SFUND RECORDS CTR
99609

EPA/600/FR-91/001
December 1991

Guidelines for Developmental Toxicity Risk Assessment

Published on December 5, 1991, Federal Register 56(234):63798-63826

Risk Assessment Forum
U.S. Environmental Protection Agency

Washington, DC

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.

Note: This document represents the final guidelines. A number of editorial corrections have been made during conversion and subsequent proofreading to ensure the accuracy of this publication.

CONTENTS

Lists of Tables and Figures v

Federal Register Preamble vi

Part A: Guidelines for Developmental Toxicity Risk Assessment

1. Introduction 1

2. Definitions and Terminology 3

3. Hazard Identification/Dose-Response Evaluation of Agents That Cause Developmental Toxicity 4
   - 3.1. Developmental Toxicity Studies: Endpoints and Their Interpretation 5
     - 3.1.1. Laboratory Animal Studies 5
       - 3.1.1.1 Endpoints of Maternal Toxicity 7
       - 3.1.1.2. Endpoints of Developmental Toxicity: Altered Survival, Growth, and Morphological Development 9
       - 3.1.1.3. Endpoints of Developmental Toxicity: Functional Deficits 13
       - 3.1.1.4. Overall Evaluation of Maternal and Developmental Toxicity 17
       - 3.1.1.5. Short-Term Testing in Developmental Toxicity 18
       - 3.1.1.6. Statistical Considerations 20
     - 3.1.2. Human Studies 21
       - 3.1.2.1. Epidemiologic Studies 22
       - 3.1.2.2. Examination of Clusters or Case Reports/Series 31
     - 3.1.3. Other Considerations 31
       - 3.1.3.1. Pharmacokinetics 31
       - 3.1.3.2. Comparisons of Molecular Structure 32
   - 3.2. Dose-Response Evaluation 32
   - 3.3. Characterization of the Health-Related Database 37
   - 3.4. Determination of the Reference Dose (RfD_{DT}) or Reference Concentration (RfC_{DT}) for Developmental Toxicity 40
   - 3.5. Summary 42

CONTENTS (continued)

4. Exposure Assessment 43

5. Risk Characterization 45
   5.1. Overview 45
   5.2. Integration of the Hazard Identification/Dose-Response Evaluation and Exposure Assessment 45
   5.3. Descriptors of Developmental Toxicity Risk 47
      5.3.1. Estimation of the Number of Individuals Exposed to Levels of Concern 47
      5.3.2. Presenting Specific Scenarios 47
      5.3.3. Risk Characterization for Highly Exposed Individuals 47
      5.3.4. Risk Characterization for Highly Sensitive or Susceptible Individuals 48
      5.3.5. Other Risk Descriptors 48
   5.4. Communicating Results 49

6. Summary and Research Needs 50

7. References 51

Part B: Response to Public and Science Advisory Board Comments

1. Introduction 63

2. Intent of the Guidelines 64

3. Basic Assumptions 64

4. Maternal/Developmental Toxicity 65

5. Functional Developmental Toxicity 65

6. Weight-of-Evidence Scheme 66

7. Applicability of the RfD_{DT} Concept and the Benchmark Dose Approach 67

LIST OF TABLES

Table 1. Endpoints of maternal toxicity 8

Table 2. Endpoints of developmental toxicity 10

Table 3. Categorization of the health-related database for hazard identification/dose-response evaluation 38

LIST OF FIGURES

Figure 1. Graphical illustration of the benchmark dose approach 35

GUIDELINES FOR DEVELOPMENTAL TOXICITY RISK ASSESSMENT
[FRL-4038-3]

AGENCY: U.S. Environmental Protection Agency (EPA).

ACTION: Final Guidelines for Developmental Toxicity Risk Assessment.

SUMMARY: The U.S. Environmental Protection Agency (EPA) is today issuing final amended guidelines for assessing the risks for developmental toxicity from exposure to environmental agents. As background information for this guidance, this notice describes the scientific basis for concern about exposure to agents that cause developmental toxicity, outlines the general process for assessing potential risk to humans because of environmental contaminants, summarizes the history of these guidelines, and addresses public and Science Advisory Board comments on the 1989 "Proposed Amendments to the Guidelines for the Health Assessment of Suspect Developmental Toxicants" [54 FR 9386-9403]. These guidelines, which have been renamed "Guidelines for Developmental Toxicity Risk Assessment" (hereafter "Guidelines"), outline principles and methods for evaluating data from animal and human studies, exposure data, and other information to characterize risk to human development, growth, survival, and function because of exposure prior to conception, prenatally, or to infants and children. These Guidelines amend and replace EPA's 1986 "Guidelines for the Health Assessment of Suspect Developmental Toxicants" [51 FR 34028-34040] by adding new guidance on the relationship between maternal and developmental toxicity, characterization of the health-related database for developmental toxicity risk assessment, use of the reference dose or reference concentration for developmental toxicity (RfD_{DT} or RfC_{DT}), and use of the benchmark dose approach. In addition, the Guidelines were reorganized to combine hazard identification and dose-response evaluation since these are usually done together in assessing risk for human health effects other than cancer.

EFFECTIVE DATE: The Guidelines will be effective December 5, 1991.

FOR FURTHER INFORMATION, CONTACT: Dr. Carole A. Kimmel, Effects Identification and Characterization Group, National Center for Environmental Assessment-Washington Division (8623D), U.S. Environmental Protection Agency, 401 M Street, SW, Washington, DC 20460, TEL: 202-564-3307, FAX: 202-565-0078.

SUPPLEMENTARY INFORMATION: The Clean Air Act (CAA), the Toxic Substances Control Act (TSCA), the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), and other statutes administered by the EPA authorize the Agency to protect public health against adverse effects from environmental pollutants. One type of adverse effect of great concern is developmental toxicity, i.e., adverse effects produced prior to conception, or during pregnancy and childhood. Exposure to agents affecting development can result in any one or more of the following manifestations of developmental toxicity: death, structural abnormality, growth alteration, and/or functional deficit. These manifestations encompass a wide array of adverse developmental endpoints, such as spontaneous abortions, stillbirths, malformations, early postnatal mortality, reduced birth weight, mental retardation, sensory loss, and other adverse functional or physical changes that are manifested postnatally.

The Role of Environmental Agents in Developmental Toxicity

Several environmental agents are established as causing developmental toxicity in humans (e.g., lead, polychlorinated biphenyls, methylmercury, ionizing radiation), while many others are suspected of causing developmental toxicity in humans on the basis of data from experimental animal studies (e.g., some pesticides, other heavy metals, glycol ethers, alcohols, and phthalates). Data for several of the agents identified as causing human developmental toxicity have been compared to the experimental animal data (Nisbet and Karch, 1983; Kimmel et al., 1984; Hemminki and Vineis, 1985; Kimmel et al., 1990a). In these comparisons, the agents causing human developmental toxicity in almost all cases were found to produce effects in experimental animal studies and, in at least one species tested, types of effects similar to those in humans were generally seen. This information provides a strong basis for the use of animal data in conducting human health risk assessments. On the other hand, a number of agents found to cause developmental toxicity in experimental animal studies have not shown clear evidence of hazard in humans, but the available human data are often too limited to evaluate a cause-and-effect relationship. The comparison of dose-response relationships is hampered by differences in route, timing, and duration of exposure. When careful comparisons have been done taking these factors into account, the minimally effective dose for the most sensitive animal species was generally higher than that for humans, usually within 10-fold of the human effective dose, but sometimes was 100 times or more higher (e.g., polychlorinated biphenyls [Tilson et al., 1990]). Thus, the experimental animal data were generally predictive of adverse developmental effects in humans, but in some cases, the administered dose or exposure level required to achieve these adverse effects was much higher than the effective dose in humans.

In most cases, the toxic effects of an agent on human development have not been fully studied, even though exposure of humans to that agent may have been established. At the same time, there are many developmental effects in humans with unknown causes and no clear link with exposure to environmental agents. The background incidence of human spontaneous abortion, for example, was estimated by Hertig (1967) to be approximately 50% of all conceptions, and more recently, Wilcox et al. (1985), using sensitive techniques for detecting pregnancy as early as 9 days postconception, observed that 35% of postimplantation pregnancies ended in an embryonic or fetal loss. Of those infants born alive, approximately 7.4% are reduced in weight at birth (i.e., below 2,500 g) (Selevan, 1981), approximately 3% are found to have one or more congenital malformations at birth, and by the end of the first postnatal year, about 3% more are found to have serious developmental defects (Shepard, 1986). Of those children born with developmental defects, it has been estimated that 20% are due to genetic transmission and 10% can be attributed to known exogenous factors (including drugs, infections, ionizing radiation, and environmental agents), leaving the remaining 70% with unknown causes (Wilson, 1977). In a recent hospital-based surveillance study (Nelson and Holmes, 1989), 50.7% of congenital malformations were estimated to be due to genetic or multifactorial causes, while 3.2% were associated with exposure to exogenous agents and 2.9% to twinning or uterine factors, leaving 43.2% to unknown causes. The proportion of the effects with unknown causes that may be attributable to environmental agents or to a combination of factors, such as environmental agents and genetic factors, nutritional deficiencies, alcohol consumption, direct or indirect exposure to tobacco smoke, use of prescribed and illicit drugs, etc., is unknown.

The social and economic impact of developmental disabilities on the population is extremely high. Close to one-half of the children in hospital wards are there because of prenatally acquired malformations (Shepard, 1980). According to the Centers for Disease Control, congenital anomalies, sudden infant death syndrome, and prematurity combined account for more than 50% of infant mortality among all races in the United States (National Center for Health Statistics, 1988). In addition, among the leading causes of estimated years of potential life lost (YPLL) due to death before the age of 65, congenital anomalies, prematurity, and sudden infant death syndrome combined rank third (Centers for Disease Control, 1988a,b). The YPLL estimates for developmental defects may actually underestimate the public health impact because the estimates do not include prenatal deaths, they are based only on those cases that die before age 65 and do not account for limited quality of life, and pregnancies may be terminated early due to prenatal diagnosis of developmental defects.

These data provide the basis for a long-standing interest by Federal agencies that deal with human health to protect against exposures to agents that cause developmental toxicity, and most of

these regulatory agencies have provisions for considering data on developmental toxicity in protecting human health. As a step in developing procedures for interpreting toxicity data in the regulatory context, the National Academy of Sciences/National Research Council, in 1983, published a framework for the risk assessment process, which EPA uses as the basis for its risk assessment guidelines and for the assessment of risk due to environmental agents.

The Risk Assessment Process and Its Application to Developmental Toxicity

Risk assessment is the process by which scientific judgments are made concerning the potential for toxicity to occur in humans. The National Research Council (1983) has defined risk assessment as including 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. In practice, however, hazard identification for developmental toxicity and other noncancer health effects is usually done in conjunction with an evaluation of dose-response relationships, since the determination of a hazard is often dependent on whether a dose-response relationship is present (Kimmel et al., 1990b). One advantage of this approach is that it reflects hazard within the context of dose, route, and duration and timing of exposure, all of which are important in comparing the toxicity information available to potential human exposure scenarios. Second, this approach avoids labeling of chemicals as developmental toxicants on a purely qualitative basis. For these reasons, the Guidelines combine hazard identification and dose-response evaluation under one section (Section 3), and characterize both hazard and dose information as part of the health-related database for risk assessment. If data are considered sufficient for risk assessment, an oral or dermal reference dose for developmental toxicity (RfD_{DT}) or an inhalation reference concentration for developmental toxicity (RfC_{DT}) is then derived for comparison with human exposure estimates. A statement of the potential for human risk and the consequences of exposure can come only from integrating the hazard identification/dose-response evaluation with the human exposure estimates in the final risk characterization. Combining hazard identification and dose-response evaluation, as well as development of the RfD_{DT} and RfC_{DT}, are revisions of the 1986 Guidelines.

Hazard identification/dose-response evaluation involves examining all available experimental animal and human data and the associated doses, routes, and timing and duration of exposures to determine if an agent causes developmental toxicity and/or maternal or paternal toxicity in that species and under what exposure conditions. The no-observed-adverse-effect-level (NOAEL) and/or the lowest-observed-adverse-effect-level (LOAEL) are determined for each study and type of effect. Based upon the hazard identification/dose-response evaluation and criteria provided in these

Guidelines, the health-related database can be characterized as sufficient or insufficient for use in risk assessment (Section 3.3). Because of the limitations associated with the use of the NOAEL, the Agency is evaluating the use of an additional approach, i.e., the benchmark dose approach (Crump, 1984), for more quantitative dose-response evaluation when sufficient data are available. The benchmark dose provides an indication of the risk associated with exposures near the NOAEL, taking into account the variability in the data and the slope of the dose-response curve.

For the determination of the RfD_{DT} or the RfC_{DT}, uncertainty factors are applied to the NOAEL (or LOAEL, if a NOAEL has not been established) to account for extrapolation from experimental animals to humans and for variability within the human population. The RfD_{DT} or RfC_{DT} is generally based on a short duration of exposure as is typically used in developmental toxicity studies in experimental animals. The use of the terms RfD_{DT} and RfC_{DT} distinguish them from the oral or dermal reference dose (RfD) and the inhalation reference concentration (RfC), which refer primarily to chronic exposure situations (U.S. EPA, 1991). Uncertainty factors may also be applied to a benchmark dose for calculating the RfD_{DT} or RfC_{DT}, but the Agency has little experience with applying this approach and is currently supporting research efforts to determine the appropriate methods. As more information becomes available, guidance will be written and published as an addendum to these Guidelines. These approaches are discussed further in Section 3.4.

The exposure assessment identifies human populations exposed or potentially exposed to an agent, describes their composition and size, and presents the types, magnitudes, frequencies, and durations of exposure to the agent. The exposure assessment provides an estimate of human exposure levels for particular populations from all potential sources.

In risk characterization, the hazard identification/dose-response evaluation and the exposure assessment for given populations are combined to estimate some measure of the risk for developmental toxicity. As part of risk characterization, a summary of the strengths and weaknesses in each component of the risk assessment is discussed along with major assumptions, scientific judgments, and, to the extent possible, qualitative and quantitative estimates of the uncertainties. Confidence in the health-related data is always presented in conjunction with information on dose-response and the RfD_{DT} or RfC_{DT}. If human exposure estimates are available, the exposure basis used for the risk assessment is clearly described, e.g., highly exposed individuals, or highly sensitive or susceptible individuals. The NOAEL may be compared to the various estimates of human exposure to calculate the margin(s) of exposure (MOE). The considerations for determining adequacy of the MOE are similar to those used in determining the appropriate size of the uncertainty factor for calculating the RfD_{DT} or RfC_{DT}.

Risk assessment is just one component of the regulatory process and defines the potential adverse health consequences of exposure to a toxic agent. The other component, risk management, combines risk assessment with statutory directives regarding socioeconomic, technical, political, and other considerations, to reach decisions about the appropriate regulation of the suspected toxic agents. Risk management is not dealt with directly in these Guidelines since the basis for decision making goes beyond scientific consideration alone, but the use of scientific information in this process is discussed in some cases. For example, the acceptability of the MOE is a risk management decision, but the scientific bases for establishing this value are discussed here.

History of These Guidelines

In 1984, the Agency published "Proposed Guidelines for the Health Assessment of Suspect Developmental Toxicants" [49 FR 46324-46331]. Following extensive scientific and public review, final guidelines were issued on September 24, 1986 [51 FR 34028-34040]. The 1986 Guidelines set forth principles and procedures to guide EPA scientists in the conduct of Agency risk assessments, to help promote high scientific quality and Agencywide consistency, and to inform Agency decision makers and the public about these scientific procedures. In publishing this guidance, EPA emphasized that one purpose of its risk assessment guidelines was to "encourage research and analysis that will lead to new risk assessment methods and data," which in turn would be used to revise and improve the guidelines, and better guide Agency risk assessors. Thus, the 1986 Guidelines were developed and published with the understanding that risk assessment is an evolving science and that continued study could lead to changes.

As expected, Agency experience with the 1986 Guidelines suggested that additional or alternate approaches should be considered for certain aspects of the guidance. Proposals to amend the guidelines were considered soon after their publication in September 1986, because of new reviews or re-evaluations that focused on some of the issues identified for research in the guidelines. Included were several workshops and symposia cited in the Introduction to these Guidelines. In addition, much experience had been gained in using the 1986 Guidelines and in instructing others in their use.

Based on this experience, amendments to the 1986 Guidelines were proposed for public comment in March 1989 [54 FR 9386-9403]. Following receipt and review of the public comments, they were collated, summarized, and reviewed by scientists within the Agency. On October 27, 1989, EPA's Science Advisory Board (SAB) met to review the Proposed Amendments and the summarized public comments, and to be briefed by Agency scientists concerning proposed responses.

During this same period, several issues with implications for health effects other than cancer were under discussion in the Agency and elsewhere. These issues included use of the benchmark dose (see Section 3.2), exposure descriptors (see Section 5.3), and risk characterization (see Section 5). Thus, generic discussions on risk assessment issues, along with comments from the public and the SAB, have influenced the structure and content of these Guidelines.

These revised Guidelines were then reviewed by a number of Agency scientists and official panels, including the Risk Assessment Forum and the Risk Assessment Council. The revised Guidelines also were presented to the SAB on March 27, 1991, for final comment. In addition, a review was conducted by the interagency Working Party on Reproductive Toxicology, Subcommittee on Risk Assessment of the Federal Coordinating Committee on Science, Engineering and Technology. Comments of these groups have been considered in the revision of these Guidelines. The full text of the final "Guidelines for Developmental Toxicity Risk Assessment" is published here.

These Guidelines were developed as part of an interoffice guidelines development program under the auspices of the Risk Assessment Forum and the Office of Health and Environmental Assessment (OHEA) in the Agency's Office of Research and Development. The Agency is continuing to study risk assessment issues raised in these Guidelines, and will revise them in line with new information as appropriate.

Following this Preamble are two parts: Part A is the Guidelines and Part B is the Response to the Public and Science Advisory Board Comments. Part B includes a summary of the issues raised by the public and the SAB, and the Agency's responses to those comments.

References, supporting documents, and comments received on the Proposed Amendments, as well as a copy of the final Guidelines, are available for inspection and copying at the Public Information Reference Unit Docket (202-260-5926), EPA Headquarters Library, 401 M Street, S.W., Washington, DC, between the hours of 8:00 a.m. and 4:30 p.m.

Dated: November 26, 1991

Signed by EPA Administrator
William K. Reilly

PART A: GUIDELINES FOR DEVELOPMENTAL TOXICITY RISK ASSESSMENT

1. INTRODUCTION

These Guidelines describe the procedures that EPA follows in evaluating potential developmental toxicity associated with human exposure to environmental agents. The Agency has sponsored or participated in several conferences that addressed issues related to such evaluations and that provide some of the scientific basis for these Guidelines (U.S. EPA, 1982a; Kimmel et al., 1982b, 1987; Hardin, 1987; Perlin and McCormack, 1988; Kimmel et al., 1989; Kimmel and Francis, 1990; Kimmel et al., 1990a). The Agency's authority to regulate substances that have the potential to interfere with human development is derived from a number of statutes that are implemented through multiple offices within the EPA. The procedures described herein are intended to promote consistency in the assessment of developmental toxic effects across program offices within the Agency.

These Guidelines provide a general format for analyzing and organizing the available data for conducting risk assessments. The Agency previously has issued testing guidelines (U.S. EPA, 1982b, 1985a, 1989a, 1991a) that provide protocols designed to determine the potential of a test substance to induce structural and/or other adverse effects during development. These risk assessment Guidelines do not change any prescribed statutory or regulatory 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) that are not routinely required, but may be encountered when reviewing data on particular agents.

Since the purpose of risk assessment is to make inferences about potential risks to human health, the most appropriate data to be used are those deriving from studies of humans. If adequate human data are not available, then it is necessary to use data obtained from other species. There are a number of unknowns in the extrapolation of data from animal studies to humans. Therefore, a number of assumptions must be made on the relevance of effects to potential human risk that are generally applied in the absence of data. These assumptions provide the inferential basis for the approaches taken to risk assessment in these Guidelines.

First, it is assumed that an agent that produces an adverse developmental effect in experimental animal studies will potentially pose a hazard to humans following sufficient exposure during development. This assumption is based on the comparisons of data for agents known to cause human developmental toxicity (Nisbet and Karch, 1983; Kimmel et al., 1984; Hemminki and Vineis, 1985;

Kimmel et al., 1990a), which indicate that, in almost all cases, experimental animal data are predictive of a developmental effect in humans.

It is assumed that all of the four manifestations of developmental toxicity (death, structural abnormalities, growth alterations, and functional deficits) are of concern. In the past, there has been a tendency to consider only malformations or malformations and death as endpoints of concern. From the data on agents that are known to cause human developmental toxicity (Nisbet and Karch, 1983; Kimmel et al., 1984; Hemminki and Vineis, 1985; Kimmel et al., 1990a), there is usually at least one experimental species that mimics the types of effects seen in humans, but in other species tested, the type of developmental perturbation may be different. Thus, a biologically significant increase in any of the four manifestations is considered indicative of an agent's potential for disrupting development and producing a developmental hazard.

It is assumed that the types of developmental effects seen in animal studies are not necessarily the same as those that may be produced in humans. This assumption is made because it is impossible to determine which will be the most appropriate species in terms of predicting the specific types of effects seen in humans. The fact that every species may not react in the same way could be due to species-specific differences in critical periods, differences in timing of exposure, metabolism, developmental patterns, placentation, or mechanisms of action.

The most appropriate species is used to estimate human risk when data are available (e.g., pharmacokinetics). In the absence of such data, it is assumed that the most sensitive species is appropriate for use, based on observations that humans are as sensitive or more so than the most sensitive animal species tested for the majority of agents known to cause human developmental toxicity (Nisbet and Karch, 1983; Kimmel et al., 1984; Hemminki and Vineis, 1985; Kimmel et al., 1990a).

In general, a threshold is assumed for the dose-response curve for agents that produce developmental toxicity. This is based on the known capacity of the developing organism to compensate for or to repair a certain amount of damage at the cellular, tissue, or organ level. In addition, because of the multipotency of cells at certain stages of development, multiple insults at the molecular or cellular level may be required to produce an effect on the whole organism.

2. 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 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 lifespan 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.

Altered growth - An alteration in offspring organ or body weight or size. Changes in one endpoint may or may not be accompanied by other signs of altered growth (e.g., changes in body weight may or may not be accompanied by changes in crown-rump length and/or 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 alterations or delays in the physiological and/or biochemical competence of an organism or organ system following exposure to an agent during critical periods of development pre- and/or postnatally.

Structural abnormalities - Structural alterations in development that include both malformations and variations.

Malformations and variations - A malformation is usually defined as a permanent structural change that may adversely affect survival, development, or function. The term *teratogenicity* is used in these Guidelines to refer only to malformations. The term *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 extremely deviant. There is no generally accepted classification of malformations and variations. Other terms that are often used, but no better defined, include anomalies, deformations, and aberrations.

3. HAZARD IDENTIFICATION/DOSE-RESPONSE EVALUATION OF AGENTS THAT CAUSE DEVELOPMENTAL TOXICITY

This section discusses the evaluation and interpretation of hazards for a variety of endpoints of developmental toxicity seen in both human and animal studies, and describes the criteria for characterizing the sufficiency of the health-related database for conducting a developmental toxicity risk assessment. It also details the use of dose-response data for determining potential hazards, and describes the calculation of the RfD_{DT} or RfC_{DT}, a dose or concentration that is assumed to be without appreciable risk of deleterious developmental effects for a given agent.

Developmental toxicity is expressed as one or more of a number of possible endpoints that may be used for evaluating the potential of an agent to cause abnormal development. Developmental toxicity generally occurs in a dose-related manner, may result from short-term exposure (including single exposure situations) or from longer term low-level exposure, may be produced by various routes of exposure, and the types of effects may vary depending on the timing of exposure because of a number of critical periods of development for various organs and functional systems.

The four major manifestations of developmental toxicity are death, structural abnormality, altered growth, and functional deficit. The relationship among these manifestations may vary with increasing dose and, especially at higher doses, death of the conceptus may preclude expression of other manifestations. Of these, all four manifestations have been evaluated in human studies, but only the first three 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 or the continuous breeding study. Although functional deficits seldom have been evaluated in routine testing studies in experimental animals, functional evaluations are beginning to be required in certain regulatory situations (U.S. EPA, 1986a, 1988a, 1989b, 1991a).

Developmental toxicity can be considered a component of reproductive toxicity, and often it is difficult to distinguish between effects mediated through the parents versus direct interaction with developmental processes. For example, developmental toxicity may be influenced by the effects of toxic agents on the maternal system when exposure occurs during pregnancy or lactation. In addition, following parental exposure prior to conception, developmental toxicity may result in their offspring and, potentially, in subsequent generations. Therefore, it is useful to consult the "Proposed Guidelines for Assessing Male Reproductive Risk" (U.S. EPA, 1988b) and the "Proposed Guidelines for Assessing Female Reproductive Risk" (U.S. EPA, 1988c) in conjunction with these Guidelines. Mutational events that occur as a result of exposure to agents that cause developmental toxicity may be difficult to

discriminate from other possible mechanisms in standard studies of developmental toxicity. When mutational events are suspected, the "Guidelines for Mutagenicity Risk Assessment" (U.S. EPA, 1986c), which specifically address the risks of heritable mutation, should be consulted.

Carcinogenic effects have occurred in humans following developmental exposures to diethylstilbestrol (Herbst et al., 1971). Several additional agents (e.g., direct-acting alkylating agents) have been shown to cause cancer following developmental exposures in experimental animals, and it appears from the data collected thus far that agents capable of causing cancer in adults may also cause transplacental or neonatal carcinogenesis (Anderson et al., 1985). Currently, there is no way to predict whether the developing offspring or adult will be more sensitive to the carcinogenic effects of an agent. At present, testing for carcinogenesis following developmental exposure is not routinely required. However, if this type of effect is reported for an agent, it is considered appropriate to use the "Guidelines for Carcinogen Risk Assessment" (U.S. EPA, 1986b) for assessing human risk.

3.1. DEVELOPMENTAL TOXICITY STUDIES: ENDPOINTS AND THEIR INTERPRETATION

3.1.1. Laboratory Animal Studies

This section discusses the endpoints examined in routinely used protocols as well as the use of other types of studies, including functional studies and short-term tests.

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 (U.S. EPA, 1982b, 1985a; Food and Drug Administration [FDA], 1966, 1970; Organization for Economic Cooperation and Development [OECD], 1981). Some studies may use exposures of one to a few days to investigate periods of particular sensitivity for induction of abnormalities in specific organs or organ systems. In addition, developmental toxicity may be evaluated in studies involving exposure to one or both parents prior to conception, to the conceptus during pregnancy and over several generations, or to offspring during the prenatal and preweaning periods (U.S. EPA, 1982b, 1985a, 1986a, 1988a, 1991a; FDA, 1966, 1970; OECD, 1981; Lamb, 1985). These Guidelines are intended to provide information for interpreting developmental effects related to any of these types of exposure.

Appropriate study designs include a number of important factors. For example, test animal selection is generally based on considerations of species, strain, age, weight, and health status. Assignment of animals to dose groups by stratified randomization (on the basis of body weight) reduces

bias and provides a basis for performing valid statistical tests. At a minimum, a high dose, a low dose, and one intermediate dose are included. The high dose is selected to produce some minimal maternal or adult toxicity (i.e., a level that at the least produces marginal but significantly reduced body weight, reduced weight gain, or specific organ toxicity, and at the most produces no more than 10% mortality). At doses that cause excessive maternal toxicity (that is, significantly greater than the minimal toxic level), information on developmental effects may be difficult to interpret and of limited value. The low dose is generally a NOAEL for adult and offspring effects, although if the low dose produces a biologically or statistically significant increase in response, it is considered a LOAEL (see Section 3.1.1.6 for a discussion of biological versus statistical significance). A concurrent control group treated with the vehicle used for agent administration is a critical component of a well-designed study.

The route of exposure in these studies is usually oral, unless the chemical or physical characteristics of the test substance or pattern of human exposure suggest a more appropriate route of administration. In the case of dermal exposure, developmental toxicity studies showing no indication of maternal or developmental toxicity are considered insufficient for risk assessment unless accompanied by absorption data (Kimmel and Francis, 1990). Dermal developmental toxicity studies in which skin irritation is too marked (moderate erythema and/or moderate edema, i.e., raised approximately 1 mm) also are considered insufficient, since excessive maternal toxicity may be produced from the irritation rather than from systemic exposure to the agent. Assessment of maternal toxicity is based on signs of systemic toxicity rather than on local effects such as skin irritation. Absorption data and limited pharmacokinetic data collected in dermal developmental toxicity studies provide very useful information in the evaluation of study design and data interpretation (Kimmel and Francis, 1990). Many of these points also are pertinent to studies by other routes of exposure.

The evaluation of specific endpoints of maternal and developmental toxicity is discussed in the next several sections. Appropriate historical control data sometimes can be very useful in the interpretation of these endpoints. Comparison of data from treated animals with concurrent study controls should always take precedence over comparison with historical control data. The most appropriate historical control data are those from the same laboratory in which studies were conducted. Even data from the same laboratory, however, should be used cautiously and examined for subtle changes over time that may result from genetic alterations in the strain or stock of the species used, changes in environmental conditions both in the breeding colony of the supplier and in the laboratory, and changes in personnel conducting studies and collecting data (Kimmel and Price, 1990). Study data should be compared with recent as well as cumulative historical data. Any change in laboratory

procedure that might affect control data should be noted and the data accumulated separately from previous data.

The next three sections (3.1.1.1, 3.1.1.2, and 3.1.1.3) discuss individual endpoints of maternal and developmental toxicity as measured in the conventional developmental toxicity study, the multigeneration study, and, when available, in postnatal studies. Other endpoints specifically related to reproductive toxicity are covered in the relevant risk assessment guidelines (U.S. EPA, 1988b, 1988c). The fourth section (3.1.1.4) 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.

3.1.1.1. *Endpoints of Maternal Toxicity*

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

Body weight and change in body weight are viewed collectively as indicators of maternal toxicity for most species, although these endpoints may not be as useful in rabbits, because body weight changes are usually more variable (Kimmel and Price, 1990) and, in some strains of rabbits, body weight is not a good indicator of pregnancy status. Body weight changes may provide more information than a daily body weight measured during treatment or during gestation. Changes in weight gain 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 gain 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 intrauterine. For example, a significant reduction in weight gain throughout gestation and in gravid uterine weight without any change in corrected maternal weight gain generally would indicate an intrauterine effect. Conversely, a change in corrected weight gain and no change in gravid uterine weight generally would suggest 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 or placental tissue, or the amniotic fluid.

Table 1. Endpoints of maternal toxicity

Mortality

Mating index [(no. with seminal plugs or sperm/no. mated) × 100]

Fertility index [(no. with implants/no. of matings) × 100]

Gestation length (useful when animals are allowed to deliver pups)

Body weight

- Day 0
- During gestation
- Day of necropsy

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 target organ toxicity and especially when supported by adverse histopathology findings)

- Absolute
- Relative to body weight
- Relative to brain weight

Food and water consumption (where relevant)

Clinical evaluations

- Types, incidence, degree, and duration of clinical signs
- Enzyme markers
- Clinical chemistries

Gross necropsy and histopathology

Changes in other endpoints may also be important. For example, changes in relative and absolute organ weights may be signs of a maternal effect, especially when an agent is suspected of causing specific organ toxicity and when such findings are supported by adverse histopathologic findings in those organs. 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 also are useful when an agent is suspected of affecting appetite, water intake, or excretory function.

Clinical evaluations of toxicity also may be used as indicators of maternal toxicity. Daily clinical observations may be useful in describing the profile of maternal toxicity and alterations in general homeostasis. 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. The minimum amount of information considered useful for evaluating maternal toxicity [as noted in the "Proceedings of the Workshop on the Evaluation of Maternal and Developmental Toxicity" (Kimmel et al., 1987)], includes morbidity or mortality, maternal body weight and body weight gain, clinical signs of toxicity, food and water consumption (especially if dosing is via food or water), and necropsy for gross evidence of organ toxicity. In a well-designed study, maternal toxicity is determined in the pregnant and/or lactating animal over an appropriate part of gestation and/or the neonatal period, and is not assumed or extrapolated from other adult toxicity studies.

3.1.1.2. *Endpoints of Developmental Toxicity: Altered Survival, Growth, and Morphological Development*

Because the maternal animal, and not the conceptus, is the individual treated during gestation, data generally are calculated as incidence per litter or as number and percent of litters with particular endpoints. Table 2 indicates the ways in which offspring and litter endpoints may be expressed.

When treatment of females begins prior to implantation, an increase in preimplantation loss could indicate an adverse effect on gamete transport, the fertilization process, uterine toxicity, 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 variability that is not treatment-related in the animals being used, but the data should be examined carefully to determine if there is a dose-response

Table 2. Endpoints 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[b]/litter
- No. and percent resorptions/litter
- No. and percent litters with resorptions
- 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
- No. and percent litters with live offspring

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 or female body weight/litter[c]
- No. and percent offspring with external, visceral, or skeletal malformations/litter
- No. and percent malformed offspring/litter
- No. and percent litters with malformed offspring
- No. and percent malformed males or females/litter
- No. and percent offspring with external, visceral, or skeletal variations/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 (type, incidence, duration, and degree)
- Gross necropsy and histopathology

[a]Important when treatment begins prior to implantation. May be difficult to assess in mice.
[b]Offspring refers both to fetuses observed prior to term and to pups following birth. The endpoints examined depend on the protocol used for each study.
[c]Measured at selected intervals until termination of the study.

relationship. If preimplantation loss is related to dose, 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 of resorptions and late fetal deaths give some indication of when the conceptus died, and the number and percent of nonlive implants per litter (postimplantation loss) is a combination of these two measures. Expression of data as the number and percent of litters showing an increased incidence for these endpoints may be less useful than incidence per litter because, in the former case, a litter is counted whether one or all implants were resorbed, dead, or nonlive.

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 (preferably from the same laboratory), since there is considerable interlitter variability in the incidence of postimplantation loss (Kimmel and Price, 1990). 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 study for risk assessment purposes.

The endpoint "affected implants" (i.e., the combination of nonlive and malformed conceptuses) 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 deaths or malformations 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 is generally 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. While there is always a question as to whether weight reduction is a permanent or transitory effect, little is known about the long-term consequences of short-term fetal or neonatal weight changes. Therefore, when significant weight reduction effects are noted, they are used as a basis to establish the NOAEL. Several other factors 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 affected 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 are generally examined for external, visceral, and skeletal malformations and variations. If only a portion of the litter is examined for one or more endpoints, then random selection of those pups examined introduces less bias in the data. An increase in the incidence of malformed offspring may be indicated by a change in one or more of the following endpoints: 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).

Other ways of examining the data include determining the incidence of external, visceral, and skeletal malformations and variations that may indicate the organs or organ systems 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 the evaluation of any single endpoint based on number or percent of offspring or litters. The incidence of individual types of malformations and variations may indicate significant changes that are masked if the data on all malformations and/or variations are pooled. Appropriate historical control data can be especially helpful in the interpretation of malformations and variations, particularly those that normally occur at a low incidence and may or may not be related 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 biological significance of an altered incidence of anatomical variations is more difficult to assess, 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.

In addition, although some investigators have considered certain of these effects to simply be associated with manifestations of maternal toxicity noted at similar dose levels (Khera, 1984, 1985, 1987), such effects are still toxic manifestations and as such are generally considered a reasonable basis for Agency regulation and/or risk assessment. On a somewhat similar note, the conclusion of participants in a "Workshop on Reproductive Toxicity Risk Assessment" (Kimmel et al., 1986) was that dose-related increases in defects that may occur spontaneously are as relevant as dose-related increases in any other developmental toxicity endpoints.

3.1.1.3. *Endpoints of Developmental Toxicity: Functional Deficits*

Developmental effects that are induced by exogenous agents are not limited to death, structural abnormalities, and altered growth. Rather, it has been demonstrated in a number of instances that 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. Sometimes, these functional defects are observed at dose levels below those at which other indicators of developmental toxicity are evident (Rodier, 1978). Such effects may be transient or reversible in nature, but generally are considered adverse. Testing for functional developmental toxicity has not been required routinely by regulatory agencies in the United States, but studies in developmental neurotoxicity are beginning to be required by the EPA when other information indicates the potential for adverse functional developmental effects (U.S. EPA, 1986a, 1988a, 1989b, 1991a). Data from postnatal studies, when available, are considered very useful for further assessment of the relative importance and severity of findings in the fetus and neonate. Often, the long-term consequences of adverse developmental outcomes noted at birth are unknown, and further data on postnatal development and function are necessary to determine the full spectrum of potential developmental effects. Useful data can also be derived from well-conducted multigeneration studies, although the dose levels used in these studies may be much lower than in studies with shorter-term exposure.

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. Recent advances in this area have been reviewed in several publications (Riley and Vorhees, 1986; Kimmel, 1988; Kimmel et al., 1990a). Several expert groups have focused on the functions that should be included in a behavioral testing battery (World Health Organization [WHO], 1984; Buelke-Sam et al., 1985; Leukroth, 1986). These include: sensory systems, neuromotor development, locomotor activity,

learning and memory, reactivity and/or habituation, and reproductive behavior. No testing battery has fully addressed all of these functions, but it is important to include as many as possible, and several testing batteries have been developed and evaluated for use in testing (Buelke-Sam et al., 1985; Tanimura, 1986; Elsner et al., 1986).

The Agency recently has developed a "generic" developmental neurotoxicity test guideline that can be used for both pesticides and industrial chemicals (U.S. EPA, 1991a). Because of its design, the developmental neurotoxicity testing protocol may be conducted as a separate study, concurrently with or as a follow-up to a developmental toxicity (Segment II) study, or be folded into a multigeneration study in the second generation. Testing is generally conducted in the rat. In the protocol for the separate study, the test agent is administered orally (other routes may be used on a case-by-case basis) to at least three treated groups and one concurrent control group of animals on day 6 of gestation through day 10 postnatally. The highest dose level is selected to induce some overt signs of maternal toxicity, but not result in more than a 20% reduction in weight gain during gestation and lactation. This dose also is selected to avoid in utero or neonatal death or malformations sufficient to preclude a meaningful evaluation of developmental neurotoxicity. At least 20 litters are required per treatment group. For behavioral tests, one female and one male pup per litter are randomly selected and assigned to one of the following tests: motor activity, auditory startle, and learning and memory in animals at weaning and as adults. Neuropathological evaluation and determination of brain weights are conducted on selected pups at postnatal day 11 and at termination of the study.

Several criteria for selecting agents for developmental neurotoxicity testing have been suggested (Buelke-Sam et al., 1985; Levine and Butcher, 1990), including: agents that cause central nervous system malformations, psychoactive drugs and chemicals, agents that cause adult neurotoxicity, hormonally active agents, and chemicals that are structurally related to others that cause developmental neurotoxicity or for which widespread exposure and/or release is expected. Data from developmental neurotoxicity studies should be evaluated in light of the data that may have triggered such testing as well as all other toxicity data available.

Less work has been done on other developing functional systems, but the assessment of postnatal renal morphological and functional development may serve as a model for the use of postnatal evaluations in the risk assessment process. As an example, standard morphological analyses of the kidneys of fetal rodents have detected treatment-related changes in the relative growth of the renal papilla versus the renal cortex, an effect considered in some cases to be a malformation (hydronephrosis), while in other cases a variation (apparent hydronephrosis, enlarged or dilated renal pelvis). While some investigators (Woo and Hoar, 1972) have provided data suggesting that the

morphological effect represents a transient developmental delay, others have shown that it can persist well into postnatal life and that physiological function is compromised in the affected individuals (Kavlock et al., 1987a, 1988; Daston et al., 1988; Couture, 1990). Thus, the biological interpretation of this effect on the basis of fetal examinations alone is tenuous (U.S. EPA, 1985b). In addition, the critical period for inducing renal morphological abnormalities extends into the postnatal period (Couture, 1990), and studies on perinatally induced renal growth retardation (Kavlock et al., 1986, 1987b; Slotkin et al., 1988; Gray et al., 1989; Gray and Kavlock, 1991) have shown that renal function is generally altered in such conditions, but that manifestation of the dysfunction is not readily predictable. Thus, both morphological and functional assessment of the kidneys after birth can provide useful and complementary information on the persistence and biological significance of expressions of developmental toxicity.

Although not as well studied, data indicate that the cardiovascular, respiratory, immune, endocrine, reproductive, and digestive systems also are subject to alterations in functional competence (Kavlock and Grabowski, 1983; Fujii and Adams, 1987) following exposure during development. Currently, there are no standard testing procedures for these functional systems; however, when data are encountered on a chemical under review, they are considered in the risk assessment process.

Direct extrapolation of functional developmental effects to humans is limited in the same way as for other endpoints of developmental toxicity, i.e., by the lack of knowledge about underlying toxicological mechanisms and their significance. In evaluations of a limited number of agents known to cause developmental neurotoxic effects in humans, Adams (1986) concluded that these agents produce similar developmental neurotoxic effects in animals and humans. This conclusion was strongly supported by the results of a recent "Workshop on the Qualitative and Quantitative Comparability of Human and Animal Developmental Neurotoxicity," sponsored by EPA and the National Institute on Drug Abuse (NIDA), at which participants critically evaluated and compared the effects of agents known to cause human developmental neurotoxicity with the effects seen in experimental animal studies (Kimmel et al., 1990a). The high degree of qualitative correlation between human and experimental animal data for the agents evaluated lends strong support for the use of experimental animals in assessing the potential risk for developmental neurotoxicity in humans. Thus, as for other endpoints of developmental toxicity, the assumption can be made that functional effects in animal studies indicate the potential for altered development in humans, although the types of developmental effects seen in experimental animal studies will not necessarily be the same as those that may be produced in humans. Thus, when data from functional developmental toxicity studies are encountered for particular agents, they should be considered in the risk assessment process.

Some guidance is provided here concerning important general concepts of study design and evaluation for functional developmental toxicity studies.

- 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 and test groups, litter generally considered the statistical unit, etc.).
- A replicate study design provides added confidence in the interpretation of data.
- A pharmacological/physiological 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.
- Functional tests with a moderate degree of background variability may be more sensitive to the effects of an agent on behavioral endpoints than are tests with low variability that may be impossible to disrupt without being life-threatening (Butcher et al., 1980).
- A battery of functional tests, in contrast to a single test, is usually needed to evaluate the full complement of organ function in an animal; tests conducted at several ages may provide more information about maturational changes and their persistence.
- 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.
- Interpretation of data from studies in which postnatal exposure is included should take into account possible interaction of the agent with maternal behavior, milk composition, pup suckling behavior, possible direct exposure of pups via dosed feed or water, etc.

Although interpretation of functional data may be limited at present, it is clear that functional effects must be evaluated in light of other toxicity data, including other forms of developmental toxicity (e.g., structural abnormalities, perinatal death, and growth retardation). The level of confidence in an adverse effect may be as important as the type of change seen, and confidence may be increased by such factors as replicability of the effect either in another study of the same function or by convergence of data from tests that purport to measure similar functions. A dose-response relationship is considered an important measure of chemical effect; in the case of functional effects, both monotonic and biphasic dose-response curves are likely, depending on the function being tested.

Finally, there are at least three general 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

effects; (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 suggest organ systems to be evaluated in exposed human populations.

3.1.1.4. *Overall Evaluation of Maternal and Developmental Toxicity*

As discussed previously, individual endpoints of maternal and developmental toxicity are evaluated in developmental toxicity studies. In order to interpret the data fully, an integrated evaluation must be performed considering all maternal and developmental endpoints.

Agents that produce developmental toxicity at a dose that is not toxic to the maternal animal are especially of concern because the developing organism is affected but toxicity is not apparent in the adult. However, the more common situation is when adverse developmental effects are produced only at doses that cause minimal maternal toxicity; in these cases, the developmental effects are still considered to represent developmental toxicity and should not be discounted as being secondary to maternal toxicity. At doses causing excessive maternal toxicity (that is, significantly greater than the minimal toxic dose), information on developmental effects may be difficult to interpret and of limited value. Current information is inadequate to assume that developmental effects at maternally toxic doses result only from maternal toxicity; rather, when the LOAEL is the same for the adult and developing organisms, it may simply indicate that both are sensitive to that dose level. Moreover, whether developmental effects are secondary to maternal toxicity or not, 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 involuntarily, since several agents are known to produce adverse developmental effects at minimally toxic doses in adult humans (e.g., smoking, alcohol, isotretinoin).

Since the final risk assessment not only takes into account the potential hazard of an agent, but also the nature of the dose-response relationship, it is important that the relationship of maternal and developmental toxicity be evaluated and described. Then, information from the exposure assessment is used to determine the likelihood of exposure to levels near the maternally toxic dose for each agent and the risk for developmental toxicity in humans.

Although the evaluation of developmental toxicity is the primary objective of standard studies within this area, maternal effects seen within the context of developmental toxicity studies should be evaluated as part of the overall toxicity profile for a given chemical. Maternal toxicity may be seen in the absence of or at dose levels lower than those producing developmental toxicity. If the maternal effect level is lower than that in other evaluations of adult toxicity, this implies that the pregnant female is

likely to be more sensitive than the nonpregnant female. Data from reproductive and developmental toxicity studies on the pregnant female should be used in the overall assessment of risk.

Approaches for ranking agents according to their relative maternal and developmental toxicity have been proposed; Schardein (1983) has reviewed several of these. Several approaches involve the calculation of ratios relating an adult toxic dose to a developmentally toxic dose (Johnson, 1981; Fabro et al., 1982; Johnson and Gabel, 1983; Brown and Freeman, 1984). Such ratios may describe in a qualitative and roughly quantitative fashion the relationship of maternal (adult) and developmental toxicity. However, at the U.S. EPA-sponsored "Workshop on the Evaluation of Maternal and Developmental Toxicity" (Kimmel et al., 1987), there was no agreement as to the validity or utility of these approaches in other aspects of the risk assessment process. This is due in part to uncertainty about factors that can affect the ratios. For example, the number and spacing of dose levels, differences in study design (e.g., route and/or timing of exposure), the relative thoroughness in the assessment of maternal and developmental endpoints examined, species differences in response, and differences in the slope of the dose-response curves for maternal and developmental toxicity can all influence the maternal and developmental effects observed and the resulting ratios (Kimmel et al., 1987; U.S. EPA, 1985b). Also, maternal and developmental endpoints used in the ratios need to be better defined to permit cross-species comparison. Until such information is available, the applicability of these approaches in risk assessment is not justified.

3.1.1.5. *Short-Term Testing in Developmental Toxicity*

The need for short-term tests for developmental toxicity has arisen from the need to establish testing priorities for 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. These approaches may be useful in making preliminary evaluations of potential developmental toxicity, for evaluating structure-activity relationships, and for assigning priorities for further, more extensive testing. Furthermore, as the risk assessment process begins to incorporate more pharmacokinetic and mechanistic data, short-term tests should be particularly useful. Kimmel (1990) has recently discussed the potential application of in vitro systems in risk assessment in a context that is broader than chemical screening. However, the Agency currently considers a short-term test as "insufficient" by itself to carry out a risk assessment (see Section 3.3).

Although short-term tests for developmental toxicity are not routinely required, such data are encountered in the review of chemicals. Two approaches are considered here in terms of their contribution to the overall testing process: an in vivo mammalian screen and in vitro test systems.

3.1.1.5.1. ***In vivo mammalian developmental toxicity tests.*** The most widely studied in vivo short-term approach is that developed by Chernoff and Kavlock (1982). 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. When originally proposed, the test substance was administered to mice over the period of major organogenesis at a single dose level that would elicit some degree of maternal toxicity. At the NIOSH "Workshop on the Evaluation of the Chernoff/Kavlock Test for Developmental Toxicity" (Hardin, 1987), use of a second lower dose level was encouraged to potentially reduce the chances of false positive results, and the recording of implantation sites was recommended to provide a more precise estimate of postimplantation loss (Kavlock et al., 1987c). In this approach, the pups are counted and weighed shortly after birth, and again after 3-4 days. Endpoints that are considered in the evaluation include: general maternal toxicity (including survival and weight gain), litter size, pup viability and weight, and gross malformations in the offspring. Several schemes have been proposed for ranking the results as a means of prioritizing agents for further testing (Chernoff and Kavlock, 1982; Brown, 1984; Schuler et al., 1984).

The mouse was chosen originally for this test because of its low cost, but the procedure has been applied to the rat as well (Wickramaratne, 1987). The test can predict the potential for developmental toxicity of an agent in the species used while extrapolation of risk to other species, including humans, has the same limitations as for other testing protocols. The EPA Office of Toxic Substances has developed testing guidelines for this procedure (U.S. EPA, 1985c), and the Office of Pesticide Programs has applied similar protocols on a case-by-case basis (U.S. EPA, 1985b). The National Toxicology Program also has developed a protocol that incorporates aspects of a range-finding study, with the intent of providing information on appropriate exposure levels should a standard developmental toxicity study be required (Morrissey et al., 1989). Although testing guidelines are available, such procedures are required on a case-by-case basis. Application of this procedure in the risk assessment process within the Office of Toxic Substances has been described (Francis and Farland, 1987), and the experiences of a number of laboratories are detailed in the proceedings of a NIOSH-sponsored workshop (Hardin, 1987).

Recently, the OECD developed a screening protocol to be used for prioritizing existing chemicals for further testing (draft as of March 22, 1990). This protocol is similar to the design of the Chernoff-Kavlock test except that it involves exposure of male and female rats 2 weeks prior to mating, throughout mating and gestation, and postnatally to day 4. Male animals are exposed following mating for a period corresponding to that of the females. Adult animals are evaluated for general toxicity and effects on reproductive organs. Pups are counted, weighed, and examined for any gross

physical or behavioral abnormalities at birth and on postnatal day 4. This protocol permits evaluation of reproductive and developmental toxicity following repeated dosing with an agent, provides an indication for the need to conduct additional studies, and provides guidance in the design of further studies. Currently, this study design is insufficient by itself to make an estimate of human risk without further studies to confirm and extend the observations.

3.1.1.5.2. ***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. Examples of such systems include isolated whole mammalian embryos in culture, tissue/organ culture, cell culture, and developing nonmammalian organisms. These systems have long been used to assess events associated with normal and abnormal development, but more recently they have been considered for their potential as screens in testing (Wilson, 1978; Kimmel et al., 1982b; Brown and Fabro, 1982). 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 endpoints for toxicity and an understanding of the system's ability to handle various test agents (Kimmel et al., 1982a; Kimmel, 1985; FDA, 1987; Brown, 1987).

While in vitro test systems can provide significant information, they are considered insufficient, by themselves, for carrying out a risk assessment (see Section 3.3). In part, this is due to limitations in the application of the data to the whole-animal situation. But it is also due to the lack of assays that have been fully validated, as has been noted in several reviews of available in vitro systems (FDA, 1987; Brown, 1987; Faustman, 1988) and at a recent workshop on in vitro teratology (Morrissey et al., 1991).

3.1.1.6. *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, and fetuses or pups within litters do not respond independently, the statistical analyses are generally designed to analyze the relevant databased on incidence per litter or on the number of litters with a particular endpoint. The analytical procedures used and the results, as well as an indication of the variance in each endpoint, should be evaluated carefully when reviewing data for risk assessment purposes. Analysis of variance (ANOVA) techniques, with litter nested within dose in the model, take the litter variable into account while allowing 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 (Haseman and Kupper, 1979; Kimmel et al., 1986).

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 to 30 litters per dose group) allow 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 among studies or laboratories, provides added strength in the use of data for the estimation of risk.

An important factor to consider 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 endpoint observed, the variability in the incidence of the endpoint, and the analysis method. As an example, Nelson and Holson (1978) 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 testing protocols being 20 rodents per dose group (U.S. EPA, 1982b, 1985a), the minimum change detectable is an increased incidence of malformations 5 to 12 times above control levels, an increase 3 to 6 times the in utero death rate, and a decrease 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 endpoints measured. Consequently, for statistical reasons only, changes in fetal weight are often observed 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 endpoint(s) evaluated.

Although statistical analyses are important in determining the effects of a particular agent, the biological significance of data is most relevant. It is important to be aware that with the number of endpoints that can be observed in standard protocols for 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 pair-wise comparisons do not indicate a statistically significant effect. This may be true especially for the incidence of malformations or in utero death because of the low power of standard study designs in which 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.

3.1.2. Human Studies

In principle, human data are preferred for risk assessment. However, the complexities of obtaining sufficient human data are such that these data are not available for many potential toxicants. The following describes the methods of generation of human data, their evaluation, and the weight they should be given in risk assessments.

The category of "human studies" includes both epidemiologic studies and other reports of individual cases or clusters of events. Greatest weight should be given to carefully designed epidemiologic studies with more precise measures of exposure, since they can best evaluate exposure-response relationships (see Section 4). Epidemiologic studies in which exposure is presumed based on occupational title or residence (e.g., some case-referent and all ecologic studies) may contribute data to qualitative risk assessments, but are of limited use for quantitative risk assessments because of the generally broad categorical groupings. Reports of individual cases or clusters of events may generate hypotheses of exposure-outcome associations, but require further confirmation with well-designed epidemiologic or laboratory studies. These reports of cases or clusters may give added support to associations suggested by other human or animal data, but cannot stand by themselves in risk assessments. Risk assessors should seek the assistance of professionals trained in epidemiology when conducting a detailed analysis.

3.1.2.1. *Epidemiologic Studies*

Good epidemiologic studies provide the most relevant information for assessing human risk. As there are many different designs for epidemiologic studies, simple rules for their evaluation do not exist.

3.1.2.1.1. ***General design considerations.*** The factors that enhance a study and thus increase its usefulness for risk assessment have been noted in a number of publications (Selevan, 1980; Bloom, 1981; U.S. EPA, 1981; Wilcox, 1983; Sever and Hessol, 1984; Axelson, 1985; Tilley et al., 1985; Kimmel et al., 1986). Some of the more prominent factors are as follows:

(a) The power of the study: The power, or ability of a study to detect a true effect, is dependent on the size of the study group, the frequency of the outcome in the general population, and the level of excess risk to be identified. In a cohort study, common outcomes, such as recognized fetal loss, require hundreds of pregnancies in order to have a high probability of detecting a modest increase

in risk (e.g., 133 in both exposed and unexposed groups to detect a doubling of background; alpha = 0.05, power = 80%), while less common outcomes, such as the total of all malformations recognized at birth, require thousands of pregnancies to have the same probability (e.g., more than 1,200 in both exposed and unexposed groups) (Bloom, 1981; Selevan, 1981; Sever and Hessol, 1984; Selevan, 1985; Stein et al., 1985; Kimmel et al., 1986). In case-referent studies, study sizes are dependent on the frequency of exposure within the source population. The confidence one has in the results of a study without positive findings is related to the power of the study to detect meaningful differences in the endpoints studied.

Power may be enhanced by combining populations from several studies using a meta-analysis (Greenland, 1987). The combined analysis would increase confidence in the absence of risk for agents with negative findings. However, care must be exercised in the combination of potentially dissimilar study groups.

A posteriori determination of power of the actual study may be useful in evaluating contradictory studies in risk assessment. Absence of positive findings in a study of low power would be given less weight than either a positive study or a null study (one with no significant differences) with high power. Positive findings from very small studies are open to question due to the instability of the risk estimates and the potential for highly selected study groups.

(b) Potential bias in data collection: Sources of bias may include selection bias and information bias (Rothman, 1986). Selection bias may occur when an individual's willingness to participate varies with certain characteristics relating to the exposure status or health status of that individual. In addition, selection bias may operate in the identification of subjects for study. For example, in studies of embryonic loss, use of hospital records to identify embryonic or early fetal loss will underascertain events, because women are not always hospitalized for these outcomes. More weight might be given in a risk assessment to a study in which a more complete list of pregnancies is obtained by, for example, collecting biological data [e.g., human chorionic gonadotropin (hCG) measurements] on pregnancy status from study members. These studies may also be affected by bias. The representativeness of these data may be affected by selection factors related to the willingness of different groups of women to continue participation over the total length of the study. Interview data result in more complete ascertainment; however, this strategy carries with it the potential for recall bias, discussed in further detail below. A second example of different levels of ascertainment of events is the use of hospital records to study congenital malformations. Hospital records contain more complete data on malformations than do birth certificates (Mackeprang et al., 1972). Consequently, birth defects registries that are based on searches of hospital records are more complete than those based on vital

records (Selevan, 1986). Thus, a study using hospital records to identify congenital malformations would be given more emphasis in a risk assessment than one using birth certificates.

Studies of working women present the potential for additional bias since some factors that influence employment status may also be associated with reproductive endpoints. For example, because of child-care responsibilities, women may terminate employment, as might women with a history of reproductive problems who wish to have children and are concerned about workplace exposures (Joffe, 1985).

Information bias may result from misclassification of characteristics of individuals or events identified for study. Recall bias, one type of information bias, may occur when respondents with specific exposures or outcomes recall information differently than those without the exposures or outcomes. Interview bias may result when the interviewer knows a priori the category of exposure (for cohort studies) or outcome (for case-referent studies) in which the respondent belongs. Use of highly structured questionnaires and/or "blinding" of the interviewer will reduce the likelihood of such bias. Studies with lower likelihood of the above-listed biases should carry more weight in a risk assessment.

When data are collected by interview or questionnaire, the appropriate respondent depends on the type of data or study. For example, a comparison of husband-wife interviews on reproduction found the wives' responses to questions on pregnancy-related events to be considerably more complete and valid than those of the husbands (Selevan, 1980). A more recent study (Schnatter, 1990) found small, nonsignificant improvements in reporting of birth weights by mothers compared to fathers, and that males who provide early fetal loss data with the aid of their wives give better data (borderline significance). Studies based on interview data from the appropriate respondent(s) would carry more weight than those from proxy respondents (e.g., the specific individual when examining exposure history and the woman or both partners when examining pregnancy history).

Data from any source may be prone to errors or bias. All types of bias are difficult to assess; however, validation with an independent data source (e.g., vital or hospital records) or use of biomarkers of exposure or outcome, where possible, may indicate the degree of bias present and increase confidence in the results of the study. Those studies with a low probability of biased data should carry more weight (Axelson, 1985; Stein and Hatch, 1987).

Differential misclassification, i.e., when certain subgroups are more likely to have misclassified data than others, may either raise or lower the risk estimate. Nondifferential misclassification will bias the results toward a finding of "no effect" (Rothman, 1986).

(c) Collection of data on other risk factors, effect modifiers, and confounders: Risk factors for reproductive and developmental toxicity include such characteristics as age, smoking, alcohol

consumption, drug use, and past reproductive history. Additionally, occupational and environmental exposures are potential risk factors for reproductive and developmental effects. Known and potential risk factors should be examined to identify those that may be effect modifiers or confounders. An effect modifier is a factor that produces different exposure-response relationships at different levels of that factor. For example, maternal age would be an effect modifier if the risk associated with a given exposure increased with the mother's age. A confounder is a variable that is a risk factor for the disease under study and is associated with the exposure under study, but is not a consequence of the exposure. A confounder may distort both the magnitude and direction of the measure of association between the exposure of interest and the outcome. For example, socioeconomic status might be a confounder in a study of the association of smoking and fertility, since socioeconomic status may be associated with both.

Studies that fail to account for effect modifiers and confounders should be given less weight in a risk assessment. Both of these important factors need to be controlled in the study design and/or analysis to improve the estimate of the effects of exposure (Kleinbaum et al., 1982). A more in-depth discussion may be found elsewhere (Epidemiology Workgroup, 1981; Kleinbaum et al., 1982; Rothman, 1986). The statistical techniques used to control for these factors require careful consideration in their application and interpretation (Kleinbaum et al., 1982; Rothman, 1986).

(d) Statistical factors: As in animal studies, pregnancies experienced by the same woman are not independent events (Kissling, 1981; Selevan, 1985). Women who have had embryo/fetal loss are reported to be more likely to have subsequent losses (Leridon, 1977). In animal studies, the litter is generally used as the unit of measure to deal with nonindependence of events. In studies of humans, pregnancies are sequential with the risk factors changing for different pregnancies, making analyses considering nonindependence of events very difficult (Epidemiology Workgroup, 1981; Kissling, 1981). If more than one pregnancy per woman is included, as is often necessary due to small study groups, the use of nonindependent observations overestimates the true size of the groups being compared, thus artificially increasing the probability of reaching statistical significance (Stiratelli et al., 1984). Biased estimates of risk might also result if family size confounds the relationship between exposure and outcome. Some approaches to deal with these issues have been suggested (Kissling, 1981; Stiratelli et al., 1984; Selevan, 1985). At this point in time, a generally accepted solution to this problem has not been developed.

3.1.2.1.2. ***Selection of outcomes for study.*** As already discussed, a number of endpoints can be considered in the evaluation of adverse developmental effects. However, some of the outcomes are not

easily observed in humans, such as early embryonic loss and reproductive capacity of the offspring. Currently, the most feasible endpoints for epidemiologic studies are reproductive history studies of some pregnancy outcomes (e.g., embryo/fetal loss, birth weight, sex ratio, congenital malformations, postnatal function, and neonatal growth and survival) and measures of fertility/infertility, which would include indirect evaluations of very early embryonic loss. Postnatal outcomes for examination could include physical growth and development, organ or system function, and behavioral effects of exposure. Factors requiring control in the design or analysis (such as effect modifiers and confounders) may vary depending on the specific outcomes selected for study.

The developmental outcomes available for epidemiologic examination are limited by a number of factors, including the relative magnitude of the exposure, because differing spectra of outcomes may occur at different exposure levels, different size and demographic characteristics of the population, and different ability to observe the developmental outcome in humans. Improved methods for identifying some outcomes such as very early embryonic loss using new hCG assays may change the spectrum of outcomes available for study (Wilcox et al., 1985; Sweeney et al., 1988).

Demographic characteristics of the population, such as marital status, age distribution, education, socioeconomic status (SES), and prior reproductive history are associated with the probability of whether couples will attempt to have children. Differences in the use of birth control would also affect the number of outcomes available for study. In addition, women with live births are more likely to terminate employment than are those with other outcomes, such as infertility or early embryonic loss. Thus, retrospective studies of female exposure that do not include terminated women workers may be of limited use in risk assessment because the level of risk for these outcomes is likely to be overestimated (Lemasters and Pinney, 1989).

In addition to the above-mentioned factors, developmental endpoints may be envisioned as effects recognized at various points in a continuum, starting at conception through death of the offspring. Thus, a malformed stillbirth would not be included in a study of defects observed at live birth, even though the etiology could be identical (Stein et al., 1975; Bloom, 1981). A shift in the patterns of outcomes could result from differences in timing or in level of exposure (Selevan and LeMasters, 1987).

3.1.2.1.3. ***Reproductive history studies.*** (a) Measures of fertility: Normally, studies of sub- or infertility would not be included in an evaluation of developmental effects. However, in humans it is difficult to identify very early embryonic loss, and distinguish it from sub- or infertility. Thus, studies that examine sub- or infertility indirectly examine loss very early in the gestational period. Infertility or

subfertility may be thought of as a nonevent: a couple is unable to have children within a specific time frame. Therefore, the epidemiologic measurement of reduced fertility is typically indirect, and is accomplished by comparing birth rates or time intervals between births or pregnancies. In these evaluations, the couple's joint ability to procreate is estimated. One method, the Standardized Birth Ratio (SBR; also referred to as the Standardized Fertility Ratio), compares the number of births observed to those expected based on the person-years of observation stratified by factors such as time period, age, race, marital status, parity, contraceptive use, etc. (Wong et al., 1979; Levine et al., 1980, 1981; Levine, 1983; Starr et al., 1986). The SBR is analogous to the Standardized Mortality Ratio (SMR), a measure frequently used in studies of occupational cohorts, and has similar limitations in interpretation (Gaffey, 1976; McMichael, 1976; Tsai and Wen, 1986).

Analysis of the time period between recognized pregnancies or live births has been suggested as another indirect measure of fertility (Dobbins et al., 1978; Baird et al., 1986; Weinberg and Gladen, 1986). Because the time interval between births increases with increasing parity (Leridon, 1977), comparisons within birth order (parity) are more appropriate. A statistical method (Cox regression) can stratify by birth or pregnancy order to help control for nonindependence of these events in the same woman.

Fertility may also be affected by alterations in sexual behavior. However, limited data are available linking toxic exposures to these alterations in humans. Moreover, such data are not easily obtained in epidemiology studies. More information on this subject is available in the proposed male and female reproductive risk assessment guidelines (U.S. EPA, 1988b, 1988c).

(b) Pregnancy outcomes: Pregnancy outcomes examined in human studies of parental exposures may include embryo/fetal loss, congenital malformations, birth weight, sex ratio at birth, and postnatal effects (e.g., physical growth and development, organ or system function, and behavioral effects of exposure). Postnatal effects are discussed in more detail in the next section. As mentioned previously, epidemiologic studies that focus on only one type of pregnancy outcome may miss a true effect of exposure because of the continuum of outcomes. Examination of individual outcomes could mask a true effect due to reduced power resulting from fewer events for study. Studies that examine multiple endpoints could yield more information, but the results may be difficult to interpret.

Evidence of a dose-response relationship is usually an important criterion in the assessment of a toxic exposure. However, traditional dose-response relationships may not always be observed for some endpoints. For example, with increasing dose, a pregnancy might end in a fetal loss rather than a live birth with malformations. A shift in the patterns of outcomes could result from differences either in level of exposure or in timing (Wilson, 1973; Selevan and Lemasters, 1987) (for a more detailed

description, see Section 3.1.2.1.5). Therefore, a risk assessment should, when possible, attempt to look at the interrelationship of different reproductive endpoints and patterns of exposure.

(c) Postnatal developmental effects: These effects may include changes in growth, behavior, organ or system function, or cancer. Studies of neurological and reproductive function are discussed here as examples. Postnatal behavioral and functional effects in humans have been examined for a small number of environmental and occupational agents (e.g., lead, PCBs, methyl mercury, alcohol). For some agents (e.g., lead and PCBs), subtle changes have been observed in groups of children at lower exposures than for other developmental effects (e.g., Bellinger et al., 1987; Needleman, 1988; Davis et al., 1990; Tilson et al., 1990). This may not be true for all toxic agents. These subtle differences would be difficult to identify in individuals, but could result in an overall shifting of mean values when comparing groups of exposed and unexposed children. Some postnatal studies have examined infants or young children using standard developmental scales (e.g., Brazelton Neonatal Behavioral Assessment Scale, Bayley Scales of Infant Development, Stanford Binet IV, and Wechsler Scales) and some biologic measure of exposure (e.g., blood lead levels). These tests are designed to examine certain endpoints and have been developed to cover certain age ranges. Certain tests examine specific aspects of development. For example, the Bayley Scales look at motor and language development, but do not examine sensory function. Batteries of tests are important for a proper evaluation because of the possibility of interrelated effects, e.g., hearing deficits and language development. Thus, batteries of tests will give a clearer indication of direct effects of exposure resulting in postnatal developmental deficits.

Factors that may influence the examination of these effects include parental education, SES, obstetrical history, and health characteristics independent of exposure that may affect functional measurement (e.g., injuries and infections). Many social and lifestyle factors may also affect scoring on these scales (e.g., neonatal-maternal interactions, SES, home environment).

Studies of premature infants carry special problems. For proper comparisons, tests keyed to age in very young children (less than 2.5 years of age) need to "correct" the age for premature infants to the age they would have been had they been born at term. In addition, premature infants or those with low birth weight for their gestational age may have problems resulting from the birth process not directly related to exposure (e.g., intraventricular hemorrhage in the brain which can then cause developmental problems). Thus, the developmental effects resulting from exposure may have their own sequelae.

Other studies may examine effects occurring at a later age (e.g., in utero exposure and cancer in young women). This long time interval typically carries with it the need for retrospective studies, with the inherent limitations in accurate determination of exposure, effect modifiers, and confounders. Risk

assessment methods for cancer are described in the "Guidelines for Carcinogen Risk Assessment" (U.S. EPA, 1986b).

Reproductive effects may result from developmental exposures. For example, environmental exposures may result in oocyte toxicity, in which a loss of primordial oocytes irreversibly affects a woman's fertility. The exposures of importance may occur during both the prenatal period and after birth. Oocyte depletion is difficult to examine directly in women because of the invasiveness of the tests required; however, it can be studied indirectly through evaluation of the age at reproductive senescence (menopause) (Everson et al., 1986). Risk assessment methods for female reproductive effects are described in the "Proposed Guidelines for Assessing Female Reproductive Risk" (U.S. EPA, 1988c).

Developmental exposures to males could affect their reproductive function (e.g., deplete stem or Sertoli cells potentially affecting sperm production) (Zenick and Clegg, 1989). If stem cell death occurs with exposure at any age, recovery is possible as long as some stem cells survive. The same is true for Sertoli cells, except that they cease multiplication before puberty. Thus, cell replication cannot compensate for Sertoli cell death after puberty. Human studies of stem and Sertoli cells would be difficult due to the invasiveness of the measure. Less direct measures, e.g., sperm count, morphology, and motility, could be evaluated, but this would not indicate what cells or stage of spermatogenesis had been affected. Risk assessment methods for male reproductive effects are described in the "Proposed Guidelines for Assessing Male Reproductive Risk" (U.S. EPA, 1988b).

In addition to the above effects, genetic damage to germ cells may result from developmental exposures. Outcomes resulting from germ-cell mutations could include reduced probability of conception as well as increased probability of embryo/fetal loss and other developmental effects. These endpoints could be studied using the approaches described above. However, a human germ-cell mutagen has not yet been demonstrated (U.S. EPA, 1986c). Based on animal studies, critical exposures are to germ cells or early zygotes. Germ-cell mutagenicity could also be expressed as genetic diseases in future generations. Unfortunately, these studies would be very difficult to conduct in human populations because of the long time lag between exposure and outcome. For more information, refer to the "Guidelines for Mutagenicity Risk Assessment" (U.S. EPA, 1986c).

3.1.2.1.4. ***Community studies/surveillance programs.*** Epidemiologic studies may also be based on broad populations such as a community, a nationwide probability sample, or surveillance programs (such as birth defects registries). Other studies have examined environmental exposures, such as toxic agents in the water system, and adverse pregnancy outcome (Swan et al., 1989; Deane et al., 1989). Unfortunately, in these studies maternally mediated effects may be difficult to distinguish from paternally

mediated effects. In addition, the presumably lower exposure levels (compared to industrial settings) may require very large groups for study. A number of case-referent studies have examined the relationship between broad classes of parental occupation in certain communities or countries and embryo/fetal loss (Silverman et al., 1985), birth defects (Hemminki et al., 1980; Kwa and Fine, 1980; Papier, 1985), and childhood cancer (Kwa and Fine, 1980; Zack et al., 1980; Hemminki et al., 1981; Peters et al., 1981). In these reports, jobs are typically classified into broad categories based on the probability of exposure to certain classes or levels of exposure (e.g., Kwa and Fine, 1980). Such studies are most helpful in the identification of topics for additional study. However, because of the broad groupings of types or levels of exposure, such studies are not typically useful for risk assessment of a particular agent.

Surveillance programs may also exist in occupational settings. In this case, reproductive histories and/or clinical evaluations could be followed to monitor for reproductive effects of exposures. Both could yield very useful data for risk assessment; however, a clinical evaluation program would be costly to maintain, and there are numerous impediments to the collection of reliable and valid information in the workplace. These might include concerns similar to those previously discussed plus potentially low participation rates due to employee sensitivities and confidentiality concerns.

3.1.2.1.5. ***Identification of exposures important for developmental effects.*** For all examinations of the relationship between developmental effects and potentially toxic exposures, the identification of the appropriate exposure is crucial. Preconceptional exposures to either parent and in utero exposures have been associated with the more commonly examined outcomes (e.g., fetal loss, malformations, birth weight, and measures of infertility). These exposures, plus postnatal exposure from breast milk, food, and the general environment, may be associated with postnatal developmental effects (e.g., changes in behavioral and cognitive function, or growth). The magnitude of exposure may affect the spectrum of outcomes observed. This issue is discussed in more detail in Sections 3.1.1.2 and 3.2.

Infants and young children may receive disproportionate levels of exposure due to their tendency to "put everything" in their mouths (pica) and the greater time they spend on the floor. Carpets may serve as a reservoir for toxic agents (e.g., pesticides and lead dust), and the air nearer the floor may have greater levels of certain airborne toxicants (e.g., mercury from latex paints).

Exposures in environmental settings are frequently lower than in industrial and agricultural settings. However, this relationship may change as exposures are reduced in workplaces, and as more is learned about environmental exposures (e.g., indoor air exposures, pesticides usage). Larger populations are necessary in settings with lower exposures (Lemasters and Selevan, 1984). Other

factors affect the identification of reproductive or developmental events with various levels of exposure. Exposed individuals may move in and out of areas with differing levels and types of exposures, affecting the number of exposed and comparison events for study. Thus, exposures can be short-term or chronic.

Data on exposure from human studies are frequently qualitative, such as employment or residence histories. More quantitative data may be difficult to obtain due to the nature of certain study designs (e.g., retrospective studies) and historical limitations in exposure measurements. Many developmental outcomes result from exposures during certain critical times. The appropriate exposure classification depends on the outcome(s) studied, the biologic mechanism affected by exposure, and the biologic half-life of the agent. The biologic half-life, in combination with the patterns of exposure (e.g., continuous or intermittent) affect the individual's body burden and consequently the "true" dose during the critical period. The probability of misclassification of exposure status may affect the ability to recognize a true effect in a study (Selevan, 1981; Hogue, 1984; Lemasters and Selevan, 1984; Sever and Hessol, 1984; Kimmel et al., 1986). As more prospective studies are done, better estimates of exposure will be developed.

3.1.2.2. *Examination of Clusters or Case Reports/Series*

The identification of cases or clusters of adverse pregnancy outcomes is generally limited to those identified by the women involved, or clinically by their physicians. Examples of outcomes more easily identified include mid-to-late fetal loss or congenital malformations. Identification of other effects, such as very early embryonic loss, may be difficult to separate from the study of sub- or infertility. Such "nonevents" (e.g., lack of pregnancies or children) are much harder to recognize than are developmental effects such as malformations resulting from in utero exposure. While case reports have been important in the recognition of some agents that cause developmental toxicity, they may be of greatest use in suggesting topics for further investigation (Hogue, 1985). Reports of clusters and case reports/series are best used in risk assessment in conjunction with strong laboratory data to suggest that effects observed in animals also occur in humans. Previous discussion of the use of human data should be taken into account wherever possible.

3.1.3. Other Considerations

Several other types of information may be considered in the evaluation and interpretation of human and animal data. Information on pharmacokinetics and structure-activity relationships may be very useful, but is often lacking for developmental toxicity risk assessments.

3.1.3.1. *Pharmacokinetics*

Extrapolation of toxicity data between species can be aided considerably by the availability of data on the pharmacokinetics of a particular agent in the species tested and, when available, in humans. Information on absorption, half-life, steady-state and/or peak plasma concentrations, placental metabolism and transfer, excretion in breast milk, comparative metabolism, and concentrations of the parent compound and metabolites 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 (Wilson et al., 1975, 1977), 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 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 (Kimmel and Young, 1983).

While human pharmacokinetic data are often lacking, absorption data in laboratory animals for studies conducted by any relevant route of exposure may assist in the interpretation of the developmental toxicity studies in the animal models for the purposes of risk assessment. Specific guidance regarding both the development and application of pharmacokinetic data was agreed upon by the participants at the "Workshop on the Acceptability and Interpretation of Dermal Developmental Toxicity Studies" (Kimmel and Francis, 1990). It was concluded that absorption data are needed both when a dermal developmental toxicity study shows no developmental effects and when developmental effects are seen. The results of a dermal developmental toxicity study showing no adverse developmental effects and without blood level data (as evidence of dermal absorption) are potentially misleading and would be insufficient for risk assessment, especially if interpreted as a "negative" study. In studies where developmental toxicity is detected, regardless of the route of exposure, absorption data can be used to establish the internal dose in maternal animals for risk extrapolation purposes.

3.1.3.2. *Comparisons of Molecular Structure*

Comparisons of the chemical or physical properties of an agent with those known to cause developmental toxicity may indicate 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.

3.2. DOSE-RESPONSE EVALUATION

The evaluation of dose-response relationships for developmental toxicity includes the evaluation of data from both human and animal studies. When quantitative dose-response data are available in humans and with sufficient range of exposure, dose-response relationships may be examined. Since data on human dose-response relationships have been available infrequently, the dose-response evaluation is usually based on the assessment of data from tests performed in laboratory animals.

Evidence for a dose-response relationship is an important criterion in the assessment of developmental toxicity, which is usually based on limited data from standard studies using three dose groups and a control group. Most agents causing developmental toxicity in humans alter development at doses within a narrow range near the lowest maternally toxic dose (Kimmel et al., 1984). Therefore, for most agents, the exposure situations of concern will be those that are potentially near the maternally toxic dose range. For those few agents that produce developmental effects at much lower levels than maternal effects, the potential for exposing the conceptus to damaging doses is much greater than when the maternal and developmental toxic doses are similar. As mentioned previously (Section 3.1.1.2), however, traditional dose-response relationships may not always be observed for some endpoints. For example, as exposure increases, embryolethal levels may be reached, resulting in an observed decrease in malformations with increasing dose (Wilson, 1973; Selevan and LeMasters, 1987). The potential for this response pattern indicates that dose-response relationships of individual endpoints as well as combinations of endpoints (e.g., dead and malformed combined) must be carefully examined and interpreted.

The evaluation of dose-response relationships includes the identification of effective dose levels as well as doses that are associated with no increased incidence of adverse effects when compared with controls. Much of the focus is on the identification of the critical effect(s) (i.e., the adverse effect(s) observed at the lowest dose level) and the LOAEL and NOAEL associated with that developmental effect, which may be any of the four manifestations of developmental toxicity. The NOAEL is defined as the highest dose at which there is no statistically or biologically significant increase in the frequency of an adverse effect in any of the possible manifestations of developmental toxicity when compared with the appropriate control group in a data base characterized as having sufficient evidence for use in a risk assessment (see Section 3.3). The LOAEL is the lowest dose at

which there is a statistically or biologically significant increase in the frequency of adverse developmental effects when compared with the appropriate control group in a database characterized as having sufficient evidence. Although a threshold is assumed for developmental effects, the existence of a NOAEL in an animal study does not prove or disprove the existence or level of a biological threshold; it only defines the highest level of exposure under the conditions of the study that is not associated with a significant increase in adverse effects.

Several limitations in the use of the NOAEL have been described (Gaylor, 1983; Crump, 1984; Kimmel and Gaylor, 1988; Gaylor, 1989; Brown and Erdreich, 1989, Kimmel, 1990): (1) Use of the NOAEL focuses only on the dose that is the NOAEL, and does not incorporate information on the slope of the dose-response curve or the variability in the data. (2) Since data variability is not taken into account (i.e., confidence limits are not used), the NOAEL will likely be higher with decreasing sample size or poor study conduct, either of which is usually associated with increasing variability in the data. (3) The NOAEL is limited to one of the experimental doses. (4) The number and spacing of doses in a study influence the dose chosen for the NOAEL. (5) Since the NOAEL is defined as a dose that does not produce an observed increase in adverse responses from control levels and is dependent on the power of the study, theoretically, the risk associated with it may fall anywhere between zero and an incidence just below that detectable from control levels (usually in the range of 7% to 10% for quantal data). Crump (1984) and Gaylor (1989) have estimated the upper confidence limit on risk at the NOAEL to be 2% to 6% for specific developmental endpoints from several data sets.

Because of the limitations associated with the use of the NOAEL (Kimmel and Gaylor, 1988; Gaylor, 1989; Kimmel, 1990), the Agency is evaluating the use of an additional approach for more quantitative dose-response evaluation when sufficient data are available, i.e., the benchmark dose (Crump, 1984). The benchmark dose is based on a model-derived estimate of a particular incidence level, such as 10% incidence. More specifically, the benchmark dose (BD) is derived by modeling the data in the observed range, selecting an incidence level within or near the observed range (e.g., the effective dose to produce a 10% increased incidence of response, the ED_{10}), and determining the upper confidence limit on the model. The upper confidence value corresponding to, for example, a 10% excess in response is used to derive the BD, which is the lower confidence limit on dose for that level of excess response, in this case the LED_{10} (see Figure 1).

Various mathematical approaches have been proposed for deriving the benchmark dose for developmental toxicity data (e.g., Crump, 1984; Rai and Van Ryzin, 1985; Kimmel and Gaylor, 1988; Faustman et al., 1989; Chen and Kodell, 1989; Kodell et al., 1991). Such models may be used to calculate the benchmark dose, and the particular model used may be less critical since estimation of the

benchmark dose is limited to the observed dose range. Since the model is only used to fit the observed data, the assumptions about the existence or nonexistence of a threshold are not as pertinent. Thus, models that fit the empirical data may well provide a reasonable estimate of the benchmark dose, although biological factors known to influence data should be incorporated into the model [e.g., intralitter correlations, correlations among endpoints (Ryan et al., 1991)]. The Agency is currently conducting studies to evaluate the application of several models to actual data sets for calculating the benchmark dose, to determine the minimum data required for modeling, and to develop methods for application to continuous data. In addition, information from these studies will be used to develop guidance for application of the benchmark dose approach to the calculation of the RfD_{DT} or the RfC_{DT}, since the Agency has limited experience with this approach (see Section 3.4 for a discussion of the RfD_{DT} and RfC_{DT}).

Using the benchmark dose approach, an LED_{10} can be calculated for each effect of an agent for which there is a database with sufficient evidence to conduct a risk assessment. In some cases, the data may be sufficient to also estimate the ED_{05} or ED_{01}, which should be closer

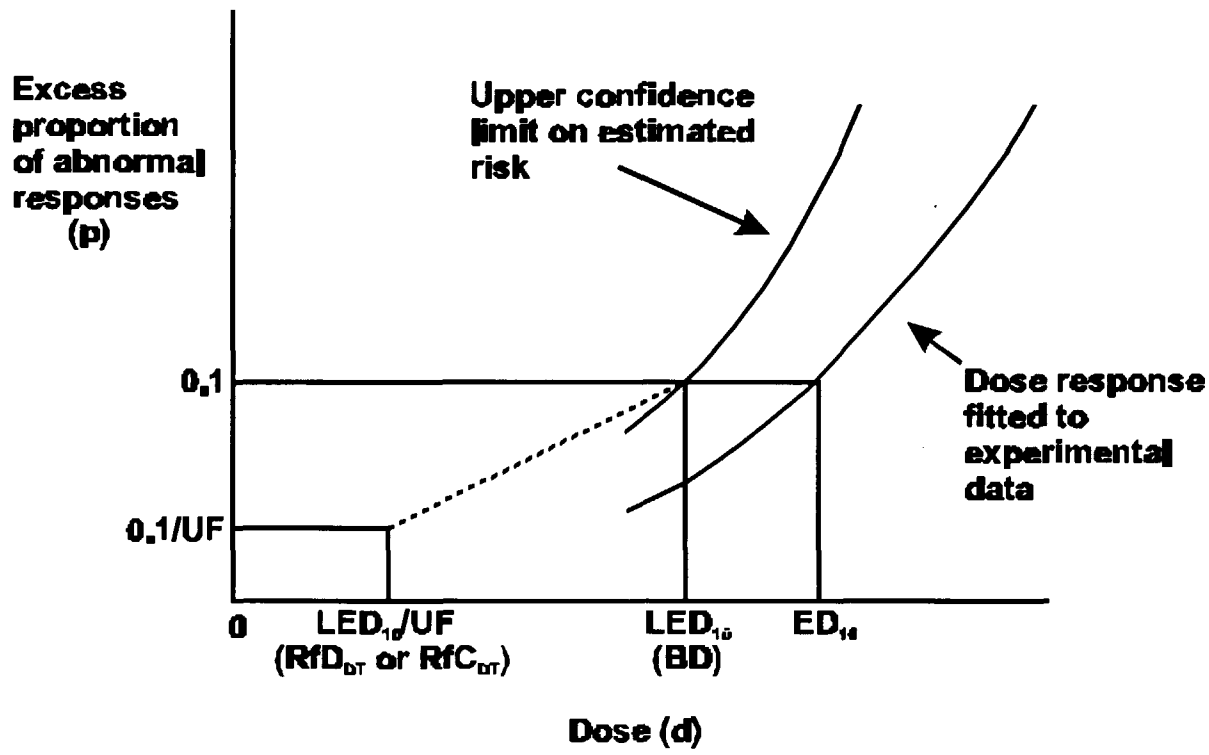

Figure 1. This graphical illustration of the benchmark dose approach is based on Crump (1984) and Kimmel and Gaylor (1988). The benchmark dose (BD) is derived by modeling the data in the observed range, selecting an incidence level within or near the observed range (e.g., the effective dose to produce a 10% increased incidence of response, the ED_{10}), and determining the upper confidence limit on the model. The upper confidence value corresponding to, for example, a 10% excess in response is used to derive the BD, which is the lower confidence limit on dose for that level of excess response, in this case the LED_{10}. The RfD_{DT} or RfC_{DT} estimated by applying uncertainty factors (UF) to the BD would be greater than or equal to the BD/UF.

to a true no-effect dose. A level between the ED_{01} and the ED_{10} usually corresponds to the lowest level of risk that can be estimated for binomial endpoints from standard developmental toxicity studies

Certain principles are especially applicable for determining the NOAEL, LOAEL, and benchmark dose for developmental toxicity studies. First, the NOAEL, LOAEL, or benchmark dose are identified for both developmental and maternal or adult toxicity, based on the information available from studies in which developmental toxicity has been evaluated. The NOAEL, LOAEL, or benchmark dose for maternal or adult toxicity should be compared with the corresponding values from other adult toxicity data to determine if the pregnant or lactating female or the paternal animal (if exposure is prior to mating) may be more sensitive to an agent than adult males or nonpregnant females in other toxicity studies that generally involve longer exposure times.

Second, for developmental toxic effects, a primary assumption is that a single exposure at a critical time in development may produce an adverse developmental effect, i.e., repeated exposure is not a necessary prerequisite for developmental toxicity to be manifested. In most cases, however, the data available for developmental toxicity risk assessment are from studies using exposures over several days of development, and the NOAEL, LOAEL, and/or benchmark dose is most often based on a daily dose, e.g., mg/kg-day. Usually, the daily dose is not adjusted for duration of exposure because appropriate pharmacokinetic data are not available. In cases where such data are available, adjustments may be made to provide an estimate of equal average concentration at the site of action for the human exposure scenario of concern. For example, inhalation studies often use 6 hr/day exposures during development. If the human exposure scenario is continuous and pharmacokinetic data indicate an accumulation with continuous exposure, appropriate adjustments can be made. If, on the other hand, the human exposure scenario of concern is very brief or intermittent, pharmacokinetic data indicating a long half-life may also require adjustment of dose. When quantitative absorption data by any route of exposure are available, the NOAEL may be adjusted accordingly; e.g., absorption of 50% of administered dose could result in a 50% reduction in the NOAEL. If absorption in the experimental species has been determined, but human absorption is not known, human absorption is generally assumed to be the same as that for the species with the greatest degree of absorption. NOAELs from inhalation exposure studies are adjusted to derive a human equivalent concentration (HEC) by taking into account known anatomical and physiological species differences (e.g., minute volume, respiratory rate, etc.) (U.S. EPA, 1991b).

In summary, the dose-response evaluation identifies the NOAEL, LOAEL, or benchmark dose, defines the range of doses for a given agent that are effective in producing developmental and maternal toxicity; the route, timing, and duration of exposure; species specificity of effects, and any

pharmacokinetic or other considerations that might influence the comparison with human exposure scenarios. This information should always accompany the characterization of the health-related database (discussed in the next section).

3.3. CHARACTERIZATION OF THE HEALTH-RELATED DATABASE

This section describes the process for evaluating the health-related database as a whole on a particular agent and provides criteria for characterizing the evidence for judging a potential developmental hazard in humans within the context of expected exposure or dose. This determination provides the basis for judging whether or not there are sufficient data for proceeding further in the risk assessment process. This section does not address the nature and magnitude of human health risks, which are discussed as part of the final characterization of risk along with estimates of potential human exposure and the relevancy of available data for estimating human risk. Characterization of hazard potential within the context of exposure or dose should assist the risk assessor in clarifying the strengths and uncertainties associated with a particular database. Because a complex interrelationship exists among study design, statistical analysis, and biological significance of the data, a great deal of scientific judgment, based on experience with developmental toxicity data and with the principles of study design and statistical analysis, may be required to adequately evaluate the database. Scientific judgment is always necessary, and in many cases, interaction with scientists in specific disciplines (e.g., developmental toxicology, epidemiology, statistics) is recommended.

A categorization scheme for characterizing the evidence for developmental toxicity is presented in Table 3. The categorization scheme contains two broad categories, sufficient evidence and insufficient evidence, which are defined in the table. Data from all available studies, whether indicative of potential hazard or not, must be evaluated and factored into a judgment as to the strength of evidence available to support a complete risk assessment for developmental toxicity. The primary considerations are the human data, if available, and the experimental animal data. The judgment of whether the data are sufficient or insufficient should consider quality of the data, power of the studies, number and types of endpoints examined, replication of effects, relevance of the test species to humans, relevance of route and timing of exposure for both human and animal studies, appropriateness of the dose selection in animal studies, and number of species examined. In addition, pharmacokinetic data and structure-activity considerations, data from other toxicity studies, as well as other factors that may affect the strength of the evidence, should be taken into account.

In general, the categorization is based on criteria that define the minimum evidence necessary to conduct a hazard identification/dose-response evaluation. Establishing the

Table 3. Categorization of the health-related database for hazard identification/dose-response evaluation

SUFFICIENT EVIDENCE

The sufficient evidence category includes data that collectively provide enough information to judge whether or not a human developmental hazard could exist within the context of dose, duration, timing, and route of exposure. This category includes both human and experimental animal evidence.

Sufficient Human Evidence: This category includes data from epidemiologic studies (e.g., case control and cohort) that provide convincing evidence for the scientific community to judge that a causal relationship is or is not supported. A case series in conjunction with strong supporting evidence may also be used. Supporting animal data may or may not be available.

Sufficient Experimental Animal Evidence/Limited Human Data: This category includes data from experimental animal studies and/or limited human data that provide convincing evidence for the scientific community to judge if the potential for developmental toxicity exists. The minimum evidence necessary to judge that a potential hazard exists generally would be data demonstrating an adverse developmental effect in a single, appropriate, well-conducted study in a single experimental animal species. The minimum evidence needed to judge that a potential hazard does not exist would include data from appropriate, well-conducted laboratory animal studies in several species (at least two) which evaluated a variety of the potential manifestations of developmental toxicity and showed no developmental effects at doses that were minimally toxic to the adult.

INSUFFICIENT EVIDENCE

This category includes situations for which there is less than the minimum sufficient evidence necessary for assessing the potential for developmental toxicity, such as when no data are available on developmental toxicity, as well as for databases from studies in animals or humans that have a limited study design (e.g., small numbers, inappropriate dose selection/exposure information, other uncontrolled factors), or data from a single species reported to have no adverse developmental effects, or databases limited to information on structure/activity relationships, short-term tests, pharmacokinetics, or metabolic precursors.

minimum sufficient human evidence necessary to do a hazard identification/dose-response evaluation is difficult, since there are often considerable variations in study designs and study group selection. The body of human data should contain convincing evidence as described in the "Sufficient Human Evidence" category. Because the human data necessary to judge whether or not a causal relationship exists are generally limited, there are currently few agents that can be classified in this category. In the case of animal data, agents that have been tested adequately in laboratory animals according to current test guidelines generally would be included in the "Sufficient Experimental Animal Evidence/Limited Human Data" category. The strength of evidence for a database increases with replication of the findings and with additional animal species tested. Information on pharmacokinetics or mechanisms, or on more than one route of exposure may reduce uncertainties in extrapolation to the human.

More evidence is necessary to judge that an agent is unlikely to pose a hazard for developmental toxicity than that required to judge a potential hazard. This is because it is more difficult, both biologically and statistically, to support a finding of no apparent adverse effect than a finding of an adverse effect. For example, to judge that a hazard for developmental toxicity could exist for a given agent, the minimum evidence necessary would be data from a single, appropriate, well-executed study in a single experimental animal species that demonstrate developmental toxicity, and/or suggestive evidence from adequately conducted clinical/epidemiologic studies. On the other hand, to judge that an agent is unlikely to pose a hazard for developmental toxicity, the minimum evidence would include data from appropriate, well-executed laboratory animal studies in several species (at least two) which evaluated a variety of the potential manifestations of developmental toxicity and showed no adverse developmental effects at doses that were minimally toxic to the adult animal. In addition, there may be human data from appropriate studies supportive of no adverse developmental effects.

If a database on a particular agent includes less than the minimum sufficient evidence (as defined in the "Insufficient Evidence" category) necessary for a risk assessment, but some data are available, this information could be used to determine the need for additional testing. In the event that a substantial database exists for a given chemical, but no single study meets current test guidelines, the risk assessor should use scientific judgment to determine whether the composite database may be viewed as meeting the "Sufficient Evidence" criteria. In some cases, a database may contain conflicting data. In these instances, the risk assessor must consider each study's strengths and weaknesses within the context of the overall database in an attempt to define the strength of evidence of the database for assessing the potential for developmental toxicity.

Judging that the health-related database is sufficient to indicate a potential developmental hazard does not mean that the agent will be a hazard at every exposure level (because of the assumption of a

threshold) or in every situation (e.g., hazard may vary significantly depending on route and timing of exposure). In the final risk characterization, the characterization of the health-related database should always be presented with information on the dose-response evaluation (e.g., LOAEL, NOAEL, and/or benchmark dose), exposure route, timing and duration of exposure, and with the human exposure estimate.

3.4. DETERMINATION OF THE REFERENCE DOSE (RfD_{DT}) OR REFERENCE CONCENTRATION (RfC_{DT}) FOR DEVELOPMENTAL TOXICITY

The RfD_{DT} or RfC_{DT} is an estimate of a daily exposure to the human population that is assumed to be without appreciable risk of deleterious developmental effects. The use of the subscript DT is intended to distinguish these terms from the reference dose (RfD) for oral or dermal exposure or the reference concentration (RfC) for inhalation exposure, terms that refer primarily to chronic exposure situations (U.S. EPA, 1991b). The RfD_{DT} or RfC_{DT} is derived by applying uncertainty factors to the NOAEL (or the LOAEL, if a NOAEL is not available), or the benchmark dose. To date, the Agency has applied uncertainty factors only to the NOAEL or LOAEL to derive an RfD_{DT} or RfC_{DT}. The Agency is planning eventually to use the benchmark dose approach as the basis for derivation of the RfD_{DT} or RfC_{DT} and will develop guidance as information is acquired and analyzed from ongoing Agency studies.

The most sensitive developmental effect (i.e., the critical effect) from the most appropriate and/or sensitive mammalian species is used for determining the NOAEL, LOAEL, or the benchmark dose in deriving the RfD_{DT} or RfC_{DT} (Section 3.2). Uncertainty factors (UFs) for developmental and maternal toxicity applied to the NOAEL generally include a 10-fold factor for interspecies variation and a 10-fold factor for intraspecies variation. In general, an uncertainty factor is not applied to account for duration of exposure.

Additional factors may be applied to account for other uncertainties or additional information that may exist in the database. For example, the standard study design for a developmental toxicity study calls for a low dose that demonstrates a NOAEL, but in some cases, the lowest dose administered may cause significant adverse effect(s), and thus be identified as the LOAEL. In circumstances where only a LOAEL is available, the use of an additional uncertainty factor of up to 10 may be required, depending on the sensitivity of the endpoints evaluated, adequacy of dose levels tested, or general confidence in the LOAEL. In addition, if a benchmark dose has been calculated, it may be used to help interpret how close the LOAEL is to a level that would not be detectable from controls (equivalent to the NOAEL), and thus the size of the uncertainty factor to be applied. Other

modifying factors (MFs) may be used depending on the characterization of the database (Section 3.3), data on pharmacokinetics, or other considerations that may alter the level of confidence in the data (U.S. EPA, 1991b). The total size of the uncertainty factor will vary from agent to agent and will require the exercise of scientific judgment, taking into account interspecies differences, variability within species, the slope of the dose-response curve, the background incidence of the effects, the route of administration, and pharmacokinetic data.

As stated above, there is little experience with the application of uncertainty factors to the benchmark dose approach for calculating the RfD_{DT} or RfC_{DT}, and there are several issues that must be addressed prior to its use for this purpose. For example, which benchmark dose (e.g., LED_{01}, LED_{05}, LED_{10}) should be used for calculating the RfD_{DT} or RfC_{DT}, and what are the appropriate uncertainty factors that should be applied to the benchmark dose for deriving the RfD_{DT} or RfC_{DT}? That is, should the uncertainty factor applied to an LED_{10} be similar to that applied to a LOAEL, or should the uncertainty factor applied to an LED_{01} be equal to or less than that applied to a NOAEL? These and other questions are being addressed in ongoing Agency studies on the calculation of the RfD_{DT} or RfC_{DT} using the benchmark dose approach. As results become available, and as further guidance is developed, this information will be published as a supplement to these Guidelines.

The total uncertainty factor selected is divided into the NOAEL or LOAEL (or the benchmark dose) for the critical effect in the most appropriate and/or sensitive mammalian species to determine the RfD_{DT} or RfC_{DT}. If the NOAEL, LOAEL, or benchmark dose for maternal toxicity is lower than that for developmental toxicity, this should be noted in the risk characterization, and this value compared with data from other studies in which adult animals are exposed.

The modeling approaches that have been proposed for developmental toxicity are, for the most part, statistical probability models that do not take into account underlying biological processes or mechanisms (e.g., Crump, 1984; Rai and Van Ryzin, 1985; Kimmel and Gaylor, 1988; Faustman et al., 1989; Chen and Kodell, 1989; Kodell et al., 1991). These models can be applied to derive dose-response curves for data in the observed dose range, but may or may not accurately predict risk at low levels of exposure. It has generally been assumed that there is a biological threshold for developmental toxicity; however, a threshold for a population of individuals may or may not exist because of other endogenous or exogenous factors that may increase the sensitivity of some individuals in the population. Thus, the addition of a toxicant may result in an increased risk for the population, but not necessarily for all individuals in the population.

Models that are more biologically based should provide a more accurate estimation of low-dose risk to humans. The development of biologically based dose-response models in developmental

toxicology has been limited by a number of factors, including a lack of understanding of the biological mechanisms underlying developmental toxicity, intra/interspecies differences in the types of developmental events, appropriate pharmacokinetic data, and the influence of maternal effects on the dose-response curve. The Agency is currently supporting several major research efforts to develop biologically based dose-response models for developmental toxicity risk assessment that include the consideration of threshold under its Research to Improve Health Risk Assessment program.

3.5. SUMMARY

In summary, the hazard identification/dose-response evaluation of developmental toxicity data is used as part of the final characterization of risk along with information on estimates of human exposure. This analysis depends on scientific judgment as to the accuracy and sufficiency of the health-related data, biological relevance of significant effects, the conditions of human exposure, and other considerations important in the extrapolation of data from animals to humans. Scientific judgment is always necessary, and in many cases, interaction with scientists in specific disciplines (e.g., developmental toxicology, epidemiology, statistics) is recommended.

4. EXPOSURE ASSESSMENT

In order to obtain quantitative estimates of risk for human populations, estimates of human exposure are required. This discussion is not intended to provide definitive guidance on exposure assessment; the "Guidelines for Estimating Exposures" have been published separately (U.S. EPA, 1986d) and will not be discussed in detail here. Rather, the issues important to developmental toxicity risk assessment are addressed. In general, the exposure assessment describes the magnitude, duration, frequency, and route(s) of exposure. This information is usually developed from monitoring data and from estimates based on various scenarios of environmental exposures.

There are several exposure considerations that are unique for developmental toxicity. For example, exposure to developing individuals is often secondary via placental transfer or through breast milk. Thus, exposure to the embryo/fetus or child may not be the same as for the pregnant or lactating mother, and measurements of an agent in maternal or cord blood and in breast milk may provide a better estimate of developmental exposure. Direct exposure of neonates and children may also occur via environmental media such as water, air and soil, and thus may require estimates of exposure from multiple sources. Duration and period of exposure also must be related to stage of development, if possible (e.g., first, second, or third trimester of pregnancy, infancy, early, middle, and late childhood, adolescence, etc.). These stages of development may have different sensitivities to agents, and exposure estimates should be derived for as many as possible. In addition, exposure to either parent prior to conception must be considered in relation to adverse developmental effects.

There is also a possibility that a single exposure may be sufficient to produce adverse developmental effects (i.e., repeated exposure is not a necessary prerequisite for developmental toxicity to be manifested, although it should be considered in cases where there is evidence of cumulative exposure or where the half-life of the agent is sufficiently long to produce an increasing body burden over time). Therefore, it is assumed that, in most cases, a single exposure at any of several developmental stages may be sufficient to produce an adverse developmental effect. Most of the data available for risk assessment involve exposures over several days of development. Thus, human exposure estimates used to calculate margins of exposure (MOE, see following section) or to compare with the RfD_{DT} or RfC_{DT} are usually based on a daily dose that is not adjusted for duration or pattern of exposure. For example, it would be inappropriate in developmental toxicity risk assessments to use time-weighted averages or adjustment of exposure over a different time frame than that actually encountered (such as the adjustment of a 6-hr inhalation exposure to account for a 24-hr exposure scenario), unless pharmacokinetic data were available to indicate an accumulation with continuous

exposure. In the case of intermittent exposures, examination of the peak exposure(s), as well as the average exposure over the time period of exposure, would be important.

It should be recognized that, based on the definition used in these Guidelines for developmental toxicity, exposure of almost any segment of the human population may lead to risk to the developing organism. This would include fertile men and women, the developing embryo and fetus, and children up to the age of sexual maturation. Although some effects of developmental exposures may be manifested while the exposure is occurring (e.g., spontaneous abortion, structural abnormality present at birth, childhood mental retardation), some effects may not be detectable until later in life, long after exposure has ceased (e.g., perinatally induced carcinogenesis, impaired reproductive function, shortened lifespan).

5. RISK CHARACTERIZATION

5.1. OVERVIEW

Risk characterization is the culmination of the risk assessment process. In this final step, risk characterization involves integration of the toxicity information from the hazard identification/dose-response evaluation with the human exposure estimates and provides an evaluation of the overall quality of the assessment, describes risk in terms of the nature and extent of harm, and communicates the results of the risk assessment to a risk manager. The risk manager can then use the risk assessment, along with other risk management elements, to make public health decisions. The following sections describe these three aspects of the risk characterization in more detail, but do not attempt to provide a full discussion of risk characterization. Rather, these Guidelines point out issues that are important to risk characterization for developmental toxicity.

5.2. INTEGRATION OF THE HAZARD IDENTIFICATION/DOSE-RESPONSE EVALUATION AND EXPOSURE ASSESSMENT

In developing the hazard identification/dose-response and exposure portions of the risk assessment, the risk assessor makes many judgments concerning human relevance of the toxicity data, including the appropriateness of the various animal models for which data are available, the route, timing, and duration of exposure relative to expected human exposure, etc. These judgments should be summarized at each stage of the risk assessment process (e.g., the biological relevance of anatomical variations may be made in the hazard identification process, or species differences in metabolic patterns in the dose-response evaluation). When data are not available to make such judgments, as is often the case, the background information and assumptions discussed in the Introduction (Section 1) provide a default position. The risk assessor must determine if some of these judgments have implications for other portions of the assessment, and whether the various components of the assessment are compatible.

The description of the relevant data should convey the major strengths and weaknesses of the assessment that arise from availability of data and the current limits of understanding of the mechanisms of toxicity. Confidence in the results of a risk assessment is a function of confidence in the results of the analysis of these elements. Each of these elements should have its own characterization as a part of it. Interpretation of data should be explained, and the risk manager should be given a clear picture of consensus or lack of consensus that exists about significant aspects of the assessment. Whenever more than one view is supported by the data and choosing between them is difficult, both views should be

presented. If one has been selected over another, the rationale should be given; if not, then both should be presented as plausible alternative results.

The risk characterization should not only examine the judgments, but also explain the constraints of available data and the state of knowledge about the phenomena studied in making them, including:

- the qualitative conclusions about the likelihood that the agent may pose a specific hazard to human health, the nature of the observed effects, under what conditions (route, dose levels, time, and duration) of exposure these effects occur, and whether the health-related data are sufficient to use in a risk assessment;
- a discussion of the dose-response patterns for the critical effect(s), data such as the shapes and slopes of the dose-response curves for the various endpoints, the rationale behind the determination of the NOAEL, LOAEL, and/or calculation of the benchmark dose, and the assumptions underlying the estimation of the RfD_{DT} or RfC_{DT}; and
- the estimates of the magnitude of human exposure, the route, duration, and pattern of the exposure, relevant pharmacokinetics, and the size and characteristics of the populations exposed.

The risk characterization of an agent should be based on data from the most appropriate species, or, if such information is not available, on the most sensitive species tested. It should also be based on the most sensitive indicator of toxicity, whether maternal, paternal, or developmental, when such data are available, and should be considered in relationship to other forms of toxicity.

If data used in characterizing risk are from a route of exposure other than the expected human exposure, then pharmacokinetic data should be used, if available, to extrapolate across routes of exposure. If such data are not available, the Agency makes certain assumptions concerning the amount of absorption likely or the applicability of the data from one route to another (U.S. EPA, 1984, 1985b).

The level of confidence in the hazard identification/dose-response evaluation should be stated to the extent possible, including determination of the appropriate category regarding sufficiency of the health-related data. A comprehensive risk assessment ideally includes information on a variety of endpoints that provide insight into the full spectrum of developmental responses. A profile that integrates both human and test species data and incorporates a broad range of developmental effects provides more confidence in a risk assessment for a given agent.

The ability to describe the nature of human exposure is important for prediction of specific outcomes and the likelihood of permanence or reversibility of the effect. An important part of this effort is a description of the nature of the exposed populations. For example, the consequences of exposure to the developing individual versus the adult can differ markedly and again can influence whether the

effects are transient or permanent. Other considerations relative to human exposures might include potential synergistic effects, increased susceptibility resulting from concurrent exposures to other agents, concurrent disease, and nutritional status.

5.3. DESCRIPTORS OF DEVELOPMENTAL TOXICITY RISK

There are a number of ways to describe risks. These include:

5.3.1. Estimation of the Number of Individuals Exposed to Levels of Concern

The RfD_{DT} or RfC_{DT} is assumed to be a level at or below which no significant risk occurs. Therefore, information on the populations at or below the RfD_{DT} or RfC_{DT} ("not likely to be at risk") and above the RfD_{DT} or RfC_{DT} ("may be at risk") may be useful information for risk managers.

This method is particularly useful to a risk manager considering possible actions to ameliorate risk for a population. If the number of persons in the "at risk" category can be estimated, then the number of persons potentially removed from the "at risk" category after a contemplated action is taken can be used as an indication of the efficacy of that action.

5.3.2. Presenting Specific Scenarios

Presenting specific scenarios in the form of "what if?" questions is particularly useful to give perspective to the risk manager, especially where criteria, tolerance limits, or media quality limits are being set. The question being asked in these cases is, "At this proposed limit, what would be the resulting risk for developmental toxicity above the RfD_{DT}?"

5.3.3. Risk Characterization for Highly Exposed Individuals

This measure and the next are examples of specific scenarios. The purpose of this measure is to describe the upper end of the exposure distribution. This allows risk managers to evaluate whether certain individuals are at disproportionately high or unacceptably high risk.

The objective of looking at the upper end of the exposure distribution is to derive a realistic estimate of a relatively highly exposed individual(s), for example, by identifying a specified upper percentile of exposure in the population and/or by estimating the exposure of the most highly exposed individual(s). Whenever possible, it is important to express the number of individuals who comprise the highly exposed group and discuss the potential for exposure at still higher levels.

If population data are absent, it will often be possible to describe a scenario representing high-end exposures using upper percentile or judgment-based values for exposure variables. In these

instances, caution should be taken not to overestimate the high-end values if a "reasonable" exposure estimate is to be achieved.

5.3.4. Risk Characterization for Highly Sensitive or Susceptible Individuals

The purpose of this measure is to quantify exposure to identified sensitive or susceptible populations to the effect of concern. Sensitive or susceptible individuals are those within the exposed population at increased risk of expressing the adverse effect. All stages of development might be considered highly sensitive or susceptible, but certain subpopulations can sometimes be identified because of critical periods for exposure; for example, pregnant or lactating women, infants, children, adolescents.

In general, not enough is understood about the mechanisms of toxicity to identify sensitive subgroups for all agents, although factors such as nutrition, personal habits (e.g., smoking, alcohol consumption, illicit drug abuse), or pre-existing disease (e.g., diabetes) may predispose some individuals to be more sensitive to the developmental effects of various agents.

5.3.5. Other Risk Descriptors

In risk characterization, dose-response information and the human exposure estimates may be combined either by comparing the RfD_{DT} or RfC_{DT} and the human exposure estimate or by calculating the margin of exposure (MOE). The MOE is the ratio of the NOAEL from the most appropriate or sensitive species to the estimated human exposure level from all potential sources (U.S. EPA, 1985b). If a NOAEL is not available, a LOAEL may be used in the calculation of the MOE, but considerations for the acceptability would be different from those when a NOAEL is used. Considerations for the acceptability of the MOE are similar to that for the uncertainty factor applied to the LOAEL, NOAEL, or the benchmark dose. The MOE is presented along with the characterization of the database, including the strengths and weaknesses of the toxicity and exposure data, the number of species affected, and the dose-response, route, timing, and duration information. The RfD_{DT} or RfC_{DT} comparison with the human exposure estimate and the calculation of the MOE are conceptually similar but are used in different regulatory situations. If the MOE is equal to or more than the uncertainty factor used as a basis for an RfD_{DT} or RfC_{DT}, then the need for regulatory concern is likely to be reduced.

The choice of approach is dependent upon several factors, including the statute involved, the situation being addressed, the database used, and the needs of the decision maker. While these methods of describing risk do not actually estimate risks per se, they give the risk manager some sense

of how close the exposures are to levels of concern. The RfD_{DT}, RfC_{DT}, and/or the MOE are considered along with other risk assessment and risk management issues in making risk management decisions, and the scientific issues that must be taken into account in establishing them have been addressed here.

5.4. COMMUNICATING RESULTS

Once the risk characterization is completed, the focus turns to communicating results to the risk manager. The risk manager uses the results of the risk characterization, other technological factors, and nontechnological social and economic considerations in reaching a regulatory decision. Because of the way in which these risk management factors may impact different cases, consistent but not necessarily identical risk management decisions must be made on a case-by-case basis. Consequently, it is entirely possible and appropriate that an agent with a specific risk characterization may be regulated differently under different statutes. These Guidelines are not intended to give guidance on the nonscientific aspects of risk management decisions.

6. SUMMARY AND RESEARCH NEEDS

These Guidelines summarize the procedures that the U.S. Environmental Protection Agency uses in evaluating the potential for agents to cause developmental toxicity. While these are the first amendments to the developmental toxicity guidelines issued in 1986, further revisions and updates will be made as advances occur in the field. These Guidelines discuss the assumptions that should be made in risk assessment for developmental toxicity because of gaps in our knowledge about underlying biological processes and how these compare across species.

Research to improve the risk assessment process is needed in a number of areas. For example, research is needed to delineate the mechanisms of developmental toxicity and pathogenesis, provide comparative pharmacokinetic data, examine the validity of short-term in vivo and in vitro tests, elucidate possible functional alterations and their critical periods of exposure to toxic agents, develop improved animal models to examine the developmental effects of exposure during the premating and early postmating periods and in neonates, further evaluate the relationship between maternal and developmental toxicity, provide insight into the concept of threshold, develop approaches for improved mathematical modeling of adverse developmental effects, and improve animal models for examining the effects of agents given by various routes of exposure. Epidemiologic studies with quantitative measures of exposure are also strongly encouraged. Such research will aid in the evaluation and interpretation of data on developmental toxicity, and should provide methods to more precisely assess risk.

7. REFERENCES

Adams, J. (1986) Clinical relevance of experimental behavioral teratology. Neurotoxicology 7:19-34.

Anderson, L.M.; Donovan, P.J.; Rice, J.M. (1985) Risk assessment for transplacental carcinogens. In: Li, A.P., ed. New approaches in toxicity testing and their application in human risk assessment. New York, NY: Raven Press, pp. 179-202.

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

Baird, D.D.; Wilcox, A.J.; Weinberg, C.R. (1986) Use of time to pregnancy to study environmental exposures. Am. J. Epidemiol. 124:470-480.

Bellinger, D.; Leviton, A.; Waternaux, C.; et al. (1987) Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. N. Engl. J. Med. 316:1037-1043.

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

Brown, J.M. (1984) Validation of an in vivo screen for the determination of embryo/fetal toxicity in mice. Prepared by SRI International for the U.S. EPA, Washington, DC, under EPA contract no. 68-01-5079.

Brown, N.A. (1987) Teratogenicity testing in vitro: status of validation studies. Arch. Toxicol. Suppl. 11:105-114.

Brown, K.G.; Erdreich, L.S. (1989) Statistical uncertainty in the no-observed-adverse-effect level. Fundam. Appl. Toxicol. 13:235-244.

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

Brown, N.A.; Freeman, S.J. (1984) Alternative tests for teratogenicity. Altern. Lab. Anim. 12:7-23.

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

Butcher, R.E.; Wootten, V.; Vorhees, C.V. (1980) Standards in behavioral teratology testing: test variability and sensitivity. Teratogen. Carcinogen. Mutagen. 1:49-61.

Centers for Disease Control. (1988a) Trends in years of potential life lost due to infant mortality and perinatal conditions, 1980-1983 and 1984-1985. Morbidity and Mortality Weekly Report 37:249-256.

Centers for Disease Control. (1988b) Premature mortality due to congenital anomalies - United States. Morbidity and Mortality Weekly Report 37:505-506.

Chen, J.J.; Kodell, R.L. (1989) Quantitative risk assessment for teratological effects. J. Amer. Statistical Assoc. 84:966-971.

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

Couture, L.A. (1990) 2,3,7,8-Tetrachlorodibenzo-p-dioxin-induced hydronephrosis: characterization of the peak period of sensitivity for placentally- and lactationally-induced renal lesions, and assessment of persistence [dissertation]. Chapel Hill, NC: University of North Carolina. Available from: University of Michigan, Dissertation Library, Ann Arbor, MI.

Crump, K.S. (1984) A new method for determining allowable daily intakes. Fundam. Appl. Toxicol. 4:854-871.

Daston, G.P.; Rehnberg, B.F.; Carver, B.A.; et al. (1988) Functional teratogens of the rat kidney. II. Nitrofen and ethylenethiourea. Fundam. Appl. Toxicol. 11:401-415.

Davis, J.M.; Otto, D.A.; Weil, D.E.; et al. (1990) The comparative developmental neurotoxicity of lead in humans and animals. Neurotoxicol. Teratol. 12:215-229.

Deane, M.; Swan, S.H.; Harris, J.A.; et al. (1989) Adverse pregnancy outcomes in relation to water contamination, Santa Clara County, CA, 1980-1981. Am. J. Epidemiol. 129:894-904.

Dobbins, J.G.; Eifler, C.W.; Buffler, P.A. (1978) The use of parity survivorship analysis in the study of reproductive outcomes. Presented at the Society for Epidemiologic Research Conference; June; Seattle, WA.

Elsner, J.; Suter, K.E.; Ulbrich, B.; et al. (1986) Testing strategies in behavioral teratology: IV. Review and general conclusions. Neurobehav. Toxicol. Teratol. 8:585-590.

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

Everson, R.B.; Sandler, D.P.; Wilcox, A.J.; et al. (1986) Effect of passive exposure to smoking on age at natural menopause. Br. Med. J. 293(6550):792.

Fabro, S.; Shull, G.; Brown, N.A. (1982) The relative teratogenic index and teratogenic potency: proposed components of developmental toxicity risk assessment. Teratogen. Carcinogen. Mutagen. 2:61-76.

Faustman, E.M. (1988) Short-term tests for teratogens. Mutat. Res. 205:355-384.

Faustman, E.M.; Wellington, D.G.; Smith, W.P.; et al. (1989) Characterization of a developmental toxicity dose-response model. Environ. Health Perspect. 79:229-241.

Food and Drug Administration. (1966) Guidelines for reproduction and studies for safety evaluation of drugs for human use. Bureau of Drugs, Rockville, MD.

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.

Food and Drug Administration. (1987) Report of the in vitro teratology task force. Environ. Health Perspect. 72:201-249.

Francis, E.Z.; Farland, W.H. (1987) Application of the preliminary developmental toxicity screen for chemical hazard identification under the Toxic Substances Control Act. Teratogen. Carcinog. Mutagen. 7:107-117.

Fujii, T.; Adams, P.M. (1987) Functional teratogenesis: functional effects on the offspring after parental drug exposure. Tokyo, Japan: Teikyo University Press.

Gaffey, W.R. (1976) A critique of the standard mortality ratio. J. Occup. Med. 18:157-160.

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

Gaylor, D.W. (1989) Quantitative risk analysis for quantal reproductive and developmental effects. Environ. Health Perspect. 79:243-246.

Gray, J.A.; Kavlock, R.J. (1991) Physiological consequences of early neonatal growth retardation: effects of a-difluoromethylornithine on renal growth and function in the rat. Teratology 43:19-26.

Gray, J.A.; Rehnberg, B.F.; Rogers, E.H.; et al. (1989) Prenatal a-difluoromethylornithine treatment: effects on postnatal growth and function in the rat. Teratology 40:105-111.

Greenland, S. (1987) Quantitative methods in the review of epidemiologic literature. Epidemiol. Rev. 9:1-30.

Hardin, B.D., ed. (1987) Evaluation of the Chernoff/Kavlock test for developmental toxicity. Teratogen. Carcinogen. Mutagen. 7:1-127.

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

Hemminki, K.; Vineis, P. (1985) Extrapolation of the evidence on teratogenicity of chemicals between humans and experimental animals: chemicals other than drugs. Teratogen. Carcinogen. Mutagen. 5:251-318.

Hemminki, K.; Mutanen, P.; Luoma, K.; et al. (1980) Congenital malformations by the parental occupation in Finland. Int. Arch. Occup. Environ. Health 46:93-98.

Hemminki, K.; Saloniemi, I.; Salonen, T.; et al. (1981) Childhood cancer and parental occupation in Finland. J. Epidemiol. Commun. Health 35:11-15.

Herbst, A.L.; Ulfelder, H.; Poskanzer, D.C. (1971) Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with appearance in young women. N. Engl. J. Med. 284:878.

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

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

Hogue, C.J.R. (1985) Developmental risks. Presented at: Symposium on epidemiology and health risk assessment; May 14; Columbia, MD.

Joffe, M. (1985) Biases in research on reproduction and women's work. Int. J. Epidemiol. 14(1):118-123.

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

Johnson, E.M.; Gabel, B.E.G. (1983) An artificial embryo for detection of abnormal developmental biology. Fundam. Appl. Toxicol. 3:243-249.

Kavlock, R.J.; Grabowski, C.T., 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.

Kavlock, R.J.; Rehnberg, B.F.; Rogers, E.H. (1986) Congenital renal hypoplasia: effects on basal renal function in the developing rat. Toxicology 40:247-258.

Kavlock, R.J.; Rehnberg, B.F.; Rogers, E.H. (1987a) The fate of adriamycin induced dilated renal pelvis in the fetal rat: physiological and morphological effects in the offspring. Teratology 36:51-58.

Kavlock, R.J.; Rehnberg, B.F.; Rogers, E.H. (1987b) Critical prenatal periods for chlorambucil induced functional teratology of the kidneys. Toxicology 43:51-64.

Kavlock, R.J.; Short, R.D., Jr.; Chernoff, N. (1987c) Further evaluation of an in vivo teratology screen. Teratogen. Carcinogen. Mutagen. 7:7-16.

Kavlock, R.J.; Hoyle, B.R.; Rehnberg, B.F.; et al. (1988) The significance of dilated renal pelvis in the nitrofen exposed fetal rat. Toxicol. Appl. Pharmacol. 94:287-296.

Khera, K.S. (1984) Maternal toxicity - a possible factor in fetal malformations in mice. Teratology 29:411-416.

Khera, K.S. (1985) Maternal toxicity: a possible etiologic factor in embryo-fetal deaths and fetal malformations in rodent-rabbit species. Teratology 31:129-153.

Khera, K.S. (1987) Maternal toxicity of drugs and metabolic disorders - a possible etiologic factor in the intrauterine death and congenital malformation: a critique on human data. CRC Crit. Rev. Toxicol. 17:345-375.

Kimmel, C.A. (1988) Current status of behavioral teratology—science and regulation. CRC Crit. Rev. Toxicol. 19(1):1-10.

Kimmel, C.A. (1990) Quantitative approaches to human risk assessment for noncancer health effects. Neurotoxicology 11:189-198.

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

Kimmel, G.L. (1990) In vitro assays in developmental toxicology: their potential application in risk assessment. In: In vitro methods in developmental toxicology: use in defining mechanisms and risk parameters. Kimmel, G.L.; Kochhar, D.M., eds. Boca Raton, FL: CRC Press, pp. 163-173.

Kimmel, C.A.; Francis, E.Z. (1990) Proceedings of the workshop on the acceptability and interpretation of dermal developmental toxicity studies. Fundam. Appl. Toxicol. 14:386-398.

Kimmel, C.A.; Gaylor, D.W. (1988) Issues in qualitative and quantitative risk analysis for developmental toxicology. Risk Anal. 8:15-20.

Kimmel, C.A.; Price, C.J. (1990) Developmental toxicity studies. In: Arnold, D.L.; Grice, H.C.; Krewski, D.R., eds. Handbook of in vivo toxicity testing. San Diego, CA: Academic Press, pp. 271-301.

Kimmel, C.A.; Young, J.F. (1983) Correlating pharmacokinetics and teratogenic endpoints. Fundam. Appl. Toxicol. 3:250-255.

Kimmel, G.L.; Smith, K.; Kochhar, D.M.; et al. (1982a) Overview of in vitro teratogenicity testing: aspects of validation and application to screening. Teratogen. Carcinogen. Mutagen. 2:221-229.

Kimmel, G.L.; Smith, K.; Kochhar, D.M.; et al. (1982b) Proceedings of the consensus workshop on in vitro teratogenesis testing. Teratogen. Carcinogen. Mutagen. 2:221-374.

Kimmel, C.A.; Holson, J.F.; Hogue, C.J.; et al. (1984) Reliability of experimental studies for predicting hazards to human development. National Center for Toxicological Research, Jefferson, AR. NCTR Technical Report for Experiment No. 6015.

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

Kimmel, G.L.; Kimmel, C.A.; Francis, E.Z., eds. (1987) Evaluation of maternal and developmental toxicity. Teratogen. Carcinogen. Mutagen. 7:203-338.

Kimmel, C.A.; Wellington, D.G.; Farland, W.; et al. (1989) Overview of a workshop on quantitative models for developmental toxicity risk assessment. Environ. Health Perspect. 79:209-215.

Kimmel, C.A.; Rees, D.C.; Francis, E.Z., eds. (1990a) Proceedings of the Workshop on the Qualitative and Quantitative Comparability of Human and Animal Developmental Neurotoxicity. Neurotoxicol. Teratol. 12(3):173-292.

Kimmel, C.A.; Kimmel, G.L.; Francis, E.Z.; et al. (1990b) An overview of the U.S. EPA's proposed amendments to the guidelines for the health assessment of suspect developmental toxicants. J. Am. Coll. Toxicol. 9:39-47.

Kissling, G. (1981) A generalized model for analysis of non-independent observations [dissertation]. Chapel Hill, NC: University of North Carolina. Available from: University Microfilms, Ann Arbor, MI.

Kleinbaum, D.G.; Kupper, L.L.; Morgenstern, H. (1982) Epidemiologic research: principles and quantitative methods. London: Lifetime Learning Publications.

Kodell, R.L.; Howe, R.B.; Chen, J.J.; et al. (1991) Mathematical modeling of reproductive and developmental toxic effects for quantitative risk assessment. Risk Anal. 11(4):583-590.

Kwa, S.-L.; Fine, L.J. (1980) The association between parental occupation and childhood malignancy. J. Occup. Med. 22:792-794.

Lamb, J.C., IV. 1985. Reproductive toxicity testing: evaluating and developing new testing systems. J. Am. Coll. Toxicol. 4:163-171.

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

Lemasters, G.K.; Pinney, S.M. (1989) Employment status as a confounder when assessing occupational exposures and spontaneous abortion. J. Clin. Epidemiol. 42:975-981.

Leridon, H. (1977) Human fertility: the basic components. Chicago, IL: The University of Chicago Press.

Leukroth, R.W., ed. (1986) Predicting neurotoxicity and behavioral dysfunction from preclinical toxicologic data. Neurotoxicol. Teratol. 9:395-471.

Levine, R.J. (1983) Methods for detecting occupational causes of male infertility: reproductive history versus semen analysis. Scand. J. Work Environ. Health 9:371-376.

Levine, T.E.; Butcher, R.E. (1990) Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity. Work group IV report: Triggers for developmental neurotoxicity testing. Neurotoxicol. Teratol. 12:281-284.

Levine, R.J.; Symons, M.J.; Balogh, S.A.; et al. (1980) A method for monitoring the fertility of workers: I. Method and pilot studies. J. Occup. Med. 22:781-791.

Levine, R.J.; Symons, M.J.; Balogh, S.A.; et al. (1981) A method for monitoring the fertility of workers: II. Validation of the method among workers exposed to dibromochloropropane. J. Occup. Med. 23:183-188.

Mackeprang, M.; Hay, S.; Lunde, A.S. (1972) Completeness and accuracy of reporting of malformations on birth certificates. HSMHA Health Reports 84:43-49.

McMichael, A.J. (1976) Standardized mortality ratios and the 'healthy worker effect': scratching beneath the surface. J. Occup. Med. 18:165-168.

Morrissey, R.E.; Harris, M.W.; Schwetz, B.A. (1989) Developmental toxicity screen: results of rat studies with diethylhexyl phthalate and ethylene glycol monomethyl ether. Teratogen. Carcinogen. Mutagen. 9:119-129.

National Center for Health Statistics. (1988) Advance report of final mortality statistics, 1986. Monthly Vital Statistics Report 37(6): Supp 1. NCHR, Hyattsville, MD. DHHS Publ. No. (PHS) 88-1120.

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

Needleman, H. (1988) The neurotoxic, teratogenic, and behavioral teratogenic effects of lead at low dose: a paradigm for transplacental toxicants. In: Transplacental effects on fetal health. New York, NY: Alan R. Liss, Inc., pp. 279-287.

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

Nelson, K.; Holmes, L.B. (1989) Malformations due to presumed spontaneous mutations in newborn infants. New Engl. J. Med. 320:19-23.

Nisbet, I.C.T.; Karch, N.J. (1983) Chemical hazards to human reproduction. Park Ridge, IL: Noyes Data Corp.

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

Papier, C.M. (1985) Parental occupation and congenital malformations in a series of 35,000 births in Israel. Prog. Clin. Biol. Res. 163:291-294.

Perlin, S.A.; McCormack, C. (1988) Using weight-of-evidence classification schemes in the assessment of non-cancer health risks. In: Proceedings of the 5th National Conference on Hazardous Wastes and Hazardous Materials (HWHM '88); April 19-21; Las Vegas, NV.

Peters, J.M.; Preston-Martin, S.; Yu, M.C. (1981) Brain tumors in children and occupational exposure of parents. Science 213:235-237.

Rai, K.; Van Ryzin, J. (1985) A dose-response model for teratological experiments involving quantal responses. Biometrics 41:1-9.

Riley, E.P.; Vorhees, C.V., eds. (1986) Handbook of behavioral teratology. New York, NY: Plenum Press.

Rodier, P.M. (1978) Behavioral teratology. In: Wilson, J.G.; Fraser, F.C., eds. Handbook of teratology, vol. 4. New York, NY: Plenum Press, pp. 397-428.

Rothman, K.J. (1986) Modern epidemiology. Boston, MA: Little, Brown and Co., pp. 83-94.

Ryan, L.M.; Catalano, P.J.; Kimmel, C.A.; et al. (1991) Relationship between fetal weight and malformation in developmental toxicity studies. Teratology 44:215-223.

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

Schnatter, A.R.L. (1990) The development of methods for implementing industry-based reproductive surveillance [dissertation]. New York, NY: Columbia University. Available from: University Microfilms, Ann Arbor, MI.

Schuler R.; Hardin, B.: Niemeyer, R.; et al. (1984) Results of testing fifteen glycol ethers in a short-term, in vivo reproductive toxicity assay. Environ. Health Perspect. 57:141-146.

Schwetz, B.A.; Morrissey, R.E.; Welsch, F.; et al. (1991) In vitro teratology. Environ. Health Perspect. 94:265-268.

Selevan, S.G. (1980) Evaluation of data sources for occupational pregnancy outcome studies [dissertation]. Cincinnati, OH: University of Cincinnati. Available from: University Microfilms, Ann Arbor, MI.

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

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

Selevan, S.G.; Hemminki, K.; Lindbohm, M-L. (1986) Linking data to study reproductive effects of occupational exposures. Occup. Med.: State of the Art Revs. 1(3):445-455.

Selevan, S.G.; Lemasters, G.K. (1987) The dose-response fallacy in human reproductive studies of toxic exposures. J. Occup. Med. 29:451-454.

Sever, L.E.; Hessol, N.A. (1984) Overall design considerations in male and female occupational reproductive studies. In: Lockey, J.E.; LeMasters, G.K.; Keye, W.R., eds. Reproduction: the new frontier in occupational and environmental research. New York, NY: Alan R. Liss, Inc. pp. 15-47.

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

Shepard, T.H. (1986) Human teratogenicity. Adv. Pediatr. 33:225-268.

Silverman, J.; Kline, J.; Hutzler, M.; et al. (1985) Maternal employment and the chromosomal characteristics of spontaneously aborted conceptions. J. Occup. Med. 27:427-438.

Slotkin, T.A.; Lau, C.; Kavlock, R.J.; et al. (1988) Role of sympathetic neurons in biochemical and functional development of the kidney: neonatal sympathectomy with 6-hydroxydopamine. J. Pharmacol. Exp. Ther. 246:427-433.

Starr, T.B.; Dalcorso, R.D.; Levine, R.J. (1986) Fertility of workers: a comparison of logistic regression and indirect standardization. Am. J. Epidemiol. 123:490-498.

Stein, Z.; Hatch, M. (1987) Biological markers in reproductive epidemiology: prospects and precautions. Environ. Health Perspect. 74:67-75.

Stein, Z.; Susser, M.; Warburton, D.; et al. (1975) Spontaneous abortion as a screening device. The effect of fetal surveillance on the incidence of birth defects. Am. J. Epidemiol. 102:275-290.

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

Stiratelli, R.; Laird, N.; Ware, J.H. (1984) Random-effects models for serial observations with binary responses. Biometrics 40:961-971.

Swan, S.H.; Shaw, G.; Harris, J.A.; et al. (1989) Congenital cardiac anomalies in relation to water contamination, Santa Clara County, CA, 1981-1983. Am. J. Epidemiol. 129:885-893.

Sweeney, A.M.; Meyer, M.R.; Aarons, J.H.; et al. (1988) Evaluation of methods for the prospective identification of early fetal losses in environmental epidemiology studies. Am. J. Epidemiol. 127:843-850.

Tanimura, T. (1986) Collaborative studies on behavioral teratology in Japan. Neurotoxicology 7:35-45.

Tilley, B.C.; Barnes, A.B.; Bergstralh, E.; et al. (1985) A comparison of pregnancy history recall and medical records: implications for retrospective studies. Am. J. Epidemiol. 121:269-281.

Tilson, H.A.; Jacobson, J.L.; Rogan, W.J. (1990) Polychlorinated biphenyls and the developing nervous system: cross-species comparisons. Neurotoxicol. Teratol. 12:239-248.

Tsai, S.P.; Wen, C.P. (1986) A review of methodological issues of the standardized mortality ratio (SMR) in occupational cohort studies. Int. J. Epidemiol. 15:8-21.

U.S. Environmental Protection Agency (1981) Spontaneous abortion and exposure during pregnancy to the herbicide 2,4,5-T: a pilot study. U.S. EPA, Washington, DC. EPA/560/6-81-006.

U.S. Environmental Protection Agency. (1982a) Assessment of risks to human reproduction and to development of the human conceptus from exposure to environmental substances, pp. 99-116. EPA/600/9-82-001. Available from: NTIS, Springfield, VA. DE82-007897.

U.S. Environmental Protection Agency. (1982b) Pesticide assessment guidelines, subdivision F. Hazard evaluation: human and domestic animals. Office of Pesticides and Toxic Substances, Washington, DC. EPA/540/9-82-025. Available from: NTIS, Springfield, VA.

U.S. Environmental Protection Agency. (1984) Pesticide assessment guidelines, subdivision K. Exposure: reentry protection. Office of Pesticides and Toxic Substances, Washington, DC. EPA/540/9-84-001. Available from: NTIS, Springfield, VA.

U.S. Environmental Protection Agency. (1985a) Toxic Substances Control Act test guidelines; final rules. Federal Register 50:39426-39428 and 39433-39434.

U.S. Environmental Protection Agency. (1985b) Hazard Evaluation Division standard evaluation procedure: teratology studies, pp. 22-23. Office of Pesticide Programs, Washington, DC. EPA/540/9-85-018.

U.S. Environmental Protection Agency. (1985c) Toxic Substances Control Act test guidelines; final rules. Federal Register 50:39428-39429.

U.S. Environmental Protection Agency. (1986a) Triethylene glycol monomethyl, monoethyl, and monobutyl ethers; proposed test rule. Federal Register 51:17883-17894.

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

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

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

U.S. Environmental Protection Agency. (1988a, Feb. 26) Diethylene glycol butyl ether and diethylene glycol butyl ether acetate; final test rule. Federal Register 53:5932-5953.

U.S. Environmental Protection Agency. (1988b) Proposed guidelines for assessing male reproductive risk. Federal Register 53:24850-24869.

U.S. Environmental Protection Agency. (1988c) Proposed guidelines for assessing female reproductive risk. Federal Register 53:24834-24847.

U.S. Environmental Protection Agency. (1989a) FIFRA accelerated reregistration phase 3 technical guidance, Appendix D. Office of Pesticides and Toxic Substances, Washington, DC. EPA No. 540/09-90-078. Available from: NTIS, Springfield, VA.

U.S. Environmental Protection Agency. (1989b) Triethylene glycol monomethyl ether; final test rule. Federal Register 54:13472-13477.

U.S. Environmental Protection Agency. (1991a) Pesticide assessment guidelines, subdivision F. Hazard evaluation: human and domestic animals. Addendum 10: Neurotoxicity, series 81, 82, and 83. Office of Pesticides and Toxic Substances, Washington, DC. EPA 540/09-91-123. Available from: NTIS, Springfield, VA. PB91-154617.

U.S. Environmental Protection Agency. (1991b) Integrated Risk Information System (IRIS). Online. Office of Health and Environmental Assessment, Washington, DC.

Weinberg, C.R.; Gladen, B.C. (1986) The beta-geometric distribution applied to comparative fecundability studies. Biometrics 42:547-560.

Wickramaratne, G.A. de S. (1987) The Chernoff-Kavlock assay: its validation and application in rats. Teratogen. Carcinogen Mutagen. 7:73-83.

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

Wilcox, A.J.; Weinberg, C.R.; Wehmann, R.E.; et al. (1985) Measuring early pregnancy loss: laboratory and field methods. Fertil. Steril. 44:366-374.

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

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

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

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

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

Wong, O.; Utidjian, H.M.D.; Karten, V.S. (1979) Retrospective evaluation of reproductive performance of workers exposed to ethylene dibromide. J. Occup. Med. 21:98-102.

Woo, D.C.; Hoar, R.M. (1972) "Apparent hydronephrosis" as a normal aspect of renal development in late gestation of rats: the effect of methyl salicylate. Teratology 6:191-196.

World Health Organization. (1984) Principles for evaluating health risks to progeny associated with exposure to chemicals during pregnancy. In: Environmental Health Criteria, vol. 30. Geneva: World Health Organization.

Zack, M.; Cannon, S.; Lloyd, D.; et al. (1980) Cancer in children of parents exposed to hydrocarbon-related industries and occupations. Am. J. Epidemiol. 3:329-336.

Zenick, H.; Clegg, E.D. (1989) Assessment of male reproductive toxicity: a risk assessment approach. In: Hayes, A.W., ed. Principles and methods of toxicology. Second ed. New York, NY: Raven Press, pp. 279-309.

PART B: RESPONSE TO PUBLIC AND SCIENCE ADVISORY BOARD COMMENTS

1. INTRODUCTION

This section summarizes the major issues raised in the public and Science Advisory Board (SAB) comments on the Proposed Amendments to the Guidelines for the Health Assessment of Suspect Developmental Toxicants published March 6, 1989 [54 FR 9386-9403]. Comments were received from 25 individuals or organizations. The Agency's initial summary of the public comments and proposed responses were presented to the Environmental Health Committee of the SAB on October 27, 1989. The report of the SAB Committee was provided to the Agency on April 23, 1990.

The SAB and public comments were diverse and addressed issues from a variety of perspectives. The majority of the comments were favorable and in support of the Proposed Amendments to the Guidelines. Many praised the Agency's efforts as being timely and well-justified. Most commentors also gave specific comments or criticisms for further consideration, clarification, or re-evaluation. For example, there was concern expressed about the Guidelines imposing further testing requirements, particularly functional testing, and many commentors felt that the Proposed Amendments discounted the role of maternal toxicity in developmental toxicity. In addition, there was concern that the proposed weight-of-evidence scheme would promote labeling of agents as causing developmental toxicity before the entire risk assessment process was completed.

The SAB Committee also indicated that the proposed revisions were adequately founded in developmental toxicology and represented a step forward for the Agency. They suggested that the Agency revisit the weight-of-evidence scheme to avoid confusion with more commonly applied uses of such classifications, and to develop a more powerful conceptual approach. Further, the SAB Committee urged that the Agency begin to move away from the current use of the no-observed-adverse-effect level (NOAEL) and lowest-observed-adverse-effect level (LOAEL) basis for calculating the reference dose for developmental toxicity to a benchmark dose and confidence limit approach tied to empirical models of dose-response relationships.

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

2. INTENT OF THE GUIDELINES

Many of the public comments indicated some misunderstanding of the intent of the Guidelines, apparently assuming that the risk assessment guidelines impose testing requirements. In particular, some commentors suggested that because the Agency was providing guidance on the interpretation of tests not required in the EPA testing guidelines, the Agency was suggesting that these tests be required in the future.

The 1986 Guidelines and the 1989 Proposed Amendments clearly state that these guidelines are not Agency testing guidelines, but rather are intended to ensure uniform interpretation of all existing, relevant data. However, to avoid any confusion, the discussion of study designs has been changed to avoid the impression that these Guidelines set testing requirements. In the evaluation of data on an agent for risk assessment, relevant data are often encountered that have been generated from nontraditional tests. In such cases, it is imperative that the Agency provide guidance so that all data considered to be relevant are included in the risk assessment and are interpreted uniformly.

3. BASIC ASSUMPTIONS

In the 1986 Guidelines, several assumptions were implicit in the approach to risk assessment, but were not explicitly stated. These assumptions were detailed in the 1989 Proposed Amendments. Comments received from the public and the SAB favored presentation of these assumptions and generally agreed with the wording, except for the fourth assumption, which concerns the use of the most relevant or most sensitive species. The 1989 Proposed Amendments stated that "it is assumed that the most sensitive species should be used to estimate human risk. When data are available (e.g., pharmacokinetic, metabolic) to suggest the most appropriate species, that species will be used for extrapolation." The SAB recommended that, for this assumption, the basic position of the Agency should be to use data from the most relevant species, and that use of data from the most sensitive species should be the default position. In addition, the SAB recommended that the threshold assumption be considered carefully in the dose-response assessment of any agent, and that the Agency develop more comprehensive approaches to risk assessment as discussed further in the following sections.

Changes have been made in the statement of the basic assumptions in line with the SAB and public comments that clarify, but do not alter, the intent of the assumptions.

4. MATERNAL/DEVELOPMENTAL TOXICITY

The 1989 Proposed Amendments stated that "when adverse developmental effects are produced only at maternally toxic doses, they are still considered to represent developmental toxicity and should not be discounted as being secondary to maternal toxicity." This statement and others concerning the interpretation of developmental toxicity in the presence of maternal toxicity were the subject of a considerable number of public comments and were also addressed by the SAB. In general, commentors were divided in their opinions on whether they supported the Agency's statements or felt that they discounted the role of maternal toxicity in developmental toxicity, but in general, the recommended changes did not significantly alter the intent of the statements. The SAB endorsed the proposed revision, and suggested that the Agency retain the statement that was made in the Proposed Amendments.

In these Guidelines, the position is further clarified by indicating that when maternal toxicity is significantly greater than the minimal maternally toxic dose, developmental effects at that dose may be difficult to interpret. This statement is added to clarify, but not to change, the intent or meaning of the statements regarding the relationship between maternal and developmental toxicity. From a risk assessment point of view, whether a developmental effect is or is not secondary to maternal toxicity does not impact on the selection of the NOAEL or other dose-response methodology.

5. FUNCTIONAL DEVELOPMENTAL TOXICITY

The 1989 Proposed Amendments provided information on the state-of-the-art in the evaluation of functional effects resulting from developmental exposures. Several commentors voiced strong objection to this section because they perceived it as indicating an imminent requirement for testing. Several indicated there are no standard methods for functional testing, some felt that functional endpoints should not be used to establish the NOAEL, and others voiced concern about the problems with using postnatal exposures in animal studies.

The final Guidelines further update this section to include a discussion of the latest changes in the requirements for functional developmental toxicity testing by the Agency, and reflect the current approach to interpretation of such data, with incorporation of information from the EPA/NIDA-sponsored "Workshop on the Qualitative and Quantitative Comparability of Human and Animal Developmental Neurotoxicity" (1990). The intent of these Guidelines, as stated above, is not to change testing requirements but to give guidance when these types of data are encountered in the risk

assessment process. The Guidelines also indicate that functional developmental toxicity endpoints will be used for establishing the NOAEL when they are found to be the adverse effect occurring at the lowest dose in appropriate, well-conducted studies. Interpretation of postnatal exposure data is a concern, and must take into consideration effects on the mother, her offspring, and possible interactions; a statement to this effect has been added. Further interpretation of data will be discussed in the guidance being developed by the Agency on neurotoxicity risk assessment.

6. WEIGHT-OF-EVIDENCE SCHEME

The 1989 Proposed Amendments described important considerations in determining the relative weight of various kinds of data in estimating the risk of developmental toxicity in humans. The intent of the proposed weight-of-evidence (WOE) scheme was that it not be used in isolation, but be used as the first step in the risk assessment process, to be integrated with dose-response information and the exposure assessment.

The WOE scheme was the subject of a considerable number of public comments, and was one of the major concerns of the SAB. The concern of public commentors was that the reference to human developmental toxicity in this scheme suggested that a chemical could be prematurely designated, and perhaps labeled, as causing developmental toxicity in humans prior to the completion of the risk assessment process. The SAB suggested that the intended use of this scheme was not consistent with the use of the term "weight of evidence" in other contexts, since WOE is usually thought of as an evaluation of the total composite of information available to make a judgment about risk. In addition, the SAB Committee proposed that the Agency consider development of a more conceptual approach using decision analytical techniques to predict the relationships among various outcomes.

In the final Guidelines, the terminology used in the WOE scheme has been completely changed and retitled "Characterization of the Health-Related Database." The intended purpose of the scheme is to provide a framework and criteria for making a decision on whether or not sufficient data are available to conduct a risk assessment. This decision is based on the available data, whether animal or human, and does not necessarily imply human hazard. This decision process is part of, but not the complete, WOE evaluation, which also takes into account the RfD_{DT} or RfC_{DT} and the human exposure information, culminating in risk characterization.

The final Guidelines also place strong emphasis on the integration of the dose-response evaluation with hazard information in characterizing the sufficiency of the health-related database. In line with this approach, the Guidelines have been reorganized to combine hazard identification and

dose-response evaluation. Finally, the SAB comments on developing a conceptual matrix provide an interesting challenge, but current data indicate that the relationships among endpoints of developmental toxicity are not consistent across chemicals or species. The Agency is currently supporting modeling efforts to further explore the relationship among various development toxicity endpoints and the development of biologically based dose-response models that consider multiple effects.

7. APPLICABILITY OF THE RfD_{DT} CONCEPT AND THE BENCHMARK DOSE APPROACH

The 1989 Proposed Amendments introduced the term "reference dose for developmental toxicity - RfD_{DT}," based on short-term exposure, to distinguish it from the reference dose (RfD), which is used for chronic exposure situations. The public comments received generally supported the RfD_{DT} approach. The SAB also agreed with the concept of the RfD_{DT} for developmental toxicity risk assessment, based on short-term exposure. In addition, the SAB urged the Agency to consider strengthening the RfD approach by moving to more quantitative alternatives to the NOAEL. In particular, the use of a benchmark dose approach to replace the NOAEL was strongly suggested.

The final Guidelines have incorporated many of the SAB Committee's suggestions concerning the development of more quantitative approaches to the RfD, and state that the Agency is beginning to use the benchmark dose approach for comparison with and interpretation of the NOAEL. That is, benchmark dose calculations may allow better interpretation of dose-response data and, in particular, what level of risk may be associated with the NOAEL. The Agency also has developed the concept of an inhalation reference concentration (RfC), and the RfC_{DT} is being calculated for inhalation concentrations based on developmental toxicity. Guidance for use of the benchmark dose in the calculation of the RfD_{DT} or RfC_{DT} is not included in the final Guidelines, because of the limited experience of the Agency with this approach. There are several issues that must be addressed prior to its use for this purpose; for example, which benchmark dose (e.g., LED_{01}, LED_{05}, LED_{10}) should be used for calculating the RfD_{DT} or RfC_{DT}, and what are the appropriate uncertainty factors that should be applied to the benchmark dose for deriving the RfD_{DT} or RfC_{DT}? Should the uncertainty factor applied to an LED_{10} be similar to that applied to a LOAEL, or should the uncertainty factor applied to an LED_{01} be equal to or less than that applied to a NOAEL? These and other questions are being addressed in ongoing Agency studies on the calculation of the RfD_{DT} or RfC_{DT} using the benchmark dose approach. As results become available, and as further guidance is developed, this information will be published as a supplement to these Guidelines.