinkel, 1958).

#### B. Exposure Assessment

In order to obtain a quantitative estimate of the risk, the results of the dose-response assessment must be combined with an estimate of the exposures to which the populations of interest are likely to be subject. While the reader is referred to the Guidelines for Estimating Exposures (U.S. EPA, 1986) for specific details, it is important to convey an appreciation of the impact of the strengths and weaknesses of exposure assessment on the overall cancer risk assessment process.

At present there is no single approach to exposure assessment that is appropriate for all cases. On a case-by-case basis, appropriate methods are selected to match the data on hand and the level of sophistication required. The assumptions, approximations, and uncertainties need to be clearly stated because, in some instances, these will have a major effect on the risk assessment.

In general, the magnitude, duration, and frequency of exposure provide fundamental information for estimating the concentration of the carcinogen to which the organism is exposed. These data are generated from monitoring information, modeling results, and/or reasoned estimates. An appropriate treatment of exposure should consider

the potential for exposure via ingestion, inhalation, and dermal penetration from relevant sources of exposures including multiple avenues of intake from the same source.

Special problems arise when the human exposure situation of concern suggests exposure regimens, e.g., route and dosing schedule that are substantially different from those used in the relevant animal studies. Unless there is evidence to the contrary in a particular case, the cumulative dose received over a lifetime, expressed as average daily exposure prorated over a lifetime, is recommended as an appropriate measure of exposure to a carcinogen. That is, the assumption is made that a high dose of a carcinogen received over a short period of time is equivalent to a corresponding low dose spread over a lifetime. This approach becomes more problematical as the exposures in question become more intense but less frequent, especially when there is evidence that the agent has shown dose-rate effects.

An attempt should be made to assess the level of uncertainty associated with the exposure assessment which is to be used in a cancer risk assessment. This measure of uncertainty should be included in the risk characterization (section III.C.) in order to provide the decision-maker with a clear understanding of the impact of this uncertainty on any final quantitative risk estimate. Subpopulations with heightened susceptibility (either because of exposure or predisposition) should, when possible, be identified.

### C. Risk Characterization

Risk characterization is composed of two parts. One is a presentation of the numerical estimates of risk; the other is a framework to help judge the significance of the risk. Risk characterization includes the exposure assessment and dose-response assessment; these are used in the estimation of carcinogenic risk. It may also consist of a unit-risk estimate which can be combined elsewhere with the exposure assessment for the purposes of estimating cancer risk.

Hazard identification and dose-response assessment are covered in sections II. and III.A., and a detailed discussion of exposure assessment is contained in EPA's Guidelines for Estimating Exposures (U.S. EPA, 1986). This section deals with the numerical risk estimates and the approach to summarizing risk characterization.

1. *Options for Numerical Risk Estimates.* Depending on the needs of the individual program offices, numerical estimates can be presented in one or more of the following three ways.

a. *Unit Risk* -- Under an assumption of low-dose linearity, the unit cancer risk is the excess lifetime risk due to a continuous constant lifetime exposure of one unit of carcinogen concentration. Typical

exposure units include ppm or ppb in food or water, mg/kg/day by ingestion, or ppm or  $\mu\text{g}/\text{m}^3$  in air.

b. *Dose Corresponding to a Given Level of Risk* -- This approach can be useful, particularly when using nonlinear extrapolation models where the unit risk would differ at different dose levels.

c. *Individual and Population Risks* -- Risks may be characterized either in terms of the excess individual lifetime risks, the excess number of cancers

[51 FR 33999]

produced per year in the exposed population, or both.

Irrespective of the options chosen, the degree of precision and accuracy in the numerical risk estimates currently do not permit more than one significant figure to be presented.

2. *Concurrent Exposure.* In characterizing the risk due to concurrent exposure to several carcinogens, the risks are combined on the basis of additivity unless there is specific information to the contrary. Interactions of cocarcinogens, promoters, and initiators with known carcinogens should be considered on a case-by-case basis.

3. *Summary of Risk Characterization.* Whichever method of presentation is chosen, it is critical that the numerical estimates not be allowed to stand alone, separated from the various assumptions and uncertainties upon which they are based. The risk characterization should contain a discussion and interpretation of the numerical estimates that affords the risk manager some insight into the degree to which the quantitative estimates are likely to reflect the true magnitude of human risk, which generally cannot be known with the degree of quantitative accuracy reflected in the numerical estimates. The final risk estimate will be generally rounded to one significant figure and will be coupled with the EPA classification of the qualitative weight of evidence. For example, a lifetime individual risk of  $2 \times 10^{-4}$  resulting from exposure to a "probable human carcinogen" (Group B2) should be designated as  $2 \times 10^{-4}$  [B2]. This bracketed designation of the qualitative weight of evidence should be included with all numerical risk estimates (i.e., unit risks, which are risks at a specified concentration or concentrations corresponding to a given risk). Agency statements, such as *FEDERAL REGISTER* notices, briefings, and action memoranda, frequently include numerical estimates of carcinogenic risk. It is recommended that whenever these numerical estimates are used, the qualitative weight-of-evidence classification should also be included.

The section on risk characterization should summarize the hazard identification, dose-response assessment, exposure assessment, and the public health risk estimates. Major assumptions, scientific judgments, and, to the extent possible, estimates of

the uncertainties embodied in the assessment are presented.

#### *IV. EPA Classification System for Categorizing Weight of Evidence for Carcinogenicity from Human and Animal Studies (Adapted from IARC)*

##### **A. Assessment of Weight of Evidence for Carcinogenicity from Studies in Humans**

Evidence of carcinogenicity from human studies comes from three main sources:

1. Case reports of individual cancer patients who were exposed to the agent(s).
2. Descriptive epidemiologic studies in which the incidence of cancer in human populations was found to vary in space or time with exposure to the agent(s).
3. Analytical epidemiologic (case-control and cohort) studies in which individual exposure to the agent(s) was found to be associated with an increased risk of cancer.

Three criteria must be met before a causal association can be inferred between exposure and cancer in humans:

1. There is no identified bias that could explain the association.
2. The possibility of confounding has been considered and ruled out as explaining the association.
3. The association is unlikely to be due to chance.

In general, although a single study may be indicative of a cause-effect relationship, confidence in inferring a causal association is increased when several independent studies are concordant in showing the association, when the association is strong, when there is a dose-response relationship, or when a reduction in exposure is followed by a reduction in the incidence of cancer.

The weight of evidence for carcinogenicity<sup>1</sup> from studies in humans is classified as:

1. Sufficient evidence of carcinogenicity, which indicates that there is a causal relationship between the agent and human cancer.
2. Limited evidence of carcinogenicity, which indicates that a causal interpretation is credible, but that alternative explanations, such as chance, bias, or confounding, could not adequately be excluded.

<sup>1</sup> For purposes of public health protection, agents associated with life-threatening benign tumors in humans are included in the evaluation.

<sup>2</sup> An increased incidence of neoplasms that occur with high spontaneous background incidence (e.g., mouse liver tumors and rat pituitary tumors in certain strains) generally constitutes "sufficient" evidence of carcinogenicity, but may be changed to "limited" when warranted by the specific information available on the agent.

<sup>3</sup> Benign and malignant tumors will be combined unless the benign tumors are not considered to have the potential to progress to the associated malignancies of the same histogenic origin.

3. Inadequate evidence, which indicates that one of two conditions prevailed: (a) there were few pertinent data, or (b) the available studies, while showing evidence of association, did not exclude chance, bias, or confounding, and therefore a causal interpretation is not credible.

4. No data, which indicates that data are not available.

5. No evidence, which indicates that no association was found between exposure and an increased risk of cancer in well-designed and well-conducted independent analytical epidemiologic studies.

##### **B. Assessment of Weight of Evidence for Carcinogenicity from Studies in Experimental Animals**

These assessments are classified into five groups:

1. Sufficient evidence<sup>2</sup> of carcinogenicity, which indicates that there is an increased incidence of malignant tumors or combined malignant and benign tumors:<sup>3</sup> (a) in multiple species or strains; or (b) in multiple experiments (e.g., with different routes of administration or using different dose levels); or (c) to an unusual degree in a single experiment with regard to high incidence, unusual site or type of tumor, or early age at onset.

Additional evidence may be provided by data on dose-response effects, as well as information from short-term tests or on chemical structure.

2. Limited evidence of carcinogenicity, which means that the data suggest a carcinogenic effect but are limited because: (a) the studies involve a single species, strain, or experiment and do not meet criteria for sufficient evidence (see section IV. B.1.c); (b) the experiments are restricted by inadequate dosage levels, inadequate duration of exposure to the agent, inadequate period of follow-up, poor survival, too few animals, or inadequate reporting; or (c) an increase in the incidence of benign tumors only.

3. Inadequate evidence, which indicates that because of major qualitative or quantitative limitations, the studies cannot be interpreted as showing either the presence or absence of a carcinogenic effect.

4. No data, which indicates that data are not available.

5. No evidence, which indicates that there is no increased incidence of neoplasms in at least two well-designed

[51 FR 34000]

and well-conducted animal studies in different species.

The classifications "sufficient evidence" and "limited evidence" refer only to the weight of the experimental evidence that these agents are carcinogenic and not to the potency of their carcinogenic action.

### C. Categorization of Overall Weight of Evidence for Human Carcinogenicity

The overall scheme for categorization of the weight of evidence of carcinogenicity of a chemical for humans uses a three-step process. (1) The weight of evidence in human studies or animal studies is summarized; (2) these lines of information are combined to yield a tentative assignment to a category (see Table 1); and (3) all relevant supportive information is evaluated to see if the designation of the overall weight of evidence needs to be modified. Relevant factors to be included along with the tumor information from human and animal studies include structure-activity relationships; short-term test findings; results of appropriate physiological, biochemical, and toxicological observations; and comparative metabolism and pharmacokinetic studies. The nature of these findings may cause one to adjust the overall categorization of the weight of evidence.

The agents are categorized into five groups as follows:

#### Group A -- Human Carcinogen

This group is used only when there is sufficient evidence from epidemiologic studies to support a causal association between exposure to the agents and cancer.

#### Group B -- Probable Human Carcinogen

This group includes agents for which the weight of evidence of human carcinogenicity based on epidemiologic studies is "limited" and also includes agents for which the weight of evidence of carcinogenicity based on animal studies is "sufficient." The group is divided into two subgroups. Usually, Group B1 is reserved for agents for which there is limited evidence of carcinogenicity from epidemiologic studies. It is reasonable, for practical purposes, to regard an agent for which there is "sufficient" evidence of carcinogenicity in animals as if it presented a carcinogenic risk to humans. Therefore, agents for which there is "sufficient" evidence from animal studies and for which there is "inadequate evidence" or "no data" from epidemiologic studies would usually be categorized under Group B2.

#### Group C -- Possible Human Carcinogen

This group is used for agents with limited evidence of carcinogenicity in animals in the absence of human data. It includes a wide variety of evidence, e.g., (a) a malignant tumor response in a single well-conducted experiment that does not meet conditions for sufficient evidence, (b) tumor responses of marginal statistical significance in studies having inadequate design or reporting, (c) benign but not malignant tumors with an agent showing no response in a variety of short-term tests for mutagenicity, and (d) responses of marginal

statistical significance in a tissue known to have a high or variable background rate.

#### Group D -- Not Classifiable as to Human Carcinogenicity

This group is generally used for agents with inadequate human and animal evidence of carcinogenicity or for which no data are available.

#### Group E -- Evidence of Non-Carcinogenicity for Humans

This group is used for agents that show no evidence for carcinogenicity in at least two adequate animal tests in different species or in both adequate epidemiologic and animal studies.

The designation of an agent as being in Group E is based on the available evidence and should not be interpreted as a definitive conclusion that the agent will not be a carcinogen under any circumstances.

### V. References

- Albert, R.E., Train, R.E., and Anderson, E. 1977. Rationale developed by the Environmental Protection Agency for the assessment of carcinogenic risks. *J. Natl. Cancer Inst.* 58:1537-1541.
- Crump, K.S., Hoel, D.G., Langley, C.H., Peto R. 1976. Fundamental carcinogenic processes and their implications for low dose risk assessment. *Cancer Res.* 36:2973-2979.
- Dedrick, R.L. 1973. Animal Scale Up. *J. Pharmacokin.* Biopharm. 1:435-461.
- Feron, V.J., Grice, H.C., Griesemer, R., Peto R., Agthe, C., Althoff, J., Arnold, D.L., Blumenthal, H., Cabral, J.R.P., Della Porta, G., Ito, N., Kimmerle, G., Kroes, R., Mohr, U., Napalkov, N.P., Odashima, S., Page, N.P., Schramm, T., Steinhoff, D., Sugar, J., Tomatis, L., Uehleke, H., and Youk, V. 1980. Basic requirements for long-term assays for carcinogenesis: a critical appraisal. IARC Monographs, Supplement 2. Lyon, France: International Agency for Research on Cancer, pp 21-83.
- Freireich, E.J., Gehan, E.A., Rall, D.P., Schmidt, L.H., and Skipper, H.E. 1966. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey and man. *Cancer Chemother. Rep.* 50:219-244.
- Haseman, J.K. 1984. Statistical issues in the design, analysis and interpretation of animal carcinogenicity studies. *Environ. Health Perspect.* 58:385-392.
- Haseman, J.K., Huff, J., and Boorman, G.A. 1984. Use of historical control data in carcinogenicity studies in rodents. *Toxicol. Pathol.* 12:126-135.
- Interagency Regulatory Liaison Group (IRLG). 1979. Scientific basis for identification of potential carcinogens and estimation of risks. *J. Natl. Cancer Inst.* 63:245-267.
- Interdisciplinary Panel on Carcinogenicity. 1984. Criteria for evidence of chemical carcinogenicity. *Science* 225:682-687.
- International Agency for Research on Cancer (IARC). 1982. IARC Monographs on the  
[51 FR34001] Evaluation of the Carcinogenic Risk of Chemicals to Humans, Supplement 4. Lyon, France: International Agency for Research on Cancer.
- Mantel, N. 1980. Assessing laboratory evidence for neoplastic activity. *Biometrics* 36:361-399.
- Mantel, N., and Haenszel, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* 22:719-748.
- National Center for Toxicological Research (NCTR). 1981. Guidelines for statistical tests for carcinogenicity in chronic bioassays. NCTR Biometry Technical Report 81-001. Available from: National Center for Toxicological Research.

TABLE 1.—ILLUSTRATIVE CATEGORIZATION OF EVIDENCE BASED ON ANIMAL AND HUMAN DATA<sup>1</sup>

| Human evidence | Animal evidence |         |            |         |             |
|----------------|-----------------|---------|------------|---------|-------------|
|                | Sufficient      | Limited | Inadequate | No data | No evidence |
| Sufficient     | A               | A       | A          | A       | A           |
| Limited        | B1              | B1      | B1         | B1      | B1          |
| Inadequate     | B2              | C       | D          | D       | D           |
| No data        | B2              | C       | D          | D       | E           |
| No evidence    | B2              | C       | D          | D       | E           |

<sup>1</sup> The above assignments are presented for illustrative purposes. There may be nuances in the classification of both animal and human data indicating that different categorizations than those given in the table should be assigned. Furthermore, these assignments are tentative and may be modified by ancillary evidence. In this regard, all relevant information should be evaluated to determine if the designation of the overall weight of evidence needs to be modified. Relevant factors to be included along with the tumor data from human and animal studies include structure-activity relationships, short-term test findings, results of appropriate physiological, biochemical, and toxicological observations, and comparative metabolism and pharmacokinetic studies. The nature of these findings may cause an adjustment of the overall categorization of the weight of evidence.

National Research Council (NRC). 1983. Risk assessment in the Federal government: managing the process. Washington, D.C.: National Academy Press.

National Toxicology Program. 1984. Report of the Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation of the National Toxicology Program, Board of Scientific Counselors. Available from: U.S. Government Printing Office, Washington, D.C. 1984-421-132:4726.

Nutrition Foundation. 1983. The relevance of mouse liver hepatoma to human carcinogenic risk: a report of the International Expert Advisory Committee to the Nutrition Foundation. Available from: Nutrition Foundation. ISBN 0-935368-37-x.

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

Peto, R., Pike, M., Day, N., Gray, R., Lee, P., Parish, S., Peto, J., Richard, S., and Wahrendorf, J. 1980. Guidelines for simple, sensitive, significant tests for carcinogenic effects in long-term animal experiments. In: Monographs on the long-term and short-term screening assays for carcinogens: a critical appraisal. IARC Monographs, Supplement 2. Lyon, France: International Agency for Research on Cancer, pp. 311-426.

Pinkel, D. 1958. The use of body surface area as a criterion of drug dosage in cancer chemotherapy. Cancer Res. 18:853-856.

Tomatis, L. 1977. The value of long-term testing for the implementation of primary prevention. In: Origins of human cancer. Hiatt, H.H., Watson, J.D., and Weinstein, J.A., eds. Cold Spring Harbor Laboratory, pp. 1339-1357.

U.S. Environmental Protection Agency (U.S. EPA). 1976. Interim procedures and guidelines for health risk and economic impact assessments of suspected carcinogens. Federal Register 41:21402-21405.

U.S. Environmental Protection Agency (U.S. EPA). 1980. Water quality criteria documents; availability. Federal Register 45:79318-79379.

U.S. Environmental Protection Agency (U.S. EPA). 1983a. Good laboratory practices standards -- toxicology testing. Federal Register 48:53922.

U.S. Environmental Protection Agency (U.S. EPA). 1983b. Hazard evaluations: humans and domestic animals. Subdivision F. Available from: NTIS, Springfield, VA. PB 83-153916.

U.S. Environmental Protection Agency (U.S. EPA). 1983c. Health effects test guidelines. Available from: NTIS, Springfield, VA. PB 83-232984.

U.S. Environmental Protection Agency (U.S. EPA). 1986, Sept.

24. Guidelines for estimating exposures. Federal Register 51 (185):34042-34054.

U.S. Food and Drug Administration (U.S. FDA). 1982.

Toxicological principles for the safety assessment of direct food additives and color additives used in food. Available from: Bureau of Foods, U.S. Food and Drug Administration.

Ward, J.M., Griesemer, R.A., and Weisburger, E.K. 1979a. The mouse liver tumor as an endpoint in carcinogenesis tests. Toxicol. Appl. Pharmacol. 51:389-397.

Ward, J.M., Goodman, D.G., Squire, R.A., Chu, K.C., and Linhart, M.S. 1979b. Neoplastic and nonneoplastic lesions in aging (Cs7Bl/6N x C57BL/6N)F<sub>1</sub> (B6C3F<sub>1</sub>) mice. J. Natl. Cancer Inst. 63:849-854.

## Part B: Response to Public and Science Advisory Board Comments

### I. Introduction

This section summarizes the major issues raised during both the public comment period on the Proposed Guidelines for Carcinogen Risk Assessment published on November 23, 1984 (49 FR 46294), and also during the April 22-23, 1985, meeting of the Carcinogen Risk Assessment Guidelines Panel of the Science Advisory Board (SAB).

In order to respond to these issues the Agency modified the proposed guidelines in two stages. First, changes resulting from consideration of the public comments were made in a draft sent to the SAB review panel prior to their April meeting. Secondly, the guidelines were further modified in response to the panel's recommendations.

The Agency received 62 sets of comments during the public comment period, including 28 from corporations, 9 from professional or trade associations, and 4 from academic institutions. In general, the comments were favorable. The commentators welcomed the update of the 1976 guidelines and felt that the proposed guidelines of

1985 reflected some of the progress that has occurred in understanding the mechanisms of carcinogenesis. Many commentators, however, felt that additional changes were warranted.

The SAB concluded that the guidelines are "reasonably complete in their conceptual framework and are sound in their overall interpretation of the scientific issues" (Report by the SAB Carcinogenicity Guidelines Review Group, June 19, 1985). The SAB suggested various editorial changes and raised some issues regarding the content of the proposed guidelines, which are discussed below. Based on these recommendations, the Agency has modified the draft guidelines.

## *II. Office of Science and Technology Policy Report on Chemical Carcinogens*

Many commentators requested that the final guidelines not be issued until after publication of the report of the Office of Technology and Science Policy (OSTP) on chemical carcinogens. They further requested that this report be incorporated into the final Guidelines for Carcinogen Risk Assessment.

The final OSTP report was published in 1985 (50 FR 10372). In its deliberations, the Agency reviewed the final OSTP report and feels that the Agency's guidelines are consistent with the principles established by the OSTP. In its review, the SAB agreed that the Agency guidelines are generally consistent with the OSTP report. To emphasize this consistency, the OSTP principles have been incorporated into the guidelines when controversial issues are discussed.

## *III. Inference Guidelines*

Many commentators felt that the proposed guidelines did not provide a sufficient distinction between scientific fact and policy decisions. Others felt that EPA should not attempt to propose firm guidelines in the absence of scientific consensus. The SAB report also indicated the need to "distinguish recommendations based on scientific evidence from those based on science policy decisions."

The Agency agrees with the recommendation that policy, judgmental, or inferential decisions should be clearly identified. In its revision of the proposed guidelines, the Agency has included phrases (e.g., "the Agency takes the position that") to more clearly distinguish policy decisions.

The Agency also recognizes the need to establish procedures for action on important issues in the absence of complete scientific knowledge or consensus. This need was acknowledged in both the National Academy of Sciences book entitled *Risk Management in the Federal Government: Managing the Process* and the OSTP report on chemical carcinogens. As the NAS report states, "Risk assessment is an analytic process that is firmly based on scientific considerations, but it also

requires judgments to be made when the available information is incomplete. These judgments inevitably draw on both scientific and policy considerations."

[51 FR 34002]

The judgments of the Agency have been based on current available scientific information and on the combined experience of Agency experts. These judgments, and the resulting guidance, rely on inference; however, the positions taken in these inference guidelines are felt to be reasonable and scientifically defensible. While all of the guidance is, to some degree, based on inference, the guidelines have attempted to distinguish those issues that depended more on judgment. In these cases, the Agency has stated a position but has also retained flexibility to accommodate new data or specific circumstances that demonstrate that the proposed position is inaccurate. The Agency recognizes that scientific opinion will be divided on these issues.

Knowledge about carcinogens and carcinogenesis is progressing at a rapid rate. While these guidelines are considered a best effort at the present time, the Agency has attempted to incorporate flexibility into the current guidelines and also recommends that the guidelines be revised as often as warranted by advances in the field.

## *IV. Evaluation of Benign Tumors*

Several commentators discussed the appropriate interpretation of an increased incidence of benign tumors alone or with an increased incidence of malignant tumors as part of the evaluation of the carcinogenicity of an agent. Some comments were supportive of the position in the proposed guidelines, i.e., under certain circumstances, the incidence of benign and malignant tumors would be combined, and an increased incidence of benign tumors alone would be considered an indication, albeit limited, of carcinogenic potential. Other commentators raised concerns about the criteria that would be used to decide which tumors should be combined. Only a few commentators felt that benign tumors should never be considered in evaluating carcinogenic potential.

The Agency believes that current information supports the use of benign tumors. The guidelines have been modified to incorporate the language of the OSTP report, i.e., benign tumors will be combined with malignant tumors when scientifically defensible. This position allows flexibility in evaluating the data base for each agent. The guidelines have also been modified to indicate that, whenever benign and malignant tumors have been combined, and the agent is considered a candidate for quantitative risk extrapolation, the contribution of benign tumors to the estimation of risk will be indicated.

## *V. Transplacental and Multigenerational Animal Bioassays*

As one of its two proposals for additions to the guidelines, the SAB recommended a discussion of transplacental and multigenerational animal bioassays for carcinogenicity.

The Agency agrees that such data, when available, can provide useful information in the evaluation of a chemical's potential carcinogenicity and has stated this in the final guidelines. The Agency has also revised the guidelines to indicate that such studies may provide additional information on the metabolic and pharmacokinetic properties of the chemical. More guidance on the specific use of these studies will be considered in future revisions of these guidelines.

#### VI. Maximum Tolerated Dose

The proposed guidelines discussed the implications of using a maximum tolerated dose (MTD) in bioassays for carcinogenicity. Many commentators requested that EPA define MTD. The tone of the comments suggested that the commentators were concerned about the uses and interpretations of high-dose testing.

The Agency recognizes that controversy currently surrounds these issues. The appropriate text from the OSTP report has been incorporated into the final guidelines which suggests that the consequences of high-dose testing be evaluated on a case-by-case basis.

#### VII. Mouse Liver Tumors

A large number of commentators expressed opinions about the assessment of bioassays in which the only increase in tumor incidence was liver tumors in the mouse. Many felt that mouse liver tumors were afforded too much credence, especially given existing information that indicates that they might arise by a different mechanism, e.g., tissue damage followed by regeneration. Others felt that mouse liver tumors were but one case of a high background incidence of one particular type of tumor and that all such tumors should be treated in the same fashion.

The Agency has reviewed these comments and the OSTP principle regarding this issue. The OSTP report does not reach conclusions as to the treatment of tumors with a high spontaneous background rate, but states, as is now included in the text of the guidelines, that these data require special consideration. Although questions have been raised regarding the validity of mouse liver tumors in general, the Agency feels that mouse liver tumors cannot be ignored as an indicator of carcinogenicity. Thus, the position in the proposed guidelines has not been changed: an increased incidence of only mouse liver tumors will be regarded as "sufficient" evidence of carcinogenicity if all other criteria, e.g., replication and malignancy, are met with the understanding that this classification could be changed to "limited" if warranted. The factors that

may cause this re-evaluation are indicated in the guidelines.

#### VIII. Weight-of-Evidence Categories

The Agency was praised by both the public and the SAB for incorporating a weight-of-evidence scheme into its evaluation of carcinogenic risk. Certain specific aspects of the scheme, however, were criticized.

1. Several commentators noted that while the text of the proposed guidelines clearly states that EPA will use all available data in its categorization of the weight of the evidence that a chemical is a carcinogen, the classification system in Part A, section IV did not indicate the manner in which EPA will use information other than data from humans and long-term animal studies in assigning a weight-of-evidence classification.

The Agency has added a discussion to Part A, section IV.C. dealing with the characterization of overall evidence for human carcinogenicity. This discussion clarifies EPA's use of supportive information to adjust, as warranted, the designation that would have been made solely on the basis of human and long-term animal studies.

2. The Agency agrees with the SAB and those commentators who felt that a simple classification of the weight of evidence, e.g., a single letter or even a descriptive title, is inadequate to describe fully the weight of evidence for each individual chemical. The final guidelines propose that a paragraph summarizing the data should accompany the numerical estimate and weight-of-evidence classification whenever possible.

3. Several commentators objected to the descriptive title E (No Evidence of Carcinogenicity for Humans) because they felt the title would be confusing to people inexperienced with the classification system. The title for Group E, No Evidence of Carcinogenicity for Humans, was thought by these commentators to suggest the absence of data. This group, however, is intended to be reserved for agents for which there exists credible data demonstrating that the agent is not carcinogenic.

Based on these comments and further discussion, the Agency has changed the  
[51 FR 34003]

title of Group E to "Evidence of Non-Carcinogenicity for Humans."

4. Several commentators felt that the title for Group C, Possible Human Carcinogen, was not sufficiently distinctive from Group B, Probable Human Carcinogen. Other commentators felt that those agents that minimally qualified for Group C would lack sufficient data for such a label.

The Agency recognizes that Group C covers a range of chemicals and has considered whether to subdivide Group C. The consensus of the Agency's



Carcinogen Risk Assessment Committee, however, is that the current groups, which are based on the IARC categories, are a reasonable stratification and should be retained at present. The structure of the groups will be reconsidered when the guidelines are reviewed in the future. The Agency also feels that the descriptive title it originally selected best conveys the meaning of the classification within the context of EPA's past and current activities.

5. Some commentators indicated a concern about the distinction between B1 and B2 on the basis of epidemiologic evidence only. This issue has been under discussion in the Agency and may be revised in future versions of the guidelines.

6. Comments were also received about the possibility of keeping the groups for animal and human data separate without reaching a combined classification. The Agency feels that a combined classification is useful; thus, the combined classification was retained in the final guidelines.

The SAB suggested that a table be added to Part A, section IV to indicate the manner in which human and animal data would be combined to obtain an overall weight-of-evidence category. The Agency realizes that a table that would present all permutations of potentially available data would be complex and possibly impossible to construct since numerous combinations of ancillary data (e.g., genetic toxicity, pharmacokinetics) could be used to raise or lower the weight-of-evidence classification. Nevertheless, the Agency decided to include a table to illustrate the most probable weight-of-evidence classification that would be assigned on the basis of standard animal and human data without consideration of the ancillary data. While it is hoped that this table will clarify the weight-of-evidence classifications, it is also important to recognize that an agent may be assigned to a final categorization different from the category which would appear appropriate from the table and still conform to the guidelines.

#### *IX. Quantitative Estimates of Risk*

The method for quantitative estimates of carcinogenic risk in the proposed guidelines received substantial comments from the public. Five issues were discussed by the Agency and have resulted in modifications of the guidelines.

1. The major criticism was the perception that EPA would use only one method for the extrapolation of carcinogenic risk and would, therefore, obtain one estimate of risk. Even commentators who concur with the procedure usually followed by EPA felt that some indication of the uncertainty of the risk estimate should be included with the risk estimate.

The Agency feels that the proposed guidelines were not intended to suggest that EPA would perform quantitative risk estimates in a rote or

mechanical fashion. As indicated by the OSTP report and paraphrased in the proposed guidelines, no single mathematical procedure has been determined to be the most appropriate method for risk extrapolation. The final guidelines quote rather than paraphrase the OSTP principle. The guidelines have been revised to stress the importance of considering all available data in the risk assessment and now state, "The Agency will review each assessment as to the evidence on carcinogenic mechanisms and other biological or statistical evidence that indicates the suitability of a particular extrapolation model." Two issues are emphasized: First, the text now indicates the potential for pharmacokinetic information to contribute to the assessment of carcinogenic risk. Second, the final guidelines state that time-to-tumor risk extrapolation models may be used when longitudinal data on tumor development are available.

2. A number of commentators noted that the proposed guidelines did not indicate how the uncertainties of risk characterization would be presented. The Agency has revised the proposed guidelines to indicate that major assumptions, scientific judgments, and, to the extent possible, estimates of the uncertainties embodied in the risk assessment will be presented along with the estimation of risk.

3. The proposed guidelines stated that the appropriateness of quantifying risks for chemicals in Group C (Possible Human Carcinogen), specifically those agents that were on the boundary of Groups C and D (Not Classifiable as to Human Carcinogenicity), would be judged on a case-by-case basis. Some commentators felt that quantitative risk assessment should not be performed on any agent in Group C.

Group C includes a wide range of agents, including some for which there are positive results in one species in one good bioassay. Thus, the Agency feels that many agents in Group C will be suitable for quantitative risk assessment, but that judgments in this regard will be made on a case-by-case basis.

4. A few commentators felt that EPA intended to perform quantitative risk estimates on aggregate tumor incidence. While EPA will consider an increase in total aggregate tumors as suggestive of potential carcinogenicity, EPA does not generally intend to make quantitative estimates of carcinogenic risk based on total aggregate tumor incidence.

5. The proposed choice of body surface area as an interspecies scaling factor was criticized by several commentators who felt that body weight was also appropriate and that both methods should be used. The OSTP report recognizes that both scaling factors are in common use. The Agency feels that the choice of the body surface area scaling factor can be

justified from the data on effects of drugs in various species. Thus, EPA will continue to use this scaling factor unless data on a specific agent suggest that a different scaling factor is justified. The uncertainty engendered by choice of scaling factor will be included in the summary of uncertainties associated with the assessment of risk mentioned in point 1, above.

In the second of its two proposals for additions to the proposed guidelines, the SAB suggested that a sensitivity analysis be included in EPA's quantitative estimate of a chemical's carcinogenic potency. The Agency agrees that an analysis of the assumptions and uncertainties inherent in an assessment of carcinogenic risk must be accurately portrayed. Sections of the final guidelines that deal with this issue have been strengthened to reflect the concerns of the SAB and the Agency. In particular, the last paragraph of the guidelines states that "major assumptions, scientific judgments, and, to the extent possible, estimates of the uncertainties embodied in the assessment" should be presented in the summary characterizing the risk. Since the assumptions and uncertainties will vary for each assessment, the Agency feels that a formal requirement for a particular type of sensitivity analysis would be less useful than a case-by-case evaluation of the particular assumptions and uncertainties most significant for a particular risk assessment.

END  
DATE  
FILMED  
5-50  
NTIS

2023708969

NTIS does not permit return of items for credit or refund. A replacement will be provided if an error is made in filling your order, if the item was received in damaged condition, or if the item is defective.

*Reproduced by NTIS  
National Technical Information Service  
U.S. Department of Commerce  
Springfield, VA 22161*

**This report was printed specifically for your order from our collection of more than 1.5 million technical reports.**

For economy and efficiency, NTIS does not maintain stock of its vast collection of technical reports. Rather, most documents are printed for each order. Your copy is the best possible reproduction available from our master archive. If you have any questions concerning this document or any order you placed with NTIS, please call our Customer Services Department at (703) 487-4660.

Always think of NTIS when you want:

- Access to the technical, scientific, and engineering results generated by the ongoing multibillion dollar R&D program of the U.S. Government.
- R&D results from Japan, West Germany, Great Britain, and some 20 other countries, most of it reported in English.

NTIS also operates two centers that can provide you with valuable information:

- The Federal Computer Products Center – offers software and datafiles produced by Federal agencies.
- The Center for the Utilization of Federal Technology – gives you access to the best of Federal technologies and laboratory resources.

For more information about NTIS, send for our **FREE NTIS Products and Services Catalog** which describes how you can access this U.S. and foreign Government technology. Call (703) 487-4650 or send this sheet to NTIS, U.S. Department of Commerce, Springfield, VA 22161. Ask for catalog, PR-827.

Name \_\_\_\_\_

Address \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Telephone \_\_\_\_\_

**NTIS.** –Your Source to U.S. and Foreign Government  
Research and Technology

2023708970

2023708971

FIRST  
CLASS  
MAIL

U.S. DEPARTMENT OF COMMERCE  
National Technical Information Service  
Springfield, VA. 22161

☐ NTIS-85 (9-79)

OFFICIAL BUSINESS

Penalty for Private Use, \$300

024100094

NTIS CONTROL NUMBER

\$ 17.00

VALUE OF SHIPMENT

DDC USER CODE CONTRACT NO.

ORDER REF RB016

PURCHASE ORDER NO.

2522852

CARD SERIAL NO.

64134 0

DEPOSIT ACCT. NO.

020

FIRST CLASS

POSTAGE AND FEES PAID  
U.S. DEPARTMENT OF COMMERCE  
COM-211



028546

PHILIP MORRIS USA  
ATTN: NANCY W PETERS  
PM USA OPERATIONS CENTER  
PO BOX 26603  
RICHMOND VA 23261

NO. OF COPIES

PAPER  
COPY  
MICRO  
FORM

001

PB90-18 353 4

XAB

ORDERED AS

FILLED AS (DOCUMENT NO.)

THIS IS NOT A BILL. IT IS YOUR RECORD OF SHIPMENT. INVOICE WILL FOLLOW FOR SHIP AND BILL.  
FOR ANY ADJUSTMENT ON THIS ORDER, PLEASE RETURN THIS CARD WITH YOUR CORRESPONDENCE.

2023708972