Federal Register

Friday
November 23, 1984

Part X

Environmental Protection Agency

Proposed Guidelines for the Health Assessment of Suspect Developmental Toxicants and Request for Comments

ENVIRONMENTAL PROTECTION AGENCY

[FRL-2706-7]

Proposed Guidelines for the Health Assessment of Suspect Developmental Toxicants

AGENCY: Environmental Protection Agency (EPA).

ACTION: Proposed Guidelines for the Health Assessment of Suspect Developmental Toxicants and Request for Comments.

SUMMARY: The U.S. Environmental Protection Agency is proposing Guidelines for the Health Assessment of Suspect Developmental Toxicants (Guidelines). These Guidelines are proposed for use within the policy and procedural framework provided by the various statutes that EPA administers to guide Agency analysis of developmental toxicity data. We solicit public comment and will take public comment into account in revising these Guidelines. The Guidelines will be reviewed by the Science Advisory Board in meetings now tentatively scheduled for April 1985.

These proposed Guidelines were developed as part of a broad guidelines development program under the auspices of the Office of Health and Environmental Assessment (OHEA), located inthe Agency's Office of Research and Development. Consonant with the role of OHEA's Reproductive Effects Assessment Group (REAG) as the Agency's senior health committee for developmental toxicity assessment, the Guidelines were developed by an Agency-wide working group chaired by the REAG.

DATE: Comments must be postmarked by January 22, 1985.

ADDRESSES: Comments may be mailed or delivered to: Dr. Carole A. Kimmel, Reproductive Effects Assessment Group (RD–689), Office of Health and Environmental Assessment, U.S. Environmental Protection Agency, 401 M Street, SW., Washington, DC 20460.

FOR FURTHER INFORMATION CONTACT: Dr. Carole A. Kimmel, telephone: 202–382–7331.

SUPPLEMENTARY INFORMATION: A preliminary draft of the Guidelines was sent for review to approximately 20 scientists in the field of developmental toxicology within government, universities in the United States, and the private sector. Comments received from these reviewers, generally favorable, were taken into account in developing the Guidelines proposed here.

References and supporting documents used in the preparation of these Guidelines as well as comments received are available for inspection and copying at the Public Information Reference Unit (202–382–5926), EPA Headquarters Library, 401 M Street, SW., Washington, DC, between the hours of 8:00 a.m. and 4:30 p.m.

Dated: November 9, 1984.

William D. Ruckelshaus,

Administrator.

Contents

I. Introduction

These Guidelines describe the procedures that the U.S. Environmental Protection Agency will follow in evaluating potential developmental toxicity associated with human exposure to environmental toxicants. In the past, the Agency has sponsored conferences and issued publications which addressed issues related to such evaluations(*1, 2, 3*). These publications provided some of the scientific basis for these risk assessment Guidelines, and testing guidelines have provided protocols designed to determine the potential of a test substance to induce structural and/or other abnormalities in the developing conceptus. The Agency's authority to regulate substances that have the potential to interfere adversely with human development is derived from a number of statutes which are implemented through multiple offices within the Agency. Because many different offices evaluate developmental toxicity, there is a need for intra-agency consistency in the approach to assess these types of effects. The procedures described here will promote consistency in the Agency's assessment of developmental toxic effects.

Approximately 50% of human conceptuses fail to reach term(*3, 4*); approximately 3% of newborn children are found to have one or more significant congenital malformations at birth, and, by the end of the first postnatal year, about 3% more are found to have serious developmental defects (*5, 6*). It is estimated that 20% of human congenital malformations are caused by mutations, 10% are attributable to known environmental factors, and the remainder result from unknown causes (*7*).

Numerous agents have been shown to be developmental toxicants in animal test systems(*8*). Several of them have also been shown to be the cause of adverse developmental effects in humans, including alcohol, aminopterin, busulfan, chlorobiphenyls, diethylstilbestrol, isotretinoin, organic mercury, thalidomide, and valproic acid (*9, 10, 11, 12*). Exposure to agents affecting development generally results in multiple manifestations (malformation, functional impairment, altered growth, and/or lethality). Therefore, assessment efforts should encompass a wide array of adverse developmental end points such as spontaneous abortions, stillbirths, malformations, and other adverse functional physical changes that occur postnatally.

The developmental toxicity assessments prepared pursuant to these Guidelines will be utilized within the requirements and constraints of the applicable statutes to arrive at regulatory decisions concerning developmental toxicity. These Guidelines provide a general format for analyzing and organizing the available data for conducting risk assessments. The Guidelines do not change any statutory or regulatory prescribed standards for the type of data necessary for regulatory action. Moreover, risk assessment is just one component of the regulatory process and defines the adverse health consequences of exposure to a toxic agent. The other component, risk management, combines risk assessment with the directives of the enabling regulatory legislation together with socio-economic, technical, political, and other considerations to reach a decision as to whether or how much to control future exposure to the suspected toxic agent. The issue of risk management will not be addressed in these Guidelines.

The National Research Council(*13*) has defined risk assessment as being comprised of some or all of the following

components: hazard identification, dose-response assessment, exposure assessment, and risk characterization. In general, the process of assessing the risk of human developmental toxicity may be adapted to this format. However, due to special considerations in assessing developmental toxicity, which will be discussed later in these Guidelines, it is not always appropriate to follow the exact standards as defined for each component.

Hazard identification is the qualitative risk assessment in which all available experimental animal and human data are used to determine if an agent is likely to cause developmental toxicity. In considering developmental toxicity, these Guidelines will address not only malformations, but also fetal wastage, growth alteration, and functional abnormalities that may result from developmental exposure to environmental agents.

The dose-response assessment defines the relationship of the dose of an agent and the occurrence of developmental toxic effects. According to the National Research Council(*13*), this component would usually include the results of an extrapolation from high doses administered to experimental animals or noted in epidemiologic studies to the low exposure levels expected for human contact with the agent in the environment. However, since at present there is no mathematical extrapolation model that is generally accepted for developmental toxicity, the Agency, for the most part, continues to use safety factors and margins of safety, which will be discussed in these Guidelines.

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

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

II. Definitions and Terminology

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

Developmental Toxicology—The field dealing with the induction of adverse effects on the developing organism occurring up to the time of puberty. The manifestations of developmental toxicity include: (1) Death of the developing organism, (2) structural abnormality (teratogenicity), (3) altered growth, and (4) functional deficiency.

Embryotoxicity and Fetotoxicity—Any toxic effect on the conceptus as a result of prenatal exposure; the distinguishing feature between the terms is the period during which the insult occurred. The terms, as used here, include malformation, altered growth, and *in utero* death.

Altered Growth—A significant alteration in fetal or neonatal organ or body weight. Body weight may or may not be accompanied by a change in crown-rump length and/or in skeletal ossification. Altered growth can be induced at any stage of development, may be reversible, or may result in a permanent change.

Functional Teratology—The field dealing with the causes, mechanisms, and manifestations of alterations or delays in functional competence of the organism or organ system following exposure to an agent during critical periods of development either pre- or postnatally.

Malformations and Variations—A malfunction is usually defined as a permanent structural deviation which generally is incompatible with or severely detrimental to normal postnatal survival or development. A variation is usually defined as a divergence beyond the usual range of structural constitution but which may not have as severe an effect on survival or health as a malformation. Distinguishing between variations and malformations is difficult since there exists a continuum of responses from the normal to the extreme deviant. There is no generally accepted classification of malformations and variations. Other terminology that is often used but no better defined, includes anomalies, deformations, and aberrations.

III. Qualitative Assessment (Hazard Identification of Developmental Toxicants)

Developmental toxicity studies provide a number of end points that are useful for evaluating the potential of an agent to produce adverse outcomes of pregnancy. The four types of effects on the conceptus that may be produced by *in utero* exposure to toxicants include death, structural abnormality, altered growth, and functional deficits. Of these, the first three effects are measured in the conventional developmental toxicity (teratogenicity) protocol (discussed below), while functional deficits are seldom evaluated in routine assessments of environmental agents. This section will discuss the format and analysis of conventional studies as well as the use of data from other types of studies, including functional studies, short-term tests, and pharmacokinetics.

A. Conventional Developmental Toxicology Protocols: End Points and Their Interpretation

The most commonly used protocol for assessing developmental toxicity involves the administration of a test substance to pregnant animals (usually mice, rats, or rabbits) during the period of major organogenesis, evaluation of maternal responses throughout pregnancy, and examination of the dam and the uterine contents just prior to term(*2, 3, 14, 15*). Other protocols may use exposure periods of one to a few days to investigate periods of particular sensitivity for induction of anomalies in specific organs or organ systems(*16*). Fetuses alive at maternal sacrifice are thoroughly evaluated for alterations in morphological development. Because the relationship of maternal and fetal toxicity is important in assessing the developmental toxicity of an agent, dose-response data are important. Ideally, study designs should include a high dose, which produces some maternal toxicity (i.e., a level that produces marginal but significantly reduced body weight or weight gain during pregnancy up to a level that produces no more than 10% maternal mortality), a low dose, which demonstrates a no observed effect level (NOEL) for maternal and/or fetal effects, and at least one intermediate dose level. Test animals should be selected based on considerations of species, strain, age, weight, and health status, and should be randomized to dose groups in order to reduce bias and provide a basis for performing valid statistical tests. Replication of the study is desirable and strengthens the confidence of data interpretation.

The next two sections discuss individual end points of maternal and developmental toxicity, respectively, as measured in the conventional developmental toxicity study. The third section deals with the integrated evaluation of all data including the relative effects of exposure on maternal animals and their offspring.

1. End Points of Maternal Toxicity

A number of end points that may be observed as indicators of maternal toxicity are listed in Table 1. Maternal

mortality is an obvious end point of maternal toxicity; however, a number of other end points can be observed which may give an indication of the subtle effects of the agent. For example, in well-conducted studies the end point, percent pregnant, indicates the general fertility rate of the animal stock used and is an important indicator of toxic effects if treatment begins prior to implantation.

Table 1.—End Points of Maternal Toxicity

- Mortality
- Percent Pregnant (includes all litters with implants)
- Body Weight
  - Treatment days (at least first, middle, and last treatment days) Sacrifice day
- Body Weight Change
  - Throughout Gestation
  - During treatment (including increments of time within treatment period)
  - Post-treatment to sacrifice
  - Corrected maternal (body weight change throughout gestation minus gravid uterine weight or litter weight at sacrifice)
- Organ Weights (in cases of suspected specific organ toxicity)
  - Absolute
  - Relative to body weight
- Food and Water Consumption (where relevant)
- Clinical Signs (on days of treatment and at sacrifice)
  - Daily weight changes during treatment
  - Types and incidence of clinical signs

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

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

Changes in other end points should also be determined. For example, changes in relative and absolute organ weights may be signs of maternal effect when an agent is suspected of causing specific organ toxicity. Food and water consumption data are useful, especially if the agent is administered in the diet or drinking water. The amount ingested (total and relative to body weight) and the dose of the agent (relative to body weight) can then be calculated, and changes in food and water consumption with treatment can be evaluated along with changes in body weight and body weight gain. Consumatory data are also useful when an agent is suspected of affecting appetite, water intake, or excretory function. Clinical signs of toxicity may also be used as indicators of maternal toxicity. Daily body weight changes during treatment along with clinical observations may be useful in describing the profile of maternal toxicity.

2. End Points of Developmental Toxicity

Because the maternal animal and not the conceptus is the individual treated during gestation, statistical analysis of the data should consider both the individual fetus and the litter. Table 2 indicates the way in which fetal and litter end points can be expressed.

Table 2.—End Points of Developmental Toxicity

All litters

- No. implantation sites/dam
- No. corpora lutea (CL)/dam*
- Percent Preimplantation loss
- No. and percent live fetuses/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 nonlife implants
- No. and percent affected (nonlive + malformed) implants/litter
- No. and percent with affected implants
- No. and percent litters with total resorptions

Litters with live fetuses

- No. and percent litters with live fetuses
- No. and percent live fetuses/litter
- No. males/litter
- No. females/litter
- No. ratio/litter
- Mean ($\bar{x}$) fetal body weight/litter
- Mean ($\bar{x}$) male body weight/litter
- Mean ($\bar{x}$) female body weight/litter
- No. and percent externally malformed fetuses/litter
- No. and percent viscerally malformed fetuses/litter
- No. and percent skeletally malformed fetuses/litter
- No. and percent malformed fetuses/litter
- No. and percent litters with malformed fetuses
- No. and percent malformed males/litter
- No. and percent malformed females/litter
- No. and percent fetuses with variations/litter
- No. and percent litters having fetuses with variations
- Types and incidence of individual malformations
- Types and incidence of individual variations
- Individual fetuses and their malformations and variations (grouped according to litter and dose)

* Only when treatment begins prior to implantation. May be difficult in mice.

When treatment begins prior to implantation, an increase in preimplatation loss could indicate an adverse effect either on the developing blastocyst or on the process of implantation itself. Further studies would be necessary to determine the cause and extent of this type of effect.

The number of live fetuses per litter, based on all litters, includes any litters that have no live implants. On the other hand, total nonlive implants (postimplantation loss), is a combination of the end points, resorptions, and late fetal deaths. An increased incidence per litter for any of the end points indicating postimplantation loss would be considered a significant toxic effect to the conceptus. The number of litters showing an increased incidence for these end points is less useful than incidence per litter, because a litter is counted whether it has one or all resorbed, dead, or nonlive implants.

A statistically significant increase in postimplantation loss following exposure to an agent is a severe form of developmental toxicity, but there is considerable interlitter variability in the incidence of postimplantation loss(*17*). If a statistically significant increase is found after exposure to an agent, the data may be compared not only with concurrent controls, but also with recent historical control data. If a given study control group exhibits an unusually high or low incidence of postimplantation loss compared to historical controls, then scientific judgment would have to be used to determine the adequacy of the studies for risk assessment purposes.

The end point for affected implants (i.e., the combination of nonlive and

malformed conceptuses) given an indication of the total intrauterine response to an agent and sometimes reflects a better dose-response relationship than each taken individually. This is especially true at the high end of the dose-response curve in cases where most implants die *in utero*. In such cases, the malformation rate may appear to decrease because only unaffected fetuses have survived to term. If the incidence of prenatal death or malformation is unchanged, then the incidence of affected implants will not provide any additional information.

The number of live fetuses per litter, based on those litters that have one or more live fetuses, may be unchanged even though the incidence of nonlive in all litters is increased. This could occur either by an increase in the number of litters with no live fetuses or by an increase in the number of implants per litter. A decrease in the number of live fetuses per litter should be accompanied by an increase in the incidence of nonlive implants per litter, unless the implant numbers differ among dose groups.

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 fetal body weight is a sensitive indicator of developmental toxicity, in part because it is a continuous variable. In some cases, fetal weight reduction may be the only indicator of developmental toxicity: if so, there is always a question remaining as to whether weight reduction is a permanent or transitory effect. When fetal weight reduction is the only indicator of developmental toxicity, data from the two-generation reproduction study(3) may be useful for evaluating these parameters. Ideally, follow-up studies to evaluate postnatal viability, growth, and survival through weaning should be conducted. There are other factors that should be considered in the evaluation of fetal weight changes. For example, in polytocous animals, fetal weight is usually inversely correlated with litter size, and the upper end of the dose-response curve may be confounded by smaller litters and increased fetal weight. Additionally, the average body weight of male fetuses is greater than that of female fetuses in the more commonly used laboratory animals.

Live fetuses should be examined for external, visceral, and skelatal malformations. If only a portion of the litter is examined, then it is preferable that those to be examined be selected on a random basis from each litter. The incidence of individual types of malformations and variations gives an indication of the types of developmental deviations produced by a particular agent. A listing of individual malformations and variations by fetus gives an indication of the pattern of developmental deviations. The incidence of external, visceral, and skeletal malformations gives an indication of which systems may be specifically affected. A significant increase in the incidence of particular malformations or of the total number of fetuses malformed per treated litter as compared with controls indicates a teratogenic effect. If variations are significantly increased in a dose-related manner, these should also be evaluated as a possible indication of developmental toxicity. The Interagency Regulatory Liaison Group noted that dose-related increases in spontaneously occurring defects are as relevant as dose-related increases in any other developmental toxicity end points(*18*). The number and percentage of litters with malformed fetuses are more reliable indicators of developmental toxicity than the number of litters with resorptions, since malformations do not occur frequently in controls. The data on the incidence of individual types of malformations and variations should be examined for significant changes which may be masked if the data on all malformations and variations are pooled. This information can also be used for comparison with historical control data. Appropriate historical control data are helpful in interpretation of major malformations, especially those that normally occur at a low incidence when seen in an individual study apparently unrelated to dose.

3. Overall Evaluation of Maternal and Developmental Toxicity

As discussed previously, individual end points are evaluated in developmental toxicity studies, but an integrated evaluation must be done considering all maternal and developmental end points in order to interpret the data fully. The overall interpretation usually consists of the evaluation of maternal toxicity and the dose levels at which it occurs, then the evaluation of developmental toxicity and the levels at which these end points occur. In general, an agent that produces changes in any of the four major classes of developmental toxicity at a dose that is minimally toxic or not toxic to the maternal animal is considered to have selective developmental effects. However, when effects are produced at maternally toxic doses by agents to which adult human exposure may occur at toxic levels (e.g., smoking, alcohol, solvents), these developmental effects should no be ignored.

Approaches for ranking agents for their selective developmental toxicity are being developed: Schardein(*9*) has reviewed several of these. Of current interest are approaches that develop ratios relating an adult toxic dose to a developmental toxic dose(*19, 20, 21*). Ratios near unity indicate that developmental toxicity occurs only at doses producing maternal toxicity; as the ratio increases, there is a greater likelihood of developmental effects occurring without maternal manifestations. Although further exploration and validation are necessary, such approaches may ultimately help in identifying those agents that pose the greatest threat and should be given priority for further testing(*22*).

B. Functional teratology

Developmental effects, which are inducible by exogenous agents, are not limited to death, structural abnormalities, and altered growth. Rather, it has been demonstrated in a number of instances that subtle alterations in the functional competence of an organ or a variety of organ systems may result from exposure during critical developmental periods that may occur between conception and puberty. Often, these functional defects are observed at dose levels below those at which gross malformations are evident(*23*). Much of the early work in this field was related to behavioral evaluations, and the term "behavioral teratology" became prominent in the mid 1970s. Less work has been done on other functional systems, but sufficient data have accumulated to indicate that the cardiopulmonary, immune, endocrine, digestive, urinary, and reproductive systems are subject to alterations in functional competence. Hence the term "functional teratology" has been applied to this general area.

The variety of systems and end points that may be evaluated is too extensive to discuss here(*24*), (*25*). At present no standard testing procedures are routinely used, and this has led to apparent discrepancies in the outcome of certain studies. Some attempts to standardize and evaluate procedures are being made(*26*). The determination of functional competence often involves highly specialized training and equipment and is not generally practical for routine test procedures. Therefore, these approaches may have their greatest application in determining the nature of a suspected alteration in terms

of its biological significance and dose-response relationship.

The means for appropriate interpretation of data from functional teratology studies is not always clear due to the lack of knowledge about the toxicological significance of specific functional alterations. However, several general concepts have arisen from the research to date which may be useful in designing studies and evaluating data.

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

2. Replication of a study strengthens the confidence of data interpretation.

3. Use of a pharmacological challenge may aid 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.

4. Choice of functional tests with a moderate degree of background variability may be more useful in detecting effects of agent exposure than tests based on functional systems with low variability that may be impossible to disrupt without being life-threatening. Butcher et al.[27] have discussed this with relation to behavioral end points.

5. A battery of functional tests is often necessary to evaluate fully the functional competence of any given system; these tests may need to be conducted at several ages to account for maturational changes.

6. Critical periods for the disruption of functional competence may include both the prenatal period to the time of puberty, and the effect is likely to vary depending on the time of exposure.

Although interpretation of functional data may be difficult at present, there are at least two days in which the data from these studies may be useful for risk assessment purposes. First, these studies can be used to indicate whether or not an agent has the potential to cause functional alterations, and whether these effects occur at doses lower than those that produce other forms of toxicity. Second, if the agent in question is already in the environment, the functional data may be used for focusing on organ systems to evaluate in exposed human populations.

C. Short-Term Testing in Development Toxicity

The need for developmental toxicity screens has arisen from the large number of agents in or entering the environment and the increased interest in reducing the number of animals used in and the expense of testing. Currently, two approaches are being considered for their applicability in the overall testing process: an *in vivo* mammalian screen and a variety of *in vitro* systems. Neither approach is seen at this time as replacing current *in vivo* developmental toxicity testing. Rather, they are being considered for their usefulness in assigning priorities for further, more extensive testing.

1. *In Vivo* Mammalian Teratology Screen

An *in vivo* approach developed by Chernoff and Kavlock[28] uses the pregnant mouse and it designed to reduce the resources required for precliminary indication of developmental toxicity. This approach is based on the hypothesis that a prenatal insult, which results in altered development, will be manifested postnatally as reduced viability and/or impaired growth. In general, the test substance is administered over the period of major organogenesis at a single dose level that will elicit some degree of maternal toxicity. After birth, the pups are counted and weighed on days 1 and 3. End points that are considered in the evaluation include: general maternal toxicity (including survival and weight gain), litter size, viability and weight of the offspring, and gross malformations. Basic priority categories for further testing have also been suggested: (1) Agents that induce perinatal death should receive highest priority, (2) agents inducing perinatal weight changes should be ranked lower in priority, and (3) agents inducing no effect should receive the lowest priority[28]. The major goal of this test is to predict the potential for developmental toxicity of an agent in the species utilized. It does not increase the ability to extrapolate risk to other species, including humans. Additional studies to evaluate the validity of this approach as a screen for developmental toxicity are currently being carried out, and a system for giving a numerical ranking to the results has been suggested to prioritize agents for further testing[29, 30].

2. *In Vitro* Teratology Screens

Test systems that fall under the general heading of "*in vitro*" include any system that employs a test subject other than the intact pregnant mammal. These systems have long been used to assess events associated with normal and abnormal development, but only recently have they been considered for their potential as screens in testing [31, 32, 33]. Many of these systems are now being evaluated for their ability to predict the developmental toxicity of various agents. This validation process requires certain considerations in study design, including defined end points for toxicity and an understanding of the system's ability to handle various test agents[32, 34]. A list of agents for use in these validation studies has been developed[35].

3. Application

When the validity of a screening system is established, it may be used to set priorities for further, more comprehensive *in vivo* testing. In many cases, a battery of two or more screening systems may be needed, employing tests with end points that collectively represent several embryologic processes. In addition, many of these systems can be applied in an attempt to answer specific questions of a dose-response, target-organ, or mechanistic nature. *In vitro* approaches may aid in establishing the effective dose that reaches the target tissue. Either the *in vivo* or *in vitro* short-term approaches may be useful in addressing structure-activity relationships and the synergistic-antagonistic potential of chemical interactions. Thus, pertinent information can be derived from these approaches and may be useful in the assessment of potential risk.

D. Pharmacokinetics

Extrapolation of data between species can be aided considerably by the availability of data on the pharmacokinetics of a particular agent in the species tested and, if possible, in humans. Information on half-lives, placental metabolism and transfer, and concentrations of the parent compound and metabolites in the maternal animal and conceptus may be useful in predicting risk for developmental toxicity. Such data may also be helpful in defining the dose-response curve, developing a more accurate comparison of species sensitivity including that of humans [36, 37], 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 once a developmental toxic effect has been produced in a give species with a particular agent. Pharmacokinetic

data for risk assessment in developmental toxicology ideally should be derived from pregnant females at the stage when developmental insults occur. Often the only data available are from males, nonpregnant females, or from pregnant females at a time unrelated to the event of interest (e.g., pharmacokinetic analyses done during the fetal period when malformations were induced early in organogenesis). The correlation of pharmacokinetic and developmental toxicity data may be useful in determining the contribution of specific pharmacokinetic parameters to the effects observed [38].

E. Human Studies

Because of the ethical considerations involved, little human testing has been or is likely to be done. Therefore, dose-effect developmental toxicity data from humans are generally not available. Human epidemiologic studies may provide the best information for assessing human risk and would reduce the problems in species-to-species extrapolation. However, interpretation of epidemiologic data must account for confounding factors, such as maternal age, parity, multiple exposures to environmental agents, difficulty in obtaining accurate estimates of exposure levels in the environment, insufficient data on background incidence of certain developmental end points, etc. When human data are available, they can be used with other supporting animal data to assess human risk.

F. Comparisons of Molecular Structure

Comparisons of the chemical or physical properties of an agent with those of known developmental toxicants may provide some indication of a potential for developmental toxicity. Such information may be useful in priority-setting of Agents for testing or for further evaluation when only minimal data are available.

G. Weight-of-Evidence Determination

Information available from studies discussed previously, whether indicative of potential concern or not, must be evaluated and factored into the assessment. The types of data may vary from chemical to chemical, and certain types of data may be more relevant than other types of data in performing developmental toxicity assessments. Therefore, all data pertinent to developmental toxicity should be examined in the determination of a chemical's potential to cause developmental toxicity in humans. Whatever evidence may exist from humans must also be factored into the assessment.

IV. Quantitative Assessment

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

A. Dose-Response Assessment

Because human dose-effect data usually are not available, other methods have been used in developmental toxicology for estimating exposure levels that are unlikely to produce adverse effects in humans. The dose-response assessment is usually based upon the evaluation of tests performed in laboratory animals. Two approaches frequently employed involve the use of safety factors and margins of safety, which in some respects are conceptually similar. However, they are computed differently and are often used in different regulatory situations. The choice of approach is dependent upon many factors, including the statute involved, the situation being addressed, the data base used, and the needs of the decision-maker.

The safety factor approach is intended to derive a calculated exposure level that is unlikely to cause any developmental toxic responses in humans. The size of the safety factor will vary from agent to agent and will require the exercise of scientific judgment[3, 39], taking into account interspecies differences, the nature and extent of human exposure, the slope of the dose-response curve, and the severity of the developmental effects observed at exposure levels below maternal toxicity in the test species. The safety factor selected is then divided into the NOEL obtained from the most appropriate and/or sensitive mammalian species examined to obtain an acceptable exposure level. Currently, there is no one laboratory animal species that can be considered most appropriate for predicting risk to humans[9]. Each agent should be considered on a case-by-case basis.

The margin of safety approach derives a ratio of the NOEL from the most sensitive species to the estimated human exposure level from all potential sources[40]. The adequacy of the margin of safety is then considered, based upon the weight of evidence, including quality of data, number of species affected, dose-response relationships, and other factors such as benefits of the agent.

As discussed earlier, the preferred study design for a developmental toxicity study includes a minimum of three doses: a high dose that produces minimal maternal toxicity, at least one intermediate dose, and a low dose that demonstrates a NOEL. Nevertheless, there may be circumstances in that there is a need to perform a risk assessment based on the results of a study in which a NOEL could not be identified, but, rather, in which the lowest dose administered caused some marginally significant effect(s). This lowest dose could be identified as the lowest observed effect level (LOEL). In circumstances where a LOEL can be identified, it may be appropriate to apply an additional safety factor. The magnitude of this additional factor is dependent upon scientific judgment. In some instances, additional studies may be needed to strengthen the confidence in this additional safety factor.

B. Exposure Assessment

The results of the dose-response assessment are combined with an estimate of human exposure in order to obtain a quantitative estimate of risk. The proposed Guidelines for Exposure assessment are being developed separately and will not be discussed in any detail here. In general, the exposure assessment describes the magnitude, duration, schedule, and route of exposure. This information is developed from monitoring data and from estimates based on modeling of environmental exposures. Unique considerations relevant to developmental toxicity are duration and period of exposure as related to stage of gestation (i.e., critical periods), and the fact that a single exposure may be sufficient to produce adverse developmental effects (i.e., chronic exposure is not necessary for developmental toxicity to be manifested).

C. Risk Characterization

There are numerous uncertainties associated with the toxicological and exposure components of risk assessment that in the past have often not been

readily apparent or consistently presented. The presentation of any qualitative or quantitative risk assessment for developmental toxicity should be accompanied by statements concerning the quality of the data, resolving power of the studies, number of end points examined, selection of doses, replication of the data, the number of species examined, pharmacokinetic considerations, and any other factors that affect the quality and precision of the assessment. The presentation of any numerical estimate should be sufficiently qualified as to the assumptions used and the accuracy of the estimates.

In the assessment of developmental toxicity, statistical considerations require special attention. For example, the power of a study (i.e., the ability to demonstrate an effect), is limited by the sample size used in the study, the background incidence of the end point observed, and the variability in the incidence of the end point. As an example, Nelson and Holson[41] have shown that the number of litters needed to detect a 5 or 10 percent 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 protocol being 20 rodents per dose group[1, 3], it is possible to detect an increased incidence of malformations in the range of 5 to 12 times above control levels, an increase of 3 to 6 times the *in utero* death rate, and a decrease of 0.15 to 0.25 times the fetal weight. Thus, even within the same study, the ability to detect a change in fetal weight is much greater than for the other end points measured. Consequently, for statistical reasons only, changes in fetal weight are often observable at doses below those producing other signs of developmental toxicity.

At present, there is no mathematical model that is generally used for estimating developmental toxicity responses below the applied dose range. This is due primarily to the lack of understanding of the biological mechanisms underlying developmental toxicity, intra/interspecies differences in the types of developmental events, the influence of maternal effects on the dose-response curve, and whether or not a threshold exists below which no effect will be produced by an agent. The assumption of a threshold is based largely on the biological rationale that the embryo is known to have some capacity for repair of the damage or insult[42], and that most developmental deviations are probably multifactorial in nature[43]. However, the existence of a no effect level cannot be proven statistically.

Discussions of risk extrapolation procedures have noted that further work is needed to improve mathematical tools for developing estimates of potential human developmental risk[39, 44]. Gaylor[45] has suggested an approach for controlling risk that combines the use of mathematical models for low-dose estimation of risk with the application of a safety factor based on a preselected level of allowable risk. This approach is similar to approaches proposed for carcinogenesis, but does not preclude the possibility of a threshold, and may provide a more quantitative approach to controlling risk. For the present, the Agency will continue to use safety factors and margins of safety as described above, where applicable. However, more appropriate models will be sought and applied if considered acceptable.

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

V. References

(1) U.S. Environmental Protection Agency. 1982. Health effects test guidelines, Chapter II. Specific organ/tissue toxicity-teratogenicity. Office of Toxic Substances. Available from: NTIS, Springfield, VA. PB82-232984.

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

(3) U.S. Environmental Protection Agency. 1982. Pesticides registration: proposed data requirements. Federal Register 47:53192–53203.

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

(5) McKeown, T., and R.C. Record. 1963. Malformations in a population observed for five years after birth. In: G.E.W. Wolstenholme and C.M. O'Conner, eds., Ciba Foundation symposium on congenital malformations. Boston, MA: Little Brown, pp 2–18.

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

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

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

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

(10) Shepard, T.H. 1984. Teratogens: an update. Hosp. Pract., Jan., pp. 191–200.

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

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

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

(14) Food and Drug Administration. 1966. Guidelines for reproduction and teratology of drugs. Bureau of Drugs.

(15) 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.

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

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

(18) Interagency Regulatory Liaison Group. 1981. Report of the Developmental Toxicity End Points Workgroup. Workshop on Reproductive Toxicity Risk Assessment. Rockville, MD. September 21–23.

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

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

(21) Fabro, S., G. Schull, and N.A. Brown. 1982. The relative teratogenic index and teratogenic potency: proposed components of

developmental toxicity risk assessment. Teratogenesis Carcinog. Mutagen. 2:61–76.

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

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

(24) Buelke-Sam, J., and C.A. Kimmel. 1979. Development and standardization of screening methods for behavioral teratology. Teratology 20:17–29.

(25) Kavlock, R.J., and C.T. Grabowski, eds. 1983. Abnormal functional development of the heart, lungs, and kidneys: approaches to functional teratology. Prog. Clin. Biol. Res., Vol. 140. New York, NY: Alan R. Liss, Inc.

(26) Kimmel, C.A., J. Buelke-Sam, J. Adams, L.W. Reiter, T.J. Sobotka, and H.A. Tilson. 1982. Design considerations in the evaluation of standardized methods in a behavioral teratology study. Teratology 25:54A.

(27) Butcher, R.E., V. Wooten, and C.V. Vorhees. 1980. Standards in behavorial teratology testing: test variability and sensitivity. Teratogenesis Carcinog. Mutagen. 1:49–61.

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

(29) Brown, J.M. 1984. Validation of an *in vivo* screen for the determination of embryo/fetal toxicity in mice. SRI International. EPA contract no. 68–01–6079.

(30) Schuler R., B. Hardin, R. Niemeyer, G. Booth, K. Hazelden, V. Piccirillo, and K. Smith. 1984. Results of testing fifteen glycol ethers in a short-term, *in vivo* reproductive toxicity assay. Environ. Health. Perspect., in press.

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

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

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

(34) Kimmel, G.L. 1984. *In vitro* tests in screening teratogens: considerations to aid the validation process. In: M. Morris, ed. Prevention of physical and mental congenital defects. Part A. New York, NY: Alan R. Liss, Inc., in press.

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

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

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

(38) Kimmel, C.A., and J.F. Young. 1983. Correlating pharmacokinetics and teratogenic end points. Fund. Appl. Toxicol. 3:250–255.

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

(40) Chitlik, L.D., Q.Q. Bui, G.J. Burin, and S.C. Dapson. 1984. Standard evaluation procedures for teratology studies (Draft). Toxicology Branch, Hazard Evaluation Division, Office of Pesticide Programs, U.S. Environmental Protection Agency.

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

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

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

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

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

[FR Doc. 84-30721 Filed 11-21-84; 8:45 am]
BILLING CODE 6560-50-M