United States
Environmental Protection
Agency

Office of Health and
Environmental Assessment
Washington DC 20460

EPA/600/8-87/045
August 1987

Research and Development

EPA

The Risk Assessment Guidelines of 1986

EPA/600/8-87/045
August 1987

THE RISK ASSESSMENT GUIDELINES OF 1986

U.S. Environmental Protection Agency
Washington, D.C.

DISCLAIMER

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

CONTENTS

Preface iv

Abstract v

GUIDELINES FOR CARCINOGEN RISK ASSESSMENT 1-1

GUIDELINES FOR MUTAGENICITY RISK ASSESSMENT 2-1

GUIDELINES FOR THE HEALTH RISK ASSESSMENT OF CHEMICAL MIXTURES 3-1

GUIDELINES FOR THE HEALTH ASSESSMENT OF SUSPECT DEVELOPMENTAL TOXICANTS 4-1

GUIDELINES FOR ESTIMATING EXPOSURES 5-1

PREFACE

On September 24, 1986, the U.S. Environmental Protection Agency (EPA) issued risk assessment guidelines relating to five areas: carcinogenicity, mutagenicity, chemical mixtures, suspect developmental toxicants, and estimating exposures (51 FR 33992-34054). The guidelines were developed to promote high technical quality and Agencywide consistency in the risk assessment process.

The guidelines were developed partly in response to a 1983 National Academy of Sciences publication entitled "Risk Assessment in the Federal Government: Managing the Process," which recommended that Federal regulatory agencies establish risk assessment guidelines. An EPA task force, convened by then Administrator William D. Ruckelshaus to study ways to improve the scientific foundation for Agency regulatory decisions, accepted the recommendation, and work on the guidelines began early in 1984.

The guidelines are products of a two-year Agency development and review process which included many scientists from the larger scientific community. They were developed as part of an interoffice guidelines development program under the auspices of the Office of Health and Environmental Assessment in the Agency's Office of Research and Development. The scientists involved were skilled in each topic, and early drafts were peer-reviewed by experts from academia, industry, public interest groups, and other governmental agencies. Subsequently, proposed guidelines were published in the *Federal Register*, reviewed by special panels of EPA's Science Advisory Board (SAB), and revised to take into account public and SAB comments. After final EPA review and Office of Management and Budget review, the guidelines were signed by EPA Administrator Lee M. Thomas on August 22, 1986, and published in the *Federal Register* on September 24, 1986.

Each of the five guidelines provides both technical information and science policy guidance relating to the conduct of EPA risk assessments and presentation of risk assessment information. The guidelines are sufficiently flexible to allow skilled scientists to make appropriate technical judgments on a case-by-case basis, giving full consideration to all relevant scientific information. The guidelines also stress that risk assessments should include a discussion of the strengths and weaknesses of each assessment by describing uncertainties, assumptions, and limitations, as well as the scientific basis and rationale for each assessment. They require risk assessors to inform Agency decisionmakers and the public about the assumptions used in and the implications of individual risk assessment conclusions, so that appropriate risk management decisions can be made and explained.

While these guidelines are published Agency documents, they should not be interpreted as static, but as the first step in the continuing process of identifying the best methods for assessing risk to environmental pollutants. Consequently, the risk assessment guidelines are constantly undergoing Agency scrutiny and will be revised in line with new methods and information, as appropriate.

This document presents the five guidelines as they originally appeared in the *Federal Register* but in a format that is easier to read.

ABSTRACT

On September 24, 1986, the U.S. Environmental Protection Agency issued risk assessment guidelines relating to five areas: carcinogenicity, mutagenicity, chemical mixtures, suspect developmental toxicants, and estimating exposures (51 FR 33992-34054). The guidelines were developed to promote high technical quality and Agencywide consistency in the risk assessment process. This document presents the five guidelines as they originally appeared in the *Federal Register* but in a format that is easier to read.

51 FR 33992

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. Thomas,

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 Catagorizing 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. Meta*bolic 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. To*xicologic 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 (Haseman 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 classificaton 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^2 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 $\mu g/m^3$ 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 inititators 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 2×10^{-4} resulting from exposure to a "probable human carcinogen" (Group B2) should be designated as 2×10^{-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[1] 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.
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[2] of carcinogenicity, which indicates that there is an increased incidence of malignant tumors or combined malignant and benign tumors:[3] (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.

[1] For purposes of public health protection, agents associated with life-threatening benign tumors in humans are included in the evaluation.

[2] 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.

[3] 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.

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. Pharmacokinet. 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 Vouk, V. 1980. Basic requirements for long-term assays for carcinogenicity. In: 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 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 FR 34001]

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 36:381-399.

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[1]

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

1 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 Winstein, 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 ($C_{57}BL/6N \times C_3H/HeN)F_1$ ($B_6C_3F_1$) 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 commentors 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 commentors, 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 commentors 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 commentors 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 commentors 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 commentors raised concerns about the criteria that would be used to decide which tumors should be combined. Only a few commentors 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 commentors requested that EPA define MTD. The tone of the comments suggested that the commentors 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 commentors 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 Catagories

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 commentors 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 commentors 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 commentors 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 commentors 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 commentors felt that the title for Group C, Possible Human Carcinogen, was not sufficiently distinctive from Group B, Probable Human Carcinogen. Other commentors 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 commentors 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 commentors 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 commentors 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 commentors 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 commentors 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 commentors 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.

51 FR 34006

GUIDELINES FOR MUTAGENICITY 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 Mutagenicity Risk Assessment.

The Guidelines for Mutagenicity Risk Assessment (hereafter "Guidelines") are intended to guide Agency analysis of mutagenicity data 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 Mutagenicity Risk Assessment published November 23, 1984 (49 FR 46314).

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. Lawrence R. Valcovic
Reproductive Effects Assessment Group
Office of Health and Environmental Assessment
(RD-689)
U. S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460
202-382-7303

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 Mutagenicity Risk Assessment

Work on the Guidelines for Mutagenicity 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 genetic toxicology from universities, environmental groups, industry, labor, and other governmental agencies. They were then proposed for public comment in the *FEDERAL REGISTER* (49 FR 46314). 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 preliminary Agency responses to those comments. 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 Mutagenicity Risk Assessment were concurred on in a letter dated September 24, 1985. 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 have no other significant adverse effects on the economy. These Guidelines were reviewed by the Office of Management

[51 FR 34007]

and Budget under Executive Order 12291.

August 22, 1986

Lee M. Thomas,

Administrator

CONTENTS

Part A: Guidelines for Mutagenicity Risk Assessment

I. Introduction

A. Concepts Relating to Heritable Mutagenic Risk
B. Test Systems

II. Qualitative Assessment (Hazard Identification)

A. Mutagenic Activity
B. Chemical Interactions in the Mammalian Gonad
C. Weight-of-Evidence Determination

III. Quantitative Assessment

A. Dose Response
B. Exposure Assessment
C. Risk Characterization

IV. References

Part B: Response to Public and Science Advisory Board Comments

Part A: Guidelines for Mutagenicity Risk Assessment

I. Introduction

This section describes the procedures that the U.S. Environmental Protection Agency will follow in evaluating the potential genetic risk associated with human exposure to chemicals. The central purpose of the health risk assessment is to provide a judgment concerning the weight of evidence that an agent is a potential human mutagen, capable of inducing transmitted genetic changes, and, if so, to provide a judgment on how great an impact this agent is likely to have on public health. Regulatory decision making involves two components: risk assessment and risk management. Risk assessment estimates the potential adverse health consequences of exposure to toxic chemicals; risk management combines the risk assessment with the directives of the enabling 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 chemicals. The issue of risk management will not be dealt with in these Guidelines.

Risk assessment is comprised of the following components: hazard identification, dose-response assessment, exposure assessment, and risk characterization (1). Hazard identification is the qualitative risk assessment, dealing with the inherent toxicity of a chemical substance. The qualitative mutagenicity assessment answers the question of how likely an agent is to be a human mutagen. The three remaining components

comprise quantitative risk assessment, which provides a numerical estimate of the public health consequences of exposure to an agent. The quantitative mutagenicity risk assessment deals with the question of how much mutational damage is likely to be produced by exposure to a given agent under particular exposure scenarios.

In a dose-response assessment, the relationship between the dose of a chemical and the probability of induction of an adverse effect is defined. The component generally entails an extrapolation from the high doses administered to experimental animals or noted in some epidemiologic studies to the low exposure levels expected from human contact with the chemical in the environment.

The exposure assessment identifies populations exposed to toxic chemicals, describes their composition and size, and presents the types, magnitudes, frequencies, and durations of exposure to the chemicals. This component is developed independently of the other components of the mutagenicity assessment and is addressed in separate Agency guidelines (2).

In risk characterization, the outputs of the exposure assessment and the dose-response assessment are combined to estimate quantitatively the mutation risk, which is expressed as either estimated increase of genetic disease per generation or per lifetime, or the fractional increase in the assumed background mutation rate of humans. In each step of the assessment, the strengths and weaknesses of the major assumptions need to be presented, and the nature and magnitude of uncertainties need to be characterized.

The procedures set forth in these Guidelines will ensure consistency in the Agency's scientific risk assessments for mutagenic effects. The necessity for a consistent approach to the evaluation of mutagenic risk from chemical substances arises from the authority conferred upon the Agency by a number of statutes to regulate potential mutagens. As appropriate, these Guidelines will apply to statutes administered by the Agency, including the Federal Insecticide, Fungicide, and Rodenticide Act; the Toxic Substances Control Act; the Clean Air Act; the Federal Water Pollution Control Act; the Safe Drinking Water Act; the Resource Conservation and Recovery Act; and the Comprehensive Environmental Response, Compensation, and Liability Act. Because each statute is administered by separate offices, a consistent Agency-wide approach for performing risk assessments is desirable.

The mutagenicity risk assessments prepared pursuant to these Guidelines will be utilized with the requirements and constraints of the applicable statutes to arrive at regulatory decisions concerning mutagenicity. The standards of the applicable statutes and regulations may dictate that additional considerations (e.g., the economic and social benefits associated with use of the chemical substance) will come into play in reaching appropriate regulatory decisions.

The Agency has not attempted to provide in the Guidelines a detailed discussion of the mechanisms of mutagenicity or of the various test systems that are currently in use to detect mutagenic potential. Background information on mutagenesis and mutagenicity test systems is available in "Identifying and Estimating the Genetic Impact of Chemical Mutagens", National Academy of Sciences (NAS) Committee on Chemical Environmental Mutagens (3), as well as in other recent publications (4, 5).

The Agency is concerned with the risk associated with both germ-cell mutations and somatic-cell mutations. Mutations carried in germ cells may be inherited by future generations and may contribute to genetic disease, whereas mutations occurring in somatic cells may be implicated in the etiology of several disease states, including cancer. These Guidelines, however, are only concerned with genetic damage as it relates to germ-cell mutations. The use of mutagenicity test results in the assessment of carcinogenic risk is described in the Guidelines for Carcinogen Risk Assessment (6).

As a result of the progress in the control of infectious diseases, increases in average human life span, and better procedures for identifying genetic disorders, a considerable heritable genetic disease burden has been recognized in the human population. It is estimated that at least 10% of all human disease is related to specific genetic abnormalities, such as abnormal composition, arrangement, or dosage of genes and chromosomes (3, 7, 8). Such genetic abnormalities can lead to structural or functional health impairments. These conditions may be expressed *in utero*; at the time of birth; or during infancy, childhood, adolescence, or adult life; they may be chronic or acute in nature. As a result, they often have a severe impact upon the affected individuals and their families in terms of physical and mental

[51 FR 34008]

suffering and economic losses, and upon society in general, which often becomes responsible for institutional care of severely affected individuals. Some examples of genetic disorders are Down and Klinefelter syndromes, cystic fibrosis, hemophilia, sickle-cell anemia, and achondroplastic dwarfism. Other commonly recognized conditions that are likely to have a genetic component include hypercholesterolemia, hypertension, pyloric stenosis, glaucoma, allergies, several types of cancer, and mental retardation. These disorders are only a few of the thousands that are at least partially genetically determined (9).

Estimation of the fraction of human genetic disorders that result from new mutations is difficult, although in certain specific cases insights are available (10). It is clear that recurring mutation is important in determining the incidence of certain genetic disorders, such as some chromosomal aberration syndromes (e.g., Down syndrome) and rare dominant and X-linked recessive diseases (e.g., achondroplasia and hemophilia A). For other single-factor disorders (e.g., sickle-cell anemia) and certain multifactorial disorders (e.g., pyloric stenosis), the contribution of new mutations to disease frequency is probably small. However, it is generally recognized that most newly-arising mutations that are phenotypically expressed are in some ways deleterious to the organism receiving them (3, 7, 8). Adverse effects may be manifested at the biochemical, cellular, or physiological levels of organization. Although mutations are the building blocks for further evolutionary change of species, it is believed that increases in the mutation rate could lead to an increased frequency of expressed genetic disorders in the first and subsequent generations.

Life in our technological society results in exposure to many natural and synthetic chemicals. Some have been shown to have mutagenic activity in mammalian and submammalian test systems, and thus may have the potential to increase genetic damage in the human population. Chemicals exhibiting mutagenic activity in various test systems have been found distributed among foods, tobacco, drugs, food additives, cosmetics, industrial compounds, pesticides, and consumer products. The extent to which exposure to natural and synthetic environmental agents may have increased the frequency of genetic disorders in the present human population and contributed to the mutational "load" that will be transmitted to future generations is unknown at this time. However, for the reasons cited above, it seems prudent to limit exposures to potential human mutagens.

A. Concepts Relating to Heritable Mutagenic Risk

These Guidelines are concerned with chemical substances or mixtures of substances that can induce alterations in the genome of either somatic or germinal cells. The mutagenicity of physical agents (e.g., radiation) is not addressed here. There are several mutagenic end points of concern to the Agency. These include point mutations (i.e., submicroscopic changes in the base sequence of DNA) and structural or numerical chromosome aberrations. Structural aberrations include deficiencies, duplications, insertions, inversions, and translocations, whereas numerical aberrations are gains or losses of whole chromosomes (e.g., trisomy, monosomy) or sets of chromosomes (haploidy, polyploidy).

Certain mutagens, such as alkylating agents, can directly induce alterations in the DNA. Mutagenic effects may also come about through mechanisms other than chemical alterations of DNA. Among these are interference with normal DNA synthesis (as caused by some metal mutagens), interference with DNA repair, abnormal DNA methylation, abnormal nuclear division processes, or lesions in non-DNA targets (e.g., protamine, tubulin).

Evidence that an agent induces heritable mutations in human beings could be derived from epidemiologic data indicating a strong association between chemical exposure and heritable effects. It is difficult to obtain such data because any specific mutation is a rare event, and only a small fraction of the estimated thousands of human genes and conditions are currently useful as markers in estimating mutation rates. Human genetic variability, small numbers of offspring per individual, and long generation times further complicate such studies. In addition, only disorders caused by dominant mutations, some sex-linked recessive mutations, and certain chromosome aberrations can be detected in the first generation after their occurrence. Conditions caused by autosomal recessive disorders (which appear to occur more frequently than dominant disorders) or by polygenic traits may go unrecognized for many generations. Therefore, in the absence of human epidemiological data, it is appropriate to rely on data from experimental animal systems as long as the limitations of using surrogate and model systems are clearly stated.

Despite species differences in metabolism, DNA repair, and other physiological processes affecting chemical mutagenesis, the virtual universality of DNA as the genetic material and of the genetic code provides a rationale for using various nonhuman test systems to predict the intrinsic mutagenicity of test chemicals. Additional support for the use of nonhuman systems is provided by the observation that chemicals causing genetic effects in one species or test system frequently cause similar effects in other species or systems. Evidence also exists that chemicals can induce genetic damage in somatic cells of exposed humans. For example, high doses of mutagenic chemotherapeutic agents have been shown to cause chromosomal abnormalities (11), sister chromatic exchange (11), and, quite probably, point mutations in human lymphocytes exposed *in vivo* (12). While these results are not in germ cells, they do indicate that it is possible to induce mutagenic events in human cells *in vivo*. Furthermore, a wide variety of different types of mutations have been observed in humans including numerical chromosome aberrations, translocations, base-pair substitutions, and frameshift mutations. Although the cause of these mutations is uncertain, it is clear from these observations that the human germ-cell DNA is subject to the same types of

mutational events that are observed in other species and test systems.

Certain test systems offer notable advantages: cost; anatomical, histological, and/or metabolic similarities to humans; suitability for handling large numbers of test organisms; a large data base; or a basis for characterizing genetic events.

B. Test Systems

Many test systems are currently available that can contribute information about the mutagenic potential of a test compound with respect to various genetic end points. These tests have recently been evaluated through the EPA Gene-Tox Programs and the results of Phase I have been published (5). The Agency's Office of Pesticides and Toxic Substances has published various testing guidelines for the detection of mutagenic effects (13, 14).

Test systems for detecting point mutations include those in bacteria, eukaryotic microorganisms, higher plants, insects, mammalian somatic cells in culture, and germinal cells of intact mammals. Data from heritable, mammalian germ-cell tests provide the best experimental evidence that a

[51 FR 34009]

chemical is a potential human germ-cell mutagen since these tests require that mutations occur in germinal cells and that they are transmitted to the next generation. To date, the most extensively used test for the induction of heritable mutation is the mouse specific-locus test which measures the induction of recessive mutations at seven loci concerned with coat color and ear morphology. While this test has a large data base compared to other germ-cell assays, it is difficult to extrapolate results to humans since recessive mutations may occur more frequently than dominants, and the impact of recessive mutations is not seen for many generations. Information on frequencies of induced mutations resulting in health disorders in the first generation may be obtained from mouse systems designed to detect skeletal abnormalities, cataracts, or general morphological abnormalities. However, these assays have been used to a relatively limited extent, and there is a need for additional studies with known, chemical germ-cell mutagens to further characterize the test systems. Because large numbers of offspring must usually be generated in the systems described above, it is not expected that many chemicals will be tested using these systems. To obtain data on a large number of environmental chemicals, it will be necessary to rely on other tests to identify and characterize hazards from gene mutations.

Test systems for detecting structural chromosome aberrations have been developed in a variety of organisms including higher plants, insects, fish, birds, and several mammalian species. Many of these assays can be performed *in vitro* or *in vivo*, and in either germ or somatic cells. Procedures available for detecting structural chromosome aberrations in mammalian germ cells include measurement of heritable translocations or dominant lethality, as well as direct cytogenetic analyses of germ cells and early embryos in rodents.

Some chemicals may cause numerical chromosome changes (i.e., aneuploidy) as their sole mutagenic effect. These agents may not be detected as mutagens if evaluated only in tests for DNA damage, gene mutations, or chromosome breakage and rearrangement. Therefore, it is important to consider tests for changes in chromosome number in the total assessment of mutagenic hazards. Although tests for the detection of variation in the chromosome number are still at an early stage of development, systems exist in such diverse organisms as fungi, Drosophila, mammalian cells in culture, and intact mammals (e.g., mouse X-chromosome loss assay). Aneuploidy can arise from disturbances in a number of events affecting the meiotic process (15, 16). Although the mechanisms by which nondisjunction occurs are not well understood, mitotic structures other than DNA may be the target molecules for at least some mechanisms of induced nondisjunction.

Other end points that provide information bearing on the mutagenicity of a chemical can be detected by a variety of test systems. Such tests measure DNA damage in eukaryotic or prokaryotic cells, unscheduled DNA synthesis in mammalian somatic and germ cells, mitotic recombination and gene conversion in yeast, and sister-chromatid exchange in mammalian somatic and germ cells. Results in these assays are useful because the induction of these end points often correlates positively with the potential of a chemical to induce mutations.

In general, for all three end points (i.e., point mutations and numerical and structural aberrations), the Agency will place greater weight on tests conducted in germ cells than in somatic cells, on tests performed *in vivo* rather than *in vitro*, in eukaryotes rather than prokaryotes, and in mammalian species rather than in submammalian species. Formal numerical weighting systems have been developed (17); however, the Agency has concluded that these do not readily accommodate such variables as dose range, route of exposure, and magnitude of response.

The Agency anticipates that from time to time somatic cell data from chemically exposed human beings will be available (e.g., cytogenetic markers in peripheral lymphocytes). When possible, the Agency will use such data in conjunction with somatic and germ cell comparisons from *in vivo* mammalian experimental systems as a component in performing risk assessments.

The test systems mentioned previously are not the only ones that will provide evidence of mutagenicity or related DNA effects. These systems are enumerated merely to demonstrate the breadth of the available techniques for characterizing mutagenic hazards, and to indicate the types of data that the Agency will consider in its evaluation of mutagenic potential of a chemical agent. Most systems possess certain limitations that must be taken into account. The selection and performance of appropriate tests for evaluating the risks associated with human exposure to any suspected mutagen will depend on sound scientific judgment and experience, and may necessitate consultation with geneticists familiar with the sensitivity and experimental design of the test system in question. In view of the rapid advances in test methodology, the Agency expects that both the number and quality of the tools for assessing genetic risk to human beings will increase with time. The Agency will closely monitor developments in mutagenicity evaluation and will refine its risk assessment scheme as better test systems become available.

II. Qualitative Assessment (Hazard Identification)

The assessment of potential human germ-cell mutagenic risk is a multistep process. The first step is an analysis of the evidence bearing on a chemical's ability to induce mutagenic events, while the second step involves an analysis of its ability to produce these events in the mammalian gonad. All relevant information is then integrated into a weight-of-evidence scheme which presents the strength of the information bearing on the chemical's potential ability to produce mutations in human germ cells. For chemicals demonstrating this potential, one may decide to proceed with an evaluation of the quantitative consequences of mutation following expected human exposure.

For hazard identification, it is clearly desirable to have data from mammalian germ-cell tests, such as the mouse specific-locus test for point mutations and the heritable translocation or germ-cell cytogenetic tests for structural chromosome aberrations. It is recognized, however, that in most instances such data will not be available, and alternative means of evaluation will be required. In such cases the Agency will evaluate the evidence bearing on the agent's mutagenic activity and the agent's ability to interact with or affect the mammalian gonadal target. When evidence exists that an agent possesses both these attributes, it is reasonable to deduce that the agent is a potential human germ-cell mutagen.

While mammalian germ-cell assays are presently primarily performed on male animals, a chemical cannot be considered to be a non-mutagen for mammalian germ cells unless it is shown to be negative in both sexes. Furthermore, because most mammalian germ-cell assays are performed in mice, it is noteworthy that the data from ionizing radiation suggest that the female mouse immature oocyte may not be an appropriate surrogate for the same stage in the human female in mutagenicity testing. However,

[51 FR 34010]

mutagenicity data on the maturing and mature oocyte of the mouse may provide a useful model for human risk assessment.

A. Mutagenic Activity

In evaluating chemicals for mutagenic activity, a number of factors will be considered: 1) genetic end points (e.g., gene mutations, structural or numerical chromosomal aberrations) detected by the test systems, 2) sensitivity and predictive value of the test systems for various classes of chemical compounds, 3) number of different test systems used for detecting each genetic end point, 4) consistency of the results obtained in different test systems and different species, 5) aspects of the dose-response relationship, and 6) whether the tests are conducted in accordance with appropriate test protocols agreed upon by experts in the field.

B. Chemical Interactions in the Mammalian Gonad

Evidence for chemical interaction in the mammalian gonad spans a range of different types of findings. Each chemical under consideration needs to be extensively reviewed since this type of evidence may be part of testing exclusive of mutagenicity per se (e.g., reproduction, metabolism, and mechanistic investigations). Although it is not possible to classify clearly each type of information that may be available on a chemical, two possible groups are illustrated.

1. *Sufficient evidence* of chemical interaction is given by the demonstration that an agent interacts with germ-cell DNA or other chromatin constituents, or that it induces such end points as unscheduled DNA synthesis, sister-chromatid exchange, or chromosomal aberrations in germinal cells.

2. *Suggestive evidence* will include the finding of adverse gonadal effects such as sperm abnormalities following acute, subchronic, or chronic toxicity testing, or findings of adverse reproductive effects such as decreased fertility, which are consistent with the chemical's interaction with germ cells.

C. Weight-of-Evidence Determination

The evidence for a chemical's ability to produce mutations and to interact with the germinal target are integrated into a weight-of-evidence judgment that the agent may pose a hazard as a potential human germ-cell mutagen. All information bearing on the subject, whether indicative of potential concern or not, must be evaluated. Whatever

evidence may exist from humans must also be factored into the assessment.

All germ-cell stages are important in evaluating chemicals because some chemicals have been shown to be positive in postgonial stages but not in gonia (18). When human exposures occur, effects on postgonial stages should be weighted by the relative sensitivity and the duration of the stages. Chemicals may show positive effects for some end points and in some test systems, but negative responses in others. Each review must take into account the limitations in the testing and in the types of responses that may exist.

To provide guidance as to the categorization of the weight of evidence, a classification scheme is presented to illustrate, in a simplified sense, the strength of the information bearing on the potential for human germ-cell mutagenicity. It is not possible to illustrate all potential combinations of evidence, and considerable judgment must be exercised in reaching conclusions. In addition, certain responses in tests that do not measure direct mutagenic end points (e.g., SCE induction in mammalian germ cells) may provide a basis for raising the weight of evidence from one category to another. The categories are presented in decreasing order of strength of evidence.

1. Positive data derived from human germ-cell mutagenicity studies, when available, will constitute the highest level of evidence for human mutagenicity.
2. Valid positive results from studies on heritable mutational events (of any kind) in mammalian germ cells.
3. Valid positive results from mammalian germ-cell chromosome aberration studies that do not include an intergeneration test.
4. Sufficient evidence for a chemical's interaction with mammalian germ cells, together with valid positive mutagenicity test results from two assay systems, at least one of which is mammalian (*in vitro* or *in vivo*). The positive results may both be for gene mutations or both for chromosome aberrations; if one is for gene mutations and the other for chromosome aberrations, both must be from mammalian systems.
5. Suggestive evidence for a chemical's interaction with mammalian germ cells, together with valid positive mutagenicity evidence from two assay systems as described under 4, above. Alternatively, positive mutagenicity evidence of less strength than defined under 4, above, when combined with sufficient evidence for a chemical's interaction with mammalian germ cells.
6. Positive mutagenicity test results of less strength than defined under 4, combined with suggestive evidence for a chemical's interaction with mammalian germ cells.
7. Although definitive proof of non-mutagenicity is not possible, a chemical could be classified operationally as a non-mutagen for human germ cells, if it gives valid negative test results for all end points of concern.
8. Inadequate evidence bearing on either mutagenicity or chemical interaction with mammalian germ cells.

III. Quantitative Assessment

The preceding section addressed primarily the processes of hazard identification, i.e., the determination of whether a substance is a potential germ-cell mutagen. Often, no further data will be available, and judgments will need to be based mainly on qualitative criteria. Quantitative risk assessment is a two-step process: determination of the heritable effect per unit of exposure (dose-response) and the relationship between mutation rate and disease incidence. The procedures that are presently accepted for the estimation of an increase in disease resulting from increased mutation have been described (3, 7, 8). Dose-response information is combined with anticipated levels and patterns of human exposure in order to derive a quantitative assessment (risk characterization).

A. Dose Response

Dose-response assessments can presently only be performed using data from *in vivo*, heritable mammalian germ-cell tests, until such time as other approaches can be demonstrated to have equivalent predictability. The morphological specific-locus and biochemical specific-locus assays can provide data on the frequencies of recessive mutations induced by different chemical exposure levels, and similar data can be obtained for heritable chromosomal damage using the heritable translocation test. Data on the frequencies of induced mutations resulting in health disorders in the first generation may be obtained from mouse systems designed to detect skeletal abnormalities, cataracts, or general morphological abnormalities. Assays that directly detect heritable health effects in the first generation may provide the best basis for predicting human health risks that result from mutagen exposure. The experimental data on induced mutation frequency are usually obtained at exposure levels much higher than those that will be experienced by human beings. An assessment of human risk is obtained by extrapolating the induced mutation frequency or the observed phenotypic

[51 FR 34011]

effect downward to the approximate level of anticipated human exposure. In performing these extrapolations, the Agency will place greater weight on data derived from exposures and exposure rates that most closely simulate those experienced by the human population under study.

The Agency will strive to use the most appropriate extrapolation models for risk analysis

and will be guided by the available data and mechanistic considerations in this selection. However, it is anticipated that for tests involving germ cells of whole mammals, few dose points will be available to define dose-response functions. The Agency is aware that for at least one chemical that has been tested for mutations in mammalian germ cells, there exist departures from linearity at low exposure and exposure rates in a fashion similar to that seen for ionizing radiation that has a low linear energy transfer (19). The Agency will consider all relevant models for gene and chromosomal mutations in performing low-dose extrapolations and will choose the most appropriate model. This choice will be consistent both with the experimental data available and with current knowledge of relevant mutational mechanisms.

An experimental approach for quantitative assessment of genetic risk, which may have utility in the future, uses molecular dosimetry data from intact mammals in conjunction with mutagenicity and dosimetry data from other validated test systems (20). The intact mammal is used primarily for relating the exposure level for a given route of administration of a chemical to germ-cell dose, i.e., the level of mutagen-DNA interactions. This information is then used in conjunction with results obtained from mutagenicity test systems in which the relationship between the induction of mutations and chemical interactions with DNA can be derived. With mutagen-DNA interactions as the common denominator, a relationship can be constructed between mammalian exposure and the induced mutation frequency. The amount of DNA binding induced by a particular chemical agent may often be determined at levels of anticipated human exposure.

For some mutagenic events, DNA may not necessarily be the critical target. Interaction of chemicals with other macromolecules, such as tubulin, which is involved in the separation of chromosomes during nuclear division, can lead to chromosomal nondisjunction. At present, general approaches are not available for dose-response assessments for these types of mutations. Ongoing research should provide the means to make future assessments on chemicals causing aneuploidy.

B. Exposure Asssessment

The exposure assessment identifies populations exposed to toxic chemicals; describes their composition and size; and presents the types, magnitudes, frequencies, and durations of exposure to the chemicals. This component is developed independently of the other components of the mutagenicity assessment (2).

C. Risk Characterization

In performing mutagenicity risk assessments, it is important to consider each genetic end point individually. For example, although certain chemical substances that interact with DNA may cause both point and chromosomal mutations, it is expected that the ratio of these events may differ among chemicals and between doses for a given chemical. Furthermore, transmissible chromosomal aberrations are recoverable with higher frequencies from meiotic and postmeiotic germ-cell stages, which have a brief life span, than in spermatogonial stem cells, which can accumulate genetic damage throughout the reproductive life of an individual. For these reasons, when data are available, the Agency, to the best extent possible, will assess risks associated with all genetic end points.

Any risk assessment should clearly delineate the strengths and weaknesses of the data, the assumptions made, the uncertainties in the methodology, and the rationale used in reaching the conclusions, e.g., similar or different routes of exposure and metabolic differences between humans and test animals. When possible, quantitative risk assessments should be expressed in terms of the estimated increase of genetic disease per generation, or the fractional increase in the assumed background spontaneous mutation rate of humans (7). Examples of quantitative risk estimates have been published (7, 8, 21); these examples may be of use in performing quantitative risk assessments for mutagens.

IV. References

1. Committee on the Institutional Means for Assessment of the Risks to Public Health. 1983. Risk assessment in the Federal government: managing the process. Washington, DC: National Academy Press.

2. U.S. Environmental Protection Agency. 1986, Sept. 24. Guidelines for estimating exposures. Federal Register 51(185): 34042-34054.

3. Committee on Chemical Environmental Mutagens, National Academy of Sciences. 1982. Identifying and estimating the genetic impact of chemical mutagens. Washington, DC: National Academy Press.

4. Committee 1 Final Report. 1983. Screening strategy for chemicals that are potential germ-cell mutagens in mammals. Mutat. Res. 114:117-177.

5. A complete reference of all Gene-Tox publications is available from the TSCA Industry Assistance Office (TS-794), Office of Toxic Substances, U.S. Environmental Protection Agency, Washington, DC 20460.

6. U.S. Environmental Protection Agency. 1986, Sept. 24. Guidelines for carcinogen risk assessment. Federal Register 51(185):33992-34003.

7. National Research Council, Advisory Committee on the Biological Effects of Ionizing Radiations. 1972. The effects on populations of exposure to low levels of ionizing radiation. (BEIR I) Washington, DC: National Academy of Sciences.

National Research Council, Advisory Committee on the Biological Effects of Ionizing Radiations. 1977. Considerations of health benefit-cost analysis for activities involving ionizing radiation exposure and alternatives. (BEIR II) Washington, DC: National Academy of Sciences.

National Research Council, Advisory Committee on the Biological Effects of Ionizing Radiations. 1980. The effects on populations of exposure to low levels of ionizing radiation. (BEIR III) Washington, DC: National Academy of Sciences.

8. United Nations General Assembly. 1958. Report of the United Nations Scientific Committee on the Effects of Atomic

Radiation. Official records of the General Assembly. Thirteenth Session. Supplement No. 17 (A/3838). New York: United Nations.

United Nations General Assembly. 1962. Report of the United Nations Scientific Committee on the Effects of Atomic Radiation. Official records of the General Assembly. Seventeenth Session. Supplement No. 16 (A/5216). New York: United Nations.

United Nations General Assembly. 1966. Report of the United Nations Scientific Committee on the Effects of Atomic Radiation. Official records of the General Assembly. Twenty-first Session. Supplement No. 14 (A/6314). New York: United Nations.

United Nations General Assembly. 1969. Report of the United Nations Scientific Committee on the effects of Atomic Radiation. Official records of the General Assembly. Twenty-fourth Session. Supplement No. 13 (A/7613). New York: United Nations.

United Nations General Assembly. 1972. Report of the United Nations Scientific Committee on the Effects of Atomic Radiation. Ionizing radiation: levels and effects. Vol. II. Official records of the General Assembly. Twenty-seventh Session. Supplement No. 25 (A/8725). New York: United Nations.

United Nations General Assembly. 1982. Report of the United Nations Scientific Committee on the Effects of Atomic Radiation. Sources and biological effects. New York: United Nations.

9. McKusick, V.A. 1983. Mendelian inheritance in man: catalogs of autosomal dominant, autosomal recessive and x-linked [51 FR 34012] phenotypes. Baltimore, MD: Johns Hopkins University Press.

10. Crow, J.F., and C. Denniston. 1981. The mutation component of genetic damage. Science 212:888-893.

11. Musilova, J., K. Michalova, and J. Urban. 1979. Sister chromatid exchanges and chromosomal breakage in patients treated with cytostatics. Mutat. Res. 67:289-294.

12. Strauss, G.H., and R.J. Albertini. 1979. Enumeration of 6- thioguanine-resistant peripheral blood lymphocytes in man as a potential test for somatic cell mutations arising *in vivo*. Mutat. Res. 61:353-379.

13. U.S. Environmental Protection Agency. 1983. OTS health effects test guidelines. EPA 560/6/82-001. Natl. Tech. Inform. Serv. PB82-232984.

14. U.S. Environmental Protection Agency. November 24, 1982. Pesticides registration; proposed data requirements. Federal Register 47:53192-53203.

15. Parker, D.R., and J.H. Williamson. 1974. Some radiation effects on segregation in Drosophila. Genetics 78:163-171.

16. Grell, R.F. 1979. Origin of meiotic nondisjunction in Drosophila females. Environ. Health Perspect. 31:33-39.

17. Russell, L.B., C.S. Aaron, F. de Serres, W.M. Generoso, K.L. Kannan, M. Shelby, J. Springer, and P. Voytek. 1984. Evaluation of existing mutagenicity bioassays for purposes of genetic risk assessment. Mutat. Res. 134:143-157.

18. Russell, L.B., R.B. Cumming, and P.R. Hunsicker. 1984. Specific locus mutation rates in the mouse following inhalation of ethylene oxide, and application of results to estimation of genetic risk. Mutat. Res. 129:381-388.

19. Russell, W.L., P.R. Hunsicker, D.A. Carpenter, C.V. Cornett, and G.M. Guin. 1982. Effect of dose fractionation on the ethylnitrosourea induction of specific-locus mutations in mouse spermatogonia. Proc. Natl. Acad. Sci. 79:3592-3593.

20. Lee, W.R. 1979. Dosimetry of chemical mutagens in eukaryote germ- cells. In. A. Hollaender and F.J. de Serres, eds. Chemical mutagens: principles and methods for their detection. Vol. 5, New York: Plenum Press, pp. 177-202.

21. Ehling, U.H., and A. Neuhauser. 1979. Procarbazine-induced specific-locus mutations in male mice. Mutat. Res. 59:245-256.

Part B: Response to Public and Science Advisory Board Comments

This section summarizes some of the issues raised in public and Science Advisory Board (SAB) comments on the Proposed Guidelines for Mutagenicity Risk Assessment published on November 23, 1984 (49 FR 46314). Unlike the other guidelines published on the same date, the Proposed Guidelines for Mutagenicity Risk Assessment contained a detailed section dealing with public comments received in response to the original proposal of 1980 (45 FR 74984). Several of the comments received in response to the proposed guidelines of 1984 were similar to those received in response to the proposed guidelines of 1980. Those comments are not addressed here because the position of the Agency on those issues has been presented in the responses included with the 1984 proposed guidelines (49 FR 46315- 46316).

A total of 44 comments were received in response to the proposed guidelines of 1984: 21 from manufacturers of regulated products, 10 from associations, 9 from government agencies, 2 from educational institutions, 1 from an individual, and 1 from a private consulting firm. The proposed guidelines and the public comments received were transmitted to the Agency's SAB prior to its public review of the proposed guidelines held April 22-23, 1985. The majority of the comments were favorable and expressed the opinion that the proposed guidelines accurately represent the existing state of knowledge in the field of mutagenesis. Several commentors offered suggestions for further clarification of particular issues, and many of the suggestions have been incorporated.

The two areas that received the most substantive comments were the sections concerning Weight-of-Evidence Determination and Dose Response. The comments on the proposed weight-of-evidence scheme ranged from suggestions for the elimination of a formal scheme to the expansion of the scheme to cover more potential data configurations. The SAB recommended an eight-level rank ordering scheme to define levels of evidence relating to human germ-cell mutagenicity. The Agency has incorporated this scheme into the Guidelines. Some commentors and the SAB suggested that the molecular dosimetry approach to dose-response data be presented as a concept that may be useful in the future rather than being available for use now. The Agency agrees that the data base at the present time is too sparse to recommend a general application of this approach to a wide range of chemical classes, and the Guidelines have been changed to reflect this. It should be noted, however, that the Agency strongly supports the development of molecular dosimetry methodologies as they relate to both an understanding of dose-response relationships and to methods for studying human exposure. A number of comments suggesting clarifications and editorial changes have been incorporated and the references have been expanded.

51 FR 34014

GUIDELINES FOR THE HEALTH RISK ASSESSMENT OF CHEMICAL MIXTURES

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 the Health Risk Assessment of Chemical Mixtures.

The Guidelines for the Health Risk Assessment of Chemical Mixtures (hereafter "Guidelines") are intended to guide Agency analysis of information relating to health effects data on chemical mixtures 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 the Health Risk Assessment of Chemical Mixtures published January 9, 1985 (50 FR 1170).

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. Richard Hertzberg
Methods Evaluation and Development Staff
Environmental Criteria and Assessment Office
U.S. Environmental Protection Agency
26 W. St. Clair Street
Cincinnati, OH 45268
513-569-7582

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 the Health Risk Assessment of Chemical Mixtures

Work on the Guidelines for the Health Risk Assessment of Chemical Mixtures 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 fields of toxicology, pharmacokinetics, and statistics from universities, environmental groups, industry, labor, and other governmental agencies. They were then proposed for public comment in the *FEDERAL REGISTER* (50 FR 1170). 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 preliminary Agency responses to those comments. 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 the Health Risk Assessment of Chemical Mixtures were concurred on in a letter dated August 16, 1985. 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 SAB requested that the Agency develop a technical support document for these Guidelines. The SAB identified the need for this type of document due to the limited knowledge on interactions of chemicals in biological systems. Because of this, the SAB commented that progress in improving risk assessment will be particularly dependent upon progress in the science of interactions.

Agency staff have begun preliminary work on the technical support document and expect it to be completed by early 1987. 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.

[51 FR 34015]

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 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. Thomas,

Administrator

CONTENTS

Part A: Guidelines for the Health Risk Assessment of Chemical Mixtures

I. Introduction

II. Proposed Approach

A. Data Available on the Mixture of Concern
B. Data Available on Similar Mixtures
C. Data Available Only on Mixture Components
1. Systemic Toxicants
2. Carcinogens
3. Interactions
4. Uncertainties
a. Health Effects
b. Exposure Uncertainties
c. Uncertainties Regarding Composition of the Mixture

III. Assumptions and Limitations

A. Information on Interactions
B. Additivity Models

IV. Mathematical Models and the Measurement of Joint Action

A. Dose Addition
B. Response Addition
C. Interactions

V. References

Part B: Response to Public and Science Advisory Board Comments

I. Introduction

II. Recommended Procedures

A. Definitions
B. Mixtures of Carcinogens and Systemic Toxicants

III. Additivity Assumption

A. Complex Mixtures
B. Dose Additivity
C. Interpretation of the Hazard Index
D. Use of Interaction Data

IV. Uncertainties and the Sufficiency of the Data Base

V. Need for a Technical Support Document

Part A: Guidelines for the Health Risk Assessment of Chemical Mixtures

I. Introduction

The primary purpose of this document is to generate a consistent Agency approach for evaluating data on the chronic and subchronic effects of chemical mixtures. It is a procedural guide that emphasizes broad underlying principles of the various science disciplines (toxicology, pharmacology, statistics) necessary for assessing health risk from chemical mixture exposure. Approaches to be used with respect to the analysis and evaluation of the various data are also discussed.

It is not the intent of these Guidelines to regulate any social or economic aspects concerning risk of injury to human health or the environment caused by exposure to a chemical agent(s). All such action is addressed in specific statutes and federal legislation and is independent of these Guidelines.

While some potential environmental hazards involve significant exposure to only a single compound, most instances of environmental contamination involve concurrent or sequential exposures to a mixture of compounds that may induce similar or dissimilar effects over exposure periods ranging from short-term to lifetime. For the purposes of these Guidelines, mixtures will be defined as any combination of two or more chemical substances regardless of source or of spatial or temporal proximity. In some instances, the mixtures are highly complex consisting of scores of compounds that are generated simultaneously as by-products from a single source or process (e.g., coke oven emissions and diesel exhaust). In other cases, complex mixtures of related compounds are produced as commercial products (e.g., PCBs, gasoline and pesticide formulations) and eventually released to the environment. Another class of mixtures consists of compounds, often unrelated chemically or commercially, which are placed in the same area for disposal or storage, eventually come into contact with each other, and are released as a mixture to the environment. The quality and quantity of pertinent information available for risk assessment varies considerably for different mixtures. Occasionally, the chemical composition of a mixture is well characterized, levels of exposure to the population are known, and detailed toxicologic data on the mixture are available. Most frequently, not all components of the mixture are known, exposure data are uncertain, and toxicologic data on the known components of the mixture are limited. Nonetheless, the Agency may be required to take action because of the number of individuals at potential risk or because of the known toxicologic effects of these compounds that have been identified in the mixture.

The prediction of how specific mixtures of toxicants will interact must be based on an understanding of the mechanisms of such interactions. Most reviews and texts that discuss toxicant interactions attempt to discuss the biological or chemical bases of the interactions (e.g., Klaassen and Doull, 1980; Levine, 1973; Goldstein et al., 1974; NRC, 1980a; Veldstra, 1956; Withey, 1981). Although different authors use somewhat different classification schemes when discussing the ways in which toxicants interact, it generally is recognized that toxicant interactions may occur during any of the toxicologic processes that take place with a single compound: absorption, distribution, metabolism, excretion, and activity at the receptor site(s). Compounds may interact chemically, yielding a new toxic component or causing a change in the biological availability of the existing component. They may also interact by causing different effects at different receptor sites.

Because of the uncertainties inherent in predicting the magnitude and nature of toxicant interactions, the assessment of health risk from chemical mixtures must include a thorough discussion of all assumptions. No single approach is recommended in these Guidelines. Instead, guidance is given for the use of several approaches depending on the nature and quality of the data. Additional mathematical details are presented in section IV.

In addition to these Guidelines, a supplemental technical support document is being developed which will contain a thorough review of all available information on the toxicity of chemical mixtures and a discussion of research needs.

II. Proposed Approach

No single approach can be recommended to risk assessments for multiple chemical exposures. Nonetheless, general guidelines can be recommended depending on the type of mixture, the known toxic effects of its components, the availability of toxicity data on the mixture or similar mixtures,

[51 FR 34016]

the known or anticipated interactions among components of the mixture, and the quality of the exposure data. Given the complexity of this issue and the relative paucity of empirical data from which sound generalizations can be constructed, emphasis must be placed on flexibility, judgment, and a clear articulation of the assumptions and limitations in any risk assessment that is developed. The proposed approach is summarized in Table 1 and Figure 1 and is detailed below. An alphanumeric scheme for ranking the quality of the data used in the risk assessment is given in Table 2.

A. Data Available on the Mixture of Concern

For predicting the effects of subchronic or chronic exposure to mixtures, the preferred approach usually will be to use subchronic or chronic health effects data on the mixture of concern and adopt procedures similar to those used for single compounds, either systemic toxicants or carcinogens (see U.S. EPA, 1986a-1986c). The risk assessor must recognize, however, that dose-response models used for single compounds are often based on biological mechanisms of the toxicity of single compounds, and may not be as well justified when applied to the mixture as a whole. Such data are most likely to be available on highly complex mixtures, such as coke oven emissions or diesel exhaust, which are generated in large quantities and associated with or suspected of causing adverse health effects. Attention should also be given to the persistence of the mixture in the environment as well as to the variability of the mixture composition over time or from different sources of emissions. If the components of the mixture are known to partition into different environmental compartments or to degrade or transform at different rates in the environment, then those factors must also be taken into account, or the confidence in and applicability of the risk assessment is diminished.

TABLE 1.-- RISK ASSESSMENT APPROACH FOR CHEMICAL MIXTURES

1. Assess the quality of the data on interactions, health effects, and exposure (see Table 2).
   a. If adequate, proceed to Step 2.
   b. If inadequate, proceed to Step 14.

2. Health effects information is available on the chemical mixture of concern.
   a. If yes, proceed to Step 3.
   b. If no, proceed to Step 4.

3. Conduct risk assessment on the mixture of concern based on health effects data on the mixture. Use the same procedures as those for single compounds. Proceed to Step 7 (optional) and Step 12.

4. Health effects information is available on a mixture that is similar to the mixture of concern.
   a. If yes, proceed to Step 5.
   b. If no, proceed to Step 7.

5. Assess the similarity of the mixture on which health effects data are available to the mixture of concern, with emphasis on any differences in components or proportions of components, as well as the effects that such differences would have on biological activity.
   a. If sufficiently similar, proceed to Step 6.
   b. If not sufficiently similar, proceed to Step 7.

6. Conduct risk assessment on the mixture of concern based on health effects data on the similar mixture. Use the same procedures as those for single compounds. Proceed to Step 7 (optional) and Step 12.

7. Compile health effects and exposure information on the components of the mixture.

8. Derive appropriate indices of acceptable exposure and/or risk on the individual components in the mixture. Proceed to Step 9.

9. Assess data on interactions of components in the mixtures.
   a. If sufficient quantitative data are available on the interactions of two or more components in the mixture, proceed to Step 10.
   b. If sufficient quantitative data are not available, use whatever information is available to qualitatively indicate the nature of potential interactions. Proceed to Step 11.

10. Use an appropriate interaction model to combine risk assessments on compounds for which data are adequate, and use an additivity assumption for the remaining compounds. Proceed to Step 11 (optional) and Step 12.

11. Develop a risk assessment based on an additivity approach for all compounds in the mixture. Proceed to Step 12.

12. Compare risk assessments conducted in Steps 5, 8, and 9. Identify and justify the preferred assessment, and quantify uncertainty, if possible. Proceed to Step 13.

13. Develop an integrated summary of the qualitative and quantitative assessments with special emphasis on uncertainties and assumptions. Classify the overall quality of the risk assessment, as indicated in Table 2. Stop.

14. No risk assessment can be conducted because of inadequate data on interactions, health effects, or exposure. Qualitatively assess the nature of any potential hazard and detail the types of additional data necessary to support a risk assessment. Stop.

Note. -- Several decisions used here, especially those concerning adequacy of data and similarity between two mixtures, are not precisely characterized and will require considerable judgment. See text.

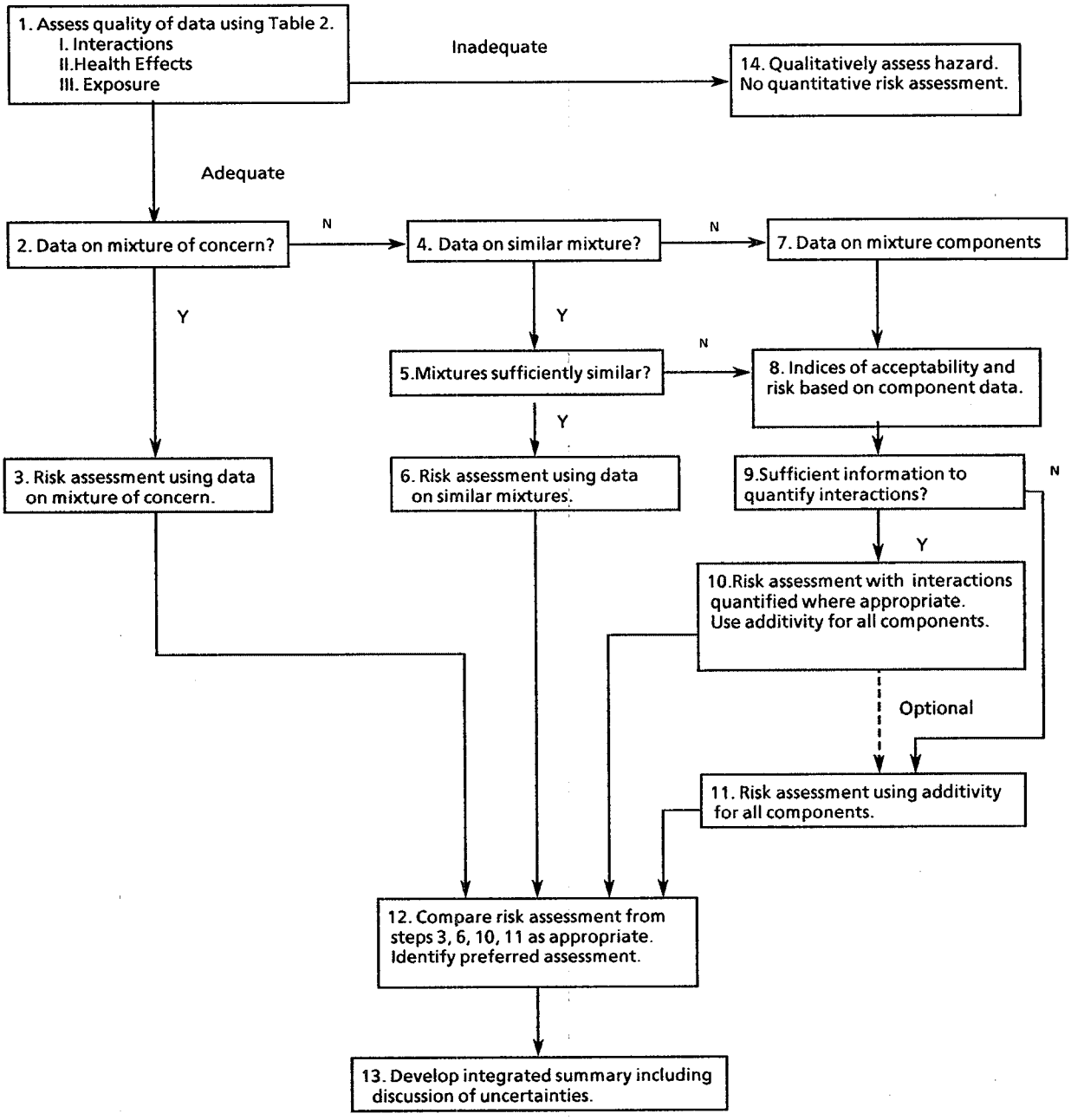

Figure 1. Flow chart of the risk assessment approach in Table 1. Note that it may be desirable to conduct all three assessments when possible (i.e., using data on the mixture, a similar mixture ,or the components) in order to make the fullest use of the available data. See text for further discussion.

TABLE 2. -- CLASSIFICATION SCHEME FOR THE QUALITY OF THE RISK ASSESSMENT OF THE MIXTURE[a]

Information on Interactions

I. Assessment is based on data on the mixture of concern.

II. Assessment is based on data on a sufficiently similar mixture.

III. Quantitative interactions of components are well characterized.

IV. The assumption of additivity is justified based on the nature of the health effects and on the number of component compounds.

V. An assumption of additivity cannot be justified, and no quantitative risk assessment can be conducted.

Health Effects Information

A. Full health effects data are available and relatively minor extrapolation is required.

B. Full health effects data are available but extensive extrapolation is required for route or duration of exposure or for species differences. These extrapolations are supported by pharmacokinetic considerations, empirical observations, or other relevant information.

C. Full health effects data are available, but extensive extrapolation is required for route or duration of exposure or for species differences. These extrapolations are not directly supported by the information available.

D. Certain important health effects data are lacking and extensive extrapolations are required for route or duration of exposure or for species differences.

E. A lack of health effects information on the mixture and its components in the mixture precludes a quantitative risk assessment.

Exposure Information[b]

1. Monitoring information either alone or in combination with modeling information is sufficient to accurately characterize human exposure to the mixture or its components.

2. Modeling information is sufficient to reasonably characterize human exposure to the mixture or its components.

3. Exposure estimates for some components are lacking, uncertain, or variable. Information on health effects or environmental chemistry suggest that this limitation is not likely to substantially affect the risk assessment.

4. Not all components in the mixture have been identified or levels of exposure are highly uncertain or variable. Information on health effects or environmental chemistry is not sufficient to assess the effect of this limitation on the risk assessment.

5. The available exposure information is insufficient for conducting a risk assessment.

B. Data Available on Similar Mixtures

If the risk assessment is based on data from a single mixture that is known to be generated with varying compositions depending on time or different emission sources, then the confidence in the applicability of the data to a risk assessment also is diminished. This can be offset to some degree if data are available on several mixtures of the same components that have different component ratios which encompass the temporal or spatial differences in composition of the mixture of concern. If such data are available, an attempt should be made to determine if significant and systematic differences exist among the chemical mixtures. If significant differences are noted, ranges of risk can be estimated based on the toxicologic data of the various mixtures. If no significant differences are noted, then a single risk assessment may be adequate, although the range of ratios of the components in the mixtures to which the risk assessment applies should also be given.

If no data are available on the mixtures of concern, but health effects data are available on a similar mixture (i.e., a mixture having the same components but in slightly different ratios, or having several common components but lacking one or more components, or having one or more additional components), a decision must be made whether the mixture on which health effects data are available is or is not "sufficiently similar" to the mixture of concern to permit a risk assessment. The determination of "sufficient similarity" must be made on a case-by-case basis, considering not only the uncertainties associated with using data on a dissimilar mixture but also the uncertainties of using other approaches such as additivity. In determining reasonable similarity, consideration should be given to any information on the components that differ or are contained in markedly different proportions between the mixture on which health effects data are available and the mixture of concern. Particular emphasis should be placed on any toxicologic or pharmacokinetic data on the components or the mixtures which would be useful in assessing the significance of any chemical difference between the similar mixture and the mixtures of concern.

Even if a risk assessment can be made using data on the mixtures of concern or a reasonably similar mixture, it may be desirable to conduct a risk assessment based on toxicity data on the

[a] See text for discussion of sufficient similarity, adequacy of data, and justification for additivity assumptions.

[b] See the Agency's Guidelines for Estimating Exposures (U.S. EPA, 1986d) for more complete information on performing exposure assessments and evaluating the quality of exposure data.

components in the mixture using the procedure outlined in section II.B. In the case of a mixture containing carcinogens and toxicants, an approach based on the mixture data alone may not be sufficiently protective in all cases. For example, this approach for a two-component mixture of one carcinogen and one toxicant would use toxicity data on the mixture of the two compounds. However, in a chronic study of such a mixture, the presence of the toxicant could mask the activity of the carcinogen. That is to say, at doses of the mixture sufficient to induce a carcinogenic effect, the toxicant could induce mortality so that at the maximum tolerated dose of the mixture, no carcinogenic effect could be observed. Since carcinogenicity is considered by the Agency to be a nonthreshold effect, it may not be prudent to construe the negative results of such a bioassay as indicating the absence of risk at lower doses. Consequently, the mixture approach should be modified to allow the risk assessor to evaluate the potential for masking, of one effect by another, on a case-by-case basis.

C. Data Available Only on Mixture Components

If data are not available on an identical or reasonably similar mixture, the risk assessment may be based on the toxic or carcinogenic properties of the components in the mixture. When little or no quantitative information is available on the potential interaction among the components, additive models (defined in the next section) are recommended for systemic toxicants. Several studies have demonstrated that dose additive models often predict reasonably well the toxicities of mixtures composed of a substantial variety of both similar and dissimilar compounds (Pozzani et al., 1959; Smyth et al., 1969, 1970; Murphy, 1980). The problem of multiple toxicant exposure has been addressed by the American Conference of Governmental Industrial Hygienists (ACGIH, 1983), the Occupational Safety and Health Administration (OSHA, 1983), the World Health Organization (WHO, 1981), and the National Research Council (NRC, 1980a, b). Although the focus and purpose of each group was somewhat different, all groups that recommended an approach elected to adopt some type of dose additive model. Nonetheless, as discussed in section IV, dose additive models are not the most biologically plausible approach if the compounds do not have the same mode of toxicologic action. Consequently, depending on the nature of the risk assessment and the available information on modes of action and patterns of joint action, the

[51 FR 34019]

most reasonable additive model should be used.

1. Systemic Toxicants. For systemic toxicants, the current risk assessment methodology used by the Agency for single compounds most often results in the derivation of an exposure level which is not anticipated to cause significant adverse effects. Depending on the route of exposure, media of concern, and the legislative mandate guiding the risk assessments, these exposure levels may be expressed in a variety of ways such as acceptable daily intakes (ADIs) or reference doses (RfDs), levels associated with various margins of safety (MOS), or acceptable concentrations in various media. For the purpose of this discussion, the term "acceptable level" (AL) will be used to indicate any such criteria or advisories derived by the Agency. Levels of exposure (E) will be estimates obtained following the most current Agency Guidelines for Estimating Exposures (U.S. EPA, 1986d). For such estimates, the "hazard index" (HI) of a mixture based on the assumption of dose addition may be defined as:

$$HI = E_1/AL_1 + E_2/AL_2 + \ldots + E_i/AL_i \qquad \text{(II-1)}$$

where:

E_i = exposure level to the ith toxicant* and
AL_i = maximum acceptable level for the ith toxicant.

Since the assumption of dose addition is most properly applied to compounds that induce the same effect by similar modes of action, a separate hazard index should be generated for each end point of concern. Dose addition for dissimilar effects does not have strong scientific support, and, if done, should be justified on a case-by-case basis in terms of biological plausibility.

The assumption of dose addition is most clearly justified when the mechanisms of action of the compounds under consideration are known to be the same. Since the mechanisms of action for most compounds are not well understood, the justification of the assumption of dose addition will often be limited to similarities in pharmacokinetic and toxicologic characteristics. In any event, if a hazard index is generated, the quality of the experimental evidence supporting the assumption of dose addition must be clearly articulated.

The hazard index provides a rough measure of likely toxicity and requires cautious interpretation. The hazard index is only a numerical indication of the nearness to acceptable limits of exposure or the degree to which acceptable exposure levels are exceeded. As this index approaches unity, concern for the potential hazard of the mixture increases. If the index exceeds unity, the concern is the same as if an individual chemical exposure exceeded its acceptable level by the same proportion. The hazard index does not define dose-response relationships, and its numerical value should not be construed to be a direct estimate of risk. Nonetheless, if sufficient data are available to derive individual acceptable levels for a spectrum of effects (e.g., MFO induction,

*See the Agency's guidelines (U.S. EPA, 1986d) for information on how to estimate this value.

variabilities of the acceptable levels are known, or if the acceptable levels are given as ranges (e.g., associated with different margins of safety), then the hazard index should be presented with corresponding estimates of variation or range. Most studies on systemic toxicity report only descriptions of the effects in each dose group. If dose-response curves are estimated for systemic toxicants, however, dose-additive or response-additive assumptions can be used, with preference given to the most biologically plausible assumption (see section IV for the mathematical details).

2. *Carcinogens.* For carcinogens, whenever linearity of the individual dose-response curves has been assumed (usually restricted to low doses), the increase in risk P (also called excess or incremental risk), caused by exposure d, is related to carcinogenic potency B, as:

$$P = d\,B \qquad \text{(II-2)}$$

For multiple compounds, this equation may be generalized to:

$$P = \Sigma d_i B_i \qquad \text{(II-3)}$$

This equation assumes independence of action by the several carcinogens and is equivalent to the assumption of dose addition as well as to response addition with completely negative correlation of tolerance, as long as $P < 1$ (see section IV). Analogous to the procedure used in equation II-1 for systemic toxicants, an index for n carcinogens can be developed by dividing exposure levels (E) by doses (DR) associated with a set level of risk:

$$HI = E_1/DR_1 + E_2/DR_2 + \ldots + E_n/DR_n \qquad \text{(II-4)}$$

Note that the less linear the dose-response curve is, the less appropriate equations II-3 and II-4 will be, perhaps even at low doses. It should be emphasized that because of the uncertainties in estimating dose-response relationships for single compounds, and the additional uncertainties in combining the individual estimate to assess response from exposure to mixtures, response rates and hazard indices may have merit in comparing risks but should not be regarded as measures of absolute risk.

3.*Interactions.* None of the above equations incorporates any form of synergistic or antagonistic interaction. Some types of information, however, may be available that suggest that two or more components in the mixture may interact. Such information must be assessed in terms of both its relevance to subchronic or chronic hazard and its suitability for quantitatively altering the risk assessment.

For example, if chronic or subchronic toxicity or carcinogenicity studies have been conducted that permit a quantitative estimation of interaction for two chemicals, then it may be desirable to consider using equations detailed in section IV, or modifications of these equations, to treat the two compounds as a single toxicant with greater or lesser potency than would be predicted from additivity. Other components of the mixture, on which no such interaction data are available, could then be separately treated in an additive manner. Before such a procedure is adopted, however, a discussion should be presented of the likelihood that other compounds in the mixture may interfere with the interaction of the two toxicants on which quantitative interaction data are available. If the weight of evidence suggests that interference is likely, then a quantitative alteration of the risk assessment may not be justified. In such cases, the risk assessment may only indicate the likely nature of interactions, either synergistic or antagonistic, and not quantify their magnitudes.

Other types of information, such as those relating to mechanisms of toxicant interaction, or quantitative estimates of interaction between two chemicals derived from acute studies, are even less likely to be of use in the quantitative assessment of long-term health risks. Usually it will be appropriate only to discuss these types of information, indicate the relevance of the information to subchronic or chronic exposure, and indicate, if possible, the nature of potential interactions, without attempting to quantify their magnitudes.

When the interactions are expected to have a minor influence on the mixture's toxicity, the assessment should indicate, when possible, the compounds most responsible for the predicted toxicity. This judgment should be based on predicted toxicity of each component,

[51 FR 34020]

based on exposure and toxic or carcinogenic potential. This potential alone should not be used as an indicator of the chemicals posing the most hazard.

4.*Uncertainties.* For each risk assessment, the uncertainties should be clearly discussed and the overall quality of the risk assessment should be characterized. The scheme outlined in Table 2 should be used to express the degree of confidence in the quality of the data on interaction, health effects, and exposure.

a. Health Effects--In some cases, when health effects data are incomplete, it may be possible to argue by analogy or quantitative structure-activity relationships that the compounds on which no health effects data are available are not likely to significantly affect the toxicity of the mixture. If a risk assessment includes such an argument, the limitations of the approach must be clearly articulated. Since a methodology has not been adopted for estimating an acceptable level (e.g., ADI) or carcinogenic potential for single compounds based either on quantitative structure-activity relationships or on the results of short-term screening tests, such methods are not at present

recommended as the sole basis of a risk assessment on chemical mixtures.

b. Exposure Uncertainties--The general uncertainties in exposure assessment have been addressed in the Agency's Guidelines for Estimating Exposures (U.S. EPA, 1986d). The risk assessor should discuss these exposure uncertainties in terms of the strength of the evidence used to quantify the exposure. When appropriate, the assessor should also compare monitoring and modeling data and discuss any inconsistencies as a source of uncertainty. For mixtures, these uncertainties may be increased as the number of compounds of concern increases.

If levels of exposure to certain compounds known to be in the mixture are not available, but information on health effects and environmental persistence and transport suggest that these compounds are not likely to be significant in affecting the toxicity of the mixture, then a risk assessment can be conducted based on the remaining compounds in the mixture, with appropriate caveats. If such an argument cannot be supported, no final risk assessment can be performed until adequate monitoring data are available. As an interim procedure, a risk assessment may be conducted for those components in the mixture for which adequate exposure and health effects data are available. If the interim risk assessment does not suggest a hazard, there is still concern about the risk from such a mixture because not all components in the mixture have been considered.

c. Uncertainties Regarding Composition of the Mixture--In perhaps a worst case scenario, information may be lacking not only on health effects and levels of exposure, but also on the identity of some components of the mixture. Analogous to the procedure described in the previous paragraph, an interim risk assessment can be conducted on those components of the mixture for which adequate health effects and exposure information are available. If the risk is considered unacceptable, a conservative approach is to present the quantitative estimates of risk, along with appropriate qualifications regarding the incompleteness of the data. If no hazard is indicated by this partial assessment, the risk assessment should not be quantified until better health effects and monitoring data are available to adequately characterize the mixture exposure and potential hazards.

III. Assumptions and Limitations

A. Information on Interactions

Most of the data available on toxicant interactions are derived from acute toxicity studies using experimental animals in which mixtures of two compounds were tested, often in only a single combination. Major areas of uncertainty with the use of such data involve the appropriateness of interaction data from an acute toxicity study for quantitatively altering a risk assessment for subchronic or chronic exposure, the appropriateness of interaction data on two component mixtures for quantitatively altering a risk assessment on a mixture of several compounds, and the accuracy of interaction data on experimental animals for quantitatively predicting interactions in humans.

The use of interaction data from acute toxicity studies to assess the potential interactions on chronic exposure is highly questionable unless the mechanism(s) of the interaction on acute exposure were known to apply to low-dose chronic exposure. Most known biological mechanisms for toxicant interactions, however, involve some form of competition between the chemicals or phenomena involving saturation of a receptor site or metabolic pathway. As the doses of the toxicants are decreased, it is likely that these mechanisms either no longer will exert a significant effect or will be decreased to an extent that cannot be measured or approximated.

The use of information from two-component mixtures to assess the interactions in a mixture containing more than two compounds also is questionable from a mechanistic perspective. For example, if two compounds are known to interact, either synergistically or antagonistically, because of the effects of one compound on the metabolism or excretion of the other, the addition of a third compound which either chemically alters or affects the absorption of one of the first two compounds could substantially alter the degree of the toxicologic interaction. Usually, detailed studies quantifying toxicant interactions are not available on multicomponent mixtures, and the few studies that are available on such mixtures (e.g., Gullino et al., 1956) do not provide sufficient information to assess the effects of interactive interference.

Concerns with the use of interaction data on experimental mammals to assess interactions in humans is based on the increasing appreciation for systematic differences among species in their response to individual chemicals. If systematic differences in toxic sensitivity to single chemicals exist among species, then it seems reasonable to suggest that the magnitude of toxicant interactions among species also may vary in a systematic manner. Consequently, even if excellent chronic data are available on the magnitude of toxicant interactions in a species of experimental mammal, there is uncertainty that the magnitude of the interaction will be the same in humans. Again, data are not available to properly assess the significance of this uncertainty.

Last, it should be emphasized that none of the models for toxicant interaction can predict the magnitude of toxicant interactions in the absence of extensive data. If sufficient data are available to

estimate interaction coefficients as described in section IV, then the magnitude of the toxicant interactions for various proportions of the same components can be predicted. The availability of an interaction ratio (observed response divided by predicted response) is useful only in assessing the magnitude of the toxicant interaction for the specific proportions of the mixture which was used to generate the interaction ratio.

The basic assumption in the recommended approach is that risk assessments on chemical mixtures are best conducted using toxicologic data on the mixture of concern or a reasonably similar mixture. While such risk

[51 FR 34021]

assessments do not formally consider toxicologic interactions as part of a mathematical model, it is assumed that responses in experimental mammals or human populations noted after exposure to the chemical mixture can be used to conduct risk assessments on human populations. In bioassays of chemical mixtures using experimental mammals, the same limitations inherent in species-to-species extrapolation for single compounds apply to mixtures. When using health effects data on chemical mixtures from studies on exposed human populations, the limitations of epidemiologic studies in the risk assessment of single compounds also apply to mixtures. Additional limitations may be involved when using health effects data on chemical mixtures if the components in the mixture are not constant or if the components partition in the environment.

B. Additivity Models

If sufficient data are not available on the effects of the chemical mixture of concern or a reasonably similar mixture, the proposed approach is to assume additivity. Dose additivity is based on the assumption that the components in the mixture have the same mode of action and elicit the same effects. This assumption will not hold true in most cases, at least for mixtures of systemic toxicants. For systemic toxicants, however, most single compound risk assessments will result in the derivation of acceptable levels, which, as currently defined, cannot be adapted to the different forms of response additivity as described in section IV.

Additivity models can be modified to incorporate quantitative data on toxicant interactions from subchronic or chronic studies using the models given in section IV or modifications of these models. If this approach is taken, however, it will be under the assumption that other components in the mixture do not interfere with the measured interaction. In practice, such subchronic or chronic interactions data seldom will be available. Consequently, most risk assessments (on mixtures) will be based on an assumption of additivity, as long as the components elicit similar effects.

Dose-additive and response-additive assumptions can lead to substantial errors in risk estimates if synergistic or antagonistic interactions occur. Although dose additivity has been shown to predict the acute toxicities of many mixtures of similar and dissimilar compounds (e.g., Pozzani et al., 1959; Smyth et al., 1969, 1970; Murphy, 1980), some marked exceptions have been noted. For example, Smyth et al. (1970) tested the interaction of 53 pairs of industrial chemicals based on acute lethality in rats. For most pairs of compounds, the ratio of the predicted LD_{50} to observed LD_{50} did not vary by more than a factor of 2. The greatest variation was seen with an equivolume mixture of morpholine and toluene, in which the observed LD_{50} was about five times less than the LD_{50} predicted by dose addition. In a study by Hammond et al. (1979), the relative risk of lung cancer attributable to smoking was 11, while the relative risk associated with asbestos exposure was 5. The relative risk of lung cancer from both smoking and asbestos exposure was 53, indicating a substantial synergistic effect. Consequently, in some cases, additivity assumptions may substantially underestimate risk. In other cases, risk may be overestimated. While this is certainly an unsatisfactory situation, the available data on mixtures are insufficient for estimating the magnitude of these errors. Based on current information, additivity assumptions are expected to yield generally neutral risk estimates (i.e., neither conservative nor lenient) and are plausible for component compounds that induce similar types of effects at the same sites of action.

IV. Mathematical Models and the Measurement of Joint Action

The simplest mathematical models for joint action assume no interaction in any mathematical sense. They describe either dose addition or response addition and are motivated by data on acute lethal effects of mixtures of two compounds.

A. Dose Addition

Dose addition assumes that the toxicants in a mixture behave as if they were dilutions or concentrations of each other, thus the true slopes of the dose-response curves for the individual compounds are identical, and the response elicited by the mixture can be predicted by summing the individual doses after adjusting for differences in potency; this is defined as the ratio of equitoxic doses. Probit transformation typically makes this ratio constant at all doses when parallel straight lines are obtained. Although this assumption can be applied to any model (e.g., the one-hit model in NRC, 1980b), it has been most often used in toxicology with the log-dose probit response model, which will be used to illustrate the assumption of dose addition.

Suppose that two toxicants show the following log-dose probit response equations:

$$Y_1 = 0.3 + 3 \log Z_1 \quad \text{(IV-1)}$$
$$Y_2 = 1.2 + 3 \log Z_2 \quad \text{(IV-2)}$$

where Y_i is the probit response associated with a dose of Z_i ($i=1,2$). The potency, p, of toxicant #2 with respect to toxicant #1 is defined by the quantity Z_1/Z_2 when $Y_1 = Y_2$ (that is what is meant by equitoxic doses). In this example, the potency, p, is approximately 2. Dose addition assumes that the response, Y, to any mixture of these two toxicants can be predicted by:

$$Y = 0.3 + 3 \log (Z_1 + pZ_2) \quad \text{(IV-3)}$$

Thus, since p is defined as Z_1/Z_2, equation IV-3 essentially converts Z_2 into an equivalent dose of Z by adjusting for the difference in potency. A more generalized form of this equation for any number of toxicants is:

$$Y = a_1 + b \log (f_1 + \Sigma f_i p_i) + b \log Z \quad \text{(IV-4)}$$

where:

a_1 = the y-intercept of the dose-response equation for toxicant #1
b = the slope of the dose-response lines for the toxicants
f_i = the proportion of the ith toxicant in the mixture
p_i = the potency of the ith toxicant with respect to toxicant #1 (i.e., Z_1/Z_i), and
Z = the sum of the individual doses in the mixture.

A more detailed discussion of the derivation of the equations for dose addition is presented by Finney (1971).

B. Response Addition

The other form of additivity is referred to as response addition. As detailed by Bliss (1939), this type of joint action assumes that the two toxicants act on different receptor systems and that the correlation of individual tolerances may range from completely negative ($r=-1$) to completely positive ($r=+1$). Response addition assumes that the response to a given concentration of a mixture of toxicants is completely determined by the responses to the components and the pairwise correlation coefficient. Taking P as the proportion of organisms responding to a mixture of two toxicants which evoke individual responses of P_1 and P_2, then

$$P = P_1 \text{ if } r = 1 \text{ and } P_1 \geq P_2 \quad \text{(IV-5)}$$
$$P = P_2 \text{ if } r = 1 \text{ and } P_1 < P_2 \quad \text{(IV-6)}$$
$$P = P_1 + P_2 (1-P_1) \text{ if } r = 0 \quad \text{(IV-7)}$$
$$P = P_1 + P_2 \text{ if } r = -1 \text{ and } P \leq 1. \quad \text{(IV-8)}$$

More generalized mathematical models for this form of joint action have been given by Plackett and Hewlett (1948).

C. Interactions

All of the above models assume no interactions and therefore do not incorporate measurements of synergistic or antagonistic effects. For measuring toxicant interactions for mixtures of two compounds, Finney (1942) proposed the

[51 FR 34022]

following modification of equation IV-4 for dose addition:

$$Y = a_1 + b \log (f_1 + pf_2 + K[pf_1f_2]^{0.5}) + b \log Z \quad \text{(IV-9)}$$

where a_1, b, f_1, f_2, p, and Z are defined as before, and K is the coefficient of interaction. A positive value of K indicates synergism, a negative value indicates antagonism, and a value of zero corresponds to dose addition as in equation IV-4. Like other proposed modifications of dose addition (Hewlett, 1969), the equation assumes a consistent interaction throughout the entire range of proportions of individual components. To account for such asymmetric patterns of interaction as those observed by Alstott et al. (1973), Durkin (1981) proposed the following modification to equation IV-9:

$$Y = a_1 + b \log (f_1 + pf_2 + K_1f_1[pf_1f_2]^{0.5} + K_2f_2[pf_1f_2]^{0.5}) + b \log Z \quad \text{(IV-10)}$$

in which $K(pf_1f_2)^{0.5}$ is divided into two components, $K_1f_1(pf_1f_2)^{0.5}$ and $K_2f_2(pf_1f_2)^{0.5}$. Since K_1 and K_2 need not have the same sign, apparent instances of antagonism at one receptor site and synergism at another receptor site can be estimated. When K_1 and K_2 are equal, equation IV-10 reduces to equation IV-9. It should be noted that to obtain a reasonable number of degrees of freedom in the estimation of K in equation IV-9 or K_1 and K_2 in equation IV-10, the toxicity of several different combinations of the two components must be assayed along with assays of the toxicity of the individual components. Since this requires experiments with large numbers of animals, such analyses have been restricted for the most part to data from acute bioassays using insects (e.g., Finney, 1971) or aquatic organisms (Durkin, 1979). Also, because of the complexity of experimental design and the need for large numbers of animals, neither equation IV-9 nor equation IV-10 has been generalized or applied to mixtures of more than two toxicants. Modifications of response-additive models to include interactive terms have also been proposed, along with appropriate statistical tests for the assumption of additivity (Korn and Liu, 1983; Wahrendorf et al., 1981).

In the epidemiologic literature, measurements of the extent of toxicant interactions, S, can be expressed as the ratio of observed relative risk to relative risk predicted by some form of additivity assumption. Analogous to the ratio of interaction in classical toxicology studies, $S = 1$ indicates no interaction, $S > 1$ indicates synergism, and $S < 1$ indicates antagonism. Several models for both

additive and multiplicative risks have been proposed (e.g., Hogan et al., 1978; NRC, 1980b; Walter, 1976). For instance, Rothman (1976) has discussed the use of the following measurement of toxicant interaction based on the assumption of risk additivity:

$$S = (R_{11} - 1)/(R_{10} + R_{01} - 2) \qquad \text{(IV-11)}$$

where R_{10} is the relative risk from compound #1 in the absence of compound #2, R_{01} is the relative risk from compound #2 in the absence of compound #1, and R_{11} is the relative risk from exposure to both compounds. A multiplicative risk model adapted from Walter and Holford (1978, equation 4) can be stated as:

$$S = R_{11}/(R_{10}R_{01}) \qquad \text{(IV-12)}$$

As discussed by both Walter and Holford (1978) and Rothman (1976), the risk-additive model is generally applied to agents causing diseases while the multiplicative model is more appropriate to agents that prevent disease. The relative merits of these and other indices have been the subject of considerable discussion in the epidemiologic literature (Hogan et al., 1978; Kupper and Hogan, 1978; Rothman, 1978; Rothman et al., 1980; Walter and Holford, 1978). There seems to be a consensus that for public health concerns regarding causative (toxic) agents, the additive model is more appropriate.

Both the additive and multiplicative models assume statistical independence in that the risk associated with exposure to both compounds in combination can be predicted by the risks associated with separate exposure to the individual compounds. As illustrated by Siemiatycki and Thomas (1981) for multistage carcinogenesis, the better fitting statistical model will depend not only upon actual biological interactions, but also upon the stages of the disease process which the compounds affect. Consequently, there is no *a priori* basis for selecting either type of model in a risk assessment. As discussed by Stara et al. (1983), the concepts of multistage carcinogenesis and the effects of promoters and cocarcinogens on risk are extremely complex issues. Although risk models for promoters have been proposed (e.g., Burns et al., 1983), no single approach can be recommended at this time.

V. References

ACGIH (American Conference of Governmental Industrial Hygienists). 1983. TLVs: threshold limit values for chemical substances and physical agents in the work environment with intended changes for 1983-1984. Cincinnati, OH, p.58.

Alstott, R.L., M.E. Tarrant, and R.B. Forney. 1973. The acute toxicities of 1-methylxanthine, ethanol, and 1-methylxanthine/ethanol combinations in the mouse. Toxicol. Appl. Pharmacol. 24:393-404.

Bliss, C.I. 1939. The toxicity of poisons applied jointly. Ann. Appl. Biol. 26:585-615.

Burns, F., R. Albert, E. Altschuler, and E. Morris. 1983. Approach to risk assessment for genotoxic carcinogens based on data from the mouse skin initiation-promotion model. Environ. Health Perspect. 50:309-320.

Durkin, P.R. 1979. Spent chlorination liquor and chlorophenolics: a study in detoxication and joint action using *Daphnia magna*. Ph.D. Thesis, Syracuse, NY: State University of New York College of Environmental Science and Forestry, p.145.

Durkin, P.R. 1981. An approach to the analysis of toxicant interactions in the aquatic environment. Proceedings of the 4th Annual Symposium on Aquatic Toxicology. American Society for Testing and Materials, p.388-401.

Finney, D.J. 1942. The analysis of toxicity tests on mixtures of poisons. Ann. Appl. Biol. 29:82-94.

Finney, D.J. 1971. Probit analysis, 3rd ed. Cambridge, Great Britain: Cambridge University Press, 333 p.

Goldstein, A., L. Aronow, and S.M. Kalman. 1974. Principles of drug action: the basis of pharmacology, 2nd ed. New York, NY: John Wiley and Sons, Inc., 854 p.

Gullino, P., M. Winitz, S.M. Birnbaum, J. Cornfield, M.C. Otey, and J.P. Greenstein. 1956. Studies on the metabolism of amino acids and related compounds *in vivo*. I. Toxicity of essential amino acids, individually and in mixtures, and the protective effect of L-arginine. Arch. Biochem. Biophys. 64:319-332.

Hammond, E.C., I.V. Selikoff, and H. Seidman. 1979. Asbestos exposure, cigarette smoking and death rates. Ann. NY Acad. Sci. 330:473-490.

Hewlett, P.S. 1969. Measurement of the potencies of drug mixtures. Biometrics. 25:477-487.

Hogan, M.D., L. Kupper, B. Most, and J. Haseman. 1978. Alternative approaches to Rothman's approach for assessing synergism (or antagonism) in cohort studies. Am. J. Epidemiol. 108(1):60-67.

Klaassen, C.D., and J. Doull. 1980. Evaluation of safety: toxicologic evaluation. In: J. Doull, C.D. Klaassen, and M.O. Amdur, eds. Toxicology: the basic science of poisons. New York, NY: Macmillan Publishing Co., Inc., p. 11-27.

Korn, E.L., and P-Y. Liu. 1983. Interactive effects of mixtures of stimuli in life table analysis. Biometrika. 70:103-110.

Kupper, L., and M.D. Hogan. 1978. Interaction in epidemiologic studies. Am. J. Epidemiol. 108(6):447-453.

Levine, R.E. 1973. Pharmacology: drug actions and reactions. Boston, MA: Little, Brown and Company, 412 p.

[51 FR 34023]

Murphy, S.D. 1980. Assessment of the potential for toxic interactions among environmental pollutants. In: C.L. Galli, S.D. Murphy, and R. Paoletti, eds. The principles and methods in modern toxicology. Amsterdam, The Netherlands: Elsevier/North Holland Biomedical Press.

NRC (National Research Council). 1980a. Drinking water and health, Vol.3. Washington, DC: National Academy Press, p.27-28.

NRC (National Research Council). 1980b. Principles of toxicological interactions associated with multiple chemical exposures. Washington, DC: National Academy Press, p.204.

OSHA (Occupational Safety and Health Administration). 1983. General Industry Standards, Subpart 2, Toxic and Hazardous Substances. Code of Federal Regulations. 40:1910.1000 (d)(2)(i). Chapter XVII Occupational Safety and Health Administration, p.667.

Plackett, R.L., and P.S. Hewlett. 1948. Statistical aspects of the independent joint action of poisons. Ann. Appl. Biol. 35:347-358.

Pozzani, U.C., C.S. Weil, and C.P. Carpenter. 1959. The toxicological basis of threshold values: 5. The experimental inhalation of vapor mixtures by rats, with notes upon the relationship between single dose inhalation and single dose oral data. Am. Ind. Hyg. Assoc. J. 20:364-369.

Rothman, K. 1976. The estimation of synergy or antagonism. Am. J. Epidemiol. 103(5):506-511.

Rothman, K. 1978. Estimation versus detection in the assessment of synergy. Am. J. Epidemiol. 108(1):9-11.

Rothman, K., S. Greenland, and A. Walker. 1980. Concepts of interaction. Am. J. Epidemiol. 112(4):467-470.

Siemiatycki, J., and D.C. Thomas. 1981. Biological models and statistical interactions: An example from multistage carcinogenesis. Int. J. Epidemiol. 10(4):383-387.

Smyth, H.F., C.S. Weil, J.S. West, and C.P. Carpenter. 1969. An exploration of joint toxic action: I. Twenty-seven industrial chemicals intubated in rats in all possible pairs. Toxicol. Appl. Pharmacol. 14:340-347.
Smyth, H.F., C.S. Weil, J.S. West, and C.P. Carpenter. 1970. An exploration of joint toxic action. II. Equitoxic versus equivolume mixtures. Toxicol. Appl. Pharmacol. 17:498-503.
Stara, J.F., D. Mukerjee, R. McGaughy, P. Durkin, and M.L. Dourson. 1983. The current use of studies on promoters and cocarcinogens in quantitative risk assessment. Environ. Health Perspect. 50:359-368.
U.S. EPA. 1986a. Guidelines for carcinogen risk assessment. Federal Register. 51(185):33992-34003.
U.S. EPA. 1986b. Guidelines for mutagenicity risk assessment. Federal Register. 51(185):34006-34012.
U.S. EPA. 1986c. Guidelines for the health assessment of suspect developmental toxicants. Federal Register 51(185): 34028-34040.
U.S. EPA. 1986d. Guidelines for estimating exposures. Federal Register 51(185):34042-34054.
Veldstra, H. 1956. Synergism and potentiation with special reference to the combination of structural analogues. Pharmacol. Rev. 8:339-387.
Wahrendorf, J., R. Zentgrof, and C.C. Brown. 1981. Optimal designs for the analysis of interactive effects of two carcinogens or other toxicants. Biometrics. 37:45-54.
Walter, S.D. 1976. The estimation and interpretation of attributable risk in health research. Biometrics. 32:829-849.
Walter, S.D., and T.R. Holford. 1978. Additive, multiplicative, and other models for disease risks. Am. J. Epidemiol. 108:341-346.
Withey, J.R. 1981. Toxicodynamics and biotransformation. In: International Workshop on the Assessment of Multichemical Contamination.Milan, Italy. (Draft copy courtesy of J.R. Withey)
WHO (World Health Organization). 1981. Health effects of combined exposures in the work environment. WHO Tech. Report Series No. 662.

Part B: Response to Public and Science Advisory Board Comments

I. Introduction

This section summarizes some of the major issues raised in public comments on the Proposed Guidelines for the Health Risk Assessment of Chemical Mixtures published on January 9, 1985 (50 FR 1170). Comments were received from 14 individuals or organizations. An issue paper reflecting public and external review comments was presented to the Chemical Mixtures Guidelines Panel of the Science Advisory Board (SAB) on March 4, 1985. At its April 22-23, 1985, meeting, the SAB Panel provided the Agency with additional suggestions and recommendations concerning the Guidelines. This section also summarizes the issues raised by the SAB.

The SAB and public commentors expressed diverse opinions and addressed issues from a variety of perspectives. In response to comments, the Agency has modified or clarified many sections of the Guidelines, and is planning to develop a technical support document in line with the SAB recommendations. The discussion that follows highlights significant issues raised in the comments, and the Agency's response to them. Also, many minor recommendations, which do not warrant discussion here, were adopted by the Agency.

II. Recommended Procedures

A. Definitions

Several comments were received concerning the lack of definitions for certain key items and the general understandability of certain sections. Definitions have been rewritten for several terms and the text has been significantly rewritten to clarify the Agency's intent and meaning.

Several commentors noted the lack of a precise definition of "mixture," even though several classes of mixtures are discussed. In the field of chemistry, the term "mixture" is usually differentiated from true solutions, with the former defined as nonhomogeneous multicomponent systems. For these Guidelines, the term "mixture" is defined as "...any combination of two or more chemicals regardless of spatial or temporal homogeneity of source" (section 1). These Guidelines are intended to cover risk assessments for any situation where the population is exposed or potentially exposed to two or more compounds of concern. Consequently, the introduction has been revised to clarify the intended breadth of application.

Several commentors expressed concern that "sufficient similarity" was difficult to define and that the Guidelines should give more details concerning similar mixtures. The Agency agrees and is planning research projects to improve on the definition. Characteristics such as composition and toxic end-effects are certainly important, but the best indicators of similarity in terms of risk assessment have yet to be determined. The discussion in the Guidelines emphasizes case-by-case judgment until the necessary research can be performed. The Agency considered but rejected adding an example, because it is not likely that any single example would be adequate to illustrate the variety in the data and types of judgments that will be required in applying this concept. Inclusion of examples is being considered for the technical support document.

B. Mixtures of Carcinogens and Systemic Toxicants

The applicability of the preferred approach for a mixture of carcinogens and systemic (noncarcinogenic) toxicants was a concern of several public commentors as well as the SAB. The Agency realizes that the preferred approach of using test data on the mixture itself may not be sufficiently protective in all cases. For example, take a simple two-component mixture of one carcinogen and one toxicant. The preferred approach would lead to using toxicity data on the mixture of the two compounds. However, it is possible to set the proportions of each component so that in a chronic bioassay of such a mixture, the presence of the toxicant could mask the activity of the carcinogen. That is to say, at doses of

the mixture sufficient for the carcinogen to induce tumors in the small

[51 FR 34024]

experimental group, the toxicant could induce mortality. At a lower dose in the same study, no adverse effects would be observed, including no carcinogenic effects. The data would then suggest use of a threshold approach. Since carcinogenicity is considered by the Agency to be a nonthreshold effect, it may not be prudent to construe the negative results of such a bioassay as indicating the absence of risk at lower doses. Consequently, the Agency has revised the discussion of the preferred approach to allow the risk assessor to evaluate the potential for masking of carcinogenicity or other effects on a case-by-case basis. Another difficulty occurs with such a mixture when the risk assessment needs to be based on data for the mixture components. Carcinogens and systemic toxicants are evaluated by the Agency using different approaches and generally are described by different types of data: response rates for carcinogens vs. effect descriptions for toxicants. The Agency recognizes this difficulty and recommends research to develop a new assessment model for combining these dissimilar data sets into one risk estimate. One suggestion in the interim is to present separate risk estimates for the dissimilar end points, including carcinogenic, teratogenic, mutagenic, and systemic toxicant components.

III. Additivity Assumption

Numerous comments were received concerning the assumption of additivity, including:

a. the applicability of additivity to "complex" mixtures;
b. the use of dose additivity for compounds that induce different effects;
c. the interpretation of the Hazard Index; and
d. the use of interaction data.

Parts of the discussion in the proposed guidelines concerning the use of additivity assumptions were vague and have been revised in the final Guidelines to clarify the Agency's intent and position.

A. Complex Mixtures

The issue of the applicability of an assumption of additivity to complex mixtures containing tens or hundreds of components was raised in several of the public comments. The Agency and its reviewers agree that as the number of compounds in the mixture increases, an assumption of additivity will become less reliable in estimating risk. This is based on the fact that each component estimate of risk or an acceptable level is associated with some error and uncertainty. With current knowledge, the uncertainty will increase as the number of components increases. In any event, little experimental data are available to determine the general change in the error as the mixture contains more components. The Agency has decided that a limit to the number of components should not be set in these Guidelines. However, the Guidelines do explicitly state that as the number of compounds in the mixture increases, the uncertainty associated with the risk assessment is also likely to increase.

B. Dose Additivity

Commentors were concerned about what appeared to be a recommendation of the use of dose additivity for compounds that induce different effects. The discussion following the dose additivity equation was clarified to indicate that the act of combining all compounds, even if they induce dissimilar effects, is a screening procedure and not the preferred procedure in developing a hazard index. The Guidelines were further clarified to state that dose (or response) additivity is theoretically sound, and therefore best applied for assessing mixtures of similar acting components that do not interact.

C.Interpretation of the Hazard Index

Several comments addressed the potential for misinterpretation of the hazard index, and some questioned its validity, suggesting that it mixes science and value judgments by using "acceptable" levels in the calculation. The Agency agrees with the possible confusion regarding its use and has revised the Guidelines for clarification. The hazard index is an easily derived restatement of dose additivity, and is, therefore, most accurate when used with mixture components that have similar toxic action. When used with components of unknown or dissimilar action, the hazard index is less accurate and should be interpreted only as a rough indication of concern. As with dose addition, the uncertainty associated with the hazard index increases as the number of components increases, so that it is less appropriate for evaluating the toxicity of complex mixtures.

D. Use of Interaction Data

A few commentors suggested that any interaction data should be used to quantitatively alter the risk assessment. The Agency disagrees. The current information on interactions is meager, with only a few studies comparing response to the mixture with that predicted by studies on components. Additional uncertainties include exposure variations due to changes in composition, mixture dose, and species differences in the extent of the interaction. The Agency is constructing an interaction data base in an attempt to answer some of these issues. Other comments concerned the use of different types of interaction data. The Guidelines restrict the use of interaction data to that obtained from whole animal bioassays of a duration appropriate to the risk assessment. Since such data are frequently lacking, at least for chronic or subchronic effects, the issue is whether to allow for the use of other information such as acute data, *in*

vitro data, or structure-activity relationships to quantitatively alter the risk assessment, perhaps by use of a safety factor. The Agency believes that sufficient scientific support does not exist for the use of such data in any but a qualitative discussion of possible synergistic or antagonistic effects.

IV. Uncertainties and the Sufficiency of the Data Base

In the last two paragraphs of section II of the Guidelines, situations are discussed in which the risk assessor is presented with incomplete toxicity, monitoring, or exposure data. The SAB, as well as several public commentors, recommended that the "risk management" tone of this section be modified and that the option of the risk assessor to decline to conduct a risk assessment be made more explicit.

This is a difficult issue that must consider not only the quality of the available data for risk assessment, but also the needs of the Agency in risk management. Given the types of poor data often available, the risk assessor may indicate that the risk assessment is based on limited information and thus contains no quantification of risk. Nonetheless, in any risk assessment, substantial uncertainties exist. It is the obligation of the risk assessor to provide an assessment, but also to ensure that all the assumptions and uncertainties are articulated clearly and quantified whenever possible.

The SAB articulated several other recommendations related to uncertainties, all of which have been followed in the revision of the Guidelines. One recommendation was that the summary procedure table also be presented as a flow chart so that all options are clearly displayed. The SAB further recommended the development of a system to express the level of confidence in the various steps of the risk assessment.

The Agency has revised the summary table to present four major options: risk assessment using data on the mixture

[51 FR 34025]

itself, data on a similar mixture, data on the mixture's components, or declining to quantify the risk when the data are inadequate. A flow chart of this table has also been added to more clearly depict the various options and to suggest the combining of the several options to indicate the variability and uncertainties in the risk assessment.

To determine the adequacy of the data, the SAB also recommended the development of a system to express the level of confidence associated with various steps in the risk assessment process. The Agency has developed a rating scheme to describe data quality in three areas: interaction, health effects, and exposure. This classification provides a range of five levels of data quality for each of the three areas. Choosing the last level in any area results in declining to perform a quantitative risk assessment due to inadequate data. These last levels are described as follows:

Interactions:
An assumption of additivity cannot be justified, and no quantitative risk assessment can be conducted.

Health effects:
A lack of health effects information on the mixture and its components precludes a quantitative risk assessment.

Exposure:
The available exposure information is insufficient for conducting a risk assessment.

Several commentors, including the SAB, emphasized the importance of not losing these classifications and uncertainties farther along in the risk management process. The discussion of uncertainties has been expanded in the final Guidelines and includes the recommendation that a discussion of uncertainties and assumptions be included at every step of the regulatory process that uses risk assessment.

Another SAB comment was that the Guidelines should include additional procedures for mixtures with more than one end point or effect. The Agency agrees that these are concerns and revised the Guidelines to emphasize these as additional uncertainties worthy of further research.

V. Need for a Technical Support Document

The third major SAB comment concerned the necessity for a separate technical support document for these Guidelines. The SAB pointed out that the scientific and technical background from which these Guidelines must draw their validity is so broad and varied that it cannot reasonably be synthesized within the framework of a brief set of guidelines. The Agency is developing a technical support document that will summarize the available information on health effects from chemical mixtures, and on interaction mechanisms, as well as identify and develop mathematical models and statistical techniques to support these Guidelines. This document will also identify critical gaps and research needs.

Several comments addressed the need for examples on the use of the Guidelines. The Agency has decided to include examples in the technical support document.

Another issue raised by the SAB concerned the identification of research needs. Because little emphasis has been placed on the toxicology of mixtures until recently, the information on mixtures is limited. The SAB pointed out that identifying research needs is critical to the risk assessment process, and the EPA should ensure that these needs are considered in the research planning process. The Agency will include a section in the

technical support document that identifies research needs regarding both methodology and data.

51 FR 34028

GUIDELINES FOR THE HEALTH ASSESSMENT OF SUSPECT DEVELOPMENTAL TOXICANTS

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 the Health Assessment of Suspect Developmental Toxicants.

The Guidelines for the Health Assessment of Suspect Developmental Toxicants (hereafter "Guidelines") are intended to guide Agency analysis of developmental toxicity data 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 the Health Assessment of Suspect Developmental Toxicants published November 23, 1984 (49 FR 46324).

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. Carole A. Kimmel
Reproductive Effects Assessment Group
Office of Health and Environmental Assessment (RD-689)
U. S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460
202-382-7331

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 the Health Assessment of Suspect Developmental Toxicants

Work on the Guidelines for the Health Assessment of Suspect Developmental Toxicants 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 developmental toxicology from universities, environmental groups, industry, labor, and other governmental agencies. They were then proposed for public comment in the *FEDERAL REGISTER* (49 FR 46324). 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 preliminary Agency responses to those comments. 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 the Health Assessment of Suspect Developmental Toxicants were concurred on in a letter dated July 26, 1985. 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 SAB suggested that the Agency pursue additional follow-up work on quantitative risk assessment. Several efforts are currently underway within the Agency on quantitative risk assessment models and procedures, the relationship of maternal and developmental toxicity, and the evaluation and interpretation of postnatal studies. In addition, a document addressing research needs is being prepared to highlight those areas that are in need of further study.

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

[51 FR 34029]

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 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. Thomas,

Administrator

CONTENTS

Part A: Guidelines for the Health Assessment of Suspect Developmental Toxicants

I. Introduction

II. Definitions and Terminology

III. Qualitative Assessment (Hazard Identification of Developmental Toxicants

A. Laboratory Animal Studies of Developmental Toxicity: End Points and Their Interpretation
1. End Points of Maternal Toxicity
2. End Points of Developmental Toxicity
3. Functional Developmental Toxicity
4. Overall Evaluation of Maternal and Developmental Toxicity
5. Short-term Testing in Developmental Toxicity
a. *In Vivo* Mammalian Developmental Toxicity Screen
b. *In Vitro* Developmental Toxicity Screens
6. Statistical Considerations

B. Human Studies

C. Other Considerations
1. Pharmacokinetics
2. Comparisons of Molecular Structure

D. Weight-of- Evidence Determination

IV. Quantitative Assessment

A. Dose-Response Assessment
B. Exposure Assessment
C. Risk Characterization

V. References

Part B: Response to Public and Science Advisory Board Comments

I. Introduction

II. Coordination With Other Guidelines

A. Other Risk Assessment Guidelines
B. Coordination With Testing Guidelines

III. Definitions

IV. Qualitative Assessment

A. Maternal and Developmental Toxicity
B. Functional Developmental Toxicity
C. Short-Term Testing
D. Comparisons of Molecular Structure

V. Quantitative Assessment

Part A: Guidelines for the Health Assessment of Suspect Developmental Toxicants

I. Introduction

These Guidelines describe the procedures that the U.S. Environmental Protection Agency will follow in evaluating potential developmental toxicity associated with human exposure to environmental toxicants. In 1980, the Agency sponsored a conference that addressed issues related to such evaluations (1) and provided some of the scientific basis for these risk assessment Guidelines. The Agency's authority to regulate substances that have the potential to interfere adversely with human development is derived from a number of statutes which are implemented through multiple offices within the Agency. Because many different offices evaluate developmental toxicity, there is a need for intra-Agency consistency in the approach to assess these types of effects. The procedures described here will promote consistency in the Agency's assessment of developmental toxic effects.

The developmental toxicity assessments prepared pursuant to these Guidelines will be utilized within the requirements and constraints of the applicable statutes to arrive at regulatory decisions concerning developmental toxicity. These Guidelines provide a general format for analyzing and organizing the available data for conducting risk assessments. The Agency previously has issued testing guidelines (2, 3) that provide protocols designed to determine the potential of a test substance to induce structural and/or other abnormalities in the developing conceptus. These risk assessment Guidelines do not change any statutory or regulatory prescribed standards for the type of data necessary for regulatory action, but rather provide guidance for the interpretation of studies that follow the testing guidelines, and in addition, provide limited information for interpretation of other studies (e.g., epidemiologic data, functional developmental toxicity studies, and short-term tests) which are not routinely required, but which may be encountered when reviewing data on particular agents. Moreover, risk assessment is just one component of the regulatory process and defines the adverse health consequences of exposure to a toxic agent. The other component, risk management, combines risk assessment with the directives of the enabling 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 agent. The issue of risk management will not be addressed in these Guidelines.

The background incidence of developmental defects in the human population is quite large. For example, approximately 50% of human conceptuses fail to reach term (4); approximately 3% of newborn children are found to have one or more significant congenital malformations at birth, and by the end of the first postnatal year, about 3% more are found to have serious developmental defects (5, 6). Of these, it is estimated that 20% of human developmental defects are of known genetic transmission, 10% are attributable to known environmental factors, and the remainder result from unknown causes (7). Approximately 7.4% of children are reduced in weight at birth (i.e., below 2500 g) (8). Exposure to agents affecting development can result in multiple manifestations (malformation, functional impairment, altered growth, and/or lethality). Therefore, assessment efforts should encompass a wide array of adverse developmental end points, such as spontaneous abortions, stillbirths, malformations, early postnatal mortality, and other adverse functional or physical changes that are manifested postnatally.

Numerous agents have been shown to be developmental toxicants in animal test systems (9). Several of them have also been shown to be the cause of adverse developmental effects in humans, including alcohol, aminopterin, busulfan, chlorobiphenyls, diethylstilbestrol, isotretinoin, organic mercury, thalidomide, and valproic acid (10, 11, 12, 13). Although a number of agents found to be positive in animal studies have not shown clear evidence of hazard in humans, usually the human data available are inadequate to determine a cause and effect relationship. Comparisons of human and animal data have been made for a limited number of agents that are positive in humans (13, 14). In these comparisons, there was almost always concordance of effects between humans and at least one species tested; also, the minimally effective dose (MED) for the most sensitive animal species was approximately 0.5 to 50 times the human

[51 FR 34030]

MED, not accounting for differences in the incidence of effect at the MED. Thus, there is some limited basis for estimating the risk of exposure to human development based on data from animal studies.

The National Research Council (15) has defined risk assessment as being comprised of some or all of the following components: hazard identification, dose-response assessment, exposure assessment, and risk characterization. In general, the process of assessing the risk of human developmental toxicity may be adapted to this format. However, due to special considerations in assessing developmental toxicity, which will be discussed later in these Guidelines, it is not always possible to follow the exact standards as defined for each component.

Hazard identification is the qualitative risk assessment in which all available experimental animal and human data are used to determine if an agent is likely to cause developmental toxicity. In considering developmental toxicity, these Guidelines will address not only malformations, but

also fetal wastage, growth alteration, and functional abnormalities that may result from developmental exposure to environmental agents.

The dose-response assessment defines the relationship of the dose of an agent and the occurrence of developmental toxic effects. According to the National Research Council (15), this component would usually include the results of an extrapolation from high doses administered to experimental animals or noted in epidemiologic studies to the low exposure levels expected for human contact with the agent in the environment. Since at present there are no mathematical extrapolation models that are generally accepted for developmental toxicity, the Agency, for the most part, uses uncertainty (safety) factors and margins of safety, which will be discussed in these Guidelines. Appropriate models are being sought by the Agency for application to data in this area.

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.

In risk characterization, the exposure assessment and the dose-response assessment are combined to estimate some measure of the risk of developmental toxicity. As part of risk characterization, a summary of the strengths and weaknesses in each component of the assessment are presented along with major assumptions, scientific judgments, and, to the extent possible, estimates of the uncertainties.

II. Definitions and Terminology

The Agency recognizes that there are differences in the use of terms in the field of developmental toxicology. For the purposes of these Guidelines the following definitions and terminology will be used.

Developmental Toxicology--The study of adverse effects on the developing organism that may result from exposure prior to conception (either parent), during prenatal development, or postnatally to the time of sexual maturation. Adverse developmental effects may be detected at any point in the life span of the organism. The major manifestations of developmental toxicity include: 1) death of the developing organism, 2) structural abnormality, 3) altered growth, and 4) functional deficiency.

Embryotoxicity and Fetotoxicity--Any toxic effect on the conceptus as a result of prenatal exposure; the distinguishing feature between the two terms is the stage of development during which the injury occurred. The terms, as used here, include malformations and variations, altered growth, and *in utero* death.

Altered Growth--An alteration in offspring organ or body weight or size. Changes in body weight may or may not be accompanied by a change in crown-rump length and/or in skeletal ossification. Altered growth can be induced at any stage of development, may be reversible, or may result in a permanent change.

Functional Developmental Toxicology--The study of the causes, mechanisms, and manifestations of alterations or delays in functional competence of the organism or organ system following exposure to an agent during critical periods of development pre- and/or postnatally.

Malformations and Variations--A malformation is usually defined as a permanent structural change that may adversely affect survival, development, or function. The term teratogenicity, which is used to describe these types of structural abnormalities, will be used in these Guidelines to refer only to structural defects. A variation is used to indicate a divergence beyond the usual range of structural constitution that may not adversely affect survival or health. Distinguishing between variations and malformations is difficult since there exists a continuum of responses from the normal to the extreme deviant. There is no generally accepted classification of malformations and variations. Other terminology that is often used, but no better defined, includes anomalies, deformations, and aberrations.

III. Qualitative Assessment (Hazard Identification of Developmental Toxicants)

Developmental toxicity is expressed as one or more of a number of possible end points that may be used for evaluating the potential of an agent to cause abnormal development. The four types of effects on the conceptus that may be produced by developmental exposure to toxicants include death, structural abnormality, altered growth, and functional deficits. Of these, the first three types of effects are traditionally measured in laboratory animals using the conventional developmental toxicity (also called teratogenicity or Segment II) testing protocol as well as in other study protocols, such as the multigeneration study. Functional deficits are seldom evaluated in routine studies of environmental agents. This section will discuss the end points examined in routinely used protocols as well as the evaluation of data from other types of studies, including functional studies and short-term tests. Transplacental carcinogenesis, another type of developmental effect, will not be discussed in detail here since, at present, it is considered more appropriate to use the Guidelines for Carcinogen Risk Assessment (16) for assessing the human risk for these types of effects. Also, mutational events may occur as part of developmental toxicity, and in practice, are difficult to discriminate from other possible mechanisms of developmental toxicity. The Guidelines for Mutagenicity Risk Assessment (17) should be consulted in cases where genetic damage is suspected.

A. Laboratory Animal Studies of Developmental Toxicity: End Points and Their Interpretation

The most commonly used protocol for assessing developmental toxicity in laboratory animals involves the administration of a test substance to pregnant animals (usually mice, rats, or rabbits) during the period of major organogenesis, evaluation of maternal responses throughout pregnancy, and examination of the dam and the uterine contents just prior to term (2, 3, 18, 19, 20). Other protocols may use exposure periods of one to a few days to investigate periods of particular sensitivity for induction of anomalies in specific organs or organ systems (21).In

[51 FR 34031]

addition, developmental toxicity may be evaluated in studies involving exposure of one or both parents prior to conception, of the conceptus during pregnancy and over several generations, or of offspring during the late prenatal and early postnatal periods. These Guidelines are intended to provide information for interpreting developmental effects related to any of these types of exposure. Since many of the end points evaluated also are related to effects on the parental reproductive systems, these Guidelines will be used in conjunction with those to be published in the future by EPA on male and female reproductive toxicity.

Study designs should include a high dose, which produces some maternal or adult toxicity (i.e., a level which at the least produces marginal but significantly reduced body weight, weight gain, or specific organ toxicity, and at the most produces no more than 10% mortality); a low dose, which demonstrates a no observed effect level (NOEL) for adult and offspring effects; and at least one intermediate dose level. A concurrent control group treated with the vehicle used for agent administration should be included. The route of exposure should be based on expected human exposure considerations, although data from other routes may sometimes be useful, especially if supported by pharmacokinetic information. Test animals should be selected based on considerations of species, strain, age, weight, and health status, and should be randomized to dose groups in order to reduce bias and provide a basis for performing valid statistical tests.

The next three sections discuss individual end points of maternal and developmental toxicity as measured in the conventional developmental toxicity study, the multigeneration study, and, on occasion, in postnatal studies. Other end points specifically related to reproductive toxicity will be covered in the relevant reproductive toxicity guidelines. The fourth section deals with the integrated evaluation of all data, including the relative effects of exposure on maternal animals and their offspring, which is important in assessing the level of concern about a particular agent.

1. End Points of Maternal Toxicity. A number of end points that may be observed as possible indicators of maternal toxicity are listed in Table 1. Maternal mortality is an obvious end point of toxicity; however, a number of other end points can be observed which may give an indication of the subtle effects of an agent. For example, in well-conducted studies, the fertility and gestation indices provide information on the general fertility rate of the animal stock used and are important indicators of toxic effects if treatment begins prior to mating or implantation. Changes in gestation length may indicate effects on the process of parturition.

Table 1.--End Points of Maternal Toxicity

- Mortality
- Fertility Index (no. with seminal plugs or sperm/no. mated)
- Gestation Index (no. with implants/no. with seminal plugs or sperm)
- Gestation Length (when allowed to deliver pups)
- Body Weight
  - Treatment days (at least first, middle, and last treatment days)
  - Sacrifice day
- Body Weight Change
  - Throughout gestation
  - During treatment (including increments of time within treatment period)
  - Post-treatment to sacrifice
  - Corrected maternal (body weight change throughout gestation minus gravid uterine weight or litter weight at sacrifice)
- Organ Weights (in cases of suspected specific organ toxicity)
  - Absolute
  - Relative to body weight
- Food and Water Consumption (where relevant)
- Clinical Evaluations (on days of treatment and at sacrifice)
  - Types and incidence of clinical signs
  - Enzyme markers
  - Clinical chemistries
- Gross Necropsy and Histopathology

Body weight and the change in body weight are viewed collectively as indicators of maternal toxicity for most species, although these end points may not be as useful in rabbits, because body weight changes in rabbits are not good indicators of pregnancy status. Body weight changes may provide more information than a daily body weight measured during treatment or during gestation. Changes in weight during treatment could occur that would not be reflected in the total weight change throughout gestation, because of compensatory weight gain that may occur following treatment but before sacrifice. For this reason, changes in weight during treatment

can be examined as another indicator of maternal toxicity.

Changes in maternal body weight corrected for gravid uterine weight at sacrifice may indicate whether the effect is primarily maternal or fetal. For example, there may be a significant reduction in weight gain throughout gestation and in gravid uterine weight, but no change in corrected maternal weight gain which would indicate primarily an intrauterine effect. Conversely, a change in corrected weight gain and no change in gravid uterine weight suggests primarily maternal toxicity and little or no intrauterine effect. An alternate estimate of maternal weight change during gestation can be obtained by subtracting the sum of the weights of the fetuses. However, this weight does not include the uterine tissue, placental tissue, or the amniotic fluid.

Changes in other end points should also be determined. For example, changes in relative and absolute organ weights may be signs of a maternal effect when an agent is suspected of causing specific organ toxicity. Food and water consumption data are useful, especially if the agent is administered in the diet or drinking water. The amount ingested (total and relative to body weight) and the dose of the agent (relative to body weight) can then be calculated, and changes in food and water consumption related to treatment can be evaluated along with changes in body weight and body weight gain. Data on food and water consumption are also useful when an agent is suspected of affecting appetite, water intake, or excretory function. Clinical evaluations of toxicity may also be used as indicators of maternal toxicity. Daily clinical observations may be useful in describing the profile of maternal toxicity. Enzyme markers and clinical chemistries may be useful indicators of exposure but must be interpreted carefully as to whether or not a change constitutes toxicity. Gross necropsy and histopathology data (when specified in the protocol) may aid in determining toxic dose levels.

2. End Points of Developmental Toxicity. Because the maternal animal, and not the conceptus, is the individual treated during gestation, data generally should be calculated as incidence per litter or as number and percent of litters with particular end points.Table 2 indicates the way in which offspring and litter end points may be expressed.

Table 2.--End Points of Developmental Toxicity.

Litters with implants

No. implantation sites/dam
No. corpora lutea (CL)/dam[a]
Percent preimplantation loss

$$\frac{(\text{CL - implantations}) \times 100}{\text{CL}}$$ [a]

No. and percent live offspring/litter
No. and percent resorptions/litter
No. and percent litters with resorptions
[51 FR 34032]
No. and percent late fetal deaths/litter
No. and percent nonlive (late fetal deaths + resorptions) implants/litter
No. and percent litters with nonlive implants
No. and percent affected (nonlive + malformed) implants/litter
No. and percent litters with affected implants
No. and percent litters with total resorptions
No. and percent stillbirths/litter

Litters with live offspring[b]

No. and percent litters with live offspring
No. and percent live offspring/litter
Viability of offspring[c]
Sex ratio/litter
Mean offspring body weight/litter[c]
Mean male body weight/litter[c]
Mean female body weight/litter[c]
No. and percent externally malformed offspring/litter
No. and percent viscerally malformed offspring/litter
No. and percent skeletally malformed offspring/litter
No. and percent malformed offspring/litter
No. and percent litters with malformed offspring
No. and percent malformed males/litter
No. and percent malformed females/litter
No. and percent offspring with variations/litter
No. and percent litters having offspring with variations
Types and incidence of individual malformations
Types and incidence of individual variations
Individual offspring and their malformations and variations (grouped according to litter and dose)
Clinical signs[c]
Gross necropsy and histopathology

[a] Important when treatment begins prior to implantation. May be difficult in mice.

[b] Offspring refers both to fetuses observed prior to term or to pups following birth. The end points examined depend on the protocol used for each study.

[c] Measured at selected intervals until termination of the study.

When treatment begins prior to implantation, an increase in preimplantation loss could indicate an adverse effect either on the developing blastocyst or on the process of implantation itself. If treatment begins around the time of implantation (i.e., day 6 of gestation in the mouse, rat, or rabbit), an increase in preimplantation loss probably reflects normal variability in the animals being used, but the data should be examined carefully to determine whether or not the effect is dose related. If preimplantation

loss is related to dose in either case, further studies would be necessary to determine the mechanism and extent of such effects.

The number and percent of live offspring per litter, based on all litters, may include litters that have no live implants. The number and percent resorptions or late fetal deaths per litter gives some indication of when the conceptus died, and the number and percent nonlive implants per litter (postimplantation loss) is a combination of resorptions and late fetal deaths. The number and percent of litters showing an increased incidence for these end points is generally useful but may be less useful than incidence per litter because, in the former case, a litter is counted whether it has one or all resorbed, dead, or nonlive implants.

If a significant increase in postimplantation loss is found after exposure to an agent, the data may be compared not only with concurrent controls, but also with recent historical control data, since there is considerable interlitter variability in the incidence of postimplantation loss (22). If a given study control group exhibits an unusually high or low incidence of postimplantation loss compared to historical controls, then scientific judgment must be used to determine the adequacy of the studies for risk assessment purposes.

The end point for affected implants (i.e., the combination of nonlive and malformed conceptuses) gives an indication of the total intrauterine response to an agent and sometimes reflects a better dose-response relationship than does the incidence of nonlive or malformed offspring taken individually. This is especially true at the high end of the dose-response curve in cases when the incidence of nonlive implants per litter is greatly increased. In such cases, the malformation rate may appear to decrease because only unaffected offspring have survived. If the incidence of prenatal death or malformation is unchanged, then the incidence of affected implants will not provide any additional dose-response information. In studies where maternal animals are allowed to deliver pups normally, the number of stillbirths per litter should also be noted.

The number of live offspring per litter, based on those litters that have one or more live offspring, may be unchanged even though the incidence of nonlive in all litters is increased. This could occur either because of an increase in the number of litters with no live offspring, or an increase in the number of implants per litter. A decrease in the number of live offspring per litter should be accompanied by an increase in the incidence of nonlive implants per litter, unless the implant numbers differ among dose groups. In postnatal studies, the viability of live born offspring should be determined at selected intervals until termination of the study.

The sex ratio per litter, as well as the body weights of males and females, can be examined to determine whether or not one sex is preferentially affected by the agent. However, this is an unusual occurrence.

A change in offspring body weight is a sensitive indicator of developmental toxicity, in part because it is a continuous variable. In some cases, offspring weight reduction may be the only indicator of developmental toxicity; if so, there is always a question remaining as to whether weight reduction is a permanent or transitory effect. A permanent weight change may be considered more severe than a transitory change, although little is known about the long-term consequences of short-term fetal or neonatal weight changes. When fetal or neonatal weight reduction is the only indicator of developmental toxicity, data from the two-generation reproduction study (2), if available, may be useful for evaluating these parameters. Ideally, follow-up studies to evaluate postnatal viability, growth, and survival through weaning should be conducted. There are other factors that should be considered in the evaluation of fetal or neonatal weight changes. For example, in polytocous animals, fetal and neonatal weights are usually inversely correlated with litter size, and the upper end of the dose-response curve may be confounded by smaller litters and increased fetal or neonatal weight. Additionally, the average body weight of males is greater than that of females in the more commonly used laboratory animals.

Live offspring should be examined for external, visceral, and skeletal malformations. If only a portion of the litter is examined, then it is preferable that those examined be randomly selected from each litter. An increase in the incidence of malformed offspring may be indicated by a change in one or more of the following end points: the incidence of malformed offspring per litter, the number and percent of litters with malformed offspring, or the number of offspring or litters with a particular malformation that appears to increase with dose as indicated by the incidence of individual types of malformations.

[51 FR 34033]

Other ways of examining the data include the incidence of external, visceral, and skeletal malformations which may indicate which general systems are affected. A listing of individual offspring with their malformations and variations may give an indication of the pattern of developmental deviations. All of these methods of expressing and examining the data are valid for determining the effects of an agent on structural development. However, care must be taken to avoid counting offspring more than once in evaluating any single end point based on number or percent of offspring or litters. The incidence of individual types of malformations and variations should be examined

for significant changes which may be masked if the data on all malformations and variations are pooled. Appropriate historical control data are helpful in the interpretation of malformations and variations, especially those that normally occur at a low incidence apparently unrelated to dose in an individual study. Although a dose-related increase in malformations is interpreted as an adverse developmental effect of exposure to an agent, the significance of anatomical variations is more difficult to determine, and must take into account what is known about developmental stage (e.g., with skeletal ossification), background incidence of certain variations (e.g., 12 or 13 pairs of ribs in rabbits), or other strain - or species-specific factors. However, if variations are significantly increased in a dose-related manner, these should also be evaluated as a possible indication of developmental toxicity. The Interagency Regulatory Liaison Group noted that dose-related increases in defects, which may occur spontaneously, are as relevant as dose-related increases in any other developmental toxicity end points (23).

3. *Functional Developmental Toxicology.* Developmental effects, which are inducible by exogenous agents, are not limited to death, structural abnormalities, and altered growth. Rather, it has been demonstrated in a number of instances that subtle alterations in the functional competence of an organ or a variety of organ systems may result from exposure during critical developmental periods that may occur between conception and sexual maturation. Often, these functional defects are observed at dose levels below those at which gross malformations are evident (24). At present, such testing is not routinely required in the United States. However, data from postnatal studies, when available, are considered very useful for the assessment of the relative importance and severity of findings in the fetus and neonate. Often, the long-term consequences of adverse developmental outcomes at birth are unknown, and further data on postnatal development and function may contribute valuable information. When regulatory statutes permit, studies designed to evaluate adverse fetal or neonatal outcomes have been requested (e.g., the Office of Pesticide Programs has sometimes requested postnatal studies where the reversibility of study findings were at issue). In some cases, useful data can be derived from well-executed multigeneration studies.

Much of the early work in functional developmental toxicology was related to behavioral evaluations, and the term "behavioral teratology" became prominent in the mid 1970s. Less work has been done on other functional systems, but sufficient data have accumulated to indicate that the cardiopulmonary, immune, endocrine, digestive, urinary, nervous, and reproductive systems are subject to alterations in functional competence (25, 26). Currently, there are no standard testing procedures, although some attempts are being made to standardize and evaluate tests and protocols (27). The functional evaluation of specific systems often involves highly specialized training and equipment. The routine use of such test procedures may not always be practical, but may be extremely important in determining the nature of a suspected alteration in terms of its biological significance and dose-response relationship.

The interpretation of data from functional developmental toxicology studies is limited due to the lack of knowledge about the underlying toxicological mechanisms and their significance. However, since such data are sometimes encountered in the risk assessment of particular agents, some guidance is provided here concerning general concepts of study design and evaluation.

a. Several aspects of study design are similar to those important in standard developmental toxicity studies (e.g., a dose-response approach with the highest dose producing minimal overt maternal or perinatal toxicity, number of litters large enough for adequate statistical power, randomization of animals to dose groups, litter generally considered the statistical unit, etc.).

b. A replicate study design provides added confidence in the interpretation of data.

c. Use of a pharmacological challenge may be valuable in evaluating function and "unmasking" effects not otherwise detectable, particularly in the case of organ systems that are endowed with a reasonable degree of functional reserve capacity.

d. Use of functional tests with a moderate degree of background variability may be more sensitive to the effects of an agent than are tests with low variability that may be impossible to disrupt without being life-threatening. Butcher et al. (28) have discussed this with relation to behavioral end points.

e. A battery of functional tests usually provides a more thorough evaluation of the functional competence of an animal; tests conducted at several ages may provide more information about maturational changes.

f. Critical periods for the disruption of functional competence include both the prenatal and the postnatal periods to the time of sexual maturation, and the effect is likely to vary depending on the time and degree of exposure.

Although interpretation of functional data may be difficult at present, there are at least three ways in which the data from these studies may be useful for risk assessment purposes: 1) to help elucidate the long-term consequences of fetal and neonatal findings; 2) to indicate the potential for an agent to cause functional alterations, and the effective doses relative to those that produce other forms of toxicity;

and 3) for existing environmental agents, to focus on organ systems to be evaluated in exposed human populations.

4. *Overall Evaluation of Maternal and Developmental Toxicity.* As discussed previously, individual end points are evaluated in developmental toxicity studies, but an integrated evaluation must be done considering all maternal and developmental end points in order to interpret the data fully. Developmental toxicity is considered to be an increase in the incidence of malformed offspring, decreased viability (prenatal or postnatal), altered growth, and/or functional deficits.

The level of concern for a developmental toxic effect is related to several issues, including the relative toxicity of an agent to the offspring versus the adult animal, and the long-term consequences of findings in the fetus or neonate. Those agents which produce developmental toxicity at a dose that is not toxic to the maternal animal are of greatest concern because the developing organism appears to be selectively affected or more sensitive than the adult. However, when developmental effects are produced only

[51 FR 34034]

at maternally toxic doses, the types of developmental effects should be examined carefully, and not discounted as being secondary to maternal toxicity. Current information is inadequate to assume that developmental effects at maternally toxic doses result only from the maternal toxicity; rather, when the lowest observed effect level is the same for the adult and developing organisms, it may simply indicate that both are sensitive to that dose level. Moreover, the maternal effects may be reversible while effects on the offspring may be permanent. These are important considerations for agents to which humans may be exposed at minimally toxic levels either voluntarily or in the workplace, since several agents are known to produce adverse developmental effects at minimally toxic doses in adult humans (e.g., smoking, alcohol).

Approaches for ranking agents for their selective developmental toxicity are being developed; Schardein (10) has reviewed several of these. Of current interest are approaches that develop ratios relating an adult toxic dose to a developmental toxic dose (29, 30, 31, 32). Ratios near unity indicate that developmental toxicity occurs only at doses producing maternal toxicity; as the ratio increases, there is a greater likelihood of developmental effects occurring without maternal manifestations. Although further exploration and validation are necessary, such approaches may ultimately help in identifying those agents that pose the greatest threat and should be given higher priority for further testing (33).

5. *Short-term Testing in Developmental Toxicity.* The need for short-term tests for developmental toxicity has arisen from the large number of agents in or entering the environment, the interest in reducing the number of animals used for routine testing, and the expense of testing. Two approaches are considered here in terms of their contribution to the overall testing process: 1) an *in vivo* mammalian screen, and 2) a variety of *in vitro* systems. Currently, neither approach is considered as a replacement for routine *in vivo* developmental toxicity testing in experimental animals, and should not be used to make the final decision as to whether an agent is a positive or negative developmental toxicant; rather, such tests may be useful as tools for assigning priorities for further, more extensive testing. Although such short-term tests are not routinely required, data are sometimes encountered in the review of chemicals; the comments are provided here for guidance in the evaluation of such data.

a. *In Vivo* Mammalian Developmental Toxicity Screen. The most widely studied *in vivo* approach is that developed by Chernoff and Kavlock (34) which uses the pregnant mouse. This approach is based on the hypothesis that a prenatal injury, which results in altered development, will be manifested postnatally as reduced viability and/or impaired growth. In general, the test substance is administered over the period of major organogenesis at a single dose level that will elicit some degree of maternal toxicity. A second lower dose level may be used which potentially will reduce the chances of false positive results. The pups are counted and weighed shortly after birth, and again after 3-4 days. End points that are considered in the evaluation include: general maternal toxicity (including survival and weight gain), litter size, and viability, weight, and gross malformations in the offspring. Basic priority-setting categories for more extensive testing have been suggested: 1) agents that induce perinatal death should receive highest priority, 2) agents inducing perinatal weight changes should be ranked lower in priority, and 3) agents inducing no effect should receive the lowest priority (34). Another scheme that has been proposed applies a numerical ranking to the results as a means of prioritizing agents for further testing (35, 36).

The mouse was chosen originally for this test because of its low cost, but the procedure should be easily applicable to other species. However, the test will only predict the potential for developmental toxicity of an agent in the species utilized and does not improve the ability to extrapolate risk to other species, including humans. The Office of Toxic Substances has developed testing guidelines for this procedure (37). Although the testing guidelines are available, such procedures are not routinely

required, and further validation is currently being carried out (38).

b. *In Vitro* Developmental Toxicity Screens. Test systems that fall under the general heading of "*in vitro*" developmental toxicity screens include any system that employs a test subject other than the intact pregnant mammal. These systems have long been used to assess events associated with normal and abnormal development, but only recently have they been considered for their potential as screens in testing (39, 40, 41). Many of these systems are now being evaluated for their ability to predict the developmental toxicity of various agents in intact mammalian systems. This validation process requires certain considerations in study design, including defined end points for toxicity and an understanding of the system's ability to handle various test agents (40, 42). A list of agents for use in such validation studies has been developed (43).

6. *Statistical Considerations*. In the assessment of developmental toxicity data, statistical considerations require special attention. Since the litter is generally considered the experimental unit in most developmental toxicity studies, the statistical analyses should be designed to analyze the relevant data based on incidence per litter or on the number of litters with a particular end point. The analytical procedures used and the results, as well as an indication of the variance in each end point, should be clearly indicated in the presentation of data. Analysis of variance (ANOVA) techniques, with litter nested within dose in the model, take the litter variable into account but allow use of individual offspring data and an evaluation of both within and between litter variance as well as dose effects. Nonparametric and categorical procedures have also been widely used for binomial or incidence data. In addition, tests for dose-response trends can be applied. Although a single statistical approach has not been agreed upon, a number of factors important in the analysis of developmental toxicity data have been discussed (23, 44).

Studies that employ a replicate experimental design (e.g., two or three replicates with 10 litters per dose per replicate rather than a single experiment with 20-30 litters per dose group) allow for broader interpretation of study results since the variability between replicates can be accounted for using ANOVA techniques. Replication of effects due to a given agent within a study, as well as between studies or laboratories, provides added strength in the use of data for the estimation of risk.

An important factor to determine in evaluating data is the power of a study (i.e., the probability that a study will demonstrate a true effect), which is limited by the sample size used in the study, the background incidence of the end point observed, the variability in the incidence of the end point, and the analysis method. As an example, Nelson and Holson (45) have shown that the number of litters needed to detect a 5 or 10% change was dramatically lower for fetal weight (a continuous variable with low variability) than for resorptions (a binomial response with high variability). With the current recommendation in

[51 FR 34035]

testing protocols being 20 rodents per dose group (2, 3), it is possible to detect an increased incidence of malformations in the range of 5 to 12 times above control levels, an increase of 3 to 6 times the *in utero* death rate, and a decrease of 0.15 to 0.25 times the fetal weight. Thus, even within the same study, the ability to detect a change in fetal weight is much greater than for the other end points measured. Consequently, for statistical reasons only, changes in fetal weight are often observable at doses below those producing other signs of developmental toxicity. Any risk assessment should present the detection sensitivity for the study design used and for the end point(s) evaluated.

Although statistical analyses are important in determining the effects of a particular agent, the biological significance of data should not be overlooked. For example, with the number of end points that can be observed in developmental toxicity studies, a few statistically significant differences may occur by chance. On the other hand, apparent trends with dose may be biologically relevant even though statistical analyses do not indicate a significant effect. This may be true especially for the incidence of malformations or *in utero* death where a relatively large difference is required to be statistically significant. It should be apparent from this discussion that a great deal of scientific judgment based on experience with developmental toxicity data and with principles of experimental design and statistical analysis may be required to adequately evaluate such data.

B. Human Studies

Because of the ethical considerations involved, studies with deliberate dosing of humans are not done. Therefore, dose-effect developmental toxicity data from humans are limited to those available from occupational, environmental, or therapeutic exposures. While animal studies provide dose-response data that can be used in the extrapolation of risk to humans, good epidemiologic data provide the best information for assessing human risk.

The category of "human studies" includes both epidemiologic studies and other reports of cases or clusters of events. While case reports have been important in identifying several human teratogens, they are potentially of greater value in identifying topics for further investigation (46). The data from case reports are often of an anecdotal or highly selected nature, and thus are of limited usefulness for risk assessment except when a unique defect is produced, as with thalidomide, or when the agent is

so potent as to greatly increase the incidence of a particular defect(s).

As there are many different designs for epidemiologic studies, simple rules for their evaluation do not exist. The assessment of epidemiologic studies requires a sophisticated level of understanding of the appropriate epidemiologic and statistical methods and interpretation of the findings. Factors that increase a study's usefulness for risk assessment include such things as the examination of multiple end points and exposure levels, the validity of the data, and proper control of other risk factors, effect modifiers, and confounders in the study design and/or analysis. A more in-depth discussion can be found elsewhere (47).

As described earlier, a single developmental toxicant can result in multiple end points (malformations, functional impairment, altered growth, and/or lethality). These end points can be thought of as sequential competing risks. For example, a malformed fetus spontaneously aborted would not be observed in a study of births with malformations (48). Very early conceptus losses may not be identified in human populations, whereas in most laboratory animal studies, all resorption sites can be identified. Many epidemiologic studies, especially of the case-control design, have focused on one end point, possibly missing a true effect of exposure. Furthermore, some studies have selected one type or class of malformations to study. Since an agent can result in different spectra of malformations following exposure at different times in the pregnancy (49), limiting a study to one class of malformation may give misleading results. Malformations can be meaningfully grouped only if there is a logical underlying teratogenic mechanism or pathogenetic pathway. As a minimum, malformations, deformations, and disruptions should be separated.

The power, or probability of a study to detect a true effect, is dependent upon the size of the study group, the frequency of the outcome in the general population, and the level of excess risk to be identified. Rarer outcomes, such as malformations, require thousands of pregnancies to have a high probability of detecting an increase in risk. More common outcomes, such as fetal loss, require hundreds of pregnancies to have the same probability (8, 23, 50, 51, 52, 53). The confidence one has in the results of a study with negative findings is directly related to the power of the study to detect clinically meaningful differences in incidence for the end points studied.

As in animal studies, pregnancies within the same family (or litter) are not independent events. In animal studies, the litter is generally used as the unit of measure. This approach is difficult in humans since the pregnancies are sequential, with the risk factors changing for the different pregnancies (23, 51, 54). If more than one pregnancy per family is included, and this is often necessary due to small study groups, the use of non-independent observations overestimates the true size of the population at risk and artificially increases the significance level (54).

Other criteria for evaluating epidemiologic studies include the following (23, 50, 52, 55, 56, 57, 58):

1. The potential for complete or relatively complete ascertainment of events for study. This can vary by outcome and by data source; for example, if hospital records are used, early fetal losses will be underascertained, but a more complete list of pregnancies could be obtained by interviewing the women. Congenital malformations can be more completely ascertained using hospital records than birth certificates. Studies with relatively complete ascertainment of events, or at least low probability of unbiased ascertainment, should carry more weight.

2. Validity (accuracy) of the data. Recall of past events in interviews may be faulty, while hospital files contain data recorded at the time of the event (but may be incomplete). Validation of interview data with an independent source, where possible, increases confidence in the results of the study.

3. Collection of data on other risk factors, effect modifiers, and confounders. Data on smoking, alcohol consumption, drug use, and environmental and occupational exposure, etc., during pregnancy should be examined and controlled for in the study design and/or analysis where appropriate. The analytic techniques used to control for these factors require careful consideration in their application and interpretation.

C. Other Considerations

1. *Pharmacokinetics*. Extrapolation of data between species can be aided considerably by the availability of data on the pharmacokinetics of a particular agent in the species tested and, if possible, in humans. Information on half-lives, placental metabolism and transfer, and concentrations of the parent compound and metabolites in the

[51 FR 34036]

maternal animal and conceptus may be useful in predicting risk for developmental toxicity. Such data may also be helpful in defining the dose-response curve, developing a more accurate comparison of species sensitivity including that of humans (59, 60), determining dosimetry at target sites, and comparing pharmacokinetic profiles for various dosing regimens or routes of exposure. Pharmacokinetic studies in developmental toxicology are most useful if conducted in pregnant animals at the stage when developmental insults occur. The correlation of pharmacokinetic parameters and developmental toxicity data may be

useful in determining the contribution of specific pharmacokinetic parameters to the effects observed (61).

2. *Comparisons of Molecular Structure.* Comparisons of the chemical or physical properties of an agent with those of known developmental toxicants may provide some indication of a potential for developmental toxicity. Such information may be helpful in setting priorities for testing of agents or for evaluation of potential toxicity when only minimal data are available. Structure/activity relationships have not been well studied in developmental toxicology, although data are available that suggest structure-activity relationships for certain classes of chemicals (e.g., glycol ethers, steroids, retinoids). Under certain circumstances (e.g., in the case of new chemicals), this is one of several procedures used to evaluate the potential for toxicity when little or no data are available.

D. Weight-of-Evidence Determination

Information available from studies discussed previously, whether indicative of potential concern or not, must be evaluated and factored into the risk assessment. The types of data may vary from chemical to chemical, and certain types of data may be more relevant than other types in performing developmental toxicity assessments. The primary considerations are the human data (which are seldom available) and the experimental animal data. The qualitative assessment for developmental toxicity should include statements concerning the quality of the data, the resolving power of the studies, the number and types of end points examined, the relevance of route and timing of exposure, the appropriateness of the dose selection, the replication of the effects, the number of species examined, and the availability of human case reports, case series, and/or epidemiologic study data. In addition, pharmacokinetic data and structure-activity considerations, as well as other factors that may affect the quality, should be taken into account. Therefore, all data pertinent to developmental toxicity should be examined in the evaluation of a chemical's potential to cause developmental toxicity in humans, and sound scientific judgment should be exercised in interpreting the data in terms of the risk for adverse human developmental health effects.

IV. Quantitative Assessment

Risk assessment involves the description of the nature and often the magnitude of potential human risk, including a description of any attendant uncertainty. In the final phase of the risk assessment (risk characterization), the results of the qualitative evaluation (hazard identification), the dose-response, and the exposure assessments are combined to give qualitative and/or quantitative estimates of the developmental toxicity risk. A summary of the strengths and weaknesses of the hazard identification, dose-response assessment, and exposure assessment should be discussed. Major assumptions, scientific judgments, and, to the extent possible, estimates of the uncertainties in the assessment also should be presented.

A. Dose-Response Assessment

When quantitative human dose-effect data are available and with sufficient range of exposure, dose-response relationships may be examined. However, such data have rarely been available; thus, other methods have been used in developmental toxicology for estimating exposure levels that are unlikely to produce adverse effects in humans. The dose-response assessment is usually based on the evaluation of tests performed in laboratory animals. Evidence for a dose-response relationship is an important criterion in the assessment of developmental toxicity, although this may be based on limited data from standard three-dose studies. As mentioned earlier (section III. A. 2.), however, traditional dose-response relationships may not always be observed for some end points. For example, as the exposure level rises, embryo/fetolethal levels may be reached, resulting in an observed decrease in malformations with increasing dose (49, 51). The potential for this relationship indicates that dose-response relationships for individual end points as well as combinations of end points (e.g., dead and malformed combined) must be carefully examined and interpreted.

Although dose-response data are important in this area, the approaches frequently employed in attempts to extrapolate to humans has involved simply the use of uncertainty (safety) factors and margins of safety, which in some respects are conceptually similar. However, uncertainty factors and margins of safety are computed differently and are often used in different regulatory situations. The choice of approach is dependent upon many factors, including the statute involved, the situation being addressed, the data base used, and the needs of the decision-maker. The final uncertainty factor used and the acceptability of the margin of safety are risk management decisions, but the scientific issues that must be taken into account are addressed here.

The uncertainty factor approach results in a calculated exposure level believed to be unlikely to cause any toxic developmental response in humans. The size of the uncertainty factor will vary from agent to agent and will require the exercise of scientific judgment (10, 62), taking into account interspecies differences, the nature and extent of human exposure, the slope of the dose-response curve, the types of developmental effects observed, and the relative dose levels for maternal and developmental toxicity in the test species. The uncertainty factor selected is then divided into the NOEL for the most sensitive end point obtained from

the most appropriate and/or sensitive mammalian species examined to obtain an acceptable exposure level. Currently, there is no one laboratory animal species that can be considered most appropriate for predicting risk to humans (10). Each agent should be considered on a case-by-case basis.

The margin of safety approach derives a ratio of the NOEL from the most sensitive species to the estimated human exposure level from all potential sources (63). The adequacy of the margin of safety is then considered, based on the weight of evidence, including the nature and quality of the hazard and exposure data, the number of species affected, dose-response relationships, and other factors such as benefits of the agent.

Although the standard study design for a developmental toxicity study calls for a low dose that demonstrates a NOEL, there may be circumstances where a risk assessment is based on the results of a study in which a NOEL for developmental toxicity could not be identified. Rather, the lowest dose administered caused significant effect(s) and was identified as the lowest observed effect level (LOEL). In circumstances where only a LOEL is available, it may be appropriate to apply

[51 FR 34037]

an additional uncertainty factor. The magnitude of this additional factor is dependent upon scientific judgment. In some instances, additional studies may be needed to strengthen the confidence in this additional uncertainty factor.

B. Exposure Assessment

The results of the dose-response assessment are combined with an estimate of human exposure in order to obtain a quantitative estimate of risk. The Guidelines for Estimating Exposures are published separately (64) and will not be discussed in detail here. In general, the exposure assessment describes the magnitude, duration, schedule, and route of exposure. This information is developed from monitoring data and from estimates based on modeling of environmental exposures. Unique considerations relevant to developmental toxicity are duration and period of exposure as related to stage of development (i.e., critical periods), and the possibility that a single exposure may be sufficient to produce adverse developmental effects (i.e., chronic exposure is not a necessary prerequisite for developmental toxicity to be manifested). Also, it should be recognized that exposure of almost any segment of the human population (i.e., fertile men and women, the conceptus, and the child up to the age of sexual maturation) may lead to risk to the developing organism.

Data on exposure to humans may be qualitative or quantitative. The qualitative data could be surrogate data, such as employment or residence histories; quantitative or dose data are frequently not available. Exposures at different stages of the reproductive process can result in different outcomes (49). In laboratory studies, these time periods can be carefully controlled. In human studies, especially retrospective ones, linking of specific time periods and specific exposures, even on a qualitative level, may be difficult due to errors of recall or record keeping (where records are available). The increased probability of misclassification of exposure status may affect the ability of a study to recognize a true effect (8, 23, 52, 65, 66).

Exposure may be defined at a specific point in time, or the cumulative lifetime exposure up to a specific point in time. Each of these definitions carries an implicit assumption about the underlying relationship between exposure and outcome. For example, a cumulative exposure measure assumes that total lifetime exposure is important, with a greater probability of effect with greater total exposure; a dichotomous exposure measure (ever exposed versus never exposed) assumes an irreversible effect of exposure; and exposure at a specific time in the reproductive process assumes that only concurrent exposure is important. The appropriate exposure depends on the outcome(s) studied, the biologic mechanism affected by exposure, and the half-life of the exposure. Unbiased misclassification of exposure, due either to poor data or to an inappropriate exposure variable, may result in missing an effect of the agent under study.

C. Risk Characterization

Many uncertainties have been pointed out in these Guidelines which are associated with the toxicological and exposure components of risk assessments in developmental toxicology. In the past, these uncertainties have often not been readily apparent or consistently presented. The presentation of any risk assessment for developmental toxicity should be accompanied by statements concerning the strength of the hazard evaluation (see section III. D. for more detail) as well as dose-response relationships, estimates of human exposure, and any other factors that affect the quality and precision of the assessment. The dose-response and exposure data are combined to estimate risk based on a NOEL for any adverse developmental effect. The uncertainty factor selected or margin of safety calculated should be sufficiently qualified as to the assumptions used and the accuracy of the estimates.

At present, there are no mathematical models that are generally accepted for estimating developmental toxicity responses below the applied dose range. This is due primarily to a lack of understanding of the biological mechanisms underlying developmental toxicity, intra/interspecies differences in the types of developmental events, the influence of maternal effects on the dose-response curve, and whether or not a threshold exists below which no effect will be

produced by an agent. Many developmental toxicologists assume a threshold for most developmental effects; this assumption is based largely on the biological rationale that the embryo is known to have some capacity for repair of the damage or insult (49), and that most developmental deviations are probably multifactorial in nature (67). The existence of a NOEL in an animal study does not prove or disprove the existence or level of a true threshold; it only defines the highest level of exposure under the conditions of the test that are not associated with a significant increase in effect. The use of NOELs and uncertainty factors or margins of safety are attempts to ensure that the allowable levels are below those that will produce a significant increase in developmental effects.

Discussions of risk extrapolation procedures have noted that further work is needed to improve mathematical tools for developing estimates of potential human developmental risk (62, 68). Gaylor (69) has suggested an approach for controlling risk that combines the use of mathematical models for low-dose estimation of risk with the application of an uncertainty factor based on a preselected level of allowable risk. This approach is similar to approaches proposed for carcinogenesis, but does not preclude the possibility of a threshold, and may provide a more quantitative approach to controlling risk. Several such approaches are being examined. For the most part, the Agency will continue to use uncertainty factors and margins of safety as described above. Other appropriate methods for expressing risk are being sought and will be applied if considered acceptable.

These Guidelines summarize the procedures that the U.S. Environmental Protection Agency will follow in evaluating the potential for agents to cause developmental toxicity. These Guidelines will be reviewed and updated as advances are made in the field, since it is evident that our ability to evaluate and predict human developmental toxicity is imprecise. Further studies that 1) delineate the mechanisms of developmental toxicity and pathogenesis, 2) provide comparative pharmacokinetic data, and 3) elucidate the functional modalities that may be altered by exposure to toxic agents will aid in the interpretation of data and interspecies extrapolation. These types of studies, along with further evaluation of the relationship between maternal and fetal toxicity and the concept of a threshold in developmental toxicity, will provide for the development of improved mathematical models to more precisely assess risk.

V. References

1. U.S. Environmental Protection Agency. 1980. Assessment of risks to human reproduction and to development of the human conceptus from exposure to environmental substances, pp. 99-116. Available from: NTIS, Springfield, VA, DE82-007897.

2. U.S. Environmental Protection Agency. 1982. Pesticides assessment guidelines, subdivision F. Hazard evaluation: human and domestic animals. EPA-540/9-82-025. Available from: NTIS, Springfield, VA.

[51 FR 34038]

3. U.S. Environmental Protection Agency. 1985. Toxic Substances Control Act Test Guidelines; Final Rules. Federal Register 50:39426-39428 and 39433-39434.

4. Hertig, A.T. 1967. The overall problem in man. In: K. Benirschke, ed. Comparative aspects of reproductive failure. New York, NY: Springer-Verlag, pp. 11-41.

5. McKeown, T., and R.C. Record. 1963. Malformations in a population observed for five years after birth. In: G.E.W. Wolstenholme and C.M. O'Conner, eds. Ciba Foundation symposium on congenital malformations. Boston, MA: Little Brown, pp. 2-16.

6. Mellin, G.W., and M. Katzenstein. 1964. Increased incidence of malformations-chance or change? J. Am. Med. Assoc. 187:570-573.

7. Wilson, J.G. 1977. Embryotoxicity of drugs in man. In: J.G. Wilson and F.C. Fraser, eds. Handbook of teratology. New York, NY: Plenum Press, pp. 309-355.

8. Selevan, S.G. 1981. Design considerations in pregnancy outcome studies of occupational populations. Scand. J. Work Environ. Health 7:76-82.

9. Shepard, T.H. 1980. Catalog of teratogenic agents. Third edition. Baltimore, MD: Johns Hopkins University Press.

10. Schardein, J.L. 1983. Teratogenic risk assessment. In: H. Kalter, ed. Issues and reviews in teratology, Vol. 1. New York, NY: Plenum Press, pp. 181-214.

11. Shepard, T.H. 1984. Teratogens: an update. Hosp. Pract., Jan., pp. 191-200.

12. Brown, N.A., and S. Fabro. 1983. The value of animal teratogenicity testing for predicting human risk. Clin. Obstet. Gynecol. 26:467-477.

13. Kimmel, C.A., J.F. Holson, C.J. Hogue, and G.L. Carlo. 1984. Reliability of experimental studies for predicting hazards to human development. NCTR Technical Report for Experiment No. 6015. NCTR, Jefferson, Arkansas.

14. Nisbet, I.C.T. and N.J. Karch. 1983. Chemical hazards to human reproduction. Park Ridge: Noyes Data Corp.

15. Committee on the Institutional Means for the Assessment of Risks to Public Health. 1983. Risk assessment in the Federal government: managing the process. Commission on Life Sciences, National Research Council. Washington, DC: National Academy Press, pp. 17-83.

16. U.S. Environmental Protection Agency. 1986, Sept. 24. Guidelines for carcinogen risk assessment. Federal Register 51(185):33992-34003.

17. U.S. Environmental Protection Agency. 1986, Sept. 24. Guidelines for mutagenicity risk assessment. Federal Register 51(185):34006-34012.

18. Food and Drug Administration. 1966. Guidelines for reproduction and teratology of drugs. Bureau of Drugs.

19. Food and Drug Administration. 1970. Advisory committee on protocols for safety evaluations. Panel on reproduction report on reproduction studies in the safety evaluation of food additives and pesticide residues. Toxicol. Appl. Pharmacol. 16:264-296.

20. Organization for Economic Cooperation and Development (OECD). 1981. Guideline for testing of chemicals teratogenicity.

21. Symposium on effects of radiation and other deleterious agents on embryonic development. 1954. J. Cell. Comp. Physiol. 43 (suppl. 1).

22. Woo, D.C., and R.M. Hoar. 1979. Reproductive performance and spontaneous malformations in control Charles River rats. A joint study for MARTA. Teratology 19:54A.

23. Kimmel, C.A., G.L. Kimmel, and V. Frankos, eds. 1986. Interagency Regulatory Liaison Group workshop on reproductive toxicity risk assessment. Environ. Health Perspect. 66:193-221.

24. Rodier, P.M. 1978. Behavorial teratology. In: J.G. Wilson and F.C. Fraser, eds. Handbook of teratology, Vol. 4. New York, NY: Plenum Press, pp. 397-428.

25. Buelke-Sam, J., and C.A. Kimmel. 1979. Development and standardization of screening methods for behavorial teratology. Teratology 20:17-29.

26. Kavlock, R.J., and C.T. Grabowski, eds. 1983. Abnormal

functional development of the heart, lungs, and kidneys: approaches to functional teratology. Prog. Clin. Biol. Res., Vol. 140. New York, NY: Alan R. Liss, Inc.

27. Buelke-Sam, J., C.A. Kimmel, and J. Adams, eds. 1985. Design considerations in screening for behavioral teratogens: Results of the Collaborative Behavioral Teratology Study. Neurobehav. Toxicol. Teratology 7(6):537-789.

28. Butcher, R.E., V. Wootten, and C.V. Vorhees. 1980. Standards in behavorial teratology testing: test variability and sensitivity. Teratogenesis Carcinog. Mutagen. 1:49-61.

29. Johnson, E.M. 1981. Screening for teratogenic hazards: nature of the problem. Annu. Rev. Pharmacol. Toxicol. 21:417-429.

30. Johnson, E.M., and B.E.G. Gabel. 1983. An artificial embryo for detection of abnormal developmental biology. Fund. Appl. Toxicol. 3:243-249.

31. Fabro, S., G. Schull, and N.A. Brown. 1982. The relative teratogenic index and teratogenic potency: proposed components of developmental toxicity risk assessment. Teratogenesis Carcinog. Mutagen. 2:61-76.

32. Brown, N.A., and S.J. Freeman. 1984. Aternative tests for teratogenicity. Alternatives Lab. Anim. 12:7-23.

33. Johnson, E.M. 1984. A prioritization and biological decision tree for developmental toxicity safety evaluations. J. Am. Coll. Toxicol. 3:141-147.

34. Chernoff, N., and R.J. Kavlock. 1982. An *in vivo* teratology screen utilizing pregnant mice. J. Toxicol. Environ. Health 10:541-550.

35. Brown, J.M. 1984. Validation of an *in vivo* screen for the determination of embryo/fetal toxicity in mice. SRI International. EPA contract no. 68-01-5079.

36. Schuler R., B. Hardin, R. Niemeyer, G. Booth, K. Hazelden, V. Piccirillo, and K. Smith. 1984. Results of testing fifteen glycol ethers in a short-term, *in vivo* reproductive toxicity assay. Environ. Health Perspect. 57:141-146.

37. U.S. Environmental Protection Agency. 1985. Toxic Substances Control Act Test Guidelines; Final Rules. Federal Register 50:39428-39429.

38. National Toxicology Program. 1984. Reproductive and Developmental Toxicology Program. Fiscal year 1984 annual plan, pp. 171-181.

39. Wilson, J.G. 1978. Survey of *in vitro* systems: their potential use in teratogenicity screening. In: J.G. Wilson and F.C. Fraser, eds. Handbook of teratology, Vol. 4. New York, NY: Plenum Press, pp. 135-153.

40. Kimmel, G.L., K. Smith, D.M. Kochhar, and R.M. Pratt. 1982. Overview of *in vitro* teratogenicity testing: aspects of validation and application to screening. Teratogenesis Carcinog. Mutagen. 2:221-229.

41. Brown, N.A., and S.E. Fabro. 1982. The *in vitro* approach to teratogenicity testing. In: K. Snell, ed. Developmental toxicology. London, England: Croom-Helm, pp. 31-57.

42. Kimmel, G.L. 1985. *In vitro* tests in screening teratogens: considerations to aid the validation process. In: M. Marois, ed. Prevention of physical and mental congenital defects, Part C. New York, NY: Alan R. Liss, Inc., pp. 259-263.

43. Smith, M.R., G.L. Kimmel, D.M. Kochhar, T.H. Shepard, S.P. Spielberg, and J.G. Wilson. 1983. A selection of candidate compounds for *in vitro* teratogenesis test validation. Teratogenesis Carcinog. Mutagen. 3:461-480.

44. Haseman, J.K. and L.L. Kupper. 1979. Analysis of dichotomous response data from certain toxicological experiments. Biometrics 35:281-293.

45. Nelson, C.J., and J.F. Holson. 1978. Statistical analysis of teratogenic data: problems and advancements. J. Environ. Pathol. Toxicol. 2:187-199.

46. Hogue, C.J.R. 1985. Developmental risks. Presented at Symposium: Epidemiology and health risk assessment. Columbia, MD, May 14, 1985.

47. Epidemiology Workgroup of the Interagency Regulatory Liaison Group. 1981. Guidelines for documentation of epidemiologic studies. Am. J. Epidemiol. 114(5):609-613.

48. Stein, Z., M. Susser, D. Warburton, J. Wittes, and J. Kline. 1975. Spontaneous abortion as a screening device. The effect of fetal surveillance on the incidence of birth defects. Am. J. Epidemiol. 102:275-290.

49. Wilson, J.G. 1973. Environment and birth defects. New York, NY: Academic Press, pp. 30-32.

50. Report of Panel II. 1981. Guidelines for reproductive studies in exposed human populations. In: Bloom, A.D., ed. Guidelines for studies of populations exposed to mutagenic and reproductive hazards. White Plains, NY: March of Dimes Birth Defects Foundation, pp. 37-110.

51. Selevan, S.G. 1985. Design of pregnancy outcome studies of industrial exposure. In: Hemminki, K., M. Sorsa, and H. Vainio, eds. Occupational Hazards and Reproduction. Washington, DC: Hemisphere Pub., pp. 219-229.

52. Sever, L.E., and N.A. Hessol. 1984. Overall design considerations in male and female occupational reproductive studies. In: Lockey, J.E., G.K. Lemasters, and W.R. Keye, eds. Reproduction: the new frontier in occupational and environmental research. New York, NY: Alan R. Liss, Inc. pp. 15-48.

53. Stein, Z., J. Kline, and P. Shrout. 1985. Power in surveillance. In: Hemminki, K., M. Sorsa, and H. Vainio, eds. Occupational hazards and reproduction. Washington, DC: Hemisphere Pub., pp. 203-208.

54. Kissling, G. 1981. A generalized model for analysis of nonindependent observations.

[51 FR 34039]

Ph.D. Dissertation. Ann Arbor, MI: University Microfilms.

55. Selevan, S.G. 1980. Evaluation of data sources for occupational pregnancy outcome studies. Ph.D. Dissertation. Ann Arbor, MI: University Microfilms.

56. Axelson, O. 1985. Epidemiologic methods in the study of spontaneous abortions: sources of data, methods, and sources of error. In: Hemminki, K., M. Sorsa, and H. Vainio, eds. Occupational hazards and reproduction. Washington, DC: Hemisphere Pub., pp. 231-236.

57. Tilley, B.C., A.B. Barnes, E. Bergstralh, D. Labarthe, K.L. Noller, T. Colton, and E. Adam. 1985. A comparison of pregnancy history recall and medical records: implications for retrospective studies. Am. J. Epidemiol. 121(2):269-281.

58. Wilcox, A.J. 1983. Surveillance of pregnancy loss in human populations. Am. J. Ind. Med. 4:285-291.

59. Wilson, J.G., W.J. Scott, E.J. Ritter, and R. Fradkin. 1975. Comparative distribution and embryotoxicity of hydroxyurea in pregnant rats and rhesus monkeys. Teratology 11:169-178.

60. Wilson, J.G., E.J. Ritter, W.J. Scott, and R. Fradkin. 1977. Comparative distribution and embryotoxicity of acetylsalicylic acid in pregnant rats and rhesus monkeys. Toxicol. Appl. Pharmacol. 41:67-78.

61. Kimmel, C.A., and J.F. Young. 1983. Correlating pharmacokinetics and teratogenic end points. Fund. Appl. Toxicol. 3:250-255.

62. Hogan, M.D., and D.G. Hoel. 1982. Extrapolation to man. In: A.W. Hayes, ed. Principles and methods of toxicology. New York, NY: Raven Press, pp. 711-731.

63. Chitlik, L.D., Q.Q. Bui, G.J. Burin, and S.C. Dapson. 1985. Standard evaluation procedures for teratology studies. Toxicology Branch, Hazard Evaluation Division, Office of Pesticide Programs, U.S. Environmental Protection Agency.

64. U.S. Environmental Protection Agency. 1986, Sept. 24. Guidelines for estimating exposures. Federal Register 51(185):34042-34054.

65. Lemasters, G.K., and S.G. Selevan. 1984. Use of exposure data in occupational reproductive studies. Scand. J. Work Environ. Health 10:1-6.

66. Hogue, C.J.R. 1984. Reducing misclassification errors through questionnaire design. In: Lockey, J.E., G.K. Lemasters, and W.R. Keye, eds. Reproduction: the new frontier in occupational and environmental health research. New York, NY: Alan R. Liss, Inc., pp. 81-97.

67. Fraser, F.C. 1977. Relation of animal studies to the problem in man. In: J.G. Wilson and F.C. Fraser, eds. Handbook of teratology, Vol. 1. New York, NY: Plenum Press, pp. 75-96.

68. Environmental Health Criteria 30. 1984. Principles for evaluating health risks to progeny associated with exposure to chemicals during pregnancy. World Health Organization, Geneva, Switzerland, pp. 111-114.

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

Part B: Response to Public and Science Advisory Board Comments

I. Introduction

This section summarizes some of the issues raised in public comments on the Proposed Guidelines for the Health Assessment of Suspect Developmental Toxicants published November 23, 1984 (49 FR 46324). Comments were received from 44 individuals or organizations. The Agency's initial summary of comments was presented to the Developmental Toxicity Guidelines Panel of the Science Advisory Board (SAB) at its organizational meeting on March 4, 1985. At its April 22-23, 1985, meeting, the Panel provided the Agency with its suggestions and recommendations concerning the Guidelines.

The SAB and public comments were diverse and addressed issues from a variety of perspectives. In general, the comments were favorable and in support of the Guidelines. The SAB Panel noted that the field of developmental toxicology is particularly weak with respect to quantitative assessment and recommended that further efforts be given to developing alternative methods for quantitative estimates of risk for developmental toxicity. They also indicated that further discussion of the relationship of maternal toxicity to fetal toxicity could be added. Concern was expressed that these Guidelines be coordinated with the reproductive toxicity guidelines which are currently being developed.

In response to the comments, the Agency has modified or clarified many sections of the Guidelines. For purposes of this discussion, only the most significant issues reflected by the public and SAB comments are discussed. Several minor recommendations, which do not warrant discussion here, were considered by the Agency in the revision of these Guidelines.

II. Coordination With Other Guidelines

A. Other Risk Assessment Guidelines

Several commentors raised concerns about aspects of developmental toxicity (e.g., paternally-mediated effects, effects of subchronic exposures, transplacental carcinogenesis, etc.) that were not covered in these Guidelines, and how these Guidelines will integrate with those on male and female reproductive toxicity which are still under development.

The Guidelines have been revised to indicate that developmental toxicity may result from several different types of exposure, including parental exposure prior to conception, acute or subacute exposure during organogenesis, perinatal and postnatal development to the time of sexual maturation, or subchronic exposure as would be the case in multigeneration studies. These Guidelines provide information for interpreting developmental effects related to any of the types of exposure mentioned above. End points of developmental toxicity, which are measured in multigeneration studies, have been added to Table 2 and discussed in the text. Transplacental carcinogenesis, although considered a developmental effect, will be evaluated and assessed in terms of human risk according to the Guidelines for Carcinogen Risk Assessment. Careful attention will be paid to integrating these developmental toxicity risk assessment Guidelines and the male and female reproductive toxicity risk assessment guidelines, which are currently being written, so that overlapping material is not in conflict, and no pertinent information is overlooked. Since the developmental and reproductive toxicity guidelines are being developed by Agency committees that have overlapping membership within the Agency, such integration will be ensured.

B. Coordination With Testing Guidelines

Several commentors indicated that these Guidelines did not make clear enough the fact that testing guidelines are already in place and that these guidelines were intended only for the purposes of risk assessment.

The Guidelines have been revised to indicate that they do not constitute any changes in current testing guidelines, but rather they are intended to provide guidance for the interpretation of studies that follow the testing guidelines. In addition, limited information is provided for interpretation of other studies (e.g., functional developmental toxicity studies and short-term tests) which are not routinely required or for which there are no current testing guidelines, but which may be encountered when reviewing data on particular agents.

III. Definitions

Several questions were raised about definitions of terminology, due to lack of clarity or inconsistency with other parts of these Guidelines or the testing guidelines.

As indicated in the Guidelines, there are differences in the use of terms in the field of developmental toxicology, and the terms have been defined so that the reader may understand how the terms are being used. Several minor changes in the definitions have been made

[51 FR 34040]

to make them more consistent. For example, the definition for developmental toxicology has been expanded to include the wide range of exposure situations that may result in developmental effects. The term functional teratology has been changed to functional developmental toxicology, and the term

teratogenicity has been discussed in the section on malformations and variations.

IV. Qualitative Assessment

A. Maternal and Developmental Toxicity

Several commentors noted the need for a better discussion of how maternal toxicity affects the evaluation of developmental toxic effects.

The Agency has taken the approach in these Guidelines of discussing in detail the individual end points of maternal and offspring toxicity, then giving guidance relating to an overall evaluation of the data in Part A, section III.A.4. This approach is consistent with the philosophy reflected in the Guidelines as follows: Those agents that cause developmental effects at doses lower than those causing maternal toxicity are of greatest concern, but developmental effects at doses that also produce maternal toxicity shoud not be discounted as secondary to maternal effects. Rather, when the lowest observed effect level (LOEL) is the same for maternal and developmental toxicity, it may indicate similar sensitivities to the agent, and maternal effects may be reversible while developmental effects may be permanent.

B. Functional Developmental Toxicity

Several commentors raised concern about the premature use of functional data in the risk assessment process. On the other hand, the SAB Panel felt that these tests were very valuable in assessing developmental toxicity.

The Agency does not routinely require such testing, and these Guidelines do not suggest requirements. However, in the review of data on existing chemicals, such data are sometimes encountered and must be evaluated by the Agency. The discussion in the Guidelines is intended to delineate the current state of the art, and to indicate to what extent the data currently may be used for risk assessment purposes.

C. Short-Term Testing

Several commentors stressed the need for further refinement, validation, and comparative testing to determine the credibility of short-term tests for developmental toxicity. The appropriateness of single dose level screens for the purpose of prioritization was endorsed by the SAB Panel with the reservation that too many false positives might occur, and that positive agents in these screens would be permanently labelled as positive developmental toxicants.

Since data from these types of test procedures may be encountered in the assessment of chemicals, the Agency felt it appropriate to give guidance as to how these should be evaluated. The Guidelines have been revised to clearly indicate that these tests are not routinely required, shoud not be considered as a replacement for routine *in vivo* developmental toxicity testing in mammals, and should not be used to make the final decision as to whether an agent is a positive or negative developmental toxicant.

D. Comparisons of Molecular Structure

Comments suggested that not much is known about structure-activity relationships for developmental toxicants, and that this procedure should not be used except in the case of hormone analogs.

A statement has been added to indicate that structure-activity relationships have not been well-studied in developmental toxicology, but under certain circumstances, e.g., in the case of the premanufacturing notice process (TSCA, section 5), the evaluation of molecular structure is one of several procedures used by the Agency to evaluate potential toxicity and to support requests for testing of new chemicals.

V. Quantitative Assessment

Most comments related to the appropriateness of using uncertainty (safety) factors, margins of safety, and no observed effect levels (NOELs). Some commentors felt that the concept of threshold was not adequately discussed in the Guidelines.

These Guidelines are intended to reflect current Agency policy and practice. Although more quantitative assessment of developmental toxicity data are desirable, and efforts are currently ongoing within the Agency to evaluate other approaches, the current practice is to use the NOEL (or the LOEL if a NOEL is not available), and to apply an uncertainty factor or to calculate the margin of safety. This practice is based in large part on the lack of understanding of the biological mechanisms involved. The uncertainty factor used or acceptability of the margin of safety are considered risk management decisions, but the scientific issues that must be taken into account are discussed in these Guidelines. An experimentally determined NOEL does not prove or disprove the existence of a threshold, although many developmental toxicologists assume a threshold for most developmental effects because of known repair capabilities in developing systems and the fact that many developmental alterations are multifactorial in nature.

51 FR 34042

GUIDELINES FOR ESTIMATING EXPOSURES

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 Estimating Exposures.

The Guidelines for Estimating Exposures (hereafter "Guidelines") are intended to guide Agency analysis of exposure assessment data 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 Exposure Assessment published November 23, 1984 (49 FR 46304).

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. Richard V. Moraski
Exposure Assessment Group
Office of Health and Environmental Assessment (RD-689)
U. S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460
202-475-8923

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 Estimating Exposures

Work on the Guidelines for Estimating Exposures 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 exposure assessment from universities, environmental groups, industry, labor, and other governmental agencies. They were then proposed for public comment in the *FEDERAL REGISTER* (49 FR 46304). 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 preliminary Agency responses to those comments. 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 Estimating Exposures were concurred on in a letter dated January 13, 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 SAB requested that the Agency develop guidelines on the principles for the measurement of pollutant concentrations in the various environmental media and for the uses of environmental measurements for exposure assessment. This effort is currently underway.

The Agency also will provide technical support documents that contain detailed technical information needed to implement the Guidelines. Two of these technical reports entitled "Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments" (available from the National Technical Information Service, PB85-242667) and "Methodology for Characterization of Uncertainty in Exposure Assessments" (available from the National Technical Information Service, PB85-240455) are currently available. Technical support documents will be revised periodically to reflect improvements in exposure assessment methods and new information or experience.

[51 FR 34043]

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 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. Thomas,

Administrator

CONTENTS

Part A: Guidelines for Estimating Exposure

I. Introduction

II. General Guidelines and Principles
- A. Exposure and Dose
- B. Decision Path to Determine Scope of the Assessment
- C. Uncertainty

III. Organization and Contents of an Exposure Assessment
- A. Overview
- B. Detailed Explanation of Outline
  1. Executive Summary
  2. Introduction (Purpose and Scope)
  3. General Information for Each Chemical or Mixture
  4. Sources
  5. Exposure Pathways and Environmental Fate
  6. Measured or Estimated Concentrations
  7. Exposed Populations
  8. Integrated Exposure Analysis
  9. References
  10. Appendices

Part B: Response to Public and Science Advisory Board Comments

I. Introduction

II. General Information
- A. Acceptable Latitude of Approach
- B. Technical Nature of Guidelines
- C. Measurements vs. Modeling

III. Data Availability and Uncertainty Analysis
- A. Information Uses
- B. Worst-Case Estimates

IV. Evaluation of Uncertainties
- A. Uncertainty Analysis
- B. Population Characterization

V. Clarification of Terminology
- A. Exposure vs. Dose
- B. Mixtures and Synergism
- C. Removal and Creation Steps

VI. Purpose, Philosophy, and Results

Part A: Guidelines for Estimating Exposures

I. Introduction

These Guidelines provide the Agency with a general approach and framework for carrying out human or nonhuman exposure assessments for specified pollutants. The Guidelines have been developed to assist future assessment activities and encourage improvement in those EPA programs that require, or could benefit from, the use of exposure assessments. The Guidelines are procedural. They should be followed to the extent possible in instances where exposure assessment is a required element in the regulatory process or where exposure assessments are carried out on a discretionary basis by EPA management to support regulatory or programmatic decisions.

This document, by laying out a set of questions to be considered in carrying out an exposure assessment, should help avoid inadvertent mistakes of omission. Ideally, exposure assessments are based on measured data. EPA recognizes that gaps in data will be common, but the Guidelines will nevertheless serve to assist in organizing the data that are available, including new data developed as part of the exposure assessment. In the absence of sufficient reliable data and the time to obtain appropriate measurements, exposure assessments may be based on validated mathematical models. Whenever possible, exposure assessments based on modeling should be complemented by reliable measurements. Furthermore, it is understood that the level of detail found in the exposure assessments depends on the scope of the assessment.

These Guidelines should also promote consistency among various exposure assessment activities that are carried out by the Agency. Consistency with respect to common physical, chemical, and biological parameters, with respect to assumptions about typical exposure situations, and with respect to the characterization of uncertainty of estimates, will enhance the comparability of results and enable the Agency to improve the state-of-the-art of exposure assessment over time through the sharing of common data and experiences.

It is recognized that the main objective of an exposure assessment is to provide reliable data and/or estimates for a risk assessment. Since a risk assessment requires the coupling of exposure information and toxicity or effects information, the exposure assessment process should be coordinated with the toxicity/effects assessment. This document provides a common approach to format, which should simplify the process of reading and evaluating exposure assessments and thereby increase their utility in assessing risk.

As the Agency performs more exposure assessments, the Guidelines will be revised to reflect the benefit of experience.

II. General Guidelines and Principles

A. Exposure and Dose

Exposure has been defined by Committee E-47, Biological Effects and Environmental Fate, of the American Society for Testing and Materials, as the contact with a chemical or physical agent. The magnitude of the exposure is determined by measuring or estimating the amount of an agent available at the exchange boundaries, i.e., lungs, gut, skin, during some specified time. Exposure assessment is the determination or estimation (qualitative or quantitative) of the magnitude, frequency, duration, and route of exposure. Exposure assessments may consider past, present, and future exposures with varying techniques for each phase, e.g., modeling of future exposures, measurements of existing exposure, and biological accumulation for past exposures. Exposure assessments are generally combined with environmental and health effects data in performing risk assessments.

In considering the exposure of a subject to a chemical agent, there are several related processes. The contact between the subject of concern and the agent may lead to the intake of some of the agent. If absorption occurs, this constitutes an uptake (or an absorbed dose). When biological tissue or fluid measurements indicate the presence of a chemical, exposures may be estimated from these data. Presence of a chemical in such biological samples is the most direct indication that an exposure has occurred. The route of exposure generally impacts the extent of absorption and should be considered in performing risk assessments.

[51 FR 34044]

B. Decision Path to Determine Scope of the Assessment

The first step in preparing an exposure assessment should be the circumscription of the problem at hand to minimize effort by use of a narrowing process. A decision path that describes this process is shown in Figure 1. As illustrated in Figure 1, the preliminary assessment and the in-depth assessment are two major phases in this logic path.

[51 FR 34046]

The preliminary assessment phase should commence by considering what risk is under study. Within this framework, a data base should be compiled from readily available scientific data and exposure information based on manufacturer, processor, and user practices. Next, the most likely areas of exposure (manufacturing, processing, consumer, distribution, disposal, water and food, etc.) should be identified. The preliminary exposure assessments should be based on data derived from environmental measurements. When a limited amount of measurement data is available, estimates may be based on modeling. Since a complete data search may not be possible, well identified

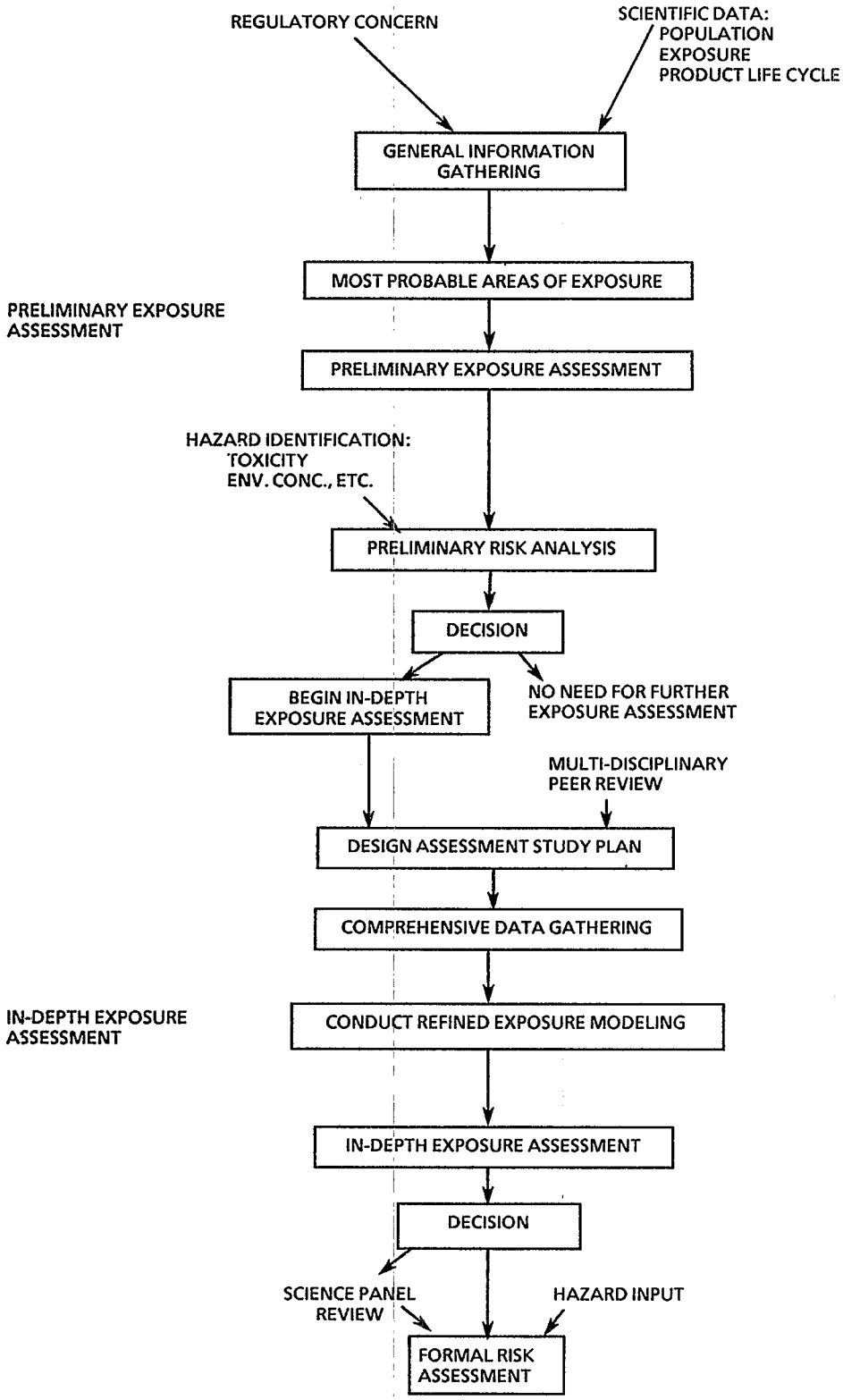

Figure 1. Decision path for exposure assessment.

assumptions and order of magnitude estimates may be used to further narrow the exposure areas of concern.

Data from this preliminary exposure assessment can then be coupled with toxicity information to perform a preliminary risk analysis. As a result of this analysis, a decision will be made that either an in-depth exposure assessment is necessary or that there is no need for further exposure information. The organization and contents of an in-depth exposure assessment are given in the following section.

In assembling the information base for either a preliminary assessment or a more detailed assessment, its adequacy should be ascertained by addressing the following considerations:

- Availability of information in every area needed for an adequate assessment;
- Quantitative and qualitative nature of the data;
- Reliability of information;
- Limitations on the ability to assess exposure.

C. Uncertainty

Exposure assessments are based on measurements, simulation model estimates, and assumptions about parameters used in approximating actual exposure conditions. Actual measurements should be used whenever possible. Both data and assumptions contain varying degrees of uncertainty which influence the accuracy of exposure assessments. Consequently, evaluation of uncertainty is an important part of all exposure assessments.

The uncertainty analyses performed will vary depending on the scope of the assessment, the quantity and quality of measurements, and the type and complexity of mathematical models used. A discussion of the types of analyses used for quantifying uncertainties in exposures is presented in the next section.

III. Organization and Contents of an Exposure Assessment

A. Overview

A suggested outline for an exposure assessment document is given in Exhibit 1. The five major topics to be addressed within most exposure assessments are as follows: Source(s), Exposure Pathways, Measured or Estimated Concentrations and Duration, Exposed Population(s), and Integrated Exposure Analysis. These five topics are appropriate for exposure assessments in general, whether the assessments are of global, national, regional, local, site specific, workplace related, or other scope. The topics are appropriate for exposure assessments on new or existing chemicals and radionuclides. They are also applicable to both single media and multimedia assessments. Since exposure assessments are performed at different levels of detail, the extent to which any assessment contains items listed in Exhibit 1 depends upon its scope. The outline is a guide to organize the data whenever they are available.

Exhibit 1.--Suggested Outline for an Exposure Assessment

1. Executive Summary
2. Introduction
   a. Purpose
   b. Scope
3. General Information for Each Chemical or Mixture
   a. Identity
      (1)Molecular formula and structure, synonyms, and Chemical Abstracts Service (CAS) number
      (2) Description of grades, contaminants, and additives
      (3) Other identifying characteristics
   b. Chemical and Physical Properties
4. Sources
   a. Characterization of Production and Distribution
   b. Uses
   c. Disposal
   d. Summary of Environmental Releases
5. Exposure Pathways and Environmental Fate
   a. Transport and Transformation
   b. Identification of Principal Pathways of Exposure
   c. Predicting Environmental Distribution
6. Measured or Estimated Concentrations
   a. Uses of Measurements
   b. Estimation of Environmental Concentrations
7. Exposed Populations
   a. Human Populations
      (1) Population size and characteristics
      (2) Population location
      (3) Population habits
   b. Nonhuman Populations (where appropriate)
      (1) Population size and characteristics
      (2) Population location
      (3) Population habits
8. Integrated Exposure Analysis
   a. Calculation of Exposure
      (1) Identification of the exposed population and critical elements of the ecosystem
      (2) Identification of pathways of exposure
   b. Human Dosimetry and Biological Measurements
   c. Development of Exposure Scenarios and Profiles
   d. Evaluation of Uncertainty
      (1) Introduction
      (2) Assessments based on limited initial data
      (3)Assessments based on subjective estimates of input variable distributions
      (4) Assessments based on data for model input variables
      (5) Assessments based on data for exposure
      (6) Summary
9. References
10. Appendices

B. Detailed Explanation of Outline

1.Executive Summary. The "Executive Summary" should be written so that it can stand on its own as a miniature report. Its main focus should be on a succinct description of the procedures used, assumptions employed, and summary tables or charts of the results. A brief discussion of the uncertainties associated with the results should be included.

2. Introduction (Purpose and Scope). This section should state the intended purpose of the exposure assessment and identify the agent being

investigated, the types of sources and exposure routes included, and the populations of concern.

3. General Information for Each Chemical or Mixture.

a. Identity.

(1) Molecular formula and structure, synonyms, and Chemical Abstracts Service number.

(2) Description of grades, contaminants, and additives.

(3) Other identifying characteristics.

b. Chemical and Physical Properties. This subsection should provide a summary description of the chemical and physical properties of the agent. Particular attention should be paid to the features that would affect its behavior in the environment.

4. Sources. The points at which a substance is believed to enter the environment should be described, along with any known rates of entry. (Points of entry may be indoors as well as outdoors; environments include indoor settings such as offices as well as outdoor environments.) A detailed exposure assessment should include a study of sources, production, uses, destruction/disposal, and environmental release of a substance. The studies

[51 FR 34047]

should include a description of human activities with respect to the substance and the environmental releases resulting from those activities. It should account for the controlled mass flow of the substance from creation to destruction and provide estimates of environmental releases at each step in this flow. Seasonal variations in environmental releases should also be examined. All sources of the substance should be accounted for with the sum of the uses, destruction, and the environmental releases. The environmental releases can be described in terms of geographic and temporal distribution and the receiving environmental media, with the form identified at the various release points.

a. Characterization of Production and Distribution. All sources of the substance's release to the environment, consistent with the scope of the assessment, should be included, such as production, extraction, processing, imports, stockpiles, transportation, accidental/ incidental production as a side reaction, and natural sources. The sources should be located, and activities involving exposure to the substance should be identified.

b. Uses. The substance should be traced from its sources through various uses (with further follow-up on the products made to determine the presence of the original material as an impurity), e.g., exports, stockpile increases, etc.

c. Disposal. This subsection should contain an evaluation of disposal sites and destruction processes, such as incineration of industrial chemical waste, incineration of the substance as part of an end-use item in municipal waste, landfilling of wastes, biological destruction, or destruction in the process of using the end product. Hazardous contaminants of the substance may be included, and products containing the substance as a contaminant may be followed from production through destruction/disposal.

d. Summary of Environmental Releases. Estimates should be made of the quantities of the substance released to the various environmental media. Sources of release to the environment include production, use, distribution/transport, natural sources, disposal, and contamination of other products. Environmental releases should be presented at a reasonable level of detail. Extremely detailed exposure estimates would attempt to specify the following information for each significant emission source: location, amount of the substance being released as a function of time to each environmental medium, physical characteristics of the emission source, and the physical and chemical form of the substance being released. Evaluation of the uncertainties associated with the emission estimates should be given. A detailed discussion of the procedures for estimating uncertainty is presented in section 8.d.

5. Exposure Pathways and Environmental Fate. The exposure pathways section should address how an agent moves from the source to the exposed population or subject. For a less detailed assessment, broad generalizations on environmental pathways and fate may be made. In the absence of data, e.g., for new substances, fate estimates may have to be predicted by analogy with data from other substances. Fate estimates may also be made by using measurements and/or models and laboratory-derived process rate coefficients. At any level of detail, certain pathways may be judged insignificant and not pursued further.

For more detailed assessments involving environmental fate, the analysis of sources described previously should provide the amount and rate of emissions to the environment, and possibly the locations and form of the emissions. The environmental pathways and fate analysis follows the substance from its point of initial environmental release, through the environment, to its ultimate fate. It may result in an estimation of the geographic and temporal distribution of concentrations of the substance in the various contaminated environmental media.

a. Transport and Transformation. The substance, once released to the environment, may be transported (e.g., convected downstream in water or on suspended sediment, through the atmosphere, etc.) or physically transformed (e.g., volatilized, melted, absorbed/desorbed, etc.); may undergo chemical transformation, such as photolysis, hydrolysis, oxidation, and reduction; may undergo biotransformation, such as biodegradation; or may accumulate in one or more media. Thus, the environmental behavior of a substance should be

evaluated before exposures are assessed. Factors that should be addressed include:

- How does the agent behave in air, water, soil, and biological media? Does it bioaccumulate or biodegrade? Is it absorbed or taken up by plants?
- What are the principal mechanisms for change or removal in each of the environmental media?
- Does the agent react with other compounds in the environment?
- Is there intermedia transfer? What are the mechanisms for intermedia transfer? What are the rates of the intermedia transfer or reaction mechanisms?
- How long might the agent remain in each environmental medium? How does its concentration change with time in each medium?
- What are the products into which the agent might degrade or change in the environment? Are any of these degradation products ecologically or biologically harmful? What is the environmental behavior of the harmful products?
- Is a steady-state concentration distribution in the environment, or in specific segments of the environment, achieved? If not, can the nonsteady-state distribution be described?
- What is the resultant distribution in the environment--for different media, different types or forms of the agent, for different geographical areas, at different times or seasons?

b. Identification of Principal Pathways of Exposure. The principal pathway analysis should evaluate the sources, locations, and types of environmental releases, together with environmental behavioral factors, to determine the significant routes of human and environmental exposure to the substance. Thus, by listing the important characteristics of the environmental release (entering media, emission rates, etc.) and the agent's behavior (intermedia transfer, persistence, etc.) after release to each of the entering media, it should be possible to follow the movement of the agent from its initial release to its subsequent fate in the environment. At any point in the environment, human or environmental exposure may occur. Pathways that result in major concentrations of the agent and high potential for human or environmental contact are the principal exposure pathways.

c. Predicting Environmental Distribution. Models may be used to predict environmental distributions of chemicals. Model estimates of environmental distribution of chemicals are based on measurements whenever feasible. In predicting environmental distributions of chemicals, available measurements must be considered.

In this section an estimation is made, using appropriate models, of representative concentrations of the agent in different environmental media, and its time-dependence in specific geographical locations (e.g., river basins, streams, etc.).

6. Measured or Estimated Concentrations.

a. Uses of Measurements. Measurements are used to identify releases (source terms) and, in the

[51 FR 34048]

exposure pathways and fate assessments, to quantitatively estimate both release rates and environmental concentrations. Some examples of uses of measurements are: sampling of stacks or discharge pipes for emissions to the environment, testing of products for chemical or radionuclide content, testing of products for chemical or radioactive releases, sampling of appropriate points within a manufacturing plant to determine releases from industrial processes or practices, sampling of potentially exposed populations using personal dosimeters, and sampling of solid waste for chemical or radionuclide content. These data should be characterized as to accuracy, precision, and representativeness. If actual environmental measurements are unavailable, concentrations can be estimated by various means, including the use of fate models (see previous section) or, in the case of new chemicals, by analogy with existing chemicals.

Measurements are a direct source of information for exposure analysis. Furthermore, reliable measurements can be used to calibrate or extrapolate models or calculations to assess environmental distributions. However, environmental pathway and fate analysis may be needed in addition to the measured data for the following reasons: for most pollutants, particularly organic and new chemicals, measurements are limited; analysis of measured data does not often yield relationships between environmental releases and environmental concentration distribution in media or geographic locations that have not been measured; analysis of measurements does not provide information on how and where biota influence the environmental distribution of a pollutant; and measured concentrations may not be traceable to individual sources.

b. Estimation of Environmental Concentrations. Concentrations of agents should be estimated for all environmental media that might contribute to significant exposures. Generally, the environmental concentrations are estimated from measurements, mathematical models, or a combination of the two. If environmental measurements are not limited by sample size or inaccuracies, then exposure assessments based on measurements have precedence over estimates based on models.

The concentrations must be estimated and presented in a format consistent with available dose-response information. In some cases an estimate of annual average concentration will be sufficient, while in other cases the temporal distribution of concentrations may be required. Future environmental concentrations resulting from current or past releases may also be projected. In some cases, both the temporal and geographic distributions of the concentration may be assessed.

Moreover, if the agent has natural sources, the contribution of these to environmental concentrations may be relevant. These "background" concentrations may be particularly important when the results of tests of toxic effects show a threshold or distinctly nonlinear dose-response.

The uncertainties associated with the estimated concentrations should be evaluated by an analysis of the uncertainties of the model parameters and input variables. When the estimates of the environmental concentrations are based on mathematical models, the model results must be compared to available measurements, and any significant discrepancies should be discussed. Reliable, analytically-determined values must be given precedence over estimated values whenever significant discrepancies are found.

7. Exposed Populations. Populations selected for study may be done *a priori*, but frequently the populations will be identified as a result of the sources and fate studies. From an analysis of the distribution of the agent, populations and subpopulations (i.e., collections of subjects) at potentially high exposure can be identified, which will then form the basis for the populations studied. Subpopulations of high sensitivity, such as pregnant women, infants, chronically ill, etc., may be studied separately.

Census and other survey data may be used to identify and describe the population exposed to various contaminated environmental media. Depending on the characteristics of available toxicological data, it may be appropriate to describe the exposed population by other characteristics such as species, subspecies-age-sex distribution, and health status.

In many cases, exposed populations can be described only generally. In some cases, however, more specific information may be available on matters such as the following:

a. Human Populations

(1) Population size and characteristics (e.g., trends, sex/age distribution)

(2) Population location

(3) Population habits-- transportation habits, eating habits, recreational habits, workplace habits, product use habits, etc.

b. Nonhuman Populations (where appropriate)

(1) Population size and characteristics (e.g., species, trends)

(2) Population location

(3) Population habits

8. Integrated Exposure Analysis. The integrated exposure analysis combines the estimation of environmental concentrations (sources and fate information) with the description of the exposed population to yield exposure profiles. Data should be provided on the size of the exposed populations; duration, frequency, and intensity of exposure; and routes of exposure. Exposures should be related to sources.

For more detailed assessments, the estimated environmental concentrations should be considered in conjunction with the geographic distribution of the human and environmental populations. The behavioral and biological characteristics of the exposed populations should be considered, and the exposures of populations to various concentration profiles should be estimated. The results can be presented in tabular or graphic form, and an estimate of the uncertainty associated with them should be provided.

a. Calculation of Exposure. The calculation of exposure involves two major aspects:

(1) Identification of the exposed population and critical elements of the ecosystem.

The estimate of environmental concentrations also should give the geographical areas and environmental media contaminated. The stated purpose of the assessment should have described the human and environmental subjects for which exposures are to be calculated. If the subjects are not listed, the contaminated geographical areas and environmental media can be evaluated to determine subject populations. The degree of detail to be used in defining the exposed population distribution depends on the concentration gradient over geographic areas.

(2) Identification of pathways of exposure.

(a) Identification and description of the routes by which the substances travel from production site, through uses, through environmental releases/sources, through transport and fate processes, to the target population.

(b) Quantitative estimates of the amounts of the chemical following each exposure pathway. Such estimates allow the various pathways to be put in the perspective of relative importance.

From the geographic and temporal distribution of environmental

[51 FR 34049]

concentrations, the exposed population, the behavioral characteristics, and the critical elements of the ecosystem, exposure distributions can be estimated. The results of exposure calculation should be presented in a format that is consistent with the requirements of the dose-response functions which may later be used in a risk assessment. For example, when health risks caused by exposure over extended durations are considered, average daily exposure over the duration of exposure usually is calculated. When lifetime risks are considered, average daily exposure over a lifetime usually is calculated. In contrast, when health risks caused by exposures over short durations are considered, exposure rates are calculated over short time intervals to ensure that peak risks are defined. Many exposure assessments are based on the average exposure occurring over the exposure period. The range of possible exposures is usually divided into intervals,

and the exposures within each interval are counted. The results can be presented in tabular form or as a histogram.

The population residing in a specific geographic area may be exposed to a substance from several exposure routes. For each exposure route, exposure of individuals in these populations may be determined by summing the contribution of all sources to the exposure route. When exposures involve more than one exposure route, the relative amounts of a substance absorbed is usually route dependent. Consequently, total absorbed dose estimates must account for these differences. Because EPA regulates sources of releases, the contribution to exposures from each type of source being considered should be displayed. Exposure estimates should be presented for each significant exposure route, and the results should be tabulated in such a way that total externally applied and absorbed dose can be determined.

b. Human Dosimetry and Biological Measurements. Biological measurements of human body fluids and tissues for substances or their metabolites can be used to estimate current or past exposure to chemicals. When analytical methods are available, chemicals that have been absorbed into the body can be measured in body tissue and fluid. Such measurements may be used to estimate human exposure if the chemical substances leave in the body reliable indicators of exposure. Furthermore, although a compound may be relatively easy to detect in body tissue, for some compounds, attributing body burdens to specific environmental releases may be difficult because of limited ability to obtain environmental measurements or appropriate metabolic data.

c. Development of Exposure Scenarios and Profiles. Depending on the scope of the exposure assessment, the total exposure may be fractionated into one or more "exposure scenarios" to facilitate quantification. As an example, Table 1 lists seven very broad scenarios: Occupational, Consumer, Transportation, Disposal, Food, Drinking Water, and Ambient. For each of the scenarios, the major topics necessary to quantify exposure include sources, pathways, measurements, and population characteristics. Investigation of only one scenario may be necessary for the scope of some assessments. For example, a pesticide application exposure assessment may consider the occupational scenario which would address the exposure to applicators and populations in the vicinity of the site. An exposure assessment around a hazardous waste site may focus on the disposal scenario. The exposure assessment

TABLE 1.--EXPOSURE ASSESSMENT INFORMATION NEEDS FOR VARIOUS EXPOSURE SCENARIOS

Exposure scenario	Sources	Fate	Population Characteristics	Measurement
Occupational (chemical production).	Site/plant locations, in-plant/on-site materials balance.	Physical and chemical properties models.	Workers, families, population around sites/plants.	In-plant/on-site releases, ambient levels surrounding site/plants; human dosimetry.
Consumer (direct use of chemical or inadvertent use).	Consumption rates, distribution pattern amounts in products.	Physical and chemical properties, shelf life release rates, models.	Consumers	Levels in products releases.
Transportation / storage/spills.	Patterns of distribution and transportation; models for spills.	Physical and chemical properties, environmental fate models.	Storage, transportation workers, general population in area.	Releases, ambient levels.
Disposal (include incineration, landfill).	Materials balance around disposal method, efficiency,releases to environment.	Fate within disposal process; environmental fate of releases; models.	Workers at site of disposal, general population around site.	Releases, levels at various points within process, ambient levels.
Food	Food chain, packaging, additives.	Food chain models, fate during preparation or processing of food.	General population, nonhuman population.	Levels in food, feedstuff; food chain sampling.
Drinking water	Groundwater, surface water, distribution system.	Leach rates from pipes, chlorination processes, fate in water; models.	General population.	Levels in drinking water, groundwater, surface water, treatment plants.
Ambient	Releases to environment; air, land, water.	Environmental fate models.	General population, nonhuman population.	Ambient air, water, soil, etc.; human dosimetry.

also may consider other scenarios. The more extensive and comprehensive the scope, the more scenarios are usually involved.

It will usually be advantageous in performing an exposure assessment to identify exposure scenarios, quantify the exposure in each scenario, and then integrate the scenarios to estimate total exposure. In this "integrated exposure analysis," the summation of independent exposures from different scenarios (keeping exposure routes separate) often will result in a breakout of exposure by subpopulations, since the individual scenarios usually treat exposure by subpopulation. Therefore, the integration of the scenarios, or integrated exposure analysis, will often result in an exposure profile.

For each exposed subpopulation, exposure profiles should include the size of the group, the make-up of the group (age, sex, etc.), the source of the agent, the exposure pathways, the frequency and the intensity of exposure by each route (dermal, inhalation, etc.), the duration of exposure, and the form of the agent when exposure occurs. Assumptions and uncertainties associated with each scenario and profile should be clearly discussed.

d. Evaluation of Uncertainty.

(1) Introduction. Often an exposure assessment progresses through several stages of refinement. The purpose of these Guidelines is to present methods appropriate for characterization of uncertainty for assessments at various stages of refinement, from assessments based on limited initial data to those based on extensive data.

The appropriate method for characterizing uncertainty for an exposure assessment depends upon the underlying parameters being estimated, the type and extent of data available, and the estimation procedures utilized.

[51 FR 34050]

The uncertainty of interest is always with regard to the population characteristic being estimated. For example, when the population distribution of exposures is being estimated, characterization of uncertainty addresses the possible differences between the estimated distribution of exposure and the true population distribution of exposure.

An exposure assessment quantifies contact of a substance with affected population members (human or nonhuman subjects). The measure of contact (e.g., environmental level or absorbed dose) depends upon what is needed to predict risk. An integrated exposure assessment quantifies this contact via all routes of exposure (inhalation, ingestion, and dermal) and all exposure pathways (e.g., occupational exposure, exposure from consumption of manufactured goods, etc.). The exposed population generally is partitioned into subpopulations such that the likely exposure of all members of a subpopulation is attributable to the same sources. The exposure for each member of a subpopulation is then the sum of exposures over a fixed set of sources and pathways. The measured or estimated exposures for members of a subpopulation are ideally used to estimate the subpopulation distribution of exposure or characteristics thereof. However, a lack of sufficient information sometimes precludes estimation of the subpopulation distributions of exposure and only summary measures of this distribution, such as the mean, minimum, maximum, etc., are estimated. In each case, characterization of uncertainty for the exposure assessment primarily addresses limitations of the data and the estimation procedures. The proportions of the population members in the individual subpopulations are usually estimated and can be used (by combining estimated distributions for the subpopulations) to estimate the distribution of exposure for the total population. Uncertainty concerning the sizes of the subpopulations should be addressed by discussing limitations of the data and estimation methods as well as by tabulating confidence interval estimates for the population sizes whenever possible.

(2) Assessments based on limited initial data. The initial exposure assessment for a substance may be based on limited data for exposure and/or input variables for an exposure prediction model (i.e., an equation that expresses exposure as a function of one or more input variables). These data might be either extant data or data produced by an initial small-scale study. The limited initial data frequently are insufficient to permit estimation of the entire distribution of exposure. Instead, summary measures of this distribution, such as the mean, minimum, and maximum, are usually estimated.

If the assessment is based on measured exposures, the methods used to characterize uncertainty depend mainly upon whether or not the data result from a probability sample for which the probability of inclusion is known for each sample member. Characterization of uncertainty for an assessment based on a probability sample of exposures is discussed later in section 8.d.(5). If the measured exposures are not based on a probability sample, acknowledgement that no strictly valid statistical inferences can be made beyond the units actually in the sample is one aspect of the characterization of uncertainty. If inference procedures are implemented, the assumptions upon which these inferences are based (e.g., treatment of the sample as if it were a simple random sample, or assumption of an underlying model) should be explicitly stated and justified. The data collection methods and inherent limitations of the data should also be discussed.

An initial exposure assessment also may be based on limited data, such as estimated ranges, for input variables for an exposure prediction model. The exposure prediction model would be derived from a postulated exposure scenario that describes the pathways from sources to contact with population members. If the data were only sufficient to support estimates of the ranges of the input variables, the exposure assessment might be limited

to a sensitivity analysis. The purpose of the sensitivity analysis would be to identify influential model input variables and develop bounds on the distribution of exposure. A sensitivity analysis would estimate the range of exposures that would result as individual model input variables were varied from their minumum to their maximum possible values with the other input variables held at fixed values, e.g., their midranges. The overall minimum and maximum possible exposures usually would be estimated also. For an exposure assessment of this type, the uncertainty would be characterized by describing the limitations of the data used to estimate possible ranges of model input variables and by discussing justification for the model. Justification of the model should include a description of the exposure scenario, choice of model input variables, and the functional form of the model. Sensitivity to the model formulation also can be investigated by replicating the sensitivity analysis for plausible alternative models.

The sensitivity analysis can be enhanced by computing the predicted exposures that result from all possible input variable combinations. If each input variable has only a finite set of possible values, the set of all possible combinations of the input variables can be formed, and the predicted exposure can be computed for each combination. These exposure predictions can be used to form a distribution of exposures by counting the number of occurrences at each exposure level or interval of exposures. This is equivalent to estimating the distribution of exposures that results from treating all input variable combinations as equally likely. This procedure can also be applied by transforming continuous input variables into discrete ones and representing them by equally spaced points. In the limit, as the equal spaces become small and the number of points becomes large, the distribution of exposure that results from counting occurrences of exposure levels is equivalent to estimating the distribution of exposures that results from statistically independent, continuous input variables with uniform distributions on the estimated ranges. This estimated distribution of exposure values can be produced by Monte Carlo simulation, one of the methods of mathematical statistics. The Monte Carlo method consists of randomly generating input variate values and using these to compute corresponding exposure levels, generating an exposure distribution via many iterations. Interpretation of statistics based on this exposure distribution would be in terms of the equally likely input variable combinations. For example, the 95th percentile of this distribution would be the exposure level exceeded by only 5% of the exposures resulting from treating all combinations of input variable values as equally likely. Although this distribution of exposures cannot be interpreted as an estimate of the population distribution (unless the input variables actually are statistically independent and uniformly distributed), it provides additional information for making regulatory decisions. Characterization of uncertainty would include a discussion of limitations of the data and justification for the model as discussed above. Sensitivity to model formulation could also be investigated by estimating the distribution of exposure that results from using the

[51 FR 34051]

same uniform input variable distributions with plausible alternative models and comparing the estimated percentiles.

(3) Assessments based on subjective estimates of input variable distributions. If a model has been formulated that expresses exposure as a function of one or more input variables, the methods of mathematical statistics, such as Monte Carlo simulation, can be used to estimate the population distribution of exposure from an estimate of the joint distribution of the model input variables. Ideally, model input variables should be represented by empirically-validated probability distributions. In some cases, it may be possible to formulate an estimate of the joint distribution of model input variables from discussions with subject matter experts (e.g., via histograms for statistically-independent input variables). The estimated population distribution of exposure will be equivalent to the distribution discussed in section 8.d.(2) for equally likely combinations of input variable values only when the input variable distributions supported are independent uniform distributions. When qualitative knowledge of input variable distributions is used to estimate the population distribution of exposure, uncertainty is characterized by discussing justification for the presumed model and input variable distributions. Alternative models and/or alternative input variable distributions also should be discussed. Sensitivity to these alternatives can be investigated by estimating the distributions of exposure that result from plausible alternatives and comparing the percentiles of the estimated exposure distributions. All available data, even if data are limited, should be used to validate the presumed input variable distributions and the predicted distribution of exposure.

(4) Assessments based on data for model input variables. The exposure assessment based on an estimate of the joint probability distribution for model input variables can be refined by collecting sample survey data for model input variables for a sample of population members. The population distribution of exposure can then be estimated by computing the expected exposure for each sample member based on the model. These expected exposures can be used to directly compute confidence interval estimates for percentiles of the exposure distribution. Alternatively, the sample survey data can be used to compute joint confidence interval estimates for percentiles of the input variable

distribution, which can then be used to generate confidence interval estimates for percentiles of the exposure distribution. In either case, the interval estimates for percentiles of the exposure distribution are a useful quantitative characterization of uncertainty.

Characterization of uncertainty for the exposure assessment would contain a thorough discussion of limitations of the data and justification for the model used to compute expected exposures. The design of the sample survey used to produce the data base should also be discussed. If a probability sample were not used, the lack of a probability sample would be an additional source of uncertainty. Any assumptions used in computing the confidence interval estimates, such as independence of model input variables, should be explicitly stated and justified. Sensitivity to model formulation can be investigated by estimating the distribution of exposure for plausible alternative models and comparing the estimated percentiles, if sample survey data have been collected for the input variables of the alternative models. Appropriate available data for exposure should be used to validate the predicted distribution of exposure. If specific probability distributions have been presumed for any model input variables, the data for these variables should be used to test for goodness of fit for these distributions.

(5) Assessments based on data for exposure. A major reduction in the uncertainty associated with an exposure assessment can be achieved by directly measuring the exposure for a sufficiently large sample of members of the affected population. This reduction in uncertainty is achieved by eliminating the use of a model to predict exposure. The measured exposure levels can be used to directly estimate the population distribution of exposure and confidence interval estimates for percentiles of the exposure distribution. Direct confidence interval estimates also can be computed for other characteristics of the exposure distribution, such as the mean exposure.

These confidence interval estimates are then the primary characterization of uncertainty for the exposure assessment. Limitations of the data and design of the sample survey used to collect the data also should be discussed. If the sample was not a probability sample, this would again be an additional source of uncertainty.

(6) Summary. A summary of the primary methods recommended for characterizing uncertainty in exposure assessments is presented in Table 2. Virtually all exposure assessments, except those based on measured exposure levels for a probability sample of population members, rely upon a model to predict exposure. The model may be any mathematical function, simple or complex, that expresses an individual's exposure as a function of one or more input variables. Whenever a model that has not been validated is used as the basis for an exposure assessment, the uncertainty associated with the exposure assessment may be substantial. The primary characterization of uncertainty is at least partly qualitative in this case, i.e., it includes a description of the assumptions inherent in the model and their justification. Plausible alternative models should be discussed. Sensitivity of the exposure assessment to model formulation can be investigated by replicating the assessment for plausible alternative models.

[51 FR 34052]

When an exposure assessment is based on directly measured exposure levels for a probability sample of population members, uncertainty can be greatly reduced and described quantitatively. In this case, the primary sources of uncertainty are measurement errors and sampling errors. The effects of these sources of error are measured quantitatively by confidence interval estimates of percentiles of the exposure distribution. Moreover, the sampling errors can be limited by taking a large sample.

Whenever it is not feasible to take a large sample, it is sometimes possible to obtain at least some data for exposure and model input variables. These data should be used to assess goodness of fit of the model and/or presumed distributions of input variables. This substantially reduces the amount of quantitative uncertainty for estimation of the distribution of exposure and is strongly recommended. It is recognized, however, that it may not be feasible to collect such data.

9.References. The references should contain a listing of all reports, documents, articles, memoranda, contacts, etc. that have been cited in the report.

10.Appendices. The appendices may contain such items as memoranda and letters that are not readily accessible, other tables of measurements, detailed lists of emission sources, detailed tables of exposures, process flow diagrams, mathematical model formulations, or any other item that may be needed to describe or document the exposure assessment.

Part B: Response to Public and Science Advisory Board Comments

I. Introduction

This section summarizes some of the issues raised in public comments on the Proposed Guidelines for Exposure Assessment published November 23, 1984 (49 FR 46304). Comments were received from 29 individuals or organizations. The Agency's initial summary of comments was presented to the Exposure Assessment Guidelines Review Group of the Science Advisory Board (SAB) on March 4, 1985. At its April 22-23, 1985, meeting, the panel provided the Agency with suggestions and recommendations concerning the Guidelines.

The SAB and public commentors expressed diverse opinions and addressed issues from a variety of perspectives. While most commentors supported the Guidelines, two urged withdrawal of the

TABLE 2.--SUMMARY OF PRIMARY METHODS FOR CHARACTERIZING UNCERTAINTY FOR ESTIMATING EXPOSURES

Type and extent of data	Population characteristic being estimated	Primary methods for characterizing uncertainty	
		Qualitative methods	Quantitative methods
Measured exposures for a large sample of population members.	Distribution of exposure.	1. Limitations of the survey design and measurement techniques.	1. Confidence interval estimates for percentiles of the exposure distribution. 2. Goodness of fit for exposure models, if any have been postulated.
Measured exposures for a small sample of population members.	Summary parameter(s) of the exposure distribution, e.g., mean or a percentile.	1. Limitations of the survey design and. measurement techniques.	1. Confidence interval estimate for the summary parameter(s). 2. Goodness of fit for exposure models, if any have been postulated.
Measured model input variables for a large sample of population members.	Distribution of exposure.	1. Limitations of the survey.design and measurement techniques. 2. Validity of the exposure model.	1. Confidence interval estimates for percentiles of the exposure distribution. 2. Goodness of fit for input variable distribution functions, if any have been postulated. 3. Estimated distribution of exposure based on alternative models.
Estimated distributions of model input variables.	Distribution of exposure.	1. Validity of the exposure model. 2. Limitations of the data. or other basis for the input variable distributions.	1. Confidence interval estimates for percentiles of the exposure distribution. 2. Goodness of fit for input variable distributions, if input variable data are available. 3. Estimated distribution of exposure based on alternative models.
Limited data for model input variables.	Minimum, maximum, and range of the exposure distribution.	1. Limitations of the data. 2. Validity of the exposure model.	If input variable data are very limited, e.g., some extant data collected for other purposes, quantitative characterization of uncertainty may not be possible.

document. The SAB Panel recommended that supplementary guidelines be written on the use of measurements in preparing exposure assessments. In addition, the Panel wished to see a greater emphasis in the current Guidelines on the use of measured data rather than models in generating exposure assessments. The Panel recommended that the technical support document entitled "Methodology for Characterization of Uncertainty in Exposure Assessments" be expanded with additional examples.

In response to the comments, the Agency has modified or clarified many sections of the Guidelines, and is planning to develop supplementary guidance in line with the SAB recommendations. The discussion that follows highlights significant issues raised in the comments, and the Agency's response to them. Also, many minor recommendations, which do not warrant discussion here, were adopted by the Agency.

II. General Information

A. Acceptable Latitude of Approach

Some commentors believe the Guidelines are too general and allow too much latitude in choice of approach and do not assure that "all" data, sources, limitations, etc. are considered before an exposure assessment is conducted. Others suggested that the Agency specify models to be used while others thought that only measured data should be allowed.

The Guidelines were developed to provide assistance in carrying out exposure assessments. The approach suggested is deliberately general in order to accommodate the development of exposure assessments with different levels of detail depending on the scope of the assessment. The Agency does not agree with the inclusion of such restrictive terminology as "in all cases." We cannot foresee all possible cases. We believe reasonable flexibility is a necessary ingredient for the proper implementation of the Guidelines while relying on uncertainty and sensitivity analyses to put the quality of the approach in perspective

B. Technical Nature of Guidelines

Some commentors believe the language of the document is too technical for the lay person to understand; one commentor expressed misgivings concerning the "state-of-the-art" methods available for conducting exposure assessments.

While the Agency recognizes that the public has an interest in the Guidelines and invites comments from the public, the Guidelines are intended for use by technical/professional people. Providing guidelines written in lay terms would result in insufficient technical specifications to the professionals in the development of scientifically acceptable exposure assessments.

The Agency believes that the suggested procedures and methods in the Guidelines are commonly accepted. The Guidelines do not suggest the use of ad hoc, untested, and unvalidated procedures, but stress the use of the best scientific methods available with maximum analysis of existing data. This is both a scientific and practical approach that reflects the level of consensus within the Agency.

[51 FR 34053]

C. Measurements vs. Modeling

Some commentors support the use of measurements alone to develop an exposure assessment. Some believed there should be no data restraints; others thought all data should be validated. Other commentors argued for the use of simulation model estimates without measurements. One commentor objected to the use of unvalidated models to perform exposure assessments. In its review, the SAB strongly encouraged the Agency to develop a supplement to the current Guidelines on the development and use of measurements for exposure assessments.

The Agency encourages the use of validated measurements when available. The Guidelines specifically state that "Reliable, analytically determined values should be given precedence over estimated values. . ." and analytically determined values ". . .can be used to calibrate. . . models. . .to assess environmental distribution." Furthermore, in practice, exposure assessments performed by the Agency use published models with varying degrees of testing and validation. It is our belief that transport process models have been adequately validated over many years in most cases.

Furthermore, the Agency has revised the Guidelines to reflect the SAB suggestions that exposure assessments based on reliable measured data are preferred over model estimates whenever feasible.

III. Data Availability and Uncertainty Analysis

A. Information Uses

Some commentors asked for guidance in the use of information that may be false and how to deal with the potential situation when different models give different results. Others asked for model selection criteria.

The Guidelines clearly state the considerations that need to be addressed when assembling information bases for exposure assessments. Two considerations are: qualitative and quantitative nature of the data and the reliability of the information. Whether the exposure assessment is based on measurements or simulation model estimates, an evaluation of uncertainties associated with the data including source data and assumptions is necessary and important.

When there is uncertainty in the scientific facts, it is Agency policy to err on the side of public safety. The Agency intends to be realistic, but will not arbitrarily select midranges of environmental distributions that may compromise human health. In addition, quality assurance is an important matter that requires detailed attention. The collection of measured data and the development of methods to collect measurements are done by another office within the EPA. These issues will be handled by the Office of Acid Deposition, Environmental Monitoring, and Quality Assurance as they develop the supplemental guidelines for measurement of exposure.

Substantial work is currently being done on the development of mathematical model selection criteria. Results of these efforts will be published as a technical support document containing detailed information to further implement the Guidelines.

B. Worst-Case Estimates

A few commentors were concerned that worst-case estimates would be used when data are nonexistent or limited. The Guidelines do not encourage the use of worst-case assessments, but rather the development of realistic assessments based on the best data available.

A technical support document and a substantial section of the Guidelines currently discuss evaluation of uncertainty in order to produce objective assessments using the best (not worst-case) estimates available either for preliminary or in-depth exposure assessments. However, the Agency will err on the side of public health when evaluating uncertainties when data are limited or nonexistent.

IV. Evaluation of Uncertainties

A. Uncertainty Analysis

Many commentors felt that the sections of the Guidelines that dealt with uncertainty needed amplification while some sections as written were confusing. Some urged that uncertainty evaluation be presented and documented for each section within a specific exposure scenario in order to judge the overall plausibility of the assessment in reaching regulatory decisions.

Since the accuracy of an exposure assessment is influenced by the degrees of uncertainty contained in both data and assumptions, the Guidelines call for the evaluation of these uncertainties. The technical support document, Methodology for

Characterization of Uncertainty in Exposure Assessments (available from the National Technical Information Service, PB85-240455), describes in detail how such analyses can be performed. The Guidelines suggest that the uncertainty characterization include a discussion of the limitations of the data and estimation procedures as the justification for the model chosen. A sensitivity analysis of the exposure assessment is appropriate if the data were only able to support the estimates of ranges of the input variables. By identifying model input variables that determine the bounds on the distribution of exposure, the range of exposure, which results as individual model input variables are varied from minimum to maximum possible values as other variables remain constant, constitutes the sensitivity analysis. Further sensitivity of model formulation can be examined by repeating the sensitivity analysis for plausible alternative models.

Nothing in the Guidelines precludes estimation of uncertainty for each specific exposure scenario. The Agency has encouraged the evaluation of uncertainty in each aspect of the exposure assessment, which could impact the total risk estimate. It is important to estimate the level of uncertainty in risk assessments so that decisions based on risk assessment will reflect total uncertainty. The information presented in the Guidelines or the technical support documents properly and adequately describes the extent and quality of appropriate uncertainty analysis. Recognizing that the basis for the decision to refine a preliminary exposure assessment involves risk management, the Agency, at the suggestion of many commentors, decided to strike from the Guidelines the paragraph beginning "If the maximum possible exposure. . . ." in section III.B.8.d.(2).

B. Population Characterization

The Guidelines state that identification of populations and subpopulations at potentially high exposure forms the basis of the populations to be studied. Separate studies of sensitive subpopulation can also be included. Population characteristics, such as age and/or sex distributions, can be derived from the use of geographic and activity-specific data. Uncertainty related to estimation of a population characteristic include a discussion of the data limitations and the estimation procedures. In addition, uncertainty in estimating sizes of sensitive subpopulations should include estimates of confidence intervals.

Some commentors suggested the inclusion of additional characteristics, such as occupational and life style factors, and the inclusion of additional guidance concerning potential pitfalls when conducting population exposure

[51 FR 34054]

assessments. Others expressed concern that the exposure of a particular subpopulation would be combined with other exposures to produce an average exposure level for the general population.

The section describing population characterization encompasses, in general terms, the many characteristics that may be available, including life style factors, to describe exposed populations. The Agency agrees that there are difficulties associated with epidemiologic studies. The relationship between exposure assessments and epidemiologic studies is currently being investigated and will be the subject of a future technical support document and the further refinement of the Guidelines.

V. Clarification of Terminology

A. Exposure vs. Dose

Commentors expressed concern with the American Society for Testing and Materials (ASTM) definition of exposure. Concern was also raised about the assertion that exposures can be estimated when biological tissues for fluid measurements indicate the presence of a chemical. Some commentors found difficulty in the wording of the last sentence in section II.A., specifically "The route of exposure. . .impacts. . .the overall exposure. . . ."

It is the Agency's opinion that the members who served on the ASTM Committee E-47 had expertise in exposure assessment. The scientists and engineers cumulatively possessed many years of experience in exposure assessment. In addition, no technical society has presented an alternate definition of exposure. The Agency will consider changing the definition if a reasonable alternate definition is written and agreed upon by the scientific community.

The Agency agrees with the commentors who were concerned that the wording provided in the Guidelines that the presence of a chemical in biological tissue can be used to estimate exposure is not correct in all cases. Consequently, the word "can" was changed to "may" to reflect the current level of understanding between tissue residue and exposure (II.A., 2nd paragraph, 4th sentence). The Agency agrees with several commentors' concerns that the route of exposure impacts the overall absorbed dose, not the overall exposure, and the Guidelines reflect this change (II.A., last sentence).

B. Mixtures and Synergism

Some commentors thought more discussion was necessary on the effect of chemical mixtures and potential synergistic effect on exposure. The Guidelines for the Health Risk Assessment of Chemical Mixtures includes a discussion of chemical synergism. The Agency recognizes the need to do further work in the area of exposure to mixtures. It is recommended that this be identified as an area requiring further research.

These Guidelines stress the need to determine the products into which the chemical might degrade or react in the environment and to determine if any of these products are ecologically or biologically harmful.

C. Removal and Creation Steps

Some commentors urged that more emphasis be placed on changes that occur once the materials have entered the ambient environment. Other commentors argued that our current understanding will not allow a comprehensive treatment, particularly for metabolic processes.

These Guidelines state the need to address how a chemical agent moves from the source to the exposed population, which may result in the estimation of geographic and temporal distributions in various environmental media. The Guidelines also state the need to know such factors as, for example, whether the chemical agent bioaccumulates or by what mechanism the agent is removed from each medium and the role of any degradation products on ecological safety. We have already stated that guidance for analysis of metabolism data is an area of ongoing research which includes consideration of metabolism data in the calculation of whole organism dose from one species to another.

VI. Purpose, Philosophy, and Results

Several commentors raised questions related to the basic style of the Guidelines. Among the issues raised were:

- the role of exposure assessment in risk assessment/risk management (many comments directed to appropriateness of Figure 1);
- statutory/regulatory authority and uses of results; and
- the need for peer review of assessments and periodic updating of Guidelines.

A deliberate effort to separate risk assessment from risk management has been made. The management of complex issues such as procedural issues, which include coordination or linkage among divisions in the Agency, are best dealt with by management and not in Guidelines.

The decision pathway (Figure 1) was included in the Guidelines at the recommendation of the SAB. It has drawn many comments. The changes suggested would include additional detail and steps that would diminish the value of the graphic. However, the figure has been truncated to remove risk management steps.

In order to remain consistent with the separation of risk assessment and risk management, any directions to consider applicable laws or regulatory decisions have been stricken from the Guidelines.

The Agency agrees that peer review is an important aspect of the assessment process. However, emergency cases may not allow peer review in preliminary assessments. All nonemergency exposure assessments have been peer reviewed and will continue to be peer reviewed. Finally, it is clearly stated in the Guidelines that periodic revision of the document will be done to reflect the benefit of experience and knowledge.

Agency Cincinnati OH 45268

BULK RATE
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Official Business
Penalty for Private Use, $300

Special Fourth-Class Rate
Book

Please make all necessary changes on the above label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE ☐; detach, or copy this cover, and return to the address in the upper left-hand corner.

EPA/600/8-87/045