

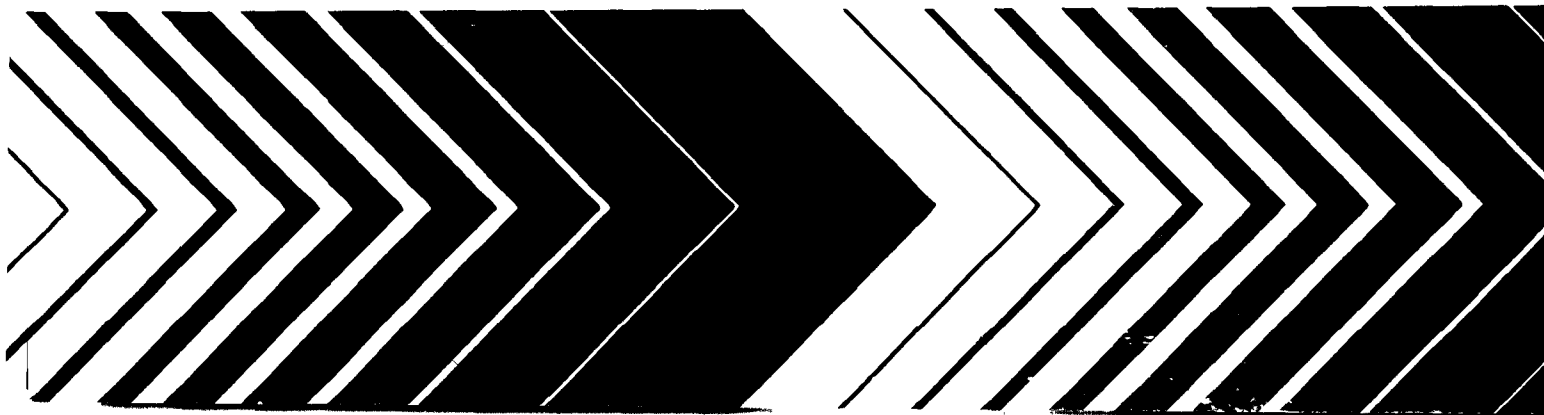


# **Guidelines for Reproductive Toxicity Risk Assessment**

**Review  
Draft**  
(Do Not  
Cite or  
Quote)

## **Notice**

This document is a preliminary draft. It has not been formally released by EPA and should not at this stage be construed to represent Agency policy. It is being circulated for comment on its technical accuracy and policy implications.



DRAFT  
DO NOT QUOTE OR CITE

EPA/600/AP-94/001  
February 1994  
External Review Draft

CLS # 2222

## **Guidelines for Reproductive Toxicity Risk Assessment**

### **NOTICE**

THIS DOCUMENT IS A PRELIMINARY DRAFT. It has not been formally released by the U.S. Environmental Protection Agency and should not at this stage be construed to represent Agency policy. It is being circulated for comment on its technical accuracy and policy implications.

Office of Health and Environmental Assessment  
Office of Research and Development  
U.S. Environmental Protection Agency  
Washington, D.C.

U.S. Environmental Protection Agency  
Region 5, Library  
77 West Jackson Boulevard, 12th Floor  
Chicago, IL 60604-3590



*Printed on Recycled Paper*

DRAFT--DO NOT QUOTE OR CITE

**DISCLAIMER**

This document is an external draft for review purposes only and does not constitute Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

DRAFT--DO NOT QUOTE OR CITE

TABLE OF CONTENTS

List of Tables .....	
Authors, Contributors, and Reviewers .....	
SUPPLEMENTAL INFORMATION .....	1
A.    REGULATORY AUTHORITY .....	1
B.    ENVIRONMENTAL AGENTS AND REPRODUCTIVE TOXICITY .....	2
C.    THE RISK ASSESSMENT PROCESS AND ITS APPLICATION TO REPRODUCTIVE TOXICITY .....	5
I.    OVERVIEW .....	9
II.   DEFINITIONS AND TERMINOLOGY .....	12
III.  HAZARD IDENTIFICATION/DOSE-RESPONSE EVALUATION OF REPRODUCTIVE TOXICANTS .....	13
A.    LABORATORY TESTING PROTOCOLS .....	13
A.1.  Introduction .....	13
A.2.  Duration of Dosing .....	13
A.3.  Length of Mating Period .....	14
A.4.  Number of Females Mated to Each Male .....	15
A.5.  Single- and Multigeneration Reproduction Tests .....	15
A.6.  Alternative Reproductive Tests .....	18
A.7.  Additional Test Protocols That May Provide Reproductive Data .....	20
B.    ENDPOINTS FOR EVALUATING MALE AND FEMALE REPRODUCTIVE TOXICITY IN TEST SPECIES .....	23
B.1.  Introduction .....	23
B.2.  Couple-mediated Endpoints .....	24
a.    Fertility and Pregnancy Outcomes .....	24
b.    Sexual Behavior .....	33
B.3.  Male-specific Endpoints .....	34
a.    Introduction .....	34
b.    Body Weight and Organ Weights .....	35
c.    Histopathologic Evaluations .....	38
d.    Sperm Evaluations .....	41
e.    Endocrine Evaluations .....	47
f.    Biochemical Tests or Markers of Toxicity to the Testes and Other Male Reproductive Organs .....	48
g.    Paternally-mediated Effects on Offspring .....	49

**TABLE OF CONTENTS (continued)**

B.4.	Female-specific Endpoints . . . . .	49
a.	Introduction . . . . .	49
b.	Body Weight, Organ Weight, Organ Morphology, and Histology .	52
c.	Oocyte Production . . . . .	58
d.	Alterations in the Female Reproductive Cycle . . . . .	60
e.	Mammary Gland and Lactation . . . . .	62
f.	Developmental and Pubertal Alterations . . . . .	63
g.	Reproductive Senescence . . . . .	64
C.	HUMAN STUDIES . . . . .	65
C.1.	Epidemiologic studies . . . . .	66
a.	General Design Considerations . . . . .	66
b.	Selection of Outcomes for Study . . . . .	70
c.	Reproductive History Studies . . . . .	74
d.	Community Studies and Surveillance Programs . . . . .	76
e.	Identification of Important Exposures for Reproductive Effects .	77
C.2.	Examination of Clusters, Case Reports, or Series . . . . .	78
D.	PHARMACOKINETIC CONSIDERATIONS . . . . .	79
E.	COMPARISONS OF MOLECULAR STRUCTURE . . . . .	81
F.	EVALUATION OF DOSE-RESPONSE RELATIONSHIPS . . . . .	82
G.	CHARACTERIZATION OF THE HEALTH-RELATED DATA BASE .	86
H.	DETERMINATION OF THE REFERENCE DOSE OR REFERENCE CONCENTRATION FOR REPRODUCTIVE TOXICITY . . . . .	94
I.	SUMMARY . . . . .	97
IV.	EXPOSURE ASSESSMENT . . . . .	97
V.	RISK CHARACTERIZATION . . . . .	102
A.	OVERVIEW . . . . .	102
B.	INTEGRATION OF HAZARD IDENTIFICATION/DOSE-RESPONSE AND EXPOSURE ASSESSMENTS . . . . .	104
C.	DESCRIPTORS OF REPRODUCTIVE RISK . . . . .	107
C.1.	Estimation of the Number/Proportion of Individuals Exposed to Levels Above the RfD or RfC . . . . .	107
C.2.	Presenting Situation-Specific Exposure Scenarios . . . . .	107
C.3.	Margin of Exposure . . . . .	108
C.4.	Risk Characterization for Highly Exposed Individuals . . . . .	108
C.5.	Risk Characterization for Highly Sensitive or Susceptible Individuals .	109
D.	SUMMARY AND RESEARCH NEEDS . . . . .	109
VI.	REFERENCES . . . . .	111

DRAFT--DO NOT QUOTE OR CITE

**LIST OF TABLES**

1.	Couple-mediated Endpoints of Reproductive Toxicity . . . . .	25
2.	Selected Indices That May Be Calculated From Endpoints of Reproductive Toxicity in Test Species . . . . .	27
3.	Male-Specific Endpoints of Reproductive Toxicity . . . . .	36
4.	Female-Specific Endpoints of Reproductive Toxicity . . . . .	51
5.	Categorization of the Health-Related Data Base Hazard Identification/Dose-Response Evaluation . . . . .	89

DRAFT--DO NOT QUOTE OR CITE

**AUTHORS AND MANAGERS**

This external review draft was prepared by an intra-Agency EPA working group chaired by Eric Clegg of the Office of Health and Environmental Assessment.

**DOCUMENT MANAGER**

Eric Clegg  
Office of Health and Environmental Assessment  
Office of Research and Development

## **DRAFT--DO NOT QUOTE OR CITE**

### **SUPPLEMENTAL INFORMATION:**

#### **A. REGULATORY AUTHORITY**

The Environmental Protection Agency is authorized by numerous statutes, including the Toxic Substances Control Act (TSCA), the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), the Clean Air Act, the Safe Drinking Water Act, and the Clean Water Act, to regulate environmental agents that have the potential to adversely affect human health, including the reproductive system. These statutes are implemented through offices within the Agency. The Office of Pesticide Programs and the Office of Pollution Prevention and Toxics within the Agency have issued testing guidelines (U.S. Environmental Protection Agency, 1982, 1985b) that provide protocols (now under review) designed to determine the potential of a test substance to produce reproductive (including developmental) toxicity in laboratory animals. The Organization for Economic Cooperation and Development (OECD) also has issued testing guidelines for reproduction studies (Organization for Economic Cooperation and Development, 1993b).

The procedures outlined here in the Guidelines for Reproductive Toxicity Risk Assessment (hereafter called Guidelines) provide guidance for interpreting, analyzing, and using the data from studies that follow the above testing guidelines. In addition, the Guidelines provide information for interpretation of other studies and endpoints (e.g., evaluations of epidemiologic data, measures of sperm production, reproductive endocrine system function, sexual behavior, female reproductive cycle normality) that have not been required routinely, but may be required subsequently or may be encountered in reviews of data on particular agents. The Guidelines will promote consistency in the Agency's assessment of toxic effects on the male and female reproductive systems, including outcomes of pregnancy and lactation, and inform others of approaches that the Agency will use in assessing those risks. Guidance is also provided by the Guidelines for Developmental Toxicity Risk Assessment (U.S. Environmental Protection Agency, 1991) and the Guidelines for Mutagenicity Risk Assessment (U.S. Environmental Protection Agency, 1986c). These three guidelines are complementary.



## **DRAFT--DO NOT QUOTE OR CITE**

Reproductive toxicity risk assessments prepared pursuant to these Guidelines will be used within the requirements and constraints of the applicable statutes to arrive at regulatory decisions concerning reproductive toxicity. These Guidelines do not change any statutory or regulatory prescribed standards, such as those specified by Agency or OECD testing guidelines for the types of data necessary for regulatory action.

The Agency has sponsored or participated in several conferences that addressed issues related to evaluations of reproductive toxicity data and that provide some of the scientific basis for these risk assessment guidelines. Numerous publications from these and other efforts are available that provide background for these Guidelines (U.S. Environmental Protection Agency, 1982, 1985b; Galbraith et al., 1983; Organization for Economic Cooperation and Development, 1983; U.S. Congress, 1985, 1988; Kimmel, C.A. et al., 1986; Francis & Kimmel, 1988; Burger et al., 1989; Sheehan et al., 1989). Also, numerous resources provide background information on the physiology, biochemistry, and toxicology of the male and female reproductive systems (Lamb & Foster, 1988; Working, 1989; Russell et al., 1990; Atterwill & Flack, 1992; Scialli & Clegg, 1992; Chapin & Heindel, 1993; Heindel & Chapin, 1993; Manson & Kang, 1994; Zenick et al., 1994; Kimmel, G.L. et al., In press). A comprehensive text on reproductive biology has been published (Knobil & Neill, 1988).

### **B. ENVIRONMENTAL AGENTS AND REPRODUCTIVE TOXICITY**

Disorders of reproduction and hazards to reproductive health have become prominent public health issues. The perception of risk leads to a corresponding perception that action is needed to protect people from hazards and prevent disorders of reproduction. Disorders of reproduction in humans include but are not limited to reduced fertility, impotence, menstrual disorders, spontaneous abortion, low birth weight and other developmental (including heritable) defects, premature reproductive senescence and various genetic diseases affecting the reproductive system and offspring.

## **DRAFT--DO NOT QUOTE OR CITE**

The prevalence of infertility, which is defined clinically as the failure to conceive after one year of unprotected intercourse, is difficult to estimate. National surveys have been conducted to obtain demographic information about infertility in the United States (Mosher & Pratt, 1990). In their 1988 survey, an estimated 4.9 million women aged 15-44 (8.4%) had impaired fertility. The proportion of married couples that was infertile was 7.9%. Of major concern is the report that human sperm concentration has declined from  $113 \times 10^6$  per ml of semen prior to 1960 to  $66 \times 10^6$  per ml subsequently (Carlsen et al., 1992). When combined with a decline in semen volume from 3.4 ml to 2.75 ml, that indicates a decline in total number of sperm of approximately 50%. Increased incidences of human male hypospadias, cryptorchidism and testicular cancer have also been indicated over the last 50 years (Giwerzman et al., 1993).

Even though not all infertile couples seek treatment, and infertility is not the only adverse reproductive effect, it is estimated that Americans spent about \$1 billion in 1986 on medical care to treat infertility alone (U. S. Congress, 1988). With the increased use of assisted reproduction techniques, that amount has increased substantially since 1986.

Disorders of the male or female reproductive system may also be manifested as adverse outcomes of pregnancy. For example, it has been estimated that approximately 50% of human conceptuses fail to reach term (Hertig, 1967). Methods that detect pregnancy as early as 9 days after conception have suggested that 35% of postimplantation pregnancies end in embryonic or fetal loss (Wilcox et al., 1985). Approximately 3% of newborn children have one or more significant congenital malformations at birth, and by the end of the first postnatal year, about 3% more are recognized to have serious developmental defects (Shepard, 1986). Of these, it is estimated that 20% are of known genetic transmission, 10% are attributable to known environmental factors, and the remaining 70% result from unknown causes (Wilson, 1977). Also, approximately 7.4% of children have low birth weight (i.e., below 2.5 kg) (Selevan, 1981).

Numerous agents have been shown to be reproductive toxicants in male and female laboratory animals and in humans (Mattison, 1985; Schrag & Dixon, 1985a, b; Waller et

## **DRAFT--DO NOT QUOTE OR CITE**

al., 1985; Lewis, 1991). For example, neonatal or peripubertal exposure to environmental compounds that possess steroidogenic (e.g., diethylstilbestrol; Steinberger & Lloyd, 1985) or antisteroidogenic (Schardein et al., 1985) activity affect the onset of puberty and reproductive function in adulthood. In adult males and females, exposure to agents of abuse such as cocaine disrupts normal reproductive function in both test species and humans (Smith, C.G. & Gilbeau, 1985). Numerous chemicals disrupt ovarian cyclicity, alter ovulation, and impair fertility in experimental animals and humans. These include agents with steroidogenic activity, certain pesticides and herbicides, and some metals (Thomas, 1981; Mattison, 1985). In males, estrogenic compounds can be testicular toxicants in rodents and humans (Colborn et al., 1993). Dibromochloropropane (DBCP) impairs spermatogenesis in both experimental animals and humans by another mechanism. These and other examples of toxicant-induced effects on reproductive function have been reviewed (Katz & Overstreet, 1981; Working, 1988).

Altered reproductive health is often manifested as an adverse effect on the reproductive success or sexual behavior of the couple even though only one of the pair may be affected directly. Often, it is difficult to discern which partner has reduced reproductive capability. For example, exposure of the male to an agent that reduces the number of normal sperm may result in reduced fertility in the couple, but without further diagnostic testing the affected partner may not be identified. Also, adverse effects on the reproductive systems of the two sexes may not be detected until a couple attempts to conceive a child.

For successful reproduction, it is critical that the biologic integrity of the human reproductive system be maintained. For example, the events in the estrous or menstrual cycle are closely interrelated; changes in one event in the cycle can alter other events. Thus, a short or inadequate luteal phase of the menstrual cycle is associated with disorders in ovarian follicular steroidogenesis, gonadotropin secretion, and endometrial integrity (Scommegna et al., 1980; Smith, S.K. et al., 1984; Sakai & Hodgen, 1987). Toxicants may interfere with luteal function by altering hypothalamic or pituitary function and by affecting ovarian response (La Bella et al., 1973a, b).

## DRAFT--DO NOT QUOTE OR CITE

Fertility of the human male is particularly susceptible to agents that reduce the number or quality of sperm produced. Compared with many other species, human males produce fewer sperm relative to the number of sperm required for fertility (Amann, 1981; Working, 1988). As a result, many men are subfertile or infertile (Amann, 1981). The incidence of infertility in men is considered to increase at sperm concentrations below  $20\text{--}40 \times 10^6$  sperm per milliliter of ejaculate. As the concentration of sperm drops below that level, the probability of a pregnancy resulting from a single ejaculation declines. If the number of normal sperm per ejaculate is sufficiently low, fertilization is unlikely, and an infertile condition exists. Toxic agents may further decrease production of sperm and increase risk of impaired fertility.

Chemical or physical agents can affect the female and male reproductive systems at any time in the life cycle, including susceptible periods in development. The reproductive system begins to form early in gestation, but structural and functional maturation is not completed until puberty. Exposure to toxicants early in development can lead to alterations that may affect reproductive function or performance well after the time of initial exposure. Adverse effects such as reduced fertility in offspring may appear as delayed consequences of in utero exposure to toxicants. Effects of toxic agents on other parameters such as sexual behavior, reproductive cycle normality, or gonadal function can also alter fertility (Chapman, 1983; Dixon & Hall, 1984; Schrag & Dixon, 1985b; U.S. Congress, 1985).

### C. THE RISK ASSESSMENT PROCESS AND ITS APPLICATION TO REPRODUCTIVE TOXICITY

Risk assessment is the process that defines the potential adverse health consequences of exposure to a toxic agent. The National Research Council (NRC) of the National Academy of Sciences defines risk assessment as comprising some or all of the following components: hazard identification, dose-response assessment, exposure assessment, and risk characterization (National Research Council, 1983). This approach was derived for carcinogens, but in general, this process is appropriate for human reproductive risk

## **DRAFT--DO NOT QUOTE OR CITE**

assessment as well. However, several significant factors lead to certain differences in risk assessment for reproductive effects for which a threshold is assumed. Below a threshold level of exposure, a chemical is not considered to be a reproductive hazard. Related factors that are significant include the process of defining an adverse effect, consideration of reversibility of effects, and consideration of target effects in the presence of other toxic effects. In practice, hazard identification for reproductive effects usually includes an evaluation of dose-response relationships, because the determination of a hazard may be dependent on whether a dose-response relationship exists and, ideally, on whether information is available on potential human exposures. A reproductive hazard for humans is defined in terms of the range of effective doses, route of exposure, timing and duration of exposure, and other relevant factors. For this reason, these Guidelines present hazard identification and dose-response evaluation together (Sections III.F and III.G.) and include both in the characterization of the health-related data as sufficient or insufficient to proceed with a quantitative risk assessment. If data are sufficient for quantitative risk assessment, a reference dose (RfD), reference concentration (RfC), or margin of exposure (MOE) for reproductive toxicity can then be derived for comparison with human exposure estimates (Section III.H.).

As discussed more fully in Section V (Risk Characterization), the components of the risk assessment process are not considered in isolation. Appreciation of the potential for human risk comes in part from integration of the hazard identification and dose-response evaluation with the human exposure estimates in the final risk characterization. The final assessment of risk depends on full consideration of all of these factors.

Hazard identification/dose-response evaluation involves the evaluation of all available human and experimental animal data on the effects of exposures as well as the associated doses, routes, durations, and patterns of exposure to determine if an agent is likely to cause reproductive toxicity. In addition to reproductive effects, all other manifestations of toxicity are examined in describing the effects for a given exposure. This description includes evaluating the relationships between endpoints at a given dose as well

## **DRAFT--DO NOT QUOTE OR CITE**

as the progression of toxicity across doses. Adequacy of the health-related data to be used further for quantitative risk assessment is characterized then, based on criteria defining a sufficient data base outlined in these Guidelines (Section III.G.).

The evaluation of dose-response relationships includes the identification of effective dose levels as well as doses associated with low or no increased incidence of adverse effects when compared with controls. Ideally, a dose-response relationship would be established from human epidemiologic data that include the expected levels of exposure. Such data are seldom available. When the data are limited to test species, the relevance of the test system to humans must be considered. The Agency typically uses a dose-response approach in which uncertainty factors are applied to the no-observed-adverse-effect level (NOAEL), or lowest-observed-adverse-effect level (LOAEL) if a NOAEL is not available, to extrapolate from experimental animals to humans and to compensate for variability within the human population. Because of limitations associated with the use of the NOAEL, the Agency also is evaluating an alternative approach to quantitative dose-response evaluation, i.e., the benchmark dose (Crump, 1984) (Section III.F.). Uncertainty factors would be applied also to a benchmark dose. In either case, the value derived is a RfD or RfC. These Guidelines discuss these approaches (Section III.H.).

The hazard identification/dose-response evaluation concludes with a determination of the sufficiency of the health-related data base to assess potential human risk (Section III.G.). This determination reflects the confidence of the risk assessor in the data base with respect to the evidence for causation and ability to estimate effects at low doses. However, it is only when the risk assessment process is carried through the remaining phases that the actual risk to humans can be estimated for the exposure conditions that human populations may experience. Thus, although an agent may cause an adverse effect in laboratory animals, the level of potential human exposure should be evaluated before the agent is considered to have potential risk to the human reproductive system. In the absence of human exposure information, potential for human toxicity may be assumed from test species results that show reproductive toxicity.

## **DRAFT--DO NOT QUOTE OR CITE**

Exposure assessment identifies and describes populations exposed or potentially exposed to an agent, and presents the type, magnitude, frequency, and duration of such exposures. Those procedures are considered separately in the Guidelines for Exposure Assessment (U.S. Environmental Protection Agency, 1992). However, unique considerations for reproductive toxicity exposure assessments are detailed in Section IV.

In risk characterization, the hazard identification/dose-response evaluation and the exposure assessment are combined to estimate the risk of human reproductive toxicity. As part of risk characterization, the strengths and weaknesses in each component of the risk assessment are summarized along with major assumptions, scientific judgments, and to the extent possible, qualitative descriptions and quantitative estimates of the uncertainties. The sufficiency of the health-related data is presented with information on dose-response, the RfD or RfC, and if available, the human exposure estimate. Here the NOAEL (or the benchmark dose if adopted for Agency use) and the estimated human exposure levels may be compared in a ratio to provide a margin of exposure (MOE). The considerations for evaluating the MOE are similar to those used in determining the appropriate size of the uncertainty factor for calculating the RfD or RfC.

Risk assessment is just one component of the regulatory process. The other component, risk management, uses the risk characterization along with the directives of the enabling regulatory legislation and other factors to decide whether to control exposure to the suspected agent and the level of control. The risk management decisions also consider socioeconomic, technical, and political factors. Risk management is not discussed directly in these guidelines.

## **DRAFT--DO NOT QUOTE OR CITE**

### **I. OVERVIEW**

These Guidelines for Reproductive Toxicity Risk Assessment (hereafter called Guidelines) describe the procedures that the U.S. Environmental Protection Agency (EPA; Agency) will follow in using existing data to evaluate the potential toxicity of environmental agents to the human male and female reproductive systems and to outcomes of pregnancy. These Guidelines focus on reproductive function as it relates to sexual behavior, fertility, pregnancy outcomes, and lactating ability, and the processes that can affect those functions directly. Included are effects on gametogenesis and gamete maturation and function, the reproductive organs, and the components of the endocrine system that directly support those functions. These Guidelines concentrate on the integrity of the male and female reproductive systems as required to ensure successful procreation. They also emphasize the importance of maintaining the integrity of the reproductive system for overall physical and psychologic health. The Guidelines for Developmental Toxicity Risk Assessment (U.S. Environmental Protection Agency, 1991) focus on effects of agents on development specifically, and should be used as a companion to these Guidelines.

In evaluating reproductive effects, it is important to consider the presence, and where possible, the contribution of other manifestations of toxicity such as developmental toxicity, mutagenicity, or carcinogenicity. The reproductive process is such that these areas overlap, and all should be considered in reproductive risk assessments.

Although the endpoints discussed in these Guidelines can detect impairment to components of the reproductive process, they may not discriminate effectively between nonmutagenic (e.g., cytotoxic) and mutagenic mechanisms. Examples of endpoints affected by either type of mechanism are sperm head morphology and preimplantation loss. If the effects seen may result from mutagenic events, then there is the potential for transmissible genetic damage. In such cases, the Guidelines for Mutagenicity Risk Assessment (U.S. Environmental Protection Agency, 1986c) should be consulted in conjunction with this document. The Guidelines for Cancer Risk Assessment (currently under review) (U.S.



## **DRAFT--DO NOT QUOTE OR CITE**

Environmental Protection Agency, 1986a) should be consulted if reproductive system or developmentally-induced cancer is detected.

For assessment of risk to the human reproductive systems, the most appropriate data are those derived from human studies. In the absence of adequate human data, our understanding of the mechanisms controlling reproduction supports the use of data from experimental animal studies to estimate the risk of reproductive effects in humans. However, some information needed for extrapolation of data from experimental animal studies to humans is not generally available. Therefore, to bridge these gaps in information, a number of assumptions are made. These assumptions should not preclude inquiry into the relevance of the data to potential human risk. These assumptions provide the inferential basis for the approaches to risk assessment in these Guidelines. Each assumption should be evaluated along with other relevant information in making a final judgement as to human risk for each agent, and that information summarized in the risk characterization.

First, an agent that produces an adverse reproductive effect in experimental animal studies is assumed to pose a potential reproductive hazard to humans. This assumption is based on comparisons of data for known human reproductive toxicants (Thomas, 1981; Nisbet & Karch, 1983; Kimmel, C.A. et al., 1984, 1990; Hemminki & Vineis, 1985; Meistrich, 1986; Working, 1988). In general, the experimental animal data indicated adverse reproductive effects that are also seen in humans.

Because similar mechanisms can be identified in the male and female of many mammalian species, effects of xenobiotics on male and female reproductive processes are assumed generally to be similar across species unless demonstrated otherwise. However, for effects on pregnancy outcomes, it is assumed that the effects seen in experimental animal studies are not necessarily the same as those produced in humans. This latter assumption is made because every species may not react in the same way because of species-specific differences in timing of exposure relative to critical periods of development, pharmacokinetics (including metabolism), developmental patterns, placentation, or mechanisms of action.

## **DRAFT--DO NOT QUOTE OR CITE**

When sufficient data are available (e.g., pharmacokinetic), the most appropriate species should be used to estimate human risk. In the absence of such data, it is assumed that the most sensitive species is most appropriate because, for the majority of known human reproductive toxicants, humans appear to be as or more sensitive than the most sensitive animal species tested (Nisbet & Karch, 1983; Kimmel, C.A. et al., 1984, 1990; Hemminki & Vineis, 1985; Meistrich, 1986; Working, 1988), based on data from studies that determined dose on a body weight or air concentration basis.

In the absence of specific information to the contrary, it is assumed that a chemical that acts as a reproductive toxicant in one sex may also adversely affect reproductive function in the other sex. This assumption for reproductive risk assessment is based on three considerations: (1) For most agents, the nature of the testing and the data available are limited, reducing confidence that the potential for toxicity to both sexes and their offspring has been examined equally; (2) Exposures of either males or females have resulted in developmental toxicity, although studies of developmental effects resulting from male exposures have been limited (Davis et al., 1992); and (3) Many of the mechanisms controlling important aspects of reproductive system function are similar in females and males, and could therefore be susceptible to the same agents. Information that could negate this assumption would demonstrate (1) that a mechanistic difference existed between the sexes that would preclude toxic action on the other sex or (2) that, on the basis of sufficient testing, an agent did not produce an adverse reproductive effect when administered to the other sex.

In general, a threshold is assumed for the dose-response curve for reproductive toxicity. This is based on the known capacity of cells, tissues, and organs of the reproductive systems and the developing organism to compensate for or to repair a certain amount of damage. Furthermore, multiple insults at the molecular or cellular level may be required to produce an adverse effect. Therefore, although the levels of exposure at which different individuals react adversely to an agent may differ, each individual should have an exposure level below which no increased risk exists.

## **DRAFT--DO NOT QUOTE OR CITE**

### **II. DEFINITIONS AND TERMINOLOGY**

The following terms are defined according to their usage in this document:

Reproductive toxicity - The occurrence of adverse effects on the reproductive systems that may result from exposure to environmental agents. The toxicity may be expressed as alterations to the female or male reproductive organs, the related endocrine system, or pregnancy outcomes. The manifestation of such toxicity may include, but not be limited to, adverse effects on onset of puberty, gamete production and transport, reproductive cycle normality, sexual behavior, fertility, gestation, parturition, lactation, pregnancy outcomes, premature reproductive senescence, or modifications in other functions that are dependent on the integrity of the reproductive systems.

Fertility - The ability to conceive and to produce offspring within a defined period of time. For litter-bearing species, the number of offspring per litter is also a measure of fertility.

Fertile - Having a level of fertility that is within or exceeds the normal range for that species.

Infertile - Lacking fertility for a specified period. The infertile condition may be temporary; permanent infertility is termed sterility.

Subfertile - Having a level of fertility that is below the normal range for that species but not infertile.

Developmental toxicity - The occurrence of adverse effects on the developing organism that may result from exposure prior to conception (either parent), during prenatal development, or postnatally to the time of sexual maturation. Adverse developmental effects may be detected at any point in the life span of the organism. The major manifestations of developmental toxicity include (1) death of the developing organism, (2) structural abnormality, (3) altered growth, and (4) functional deficiency (U.S. Environmental Protection Agency, 1991).

## **DRAFT--DO NOT QUOTE OR CITE**

### **III. HAZARD IDENTIFICATION/DOSE-RESPONSE EVALUATION OF REPRODUCTIVE TOXICANTS**

This section presents the traditional testing protocols for rodents and endpoints used to evaluate male and female reproductive toxicity along with evaluation of their strengths and limitations. Because many endpoints are common to multiple protocols, endpoints are considered separately from the discussion of the overall protocol structures.

#### **III.A. LABORATORY TESTING PROTOCOLS**

##### **III.A.1. Introduction**

Testing protocols describe the procedures to be used to provide data for risk assessments. The quality and usefulness of those data are dependent on the design and conduct of the tests, including endpoint selection and resolving power. A single protocol is unlikely to provide all of the information that would be optimal for conducting a comprehensive risk assessment. For example, the test design to study reversibility of adverse effects or mechanism of toxic action may be different from that needed to determine time of onset of an effect or for calculation of a safe level for repeated exposure over a long term. Ideally, results from several different types of tests should be available when performing a risk assessment. Typically, only limited data are available. Under those conditions, the limited data should be used to the extent possible to assess risk.

An integral part of the hazard identification and dose-response process includes evaluation of the protocols from which data are available and the quality of the resulting data. In this section, design factors that are of particular importance in reproductive toxicity testing are discussed. Then, standardized protocols that may provide useful data for reproductive risk assessments are described.

##### **III.A.2. Duration of Dosing**

To evaluate adequately the potential effects of an agent on the reproductive systems, a prolonged treatment period is needed. For example, damage to spermatogonial stem cells

## **DRAFT--DO NOT QUOTE OR CITE**

will not appear in samples from the cauda epididymis or in ejaculates for 8 to 14 weeks, depending on the test species. With some chemical agents that bioaccumulate, the full impact on a given cell type could be further delayed, as could the impact on functional endpoints such as fertility. In such situations, adequacy of the dosing duration is a critical factor in the risk assessment.

Conversely, adaptation may occur that allows tolerance to levels of a chemical that initially caused an effect that could be considered adverse. An example is interference with ovulation by chlordimeform (Goldman et al., 1991); an effect for which a compensatory mechanism is available. Thus, with continued dosing, the compensatory mechanism can be activated so that the initial adverse effect is masked.

In these situations, knowledge of the relevant pharmacokinetic and pharmacodynamic data can facilitate selection of dose levels and treatment duration (see also section on Exposure Assessment). Equally important is proper timing of examination of treated animals relative to initiation and termination of exposure to the agent.

### **III.A.3. Length of Mating Period**

Traditionally, pairs of rats or mice are allowed to cohabit for periods ranging from several days to 3 weeks. Given a 4- or 5-day estrous cycle, each female that is cycling normally should be in estrus four or five times during a 21-day mating period. Therefore, information on the interval or the number of cycles needed to achieve pregnancy may provide evidence of reduced fertility that is not available from fertility data. Additionally, during each period of behavioral estrus, the male has the opportunity to copulate a number of times, resulting in delivery of many more sperm than are required for fertilization. When an unlimited number of matings is allowed in fertility testing, a large effect on sperm production is necessary before an effect on fertility can be detected.

## **DRAFT--DO NOT QUOTE OR CITE**

### **III.A.4. Number of Females Mated to Each Male**

Current EPA test guidelines prepared pursuant to FIFRA and TSCA specify the use of 20 males and enough females to produce at least 20 pregnancies for each dose group in each generation in the multigeneration reproduction test (U.S. Environmental Protection Agency, 1982, 1985b). However, in some tests that were not designed to conform to EPA test guidelines (Organization for Economic Cooperation and Development, 1983), 20 pregnancies may have been achieved by mating two females with each male and using fewer than 20 males per treatment group. In such cases, the statistical treatment of the data should be examined carefully. With multiple females mated to each male, the degree of independence of the observations for each female may not be known. In that situation, when the cause of the adverse effect cannot be assigned with confidence to only one sex, dependence should be assumed and the male used as the experimental unit in statistical analyses. Using fewer males as the experimental unit reduces ability to detect an effect.

### **III.A.5. Single- and Multigeneration Reproduction Tests**

Reproductive toxicity studies in laboratory animals generally involve continuous exposure to a test substance for one or more generations. The objective is to detect effects on the integrated reproductive process as well as to study effects on the individual reproductive organs. Test guidelines for the conduct of single- and multigeneration reproduction protocols have been published by the Agency pursuant to FIFRA and TSCA and by OECD (U.S. Environmental Protection Agency, 1982, 1985b; Galbraith et al., 1983; Organization for Economic Cooperation and Development, 1983).

The single-generation reproduction test evaluates effects of subchronic exposure of peripubertal and adult animals. In the multigeneration reproduction protocol, F<sub>1</sub> and F<sub>2</sub> offspring are exposed continuously in utero from conception until birth and during the preweaning period. This allows detection of effects that occur from exposures throughout development, including the peripubertal and young adult phases. Because the parental and subsequent filial generations have different exposure histories, reproductive effects seen in

## **DRAFT--DO NOT QUOTE OR CITE**

any particular generation are not necessarily comparable with those of another generation. Also, successive litters from the same parents cannot be considered as replicates because of factors such as continuing exposure of the parents, increased parental age, sexual experience, and parity of the females.

In a single- or multigeneration reproduction test, rats are used most often. In a typical reproduction test, dosing is initiated at 5 to 8 weeks of age and continued for 8 to 10 weeks prior to mating to allow effects on gametogenesis to be expressed in ejaculates. Three dose levels plus one or more control groups are usually included. Enough males and females are mated to ensure 20 pregnancies per dose group for each generation. Animals producing the first generation of offspring should be considered the parental (P) generation, and all subsequent generations should be designated filial generations (e.g., F<sub>1</sub>, F<sub>2</sub>). Only the P generation is mated in a single-generation test, while both the P and F<sub>1</sub> generations are mated in a two-generation reproduction test.

In the P generation, both females and males are treated prior to and during mating, with treatment usually beginning around puberty. Cohabitation is allowed for up to 3 weeks, during which the females are monitored for evidence of mating. Females continue to be exposed during gestation and lactation.

In the two-generation reproduction test, randomly selected F<sub>1</sub> offspring continue to be exposed after weaning (day 21) and then are mated at 11 to 13 weeks of age. Treatment of mated F<sub>1</sub> females is continued throughout gestation and lactation. More than one litter may be produced from either P or F<sub>1</sub> animals. Depending on the route of exposure of lactating females, it is important to consider that offspring may be exposed additionally to a chemical by ingestion of maternal feed or water (diet or drinking water studies), by licking of exposed fur (inhalation study), by contact with treated skin (dermal study), or by coprophagia.

In single- and multigeneration reproduction tests, reproductive endpoints evaluated in P and F generations usually include visual examination of the reproductive organs. Weights and histopathology of the testes, epididymides, and accessory sex glands may be available

## **DRAFT--DO NOT QUOTE OR CITE**

from males, and histopathology of the vagina, uterus, cervix, ovaries, and mammary glands from females. Uterine and ovarian weights are also often available. Male and female mating and fertility indices (Section III.B.2.a.) are usually presented. In addition, litters (and often individual pups) are weighed at birth and examined for number of live and dead offspring, gender, gross abnormalities, and growth and survival to weaning. Maturation and behavioral testing may also be performed on the pups.

If effects on fertility or pregnancy outcome are the only adverse effects observed in a study using one of these protocols, the contributions of male- and female-specific effects often cannot be distinguished. If testicular histopathology or sperm evaluations have been included, it may be possible to characterize a male-specific effect. Similarly, ovarian and reproductive tract histology or changes in estrous cycle normality may be indicative of female-specific effects. However, identification of effects in one sex does not exclude the possibility that both sexes may have been affected adversely. Data from matings of treated males with untreated females and vice versa (crossover matings) are necessary to separate sex-specific effects.

An EPA workshop has considered the relative merits of one- versus two-generation reproductive effects studies (Francis & Kimmel, 1988). The participants concluded that a one-generation study is insufficient to identify all potential reproductive toxicants, because it would exclude detection of effects caused by prenatal and postnatal exposures (including the prepubertal period) as well as effects on germ cells that could be transmitted to and expressed in the next generation. A one-generation test might also miss adverse effects with delayed or latent onset because of the shorter duration of exposure for the P generation. Therefore, a comprehensive reproductive risk assessment should include results from a two-generation test. A further recommendation from that workshop was to include sperm analyses and estrous cycle normality as endpoints in reproductive effects studies.

In studies where parental and offspring generations are evaluated, there are additional risk assessment issues regarding the relationships of reproductive outcomes across generations. Increasing vulnerability of subsequent generations is often, but not



## **DRAFT--DO NOT QUOTE OR CITE**

always, observed. Qualitative predictions of increased risk of the filial generations could be strengthened by knowledge of the reproductive effects in the adult, the likelihood of bioaccumulation of the agent, and the potential for increased sensitivity resulting from exposure during critical periods of development (Gray, 1991).

Occasionally, the severity of effects may be static or decrease with succeeding generations. When a decrease occurs, one explanation may be that the animals in the F<sub>1</sub> and F<sub>2</sub> generations represent "survivors" who are (or become) more resistant to the agent than the average of the P generation. If such selection exists, then subsequent filial generations may show a reduced toxic response. Thus, significant adverse effects in any generation may be cause for concern regardless of results in other generations unless inconsistencies in the data indicate otherwise.

### **III.A.6. Alternative Reproductive Tests**

A number of alternative test designs have appeared in the literature (Lamb, 1985; Lamb & Chapin, 1985; Gray et al., 1988, 1989, 1990; Morrissey et al., 1989). Although not necessarily viewed as replacements for the standard two-generation reproduction tests, data from these protocols may be used on a case-by-case basis depending on what is known about the test agent in question. When mutually agreed on by the testing organization and the Agency, such alternative protocols may offer an expanded array of endpoints and increased flexibility (Francis & Kimmel, 1988).

A continuous breeding protocol, Fertility (or Reproductive) Assessment by Continuous Breeding (FACB or RACB), has been developed by the National Toxicology Program (Lamb & Chapin, 1985; Morrissey et al., 1989; Gulati et al., 1991). As originally described, this protocol was a one-generation test. However, dosing can be extended into the F<sub>1</sub> generation to make it compatible with the EPA workshop recommendations for a two-generation design (Francis & Kimmel, 1988). The RACB protocol is being used with both mice and rats. A distinctive feature of this protocol is the continuous cohabitation of male-female pairs (in the P generation) for 14 weeks. Up to five litters can be produced

## DRAFT--DO NOT QUOTE OR CITE

with the pups removed soon after birth. This protocol provides information on changes in the spacing, number, and size of litters over the 14 week dosing interval. Treatment (three dose levels plus controls) is initiated in postpubertal males and females (11 weeks of age) seven days before cohabitation and continues throughout the test. Offspring that are removed from the dam soon after birth are counted and examined for viability, litter and/or pup weight, sex, and external abnormalities. The last litter may remain with the dam until weaning to study the effects of in utero as well as perinatal and postnatal exposures. If effects on fertility are observed in the P or F generations, additional reproductive evaluations may be conducted, including fertility studies and crossover matings to define the affected gender and site of toxicity.

The sequential production of litters from the same adults allows observation of the timing of onset of an adverse effect on fertility. In addition, it may improve ability to detect subfertility due to the potential to produce larger numbers of pregnancies and litters than in a standard single- or multigeneration reproduction study. With continuous treatment, a cumulative effect could increase the incidence or extent of expression with subsequent litters. However, unless offspring are allowed to grow and reproduce (as they are in the more recent version of the RACB protocol) (Gulati et al., 1991), little or no information will be available on postnatal development or reproductive capability of a second generation.

Sperm measures (including sperm number, morphology, and motility) and vaginal smear cytology to detect changes in estrous cyclicity have been added to the RACB protocol at the end of the test period (although not at all dose levels) and their utility has been examined using model compounds in the mouse (Morrissey et al., 1989).

Another test method under development combines the use of multiple endpoints in both sexes of rats with initiation of treatment at weaning (Gray et al., 1988). Thus, morphologic and physiologic changes associated with puberty are included as endpoints. Both P sexes are treated (at least three dose levels plus controls) continuously through breeding, pregnancy, and lactation. The F<sub>1</sub> generation is mated in a continuous breeding

## **DRAFT--DO NOT QUOTE OR CITE**

protocol. Vaginal smears are recorded daily throughout the test period to evaluate estrous cycle normality and confirm breeding and pregnancy (or pseudopregnancy). Pregnancy outcome is monitored in both the P and F<sub>1</sub> generations at all doses, and terminal studies on both generations include comprehensive assessment of sperm measures (numbers, morphology, motility) as well as organ weights, histopathology, and the serum and tissue levels of appropriate reproductive hormones. As with the RACB, crossover mating studies may be conducted to identify the affected sex as warranted. This protocol combines the advantages of a continuous breeding design with acquisition of sex-specific multiple endpoint data at all doses. In addition, identification of pubertal effects makes this protocol particularly useful for detecting compounds with hormone-mediated actions such as environmental estrogens.

### **III.A.7. Additional Test Protocols That May Provide Reproductive Data**

Several shorter-term reproductive toxicity screening tests have been developed. Among those are the Reproductive/Developmental Toxicity Screening Test, which is part of the OECD's Screening Information Data Set (SIDS) protocol (Scala et al., 1992; Tanaka et al., 1992; Organization for Economic Cooperation and Development, 1993a), a draft protocol from the International Conference on Harmonization (Manson, 1994), and the National Toxicology Program's Short-Term Reproductive and Developmental Toxicity Screen (Harris et al., 1992). These protocols have been developed for setting priorities for further testing and should not be considered sufficient by themselves to establish regulatory exposure levels. Their limited exposure periods do not allow assessment of certain aspects of the reproductive process, such as developmentally-induced effects on the reproductive systems of offspring, that are covered by the multi-generation reproduction protocols.

The dominant lethal test was designed to detect mutagenic effects in the male spermatogenic process that are lethal to the offspring. A review of this test has been published as part of the EPA's Gene-Tox program (Green et al., 1985). Dominant lethal protocols may use acute dosing (1 to 5 days) followed by serial matings with one or two

## **DRAFT--DO NOT QUOTE OR CITE**

females per male per week for the duration of the spermatogenic process. An alternative protocol may use subchronic dosing for the duration of the spermatogenic process followed by mating. Females are monitored for evidence of mating, killed at approximately midgestation, and examined for incidence of pre- and postimplantation loss (see Section III.B.2. for discussions of these endpoints).

Pre- or postimplantation loss in the dominant lethal test is often considered evidence that the agent has induced mutagenic damage to the male germ cell (U.S. Environmental Protection Agency, 1986c). A genotoxic basis for a substantial portion of postimplantation loss is accepted widely. However, methods used to assess preimplantation loss do not distinguish between contributions of mutagenic events that cause embryo death and nonmutagenic factors that result in failure of fertilization or early embryo mortality (e.g., inadequate number of normal sperm, failure in sperm transport or ovum penetration). Similar effects (fertilization failure, early embryo death) could also be produced indirectly by effects that delay the timing of fertilization relative to time of ovulation. Such distinctions are important because cytotoxic effects on gametogenic cells do not imply the potential for transmittable genetic damage that is associated with mutagenic events. The interpretation of an increase in preimplantation loss may require additional data on the agent's mutagenic and gametotoxic potential if genotoxicity is to be factored into the risk assessment. Regardless, significant effects may be observed in a dominant lethal test that are considered reproductive in nature.

An acute exposure protocol, combined with serial mating, may allow identification of the spermatogenic cell types that are affected by treatment. However, acute dosing may not produce adverse effects at levels as low as with subchronic dosing because of factors such as bioaccumulation. Conversely, if tolerance to an agent is developed with longer exposure, an effect may be observed after acute dosing that is not detected after longer term dosing.

Subchronic toxicity tests may have been conducted before a detailed reproduction study is initiated. In the subchronic toxicity test with rats, exposure usually begins at six to

## **DRAFT--DO NOT QUOTE OR CITE**

eight weeks of age and is continued for 90 days (U.S. Environmental Protection Agency, 1982, 1985b). Initiation of exposure at eight weeks of age (compared with six) and exposure for approximately 90 days allows the animals to reach a more mature stage of sexual development and assures an adequate length of dosing for observation of effects on the reproductive organs with most agents. The route of administration is often oral or by gavage but may be dermal or by inhalation. Animals are monitored for clinical signs throughout the test and are necropsied at the end of dosing.

The endpoints that are usually evaluated for the male reproductive system include visual examination of the reproductive organs, plus weights and histopathology for the testes, epididymides, and accessory sex glands. For the female, endpoints may include visual examination of the reproductive organs, uterine and ovarian weights, and histopathology of the vagina, uterus, cervix, ovaries, and mammary glands.

This test may be useful to identify an agent as a potential reproductive hazard, but usually does not provide information about the integrated function of the reproductive systems (sexual behavior, fertility and pregnancy outcomes), nor does it include effects of the agent on immature animals.

Chronic toxicity tests provide an opportunity to evaluate toxic effects of long-term exposures. Oral, inhalation or dermal exposure is initiated soon after weaning and is usually continued for 12 to 24 months. Because of the extended treatment period, interim sacrifices may be available to provide useful information regarding the onset and sequence of toxicity. In males, the reproductive organs are examined visually, testes are weighed, and histopathologic examination is done on the testes and accessory sex glands. In females, the reproductive organs are examined visually, uterine and ovarian weights may be obtained, and histopathologic evaluation of the reproductive organs is done. The incidence of abnormalities is often increased in the reproductive tracts of aged control animals. Therefore, findings should be interpreted carefully.

## **DRAFT--DO NOT QUOTE OR CITE**

### **III.B. ENDPOINTS FOR EVALUATING MALE AND FEMALE REPRODUCTIVE TOXICITY IN TEST SPECIES**

#### **III.B.1. Introduction**

The following discussion emphasizes endpoints that measure characteristics that are necessary for successful sexual performance and procreation. Other areas that are related less directly to reproduction are beyond the scope of these Guidelines. For example, adverse health effects that may result from toxicity to the reproductive organs (e.g., osteoporosis or altered immune function), although important, are not included.

In these Guidelines, the endpoints of reproductive toxicity are separated into three categories: couple-mediated, female-specific, and male-specific. Couple-mediated endpoints are those in which both sexes have a contributing role. Thus, an effect on either sex or both sexes may result in an effect on that endpoint.

The discussions of endpoints and the factors influencing results that are presented in this section are directed to evaluation and interpretation of results with test species. Many of those endpoints require invasive techniques that preclude routine use with humans. However, in some instances (e.g., Tables 3 and 4), related endpoints that can be used with humans are identified. Information that is specific for evaluation of effects on humans is presented in Section III.C.

Although statistical analyses are important in determining the effects of a particular agent, the biological significance of data is most important. It is important to be aware that when many endpoints are investigated, 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. In each section, endpoints are identified in which significant changes may be considered adverse. However, concordance of results and known biology should be considered in interpreting all results. Results should be evaluated on a case-by-case basis with all of the evidence considered. Scientific judgment should be used extensively. All effects that may be

## **DRAFT--DO NOT QUOTE OR CITE**

considered as adverse are appropriate for use in establishing a NOAEL, LOAEL, or benchmark dose.

### **III.B.2. Couple-mediated Endpoints**

Data on fertility potential and associated reproductive outcomes provide the most comprehensive and direct insight into reproductive capability. As noted previously, most protocols specify cohabitation of exposed males with exposed females. This complicates the resolution of gender-specific influences. Conclusions may need to be restricted to noting that the "couple" is at reproductive risk when one or both parents are potentially exposed.

#### **III.B.2.a. Fertility and Pregnancy Outcomes**

Breeding studies with test species are a major source of data on reproductive toxicants. Evaluations of fertility and pregnancy outcomes provide measures of the functional consequences of reproductive injury. Measures of fertility and pregnancy outcome that are often obtained from multigeneration reproduction studies are presented in Table 1. Many endpoints that are pertinent for developmental toxicity are also listed and discussed in the Agency Guidelines for Developmental Toxicity Risk Assessment (U.S. Environmental Protection Agency, 1991). Also included in Table 1 are measures that may be obtained from other types of studies (e.g., single-generation reproduction studies, developmental toxicity studies, dominant lethal studies) in which offspring are not retained to evaluate subsequent reproductive performance. Significant detrimental effects on any of those endpoints should be considered adverse. Whether effects are on the female reproductive system or directly on the embryo or fetus is often not distinguishable, but the distinction may not be important because all of these effects should be cause for concern.

**DRAFT--DO NOT QUOTE OR CITE**

**TABLE 1**

**Couple-mediated Endpoints of Reproductive Toxicity**

<u>Multigeneration studies</u>	<u>Other reproductive endpoints</u>
Mating rate, time to mating (Time to pregnancy*)	Ovulation rate
Pregnancy rate*	Fertilization rate
Delivery rate*	Preimplantation loss
Gestation length*	Implantation number
Litter size (total and live)	Postimplantation loss*
Number of live and dead offspring (Fetal death rate*)	Internal malformations and variations*
Offspring gender*	Postnatal structural and functional development*
Birth weight*	
Postnatal weights*	
Offspring survival*	
External malformations and variations*	
Offspring reproduction*	

\* Endpoints that can be obtained with humans



## DRAFT--DO NOT QUOTE OR CITE

Some of the endpoints identified above are used to calculate ratios or indices (National Research Council, 1977; Collins, 1978; Schwetz et al., 1980; U.S. Environmental Protection Agency, 1982, 1985b; Dixon & Hall, 1984; Lamb et al., 1985; Thomas, 1991). While the presentation of such indices is not discouraged, the measurements used to calculate those indices should also be available for evaluation. Definitions of some of these indices in published literature vary substantially. Also, the calculation of an index may be influenced by the test design. Therefore, it is important that the methods used to calculate indices be specified. Some commonly reported indices are in Table 2.

Mating rate may be reported for the mated pairs, males only or females only. Evidence of mating may be direct observation of copulation, observation of copulatory plugs, or observation of sperm in the vaginal fluid (vaginal lavage). The mating rate may be influenced by the number of estrous cycles allowed or required for pregnancy to occur. The most meaningful measure is derived from the occurrence of mating during the first estrous cycle after initiation of cohabitation. Evidence of mating does not necessarily mean successful impregnation.

A useful indicator of impaired reproductive function may be the length of time required for each pair to mate (time to mating). An increased interval between initiation of cohabitation and evidence of mating suggests abnormal estrous cyclicity in the female or impaired sexual behavior in one or both partners.

The time to mating for normal pairs (rat or mouse) could vary by 3 or 4 days depending on the stage of the estrous cycle at which they were paired. If the stage of the estrous cycle at time of cohabitation is known, the component of the variance due to variation in stage at cohabitation can be removed in the statistical analysis.

Data on fertilization rate, the proportion of available ova that were fertilized, are seldom available because the measurement requires necropsy very early in gestation. Pregnancy rate is the proportion of mated pairs that have produced at least one pregnancy within a fixed period where pregnancy is determined by the earliest available evidence that

**DRAFT--DO NOT QUOTE OR CITE**

TABLE 2

Selected Indices That May Be Calculated From  
Endpoints of Reproductive Toxicity in Test Species

MATING INDEX

$$\frac{\text{Number of males or females mating}}{\text{Number of males or females cohabited}} \times 100$$

Note: Mating is used to indicate that evidence of copulation (observation or other evidence of ejaculation such as vaginal plug or sperm in vaginal smear) was obtained.

FERTILITY INDEX

$$\frac{\text{Number of cohabited females becoming pregnant}}{\text{Number of nonpregnant couples cohabited}} \times 100$$

Note: Because both sexes are often exposed to an agent, distinction between sexes is often not possible. If responsibility for an effect can be clearly assigned to one sex (as when treated animals are mated with controls), then a female or male fertility index could be useful.

GESTATION INDEX

$$\frac{\text{Number of females delivering live young}}{\text{Number of females with evidence of pregnancy}} \times 100$$

LIVE BIRTH INDEX

$$\frac{\text{Number of live offspring}}{\text{Number of offspring delivered}} \times 100$$

SEX RATIO

$$\frac{\text{Number of male offspring}}{\text{Number of female offspring}}$$

4-DAY SURVIVAL INDEX (VIABILITY INDEX)

$$\frac{\text{Number of live offspring at lactation day 4}}{\text{Number of live offspring delivered}} \times 100$$

**DRAFT--DO NOT QUOTE OR CITE**

TABLE 2 (continued)

Note: This definition assumes that no standardization of litter size is done until after the day 4 determination is completed.

**LACTATION INDEX (WEANING INDEX)**

$$\frac{\text{Number of live offspring at day 21}}{\text{Number of live offspring born}} \times 100$$

Note: If litters were standardized to equalize numbers of offspring per litter, number of offspring after standardization should be used instead of number born alive. When no standardization is done, measure is called weaning index. When standardization is done, measure is called lactation index.

**PREWEANING INDEX**

$$\frac{\text{Number of live offspring born} - \text{Number of offspring weaned}}{\text{Number of live offspring born}} \times 100$$

Note: If litters were standardized to equalize numbers of offspring per litter, then number of offspring remaining after standardization should be used instead of number born.

## **DRAFT--DO NOT QUOTE OR CITE**

fertilization has occurred. Generally, a more meaningful measure of fertility results when the mating opportunity was limited to one mating couple and to one estrous cycle (see Sections III.A.3. and III.A.4.).

The timing and integrity of gamete and zygote transport are important to fertilization and embryo survival and are quite susceptible to chemical perturbation. Disruption of the processes that contribute to a reduction in fertilization rate and early embryo loss are usually identified simply as preimplantation loss. Additional studies using direct assessments of fertilized ova and early embryos would be necessary to identify the cause of increased preimplantation loss (Cummings & Perreault, 1990). Preimplantation loss (described below) occurs in untreated as well as treated rodents and contributes to the variation in litter size.

After mating, uterine and oviductal contractions are critical in the transport of spermatozoa from the vagina. In rodents, sufficient stimulation during mating is necessary for initiation of those contractions. Thus, impaired mating behavior may affect sperm transport and fertilization rate. Exposure of the female to estrogenic compounds is known to alter gamete transport. In women, low doses of exogenous estrogens may accelerate ovum transport to a detrimental extent, whereas high doses of estrogens or progestins delay transport and increase the incidence of ectopic pregnancies.

Mammalian ova are surrounded by investments that the sperm must penetrate before fusing with ova. Chemicals may block fertilization by preventing this passage. Other agents may impair fusion of the sperm with the ovum plasma membrane, transformations of the sperm or ovum chromatin into the male and female pronuclei, fusion of the pronuclei, or the subsequent cleavage divisions. Carbendazim, an inhibitor of microtubule synthesis, is an example of a chemical that can interfere with fertilization (Perreault et al., 1992). The early zygote is also susceptible to detrimental effects of mutagens such as ethylene oxide (Generoso et al., 1987).

Fertility assessments in test animals have limited sensitivity as measures of reproductive injury. Therefore, results demonstrating no treatment-related effect on fertility

## **DRAFT--DO NOT QUOTE OR CITE**

may be given less weight than other endpoints that are more sensitive. Unlike humans, normal males of most test species produce sperm in numbers that greatly exceed the minimum requirements for fertility, particularly as evaluated in protocols that allow multiple matings (Amann, 1981; Working, 1988). In some strains of rats and mice, production of normal sperm can be reduced by up to 90% or more without compromising fertility (Aafjes et al., 1980; Meistrich, 1982; Robaire et al., 1984; Working, 1988). However, less severe reductions can cause reduced fertility in human males who appear to function closer to the threshold for the number of normal sperm needed to ensure full reproductive competence (see Supplemental Information). This difference between test species and humans means that negative results with test species in a study that was limited to endpoints that examined only fertility and pregnancy outcomes would provide insufficient information to conclude that the test agent poses no reproductive hazard in humans. It is unclear whether a similar consideration is applicable for females for some mechanisms of toxicity.

The limited sensitivity of fertility measures in rodents also suggests that a NOAEL, LOAEL, or benchmark dose (see Section III.H.) based on fertility may not reflect completely the extent of the toxic effect. In such instances, data from additional reproductive endpoints might indicate that an adverse effect could occur at a lower dose level. In the absence of such data, the margin of exposure or uncertainty factor applied to the NOAEL, LOAEL, or benchmark dose may need to be adjusted to reflect the additional uncertainty (see Section III.H.).

Both the blastocyst and the uterus must be ready for implantation, and their synchronous development is critical (Cummings & Perreault, 1990). The preparation of the uterine endometrium for implantation is under the control of sequential estrogen and progesterone stimulation. Treatments that alter the internal hormonal environment, inhibit protein synthesis, or inhibit mitosis or cell differentiation can block implantation and cause embryo death.

## **DRAFT--DO NOT QUOTE OR CITE**

Gestation length can be determined in test animals from data on day of mating (observation of vaginal plug or sperm-positive vaginal lavage) and day of parturition. Significant shortening of gestation can lead to adverse outcomes of pregnancy such as decreased birth weight and offspring survival. Significantly longer gestation may be caused by failure of the normal mechanism for parturition and may result in death or impairment of offspring if dystocia (difficulty in parturition) occurs. Dystocia also constitutes a maternal health threat. Lengthened gestation may result in higher birth weight; an effect that could mask a slower growth rate in utero because of exposure to a toxic agent. Comparison of offspring weights based on conceptional age may allow insight, although this comparison is complicated by generally faster growth rates postnatally than in utero.

Litter size is the number of offspring delivered and is measured at or soon after birth. Unless this observation is made soon after parturition, the number of offspring observed may be less than the actual number delivered because of cannibalism by the dam. Litter size is affected by the number of ova available for fertilization (ovulation rate), fertilization rate, implantation rate, and the proportion of the implanted embryos that survives to parturition. Litter size may include dead as well as live offspring, therefore data on the numbers of live and dead offspring should be available also.

When pregnant animals are examined by necropsy in mid- to late gestation, pregnancy status, including pre- and postimplantation losses can be determined. Preimplantation loss is the number of corpora lutea minus number of implantation sites/number of corpora lutea. Postimplantation loss is the (number of implantation sites minus number of live pups)/number of implantation sites.

Offspring gender is determined by the male through fertilization of an ovum by a Y- or an X-chromosome-bearing sperm. Therefore, selective impairment in the production, transport, or fertilizing ability of either of these sperm types can produce an alteration in the sex ratio. An agent may also induce selective loss of male or female fetuses. Although not examined routinely, these factors provide the most likely explanations for alterations in the sex ratio.

## DRAFT--DO NOT QUOTE OR CITE

Birth weight should be measured on the day of parturition. Often data from individual pups as well as the entire litter (litter weight) are provided. Birth weights are influenced by intrauterine growth rates, litter size, and gestation length. Growth rate in utero is influenced by the normality of the fetus, the maternal environment, and gender, with females tending to be smaller than males. Individual pups tend to be smaller in larger litters than individual pups in smaller litters. Thus, reduced birth weights that can be attributed to large litter size should not be considered an adverse effect unless the increased litter size is treatment related and the subsequent ability of the offspring to survive or develop is compromised. When litter weights only are reported, the increased numbers of offspring and the lower weights of the individuals tend to offset each other. When prenatal or postnatal growth is impaired by an acute exposure, compensatory growth after cessation of dosing could obscure the earlier effect.

Postnatal weights are dependent on birth weight, sex and normality of the individual, as well as the litter size, lactational ability of the dam, and suckling ability of the offspring. With large litters, small or weak offspring may not compete successfully for milk and show impaired growth. Because it is not possible usually to determine whether the effect was due solely to the increased litter size, growth retardation or decreased survival rate should be considered adverse in the absence of information to the contrary. Also, offspring weights may appear normal in very small litters and should be considered carefully in relation to controls.

Offspring survival is dependent on the same factors as postnatal weight, although more severe effects are necessary usually to affect survival. All weight and survival endpoints can be affected by toxicity of an agent, either by direct effects on the offspring or indirectly through effects on the ability of the dam to support the offspring.

Measures of malformations and variations, as well as postnatal structural and functional development, are presented in the Guidelines for Developmental Toxicity Risk Assessment (U.S. Environmental Protection Agency, 1991). That document should be consulted for additional information on those parameters.

## **DRAFT--DO NOT QUOTE OR CITE**

### **III.B.2.b. Sexual Behavior**

Sexual behavior reflects complex neural, endocrine and reproductive organ interactions and is therefore susceptible to disruption by a variety of toxic agents, diseases, and pathologic conditions. Interference with sexual behavior in either sex by environmental agents represents a potentially significant human reproductive problem. Most human information comes from clinical reports in which the detection of exposure-effect associations is unlikely. Data on sexual behavior are usually not available from studies of human populations that were exposed occupationally or environmentally to potentially toxic agents, nor are such data obtained routinely in studies of environmental agents with test species.

In the absence of human data, the perturbation of sexual behavior in test species may suggest the potential for similar effects on humans. Consistent with this position are data showing that central nervous system effects can disrupt sexual behavior in both test species and humans (Rubin & Henson, 1979; Waller et al., 1985). Although the functional components of sexual performance can be quantified in most test species, no direct evaluation of this behavior is done in most breeding studies. Rather, copulatory plugs or sperm-positive vaginal lavages are taken as evidence of sexual receptivity and successful mating. However, these markers do not demonstrate whether male performance resulted in adequate sexual stimulation of the female. Failure of the male to provide adequate stimulation to the female may impair sperm transport in the genital tract of female rats, thereby reducing the probability of successful impregnation (Adler & Toner, 1986). Such a "mating" failure would be reflected in the calculated fertility index as reduced fertility and could be attributed erroneously to an effect on the spermatogenic process in the male or on fertility of the female.

In the rat, a direct measure of female sexual receptivity is the occurrence of lordosis. Sexual receptivity of the female rat is normally cyclic, with receptivity commencing during the late evening of vaginal proestrus. Agents that interfere with normal estrous cyclicity also could cause absence of or abnormal sexual behavior. In the male, measures include



## **DRAFT--DO NOT QUOTE OR CITE**

latency periods to first mount, mount with intromission, and first ejaculation, number of mounts with intromission to ejaculation, and the post-ejaculatory interval (Beach, 1979).

Direct evaluation of sexual behavior is not warranted for all agents being tested for reproductive toxicity. Some likely candidates may be agents reported to exert neurotoxic effects. Chemicals possessing or suspected to possess androgenic or estrogenic properties (or antagonistic properties) also merit consideration as potentially causing adverse effects on sexual behavior concomitant with effects on the reproductive organs. Effects on sexual behavior (within the limited definition of these Guidelines) should be considered as adverse reproductive effects.

### **III.B.3. Male-specific Endpoints**

#### **III.B.3.a. Introduction**

The following sections (III.B.3. and III.B.4.) describe various male-specific and female-specific endpoints of reproductive toxicity that can be obtained. Included are endpoints for which data are obtained routinely by the Agency and other endpoints for which data may be encountered in the review of chemicals. Guidance is presented for interpretation of results involving these endpoints and their use in risk assessment. Effects are identified that should be considered as adverse reproductive effects if significantly different from controls. Because of substantial overlap between the sexes in discussion of effects on the reproductive systems during development, that topic is presented for both sexes in Section III.B.4.f., Developmental and Pubertal Alterations.

The Agency may obtain data on the potential male reproductive toxicity of an agent from many sources including, but not limited to, studies done according to Agency test guidelines. These may include acute, subchronic, and chronic testing and reproduction and fertility studies. Male-specific endpoints that may be encountered in such studies are identified in Table 3.

**DRAFT--DO NOT QUOTE OR CITE**

**III.B.3.b. Body Weight and Organ Weights**

Monitoring body weight during treatment provides an index of the general health status of the animals, and such information may be important for the interpretation of reproductive effects (see also Section III.B.2.). Depression in body weight or reduction in weight gain may reflect a variety of responses, including rejection of chemical-adulterated food or water because of reduced palatability, treatment-induced anorexia, or systemic toxicity. Less than severe reductions in adult body weight may have little effect on the male reproductive organs or on male reproductive function (Chapin et al., 1993a, b). When a meaningful, biologic relationship between a body weight decline and a significant effect on the male reproductive system is not apparent, it is not appropriate to dismiss significant alteration of the male reproductive system as secondary to the occurrence of non-reproductive toxicity. Unless additional data provide the needed clarification, alteration in a reproductive measure that would otherwise be considered adverse should still be considered as an adverse male reproductive effect in the presence of mild to moderate body weight changes. In the presence of severe body weight depression, it should be noted that an adverse effect on a reproductive endpoint occurred but that the effect may have resulted from another, non-reproductive effect. Regardless, adverse effects would have been

**DRAFT--DO NOT QUOTE OR CITE**

**TABLE 3**

**Male-Specific Endpoints of Reproductive Toxicity**

Organ weights	Testes, epididymides, seminal vesicles, prostate, pituitary
Visual examination & histopathology	Testes, epididymides, seminal vesicles, prostate, pituitary
Sperm evaluation*	Sperm number (count) and quality (morphology, motility)
Hormone levels*	Luteinizing hormone, follicle stimulating hormone, testosterone, estrogen, prolactin
Developmental	Testis descent*, preputial separation, sperm production*, ano-genital distance, normality of external genitalia*

\* Reproductive endpoints that can be obtained or estimated relatively noninvasively with humans

## **DRAFT--DO NOT QUOTE OR CITE**

observed in that situation and a risk assessment should be pursued if sufficient data are available.

The male reproductive organs for which weights may be useful for reproductive risk assessment include the testes, epididymides, pituitary gland, seminal vesicles (with coagulating glands), and prostate. Organ weight data may be presented as both absolute weights and as relative weights (i.e., organ weight to body weight ratios). Organ weight data may also be reported relative to brain weight since, subsequent to development, the weight of the brain remains quite stable (Stevens & Gallo, 1989). Evaluation of data on absolute organ weights is important, because a decrease in a reproductive organ weight may occur that was not necessarily related to a reduction in body weight gain. The organ weight-to-body weight ratio may show no significant difference if both body weight and organ weight change in the same direction, masking a potential organ weight effect.

Normal testis weight varies only modestly within a given test species (Schwetz et al., 1980; Blazak et al., 1985). This relatively low interanimal variability suggests that testis weight should be a precise indicator of gonadal injury. However, damage to the testes may be detected as a weight change only at doses higher than those required to produce significant effects in other measures of gonadal status (Berndtson, 1977; Foote et al., 1986). This contradiction may arise from several factors, including a delay before cell deaths are reflected in a weight decrease (due to edema and inflammation, cellular infiltration) or Leydig cell hyperplasia. Blockage of the efferent ducts by cells sloughed from the germinal epithelium or the efferent ducts themselves can lead to an increase in testis weight due to fluid accumulation (Hess et al., 1991; Nakai et al., 1993), an effect that could offset the effect of depletion of the germinal epithelium on testis weight. Thus, testis weight measurements do not indicate the nature of an effect, but a significant increase or decrease is indicative of an adverse effect.

Pituitary gland weight can provide valuable insight into the reproductive status of the animal. However, the pituitary contains cell types that are responsible for the regulation of a variety of physiologic functions including some that are separate from reproduction.

## **DRAFT--DO NOT QUOTE OR CITE**

Thus, changes in pituitary weight may not necessarily reflect reproductive impairment. If weight changes are observed, gonadotroph-specific histopathologic evaluations may be useful in identifying the affected cell types. This information may then be used to judge whether the observed effect on the pituitary is related to reproductive system function and therefore an adverse reproductive effect. Prostate and seminal vesicle weights are androgen-dependent and may reflect changes in the animal's endocrine status or testicular function. Separation of the seminal vesicles and coagulating gland (dorsal prostate) is difficult in rodents. However, the seminal vesicle and prostate can be separated and results may be reported for these glands separately or together. Because the seminal vesicles and prostate may respond differently to an agent (endocrine dependency and developmental susceptibility differ), more information may be gained if the weights were examined separately.

Significant changes in absolute or relative male reproductive organ weights may constitute an adverse reproductive effect. Such changes also may provide a basis for obtaining additional information on the reproductive toxicity of that agent. However, significant changes in other important endpoints that are related to reproductive function may not be reflected in organ weight data. Therefore, lack of an organ weight effect should not be used to negate significant changes in other endpoints that may be more sensitive.

### **III.B.3.c. Histopathologic Evaluations**

Histopathologic evaluations of test animals have a prominent role in male reproductive risk assessment. Organs that are often evaluated include the testes, epididymides, prostate, seminal vesicles (often including coagulating glands), and pituitary. Tissues from lower dose exposures are often not examined histologically if the high dose produced no difference from controls. Histologic evaluations can be especially useful by 1) providing a relatively sensitive indicator of damage; 2) providing information on toxicity

## **DRAFT--DO NOT QUOTE OR CITE**

from a variety of protocols; 3) with short-term dosing, providing information on site (including target cells) and extent of toxicity; and 4) indicating the potential for recovery.

The quality of the information presented from histologic analyses of spermatogenesis is improved by proper fixation and embedding of testicular tissue. With adequately prepared tissue (Chapin, 1988; Russell et al., 1990; Hess & Moore, 1993), a description of the nature and background level of lesions in control tissue, whether preparation induced or otherwise, can facilitate interpreting the nature and extent of the lesions observed in tissues obtained from exposed animals. Many histopathologic evaluations of the testis only detect lesions if the germinal epithelium is severely depleted, degenerating cells are obvious, or sloughed cells are present in the tubule lumen. More subtle lesions that can significantly affect the number of sperm being released normally into the tubule lumen may not be detected when less adequate methods of tissue preparation are used. Also, familiarity with the detailed morphology of the testis and the kinetics of spermatogenesis of each test species can assist in the identification of less obvious lesions that may accompany lower dose exposures or lesions that result from short-term exposure (Russell et al., 1990). Several approaches for qualitative or quantitative assessment of testicular tissue are available that can assist in the identification of less obvious lesions that may accompany lower-dose exposures, including use of the technique of "staging." A text has been prepared (Russell et al., 1990) that provides extensive information on tissue preparation, examination, and interpretation of observations for normal and high resolution histology of the germinal epithelium of rats, mice, and dogs. Included is guidance for identification and quantification of the various cell types and associations for each stage of the spermatogenic cycle. Also, a decision-tree scheme for staging with the rat has been published (Hess, 1990).

The basic morphology of other male reproductive organs (e.g., epididymides, accessory sex glands, and pituitary) has been described as well as the histopathologic alterations that may accompany certain disease states (Fawcett, 1986; Jones et al., 1987; Haschek & Rousseaux, 1991). Compared with the testes, less is known about structural

## **DRAFT--DO NOT QUOTE OR CITE**

changes in these tissues that are associated with exposure to toxic agents. With the epididymides and accessory sex glands, histologic evaluation is usually limited to the height and possibly the integrity of the secretory epithelium. Presence of debris and sloughed cells in the epididymal lumen are valuable indicators of damage to the germinal epithelium or the excurrent ducts. Information from examinations of the pituitary should include evaluation of the morphology of the cell types that produce the gonadotropins and prolactin.

The degree to which histopathologic effects are quantified is usually limited to classifying animals, within dose groups, as either affected or not affected by qualitative criteria. Little effort has been made to quantify the extent of injury, and procedures for such classifications are not applied uniformly (Linder et al., 1990). Evaluation procedures would be facilitated by adoption of more uniform approaches for quantifying the extent of histopathologic damage per individual. In the absence of a standardized quantification system, the evaluation of histopathologic data would be facilitated by the presentation of the evaluation criteria and the manner in which the level of lesions in exposed individuals was judged to be in excess of controls.

If properly obtained (i.e., proper preparation and analysis of tissue), data from histopathologic evaluations may provide a relatively sensitive tool that is useful for detection of low-dose effects. This approach may also provide insight into sites and mechanisms of action for the agent on that reproductive organ. When similar targets or mechanisms exist in humans, the basis for interspecies extrapolation is strengthened. Depending on the experimental design, information can also be obtained that may allow prediction of the eventual extent of injury and degree of recovery in that species and humans (Russell, 1983).

Significant histopathologic damage in excess of the level seen in control tissue of any of the male reproductive organs should be considered an adverse reproductive effect. Significant histopathologic damage in the pituitary should be considered as an adverse effect but should be shown to involve cells that control gonadotropin or prolactin production to be called a reproductive effect. Although thorough histopathologic evalu-

## **DRAFT--DO NOT QUOTE OR CITE**

ations that fail to reveal any treatment-related effects may be quite convincing, consideration should be given to the possible presence of other testicular or epididymal effects that are not detected histologically (e.g., genetic damage to the germ cell, decreased sperm motility), but may affect reproductive function.

### **III.B.3.d. Sperm Evaluations**

The parameters that are important for sperm evaluations are sperm number, sperm morphology, and sperm motility. Data on those parameters allow more adequate estimation of the number of "normal" sperm; a parameter that is likely to be more informative than sperm number alone. Although effects on sperm production can be reflected in other measures such as spermatid count or cauda epididymal weight, no surrogate measures are adequate to reflect effects on sperm morphology or motility. Similar data can be obtained noninvasively from human ejaculates, enhancing the ability to confirm effects seen in test species or to detect effects directly in humans. Brief descriptions of these measures are provided below, followed by discussion of use of the various sperm measures in male reproductive risk assessment.

#### **Sperm number**

Measures of sperm concentration (count) have been the most frequently reported semen variable in the literature on humans (Wyrobek et al., 1983a). Sperm number or sperm concentration from test species may be derived from ejaculated, epididymal, or testicular samples. Of the common test species, ejaculates can only be obtained readily from rabbits or dogs. Ejaculates can be recovered from the reproductive tracts of mated females of other species (Zenick et al., 1984). Measures of human sperm production are usually derived from ejaculates, but could also be obtained from spermatid counts or quantitative histology using testicular biopsy tissue samples. With ejaculates, both sperm concentration (number of sperm/ml of ejaculate) and total sperm per ejaculate (sperm concentration x volume) should be evaluated.



## **DRAFT--DO NOT QUOTE OR CITE**

Ejaculated sperm number from any species is influenced by several variables, including the length of abstinence and the ability to obtain the entire ejaculate. Intra- and interindividual variabilities are often high, but are reduced somewhat if ejaculates were collected at regular intervals from the same male (Williams et al., 1990). Such a longitudinal study design has improved detection sensitivity and thus requires a smaller number of subjects (Wyrobek et al., 1984). In addition, if a pre-exposure baseline is obtained for each male (test animal or human studies), then changes during exposure or recovery can be better defined.

Epididymal sperm evaluations with test species usually use sperm from only the cauda portion of the epididymis. It has been customary to express the sperm count in relation to the weight of the cauda epididymis. However, because sperm contribute to epididymal weight, expression of the data as a ratio may actually mask declines in sperm number. The inclusion of data on absolute sperm counts can improve resolution. As is true for ejaculated sperm counts, epididymal sperm counts are influenced directly by level of sexual activity (Amann, 1981; Hurtt & Zenick, 1986).

Sperm production data may be derived from counts of the distinctive elongated spermatid nuclei that remain after homogenization of testes in a detergent-containing medium (Amann, 1981; Meistrich, 1982; Cassidy et al., 1983; Blazak et al., 1993). The elongated spermatid counts are a measure of sperm production from the stem cells and their ensuing survival through spermatocytogenesis and spermiogenesis (Meistrich, 1982; Meistrich & van Beek, 1993). If evaluation was conducted when the effect of a lesion would be reflected adequately in the spermatid count, spermatid count may serve as a substitute for quantitative histologic analysis of sperm production (Russell et al., 1990). However, spermatid counts may be misleading when duration of exposure is shorter than the time required for a lesion to be fully expressed in spermatid count. Also, spermatid counts reported from some laboratories have large coefficients of variation that may reduce the usefulness of that measure.

## **DRAFT--DO NOT QUOTE OR CITE**

The ability to detect a decrease in testicular sperm production may be enhanced if spermatid counts are available. However, spermatid enumerations only reflect the integrity of spermatogenic processes within the testes. Posttesticular effects or toxicity expressed as alterations in motility, morphology, viability, fragility, and other properties of sperm can be determined only from epididymal or ejaculated samples.

### **Sperm morphology**

Sperm morphology refers to structural aspects of sperm and can be evaluated in cauda epididymal, vas deferens, or ejaculated samples. A thorough morphologic evaluation identifies abnormalities in the sperm head and flagellum. Because of the suggested correlation between an agent's mutagenicity and its ability to induce abnormal sperm, sperm head morphology has been a frequently reported sperm variable in toxicologic studies on test species (Wyrobek et al., 1983b). The tendency has been to conclude that increased incidence of sperm head malformations reflects germ cell mutagenicity. However, not every mutagen induces sperm head abnormalities, and other nonmutagenic chemicals may alter sperm head morphology. For example, microtubule poisons may cause increases in abnormal sperm head incidence, presumably by interfering with spermiogenesis; a microtubule-dependent process (Russell et al., 1981). Sperm morphology may be altered also due to degeneration subsequent to cell death.

An increase in abnormal sperm morphology has been considered evidence that the agent has gained access to the germ cells (U.S. Environmental Protection Agency, 1986c). Exposure of males to toxic agents may lead to sperm abnormalities in their progeny (Wyrobek & Bruce, 1978; Hugenholtz & Bruce, 1983). However, transmissible germ-cell mutations might exist in the absence of any warning morphologic indicator such as abnormal sperm. The relationships between these morphologic alterations and other karyotypic changes remains uncertain (de Boer et al., 1976).

The traditional approach to characterizing morphology in toxicologic testing has relied on subjective categorization of sperm head shape from examination of stained slides at the light microscopic level (Filler, 1993). Such an approach may be adequate for mice

## **DRAFT--DO NOT QUOTE OR CITE**

and rats with their distinctly angular head shapes. However, the observable heterogeneity of structure in human sperm and in nonrodent species makes it difficult for the morphologist to define clearly the limits of normality. More systematic, quantitative, and automated approaches have been offered that can be used with humans and test species (Katz et al., 1982; Wyrobek et al., 1984). Data that identify the types of abnormalities observed and the frequencies of their occurrences are preferred to estimation of overall proportion of abnormal sperm. Objective, quantitative approaches that are done properly should result in a higher level of confidence than more subjective measures.

Sperm morphology profiles are relatively stable and characteristic of a normal individual (and a strain within a species) over time. Sperm morphology is one of the least variable sperm measures in normal individuals, which may enhance its use in the detection of spermatotoxic events (Zenick & Clegg, 1986). However, the reproductive implications of the various types of abnormal sperm morphology need to be delineated more fully. The majority of studies in test species and humans have suggested that abnormally shaped sperm may not reach the oviduct or participate in fertilization (Nestor & Handel, 1984; Redi et al., 1984). The implication is that the greater the number of abnormal sperm in the ejaculate, the greater the probability of reduced fertility.

### **Sperm motility**

The biochemical environments in the testes and epididymides are highly regulated to assure the proper development and maturation of the sperm and the acquisition of critical functional characteristics, i.e., progressive motility and the potential to fertilize. With chemical exposures, perturbation of this balance may occur, producing alterations in sperm properties such as motility. Chemicals (e.g., epichlorohydrin) have been identified that selectively affect sperm motility and also reduce fertility. Studies have examined rat sperm motility as a reproductive endpoint (Toth et al., 1989b, 1991), and sperm motility assessments are an integral part of some reproductive toxicity tests.

Motility estimates may be obtained on ejaculated, vas deferens, or cauda epididymal samples. Standardized methods are needed because motility is influenced by a number of

## **DRAFT--DO NOT QUOTE OR CITE**

experimental variables, including abstinence interval, the elapsed time and temperature history of the sample, dilution, diluent, or sample chamber (Cassidy et al., 1983).

Historically, motility has been measured using subjective, microscopic evaluations. Estimates of percent motile sperm can be made, and a scaling system used to describe the quality of motility (i.e., the degree to which sperm show progressive, linear motility). More quantitative approaches have been taken in test species (Linder et al., 1986; Toth et al., 1989a; Slott et al., 1990) and humans (Boyers et al., 1989), including enumeration of motile and immotile sperm from videotapes. Videotaping has the advantage that a record can be retained.

Computer-assisted methods for evaluation of sperm motility allow measurement of an extensive array of motion parameters. Included may be measures such as linear velocity, curvilinear velocity, lateral sperm head displacement, and linearity of motion. This technology has been applied in a limited number of toxicologic studies using test species (Working & Hurtt, 1987; Toth et al., 1989a; Slott et al., 1990, 1991). The ability of some of these measures to predict a potential effect on ability of sperm to fertilize has not been established. Therefore, judgments concerning adversity should consider the status of knowledge on those measures. Significant reductions in proportion of sperm that are progressively motile should be considered adverse. Reductions in velocity or pattern of motility may indicate a need for further testing. For example, a reduction in velocity could indicate potential for reduced ability of the sperm to survive or fertilize. Efforts are in progress to validate and standardize these automated techniques for application in reproductive toxicity studies and to determine the relationships between these motility endpoints and fertility.

### In vitro tests of reproductive function

Numerous in vitro tests are available that can measure effects on different aspects of the reproductive systems of males and females. These include in vitro fertilization, whole organ (e.g., testis, ovary) perfusion, culture of cell populations, and incubations of subcellular fractions or cytosol from specific cell types. Tests of sperm properties and

## **DRAFT--DO NOT QUOTE OR CITE**

function include sperm-cervical mucus penetration, in vitro sperm capacitation, in vitro fertilization using zona pellucida-free hamster ova and the hemizona penetration assay (Overstreet, 1984; Franken et al., 1990).

The diagnostic information obtained from such tests may help to identify potential effects on the reproductive systems. However, each test bypasses essential components of the intact animal system and therefore, by itself, is not capable of predicting exposure levels that would result in toxicity in intact animals. While it is desirable to replace whole animal testing to the extent possible with in vitro tests, the use of such tests currently is to screen for toxicity potential and to study mechanisms of action and metabolism (Perreault, 1989; Holloway et al., 1990a, b).

### Use of sperm evaluations in risk assessment

The relationships between the various endpoints that measure effects on spermatogenesis or sperm maturation in humans and test species have not been evaluated adequately. Thus, how toxicity that is reflected in one such measure may influence other measures is not always clear. The quantitative relationships between these measures and fertility also are not well characterized for any species. Certain qualitative and quantitative standards must be met to ensure full fertility, but the lower limits of these standards have not been delineated adequately. For instance, the distributions of sperm counts for fertile and infertile men overlap, with the mean for fertile men being higher (Meistrich & Brown, 1983). However, observations with farm species (cattle, sheep, swine) have shown more clearly that reductions in the parameters of sperm quality become important when either the total number of sperm or the number of apparently normal sperm is reduced below a certain level. Additional research is needed to quantify the biologic consequences of reductions in sperm number and quality in the laboratory animal species and humans.

Human male fertility is generally lower than that of test species and may be more susceptible to damage from toxicants (see Supplemental Information). Therefore, the conservative view should be taken that, within the limits indicated in the sections on those

## **DRAFT--DO NOT QUOTE OR CITE**

parameters, significant changes in measures of sperm count, morphology or motility as well as number of normal sperm should be considered adverse effects.

### **III.B.3.e. Endocrine Evaluations**

Measurement of the reproductive hormones in males offers useful supplemental information in assessing potential reproductive toxicants for test species (Sever & Hessol, 1984; Heywood & James, 1985; National Research Council, 1989). However, such measurements have increased importance with humans where invasiveness of approaches must be limited. The reproductive hormones measured often are luteinizing hormone (LH), follicle stimulating hormone (FSH), and testosterone. Other useful measures that may be available include prolactin, inhibin, and androgen binding protein levels. In addition, challenge tests with exogenous agents (e.g., gonadotropin releasing hormone, LH, or human chorionic gonadotropin) may provide insight into the functional responsiveness of the pituitary or Leydig cells.

Toxic agents can alter endocrine function by affecting any part of the hypothalamic-pituitary-gonadal axis. If a compound affects the hypothalamus or pituitary, then serum LH and FSH may be decreased, leading to decreased testosterone levels. On the other hand, severe interference with Sertoli cell function or spermatogenesis would be expected to elevate serum FSH levels.

A toxicant having antiandrogenic activity might cause endocrine and morphologic changes that differ from those described previously. In adult male rats, exposure to an antiandrogen might elevate serum LH and testosterone. Testis weight might be unaffected, while the weight and size of the accessory sex glands may be altered. The profile presented by specific antiandrogens can differ markedly because of differences in tissue specificity and receptor kinetics.

Interpretation of endocrine effects is facilitated if information is available on a battery of hormones. However, in evaluating such data, it is important to consider that serum hormones such as FSH, LH, prolactin, and androgens exhibit cyclic variations within

## **DRAFT--DO NOT QUOTE OR CITE**

a 24-hour period (Fink, 1988). Thus, the time of sampling should be controlled rigorously to avoid excessive variability. Sequential sampling can allow detection of treatment-related changes in circadian and pulsatile rhythms.

In the absence of endocrine data, significant effects on pituitary or accessory sex gland weights or histopathology or on Leydig cell histopathology may suggest disruption of the endocrine system. In those instances, additional testing for endocrine effects may be indicated. Significant alterations in circulating levels of testosterone, LH, or FSH may be indicative of existing pituitary or gonadal injury. When significant alterations from control levels are observed in those hormones, the changes should be considered cause for concern because they are likely to affect, or occur in concert with, alterations in spermatogenesis, sperm maturation, mating ability, or fertility. Such effects, if compatible with other available information, may be considered adverse and may be used to establish a NOAEL or LOAEL. Furthermore, endocrine data may facilitate identification of sites or mechanisms of toxicant action, especially when obtained after short term exposures.

### **III.B.3.f. Biochemical Tests or Markers of Toxicity to the Testes and Other Male Reproductive Organs**

Numerous biomarkers and biochemical tests exist that are related less directly to integrated reproductive function than the endpoints discussed previously in this section. Currently, the value of such tests as endpoints of reproductive toxicity remains to be demonstrated, although a number of potential chemical markers are available (National Research Council, 1989; Kimmel, G.L. et al., In press). However, the results of such measurements may suggest the presence of an effect that should be investigated further. Another valuable role for these tests may be in delineating the target or mechanism of action for a given agent. Such data may be of use in the design of subsequent tests, interspecies extrapolation, and in estimating the potential for reversibility (Scialli & Clegg, 1992).

## **DRAFT--DO NOT QUOTE OR CITE**

### **III.B.3.g. Paternally-mediated Effects on Offspring**

The concept is well accepted that exposure of a female to toxic chemicals during gestation or lactation may produce death, structural abnormalities, growth alteration, or postnatal functional deficits in her offspring. Sufficient data now exist with a variety of agents to conclude that male-only exposure also can produce deleterious effects in offspring (Hood, 1989; Nagao & Fujikawa, 1990; Davis et al., 1992).

These effects may be the result of direct damage to the sperm. However, xenobiotics present in seminal plasma or bound to the fertilizing sperm could be introduced into the female genital tract, or even the oocyte directly, and might also interfere with fertilization or early development. With humans, the possibility also exists that a parent could transport the toxic agent from the work environment to the home (e.g., on work clothes), exposing other adults or children. Further work is needed to clarify the extent to which paternal exposures may be associated with adverse effects on offspring. Regardless, if an agent is identified in test species as causing a paternally-mediated adverse effect on offspring, the effect should be considered an adverse reproductive effect.

### **III.B.4. Female-specific Endpoints**

#### **III.B.4.a. Introduction**

The reproductive life cycle of the female may be divided into phases that include fetal, prepubertal, cycling adult, pregnant, lactating, and reproductively senescent. Detailed descriptions of all phases are available (Knobil & Neill, 1988). It is important to detect adverse effects occurring in any of these stages. Traditionally, the endpoints that have been used have emphasized ability to become pregnant, pregnancy outcome, and offspring survival and development. Although reproductive organ weights may be obtained and these organs examined histologically in test species, these measures do not necessarily detect abnormalities in dynamic processes such as estrous cyclicity or follicular atresia unless degradation is severe. Similarly, toxic effects on onset of puberty have not been examined, nor have the long-term consequences of exposure on reproductive senescence. Thus, the



## **DRAFT--DO NOT QUOTE OR CITE**

amount of information obtained routinely to detect toxic effects on the female reproductive system is limited.

The consequences of impairment in the nonpregnant female reproductive system are equally important, and endpoints to detect adverse effects on the nonpregnant reproductive system, when available, can be useful in evaluating reproductive toxicity. Such measures may also provide additional interrelated endpoints and information on mechanism of action.

Alterations in the nonpregnant female reproductive system have been observed at dose levels below those that result in reduced fertility or produce other overt effects on pregnancy or pregnancy outcomes (Le Vier & Jankowiak, 1972; Barsotti et al., 1979; Sonawane & Yaffe, 1983; Cummings & Gray, 1987). In contrast to the male reproductive system, the status of the normal female system fluctuates in adults. Thus, in nonpregnant animals (including humans), the ovarian structures and other reproductive organs change throughout the estrous or menstrual cycle. Although not cyclic, normal changes also accompany the progression of pregnancy, lactation, and return to cyclicity during or after lactation. These normal fluctuations may affect the endpoints used for evaluation. Therefore, knowledge of the reproductive status of the female at necropsy, including the stage of the estrous cycle, can facilitate detection and interpretation of effects with endpoints such as uterine weight and histopathology of the ovary and uterus. Necropsy of all test animals at the same stage of the estrous cycle can reduce the variance of test results with such measures.

A variety of measures to evaluate the integrity of the female reproductive system has been used in toxicity studies. With appropriate measures, a comprehensive evaluation of the reproductive process can be achieved, including identification of target organs and possible elucidation of the mechanisms involved in the toxicant's effect. Areas that may be examined in evaluations of the female reproductive system are listed in Table 4.

**DRAFT--DO NOT QUOTE OR CITE**

TABLE 4

Female-Specific Endpoints of Reproductive Toxicity

Body weight	
Organ weights	Ovary, uterus, vagina, pituitary
Visual examination & histopathology	Ovary, uterus, vagina, pituitary, oviduct, mammary gland
Estrous (menstrual*) cycle normality	Vaginal smear cytology
Hormone levels*	LH, FSH, estrogen, progesterone, prolactin
Lactation*	Offspring growth
Development	Normality of external genitalia*, vaginal opening, vaginal smear cytology, onset of estrus behavior (menstruation*)
Senescence (menopause*)	Vaginal smear cytology, ovarian histology

\* Endpoints that can be obtained relatively noninvasively with humans

The female reproductive system is also primarily controlled by the endocrine system. This control is accomplished through complex interactions involving the central nervous system (e.g. hypothalamus), pituitary, ovaries, the reproductive tract, and the secondary sexual organs. Other non-gonadotrophic components of the endocrine system may also modulate reproductive system function. Because it is difficult to measure certain important aspects of female reproductive function (e.g., increased rate of follicular atresia, ovulation

## **DRAFT--DO NOT QUOTE OR CITE**

failure), assessment of the endocrine status may provide needed insight that is not otherwise available.

To understand the significance of effects on the reproductive endpoints, it is critical that the relationships between the various reproductive hormones and the female reproductive organs be understood. Although certain effects may be identified routinely as adverse, all of the results should be considered in the context of the known biology.

The format for presentation of the female reproductive endpoints is altered from that used for the male to allow examination of events that are linked and that fluctuate with the changing endocrine status. Particularly, the organ weight, gross morphology, and histology are combined for each organ. Endpoints and endocrine factors for the individual female reproductive organs are discussed, with emphasis on the nonpregnant animal. This is followed by examination of measures of cyclicity and their interpretation. Then, considerations relevant to prepubertal, pregnant, lactating, and aging females are presented.

### **III.B.4.b. Body Weight, Organ Weight, Organ Morphology, and Histology**

#### **III.B.4.b.1. Body weight**

Toxicologists are often concerned about how a change in body weight may affect reproductive function. In females, an important consideration is that weights may fluctuate normally with the physiologic state of the animal because estrogen and progesterone are known to influence food intake and energy expenditure to an important extent (Wang, 1923; Wade, 1972). Water retention and fat deposition rates are also affected (Gelletti & Klopper, 1964; Hervey & Hervey, 1967). Food consumption is elevated during pregnancy, in part because of the elevated serum progesterone level. One of the most sensitive indicators of a compound with estrogenic action in the female rat is a reduction in food intake and body weight. Also, growth retardation induced by effects on extragonadal hormones (e.g., thyroid or growth hormone) can cause a delay in pubertal development, and induce acyclicity and infertility. Because of these endocrine-related fluctuations, the weights of the reproductive organs are poorly correlated with body weight, except in

## **DRAFT--DO NOT QUOTE OR CITE**

extreme cases. Thus, actual organ weight data, rather than organ to body weight ratios, should be reported and evaluated for the female reproductive system.

Chapin et al. (1993a, b) have studied the influence of food restriction on female Sprague-Dawley rats and Swiss CD-1 mice when body weights were 90, 80 or 70% of controls. Female rats were resistant to effects on reproductive function at 80% of control weight whereas mice showed adverse effects at 80%. These results indicate that differences exist between species (and probably between strains) in the response of the female rodent reproductive system to reduced food intake or body weight reduction.

### **III.B.4.b.2. Ovary**

The ovary serves a number of functions that are critical to reproductive activity, including production and ovulation of oocytes. Estrogen is produced by developing follicles and progesterone is produced by corpora lutea that are formed after ovulation.

#### **Ovarian weight**

Significant increases or decreases in ovarian weight compared with controls should be considered an indication of female reproductive toxicity. Although ovarian function shifts throughout the estrous cycle, ovarian weight in the normal rat does not show significant fluctuations. Still, oocyte and follicle depletion, persistent polycystic ovaries, inhibition of corpus luteum formation, luteal cyst development, reproductive aging, and altered hypothalamic-pituitary function may all be associated with changes in ovarian weight. Therefore, it is important that ovarian gross morphology and histology also be examined to allow correlation of alterations in those parameters with changes in ovarian weight. However, not all adverse histologic alterations in the ovary are concurrent with changes in ovarian weight. Therefore, a lack of effect on organ weights does not preclude the need for histologic evaluation.

#### **Histopathology**

Histologic evaluation of the three major compartments of the ovary (i.e., follicular, luteal, and interstitial) plus the epithelial capsule and ovarian stroma may indicate ovarian

## **DRAFT--DO NOT QUOTE OR CITE**

toxicity. A number of pathologic conditions can be detected by ovarian histology (Kurman & Norris, 1978; Langley & Fox, 1987). Methods are available to quantify the number of follicles and their stages of maturation (Plowchalk et al., 1993). These techniques may be useful when a compound depletes the pool of primordial follicles or alters their subsequent development and recruitment during the events leading to ovulation.

### **Adverse effects**

Significant changes in the ovaries in any of the following effects should be considered adverse:

- \* Increase or decrease in ovarian weight
- \* Increased incidence of follicular atresia
- \* Decreased number of primary follicles
- \* Decreased number or lifespan of corpora lutea
- \* Evidence of abnormal folliculogenesis or luteinization, including cystic follicles, luteinized follicles, and failure of ovulation
- \* Evidence of altered puberty or premature reproductive senescence

### **III.B.4.b.3. Uterus**

#### **Uterine weight**

An alteration in the weight of the uterus may be considered an indication of female reproductive organ toxicity. Compounds that inhibit cyclicity can dramatically reduce the weight of the uterus so that it appears atrophic and small. However, uterine weight fluctuates 3 to 4 fold throughout the estrous cycle, peaking at proestrus when, in response to increased estrogen secretion, the uterus is fluid filled and distended. This increase in uterine weight has been used as a basis for comparing relative potency of estrogenic compounds in bioassays (Kupfer, 1987). As a result of the wide fluctuation because of the influence of estrogenic compounds, uterine weights taken from cycling animals have a high variance, and large compound-related effects are required to demonstrate a significant effect

## **DRAFT--DO NOT QUOTE OR CITE**

unless interpreted relative to that animal's estrous cycle stage. A number of environmental compounds (e.g., pesticides such as methoxychlor and chlordecone, mycotoxins, polychlorinated biphenyls, and phytoestrogens) possess varying degrees of estrogenic activity and have the potential to stimulate the female reproductive tract (Barlow & Sullivan, 1982; Bulger & Kupfer, 1985; Hughes, 1988).

When pregnant or postpartum animals are examined, the numbers of implantation sites or implantation scars should be counted. This information, along with corpus luteum counts, can be used to calculate pre- and postimplantation losses.

### Histopathology

The histologic appearance of the normal uterus fluctuates with stage of the estrous cycle and pregnancy. The uterine endometrium is sensitive to influences of estrogens and progestogens (Warren et al., 1967), potentially leading to hypertrophy and hyperplasia. Conversely, interference with ovarian activity results in endometrial hypoplasia and atrophy. Effects induced during development may delay or prevent puberty, resulting in persistence of infantile genitalia.

### Adverse effects

Effects on the uterus that may be considered adverse include significant dose-related alteration of weight, as well as gross anatomic or histologic abnormalities. In particular, any of the following effects should be considered as adverse.

- \* Infantile or malformed uterus or cervix
- \* Decreased or increased uterine weight
- \* Endometrial hyperplasia, hypoplasia, or aplasia
- \* Decreased number of implantation sites

### III.B.4.b.4. Oviducts

Typically, the oviducts are not weighed or examined histologically in tests for reproductive toxicity. However, information from visual and histologic examinations is of value to detect morphologic anomalies. Descriptions of pathologic effects within the

## **DRAFT--DO NOT QUOTE OR CITE**

oviducts of animals other than humans are not common. Hypoplasia of otherwise well-formed oviducts results most commonly from a lack of estrogen stimulation, and for this reason, this condition may not be recognized until after puberty. Hyperplasia of the oviductal epithelium results from prolonged estrogenic stimulation. Anomalies induced during development have also been described, including agenesis, segmental aplasia, and hypoplasia.

Anatomic anomalies in the oviduct occurring in excess of control incidence should be considered as adverse effects. Hypoplasia or hyperplasia of the oviductal epithelium may be considered as an adverse effect, particularly if that result is consistent with observations in the uterine histology.

### **III.B.4.b.5. Vagina and external genitalia**

#### **Vaginal weight**

Vaginal weight changes should parallel those seen in the uterus during the estrous cycle, although the magnitude of the changes is smaller.

#### **Histopathology**

In rodents, cytologic changes in the vaginal epithelium (vaginal smear) may be used to identify the different stages of the estrous cycle (see Section III.B.4.d.). The vaginal smear pattern may be useful to identify conditions that would delay or preclude fertility, or affect sexual behavior. Other histologic alterations that may be observed include aplasia, hypoplasia, and hyperplasia of the vaginal epithelial cell lining.

#### **Developmental effects**

Developmental abnormalities, either genetic or related to prenatal exposure to compounds that disrupt the endocrine balance, include agenesis, hypoplasia, and dysgenesis. Hypoplasia of the vagina may be concomitant with hyperplasia of the external genitalia and can be induced by gonadal or adrenal steroid exposure. In rodents, malpositioning of the vaginal and urethral ducts is common in steroid-treated females. Such developmentally induced lesions are irreversible.

## **DRAFT--DO NOT QUOTE OR CITE**

The sex ratio observed at birth may be affected by developmental exposure to androgens. In cases of incomplete sex reversal because of such exposures, female rodents may appear more male-like and have an increased anogenital distance.

At puberty, the opening of the vaginal orifice normally provides a simple and useful developmental marker. However, estrogenic or antiestrogenic chemicals can act directly on the vaginal epithelium and alter the age at which vaginal patency is lost without truly affecting puberty (see Section III.B.4.f.).

### Adverse effects

Significant effects on the vagina that may be considered adverse include the following:

- \* Increases or decreases in weight
- \* Infantile or malformed vagina or vulva, including masculinized vulva  
or increased anogenital distance
- \* Vaginal hypoplasia or aplasia
- \* Altered timing of vaginal opening
- \* Abnormal vaginal smear cytology pattern

### III.B.4.b.6. Pituitary

#### Pituitary weight

Alterations in weight of the pituitary gland may be considered an adverse effect. The discussion on pituitary weight and histology for males (see Section III.B.3.) is pertinent also for females. Pituitary weight increases normally with age, as well as during pregnancy and lactation. Changes in pituitary weight can occur also as a consequence of chemical stimulation. Increased pituitary weight often precedes tumor formation, particularly in response to treatment with estrogenic compounds. Increased pituitary size associated with estrogen treatment may be accompanied by hyperprolactinemia and constant vaginal estrus. Decreased pituitary weight is less common but may result from decreased estrogenic stimulation (Cooper et al., 1989).



## **DRAFT--DO NOT QUOTE OR CITE**

### Histopathology

In histologic evaluations with rats and mice, the relative size of cell types in the anterior pituitary (acidophils and basophils) has been reported to vary with the stages of the reproductive cycle and in pregnancy (Holmes & Ball, 1974). Therefore, the relationship of morphologic pattern to estrous or menstrual cycle stage or pregnancy status should be considered in interpreting histologic observations on the female pituitary.

### Adverse effects

A significant increase or decrease in pituitary weight should be considered an adverse effect. Significant histopathologic damage in the pituitary should be considered an adverse effect, but should be shown to involve cells that control gonadotropin or prolactin production to be called a reproductive effect.

### III.B.4.c Oocyte Production

#### III.B.4.c.1. Folliculogenesis

In normal females, all of the follicles (and the resident oocytes) are present at or soon after birth. The large majority of these follicles undergo atresia and are not ovulated. Therefore, the ovaries from control animals have a background rate of follicular atresia that must be distinguished from an increased rate resulting from exposure to toxicants (Smith, B.J. et al., 1991; Heindel & Chapin, 1993). If the population of follicles is depleted, it cannot be replaced and the female will be rendered infertile. In humans, depletion of oocytes can lead to premature menopause.

In rodents, lead, mercury, cadmium, and polycyclic aromatic hydrocarbons have all been implicated in the arrest of follicular growth at various stages of the life cycle (Mattison & Thomford, 1989). Susceptibility to oocyte toxicity varies considerably between species (Mattison & Thorgeirsson, 1978).

Environmental toxicants that affect gonadotropin-mediated steroidogenesis or follicular maturation can prolong the follicular phase of the estrous or menstrual cycle and cause atresia of follicles that would otherwise ovulate. Estrogenic as well as antiestrogenic

## **DRAFT--DO NOT QUOTE OR CITE**

agents can produce that effect. Also, normal follicular maturation is essential for normal formation and function of the corpus luteum formed after ovulation (McNatty, 1979).

### **III.B.4.c.2. Ovulation**

Chemicals can delay or block ovulation by disrupting the ovulatory surge of LH or by interfering with the ability of the maturing follicle to respond to that gonadotropic signal. Examples for rats are the pesticides chlordimeform and amitraz (Goldman et al., 1990) and compounds that interfere with normal central norepinephrine receptor stimulation (Drouva et al., 1982). Compounds that increase central opioid receptor stimulation also decrease serum LH and inhibit ovulation in monkeys and rats (Pang et al., 1977; Smith, C.G., 1983). Delayed ovulation can alter oocyte viability and cause trisomy and polyploidy in the conceptus (Fugo & Butcher, 1966; Butcher & Fugo, 1967; Butcher et al., 1969, 1975; Na et al., 1985).

### **III.B.4.c.3. Corpus luteum**

The corpus luteum arises from the ruptured follicle and secretes progesterone, which has an important role in the estrous or menstrual cycle. It also serves as the principal source of progesterone required for the maintenance of early pregnancy in the human (Csapo & Pulkkinen, 1978). Therefore, establishment and maintenance of normal corpora lutea are essential to normal reproductive function. However, with the exception of histopathologic evaluations that may establish only their presence or absence, these structures are not evaluated in routine testing. Additional research is needed to determine the importance of incorporating endpoints that examine direct effects on luteal function in routine toxicologic testing.

#### **Adverse effects**

Increased rates of follicular atresia and oocyte toxicity can lead to premature menopause. Altered follicular development, ovulation failure, or altered corpus luteum formation and function can result in disruption of cyclicity, reduced fertility, and, in non-

## **DRAFT--DO NOT QUOTE OR CITE**

primates, interference with normal sexual behavior. Therefore, significant increases in rate of follicular atresia, evidence of oocyte toxicity, interference with ovulation, or altered corpus luteum formation or function should be considered adverse effects.

### **III.B.4.d. Alterations in the Female Reproductive Cycle**

The pattern of events in the estrous cycle provides a useful indicator of the normality of reproductive neuroendocrine and ovarian function in the nonpregnant female. It also provides a means to interpret hormonal, histologic, and morphologic measurements relative to stage of the cycle, and can be useful to monitor the status of mated females. Estrous cycle normality can be monitored in the rat and mouse by observing the changes in the vaginal smear cytology (Long & Evans, 1922; Cooper et al., 1993). To be most useful with cycling females, vaginal smear cytology should be examined daily for at least three normal estrous cycles prior to treatment, after onset of treatment, and before necropsy (Kimmel, G.L. et al., In press).

Daily vaginal smear data from rodents can provide useful information on (1) cycle length, (2) occurrence or persistence of estrus, (3) duration or persistence of diestrus, (4) incidence of spontaneous pseudopregnancy, (5) distinguishing pregnancy from pseudopregnancy (based on the number of days smear remains leukocytic), and (6) indications of fetal death and resorption by the presence of blood in the smear after day 12 of gestation. The technique also can detect onset of reproductive senescence in rodents (LeFevre & McClintock, 1988). It is useful further to detect the presence of sperm in the vagina as an indication of mating.

In nonpregnant females, repetitive occurrence of the four stages of the estrous cycle at regular, normal intervals suggests that neuroendocrine control of the cycle and ovarian responses to that control are normal. Occasionally, even normal, control animals show an irregular cycle. However, a significant alteration compared with controls in the interval between occurrence of estrus smears for a treatment group is cause for concern. Generally, the cycle will be lengthened or terminated. Lengthening of the cycle may be a result of

## **DRAFT--DO NOT QUOTE OR CITE**

increased duration of either the estrus or diestrus phase. Knowing the affected phase can provide direction for further investigation.

The persistence of regular vaginal cycles after treatment does not necessarily indicate that ovulation occurred, because luteal tissue may form in follicles that have not ruptured. This effect has been observed after treatment with anti-inflammatory agents (Walker et al., 1988). However, that effect should be reflected in reduced fertility. Conversely, subtle alterations of cyclicity can occur at doses below those that alter fertility (Gray et al., 1989).

Irregular cycles may reflect impaired ovulation. Extended vaginal estrus usually indicates that the female cannot spontaneously achieve the ovulatory surge of LH (Huang & Meites, 1975). A number of compounds have been shown to alter the characteristics of the LH surge including anesthetics (Nembutal), neurotransmitter receptor binding agents (Drouva et al., 1982), and the pesticides chlordimeform and lindane (Cooper et al., 1989; Morris et al., 1990). The female in constant estrus may be sexually receptive and ovulation may be induced by mating (Brown-Grant et al., 1973; Smith, E.R. & Davidson, 1974), but the fertility of such matings has not been evaluated thoroughly. Significant delays in ovulation can result in increased embryonic abnormalities and loss (Fugo & Butcher, 1966).

Persistent diestrus indicates cessation of follicular development and ovulation and thus infertility. Prolonged vaginal diestrus, or anestrus, may be indicative of agents (e.g., polyaromatic hydrocarbons) that interfere with follicular development or deplete the pool of primordial follicles (Mattison & Nightingale, 1980). The ovaries of the anestrus female are atrophic, with few primary follicles and an unstimulated uterus (Huang & Meites, 1975). Serum estradiol and progesterone are minimal.

### Adverse effects

Significant evidence that the estrous cycle (or menstrual cycle in primates) has been disrupted should be considered an adverse effect. Included should be evidence of abnormal cycle length or pattern, ovulation failure, or abnormal menstruation.

## **DRAFT--DO NOT QUOTE OR CITE**

### **III.B.4.e. Mammary Gland and Lactation**

The mammary glands of normal adults change dramatically during the period around parturition because of the sequential effects of a number of gonadal and extragonadal hormones. Milk letdown is dependent on the suckling stimulus and the release of oxytocin from the posterior pituitary. Thus, mammary tissue is highly endocrine-dependent for development and function.

Mammary gland size, milk production and release, and histology can be affected adversely by toxic agents. Reduced growth of young could be caused by reduced milk availability, by ingestion of a toxic agent secreted into the milk, or by other factors unrelated to lactational ability (e.g., reduced growth could be caused by poor suckling ability). Perinatal exposure to steroid hormones and other chemicals can alter mammary gland morphology and tumor potential in adulthood. Because of the tendency for mobilization of lipids from adipose tissue and secretion of those lipids into milk by lactating females, milk may contain lipophilic agents at concentrations equal to or higher than those present in the blood or organs of the dam. Thus, suckling offspring may be exposed to elevated levels of such toxicants.

During lactation, the mammary glands can be dissected and weighed only with difficulty. This provides a measure of milk production by the dam. A simple estimate of milk production may be obtained by measuring litter weights taken after one or two hours of nursing by milk-deprived pups (6 hours). Milk from the stomachs of pups can also be weighed. Cleared and stained whole mounts of the mammary gland can be prepared at necropsy for histologic examination. In addition, the DNA, RNA, and lipid content of the mammary gland and the composition of the milk have been measured following toxicant administration as indicators of toxicity to this target organ.

Significant reductions in milk production or negative effects on milk quality, whether measured directly or reflected in impaired development of young, should be considered adverse reproductive effects.

## DRAFT--DO NOT QUOTE OR CITE

### III.B.4.f. Developmental and Pubertal Alterations

#### Developmental effects

Alterations of reproductive differentiation and development can result in infertility, functional and morphologic alterations of the reproductive system, and cancer (Steinberger & Lloyd, 1985; Gray, 1991). Prenatal and postnatal exposure to toxicants can produce changes that may not be predicted from effects seen in adults, and those effects are often irreversible. Adverse developmental outcomes in either sex can result from exposure to toxicants in utero, through contact with exposed dams, or in milk. Dosing of dams during lactation also can result in developmental toxicity through impaired nursing capability of the dams.

Effects observed following exposure to agents in rodents include alterations in the genitalia (including ano-genital distance), impaired sexual behavior, delay or acceleration of the onset of puberty, and reduced fertility. Many of these effects have been detected in human females and males exposed prenatally to diethylstilbestrol (DES), other estrogens, progestins, and androgens. Accelerated reproductive aging and tumors of the reproductive tract have been observed in laboratory animal and human females. Generally, the type of effect seen may differ depending on the stage at which the exposure occurred. Effects may include anomalies in sexual behavior or ability to produce normal gametes that are not observed until after puberty. Hepatic enzyme systems for steroid metabolism that are imprinted during development may be altered in males. Testis descent from the abdominal cavity into the scrotum may be delayed or may not occur. Other agents, such as busulfan, that do not have endocrine activity may act via different mechanisms during critical periods of development to cause similar effects.

#### Effects on puberty

In female rats and mice, the age at vaginal opening is the most commonly measured marker of puberty. This event results from increases in the levels of estradiol in the blood. The ages and weights of females at the first cornified (estrus) vaginal smear, the first diestrus smear, and the onset of vaginal cycles have also been used as endpoints for onset

## **DRAFT--DO NOT QUOTE OR CITE**

of puberty. In males, preputial separation or appearance of sperm in expressed urine or ejaculates can serve as markers of puberty. Body weight at puberty may provide a means to separate specific delays in puberty from those that are related to general delays in development. Agents may differentially affect the endpoints related to puberty onset, so it is useful to have information on more than one marker.

Puberty can be accelerated or delayed by exogenous agents, and both types of effects may be adverse. For example, an acceleration of vaginal opening may be associated with a delay in the onset of cyclicity, infertility, and with accelerated reproductive aging (Gorski, 1979). Delays in pubertal development in rodents are usually related to delayed maturation or inhibition of function of the hypothalamic-pituitary axis. Adverse reproductive outcomes have been reported in rodents when puberty is altered by a week or more, but the biologic relevance of a change in these measures of a day or two is unknown (Gray, 1991).

### Adverse effects

Effects induced or observed during the perinatal period should be judged using guidance from the Guidelines for Assessment of Developmental Risk (U.S. Environmental Protection Agency, 1991) as well as from these Guidelines. Significant effects on age at puberty, either early and delayed, should be considered adverse as should malformations of the internal or external genitalia. Included as adverse effects for females should be effects on age at vaginal opening, onset of cyclic vaginal smears, onset of menstruation, or onset of an endocrine or behavioral pattern consistent with estrous or menstrual cyclicity. Included as adverse effects for males should be delay or failure of testis descent, as well as delays in age at preputial separation or appearance of sperm in expressed urine or ejaculates.

### III.B.4.g. Reproductive Senescence

With advancing age, there is a loss of the regular ovarian cycles and associated normal cyclical changes in the uterine and vaginal epithelium that are typical of the young-adult female rat (Cooper & Walker, 1979). Although the mechanisms responsible for this loss of cycling are not thoroughly understood, age-dependent changes within the hypothalamic-

## **DRAFT--DO NOT QUOTE OR CITE**

pituitary control of ovulation are impaired (Cooper et al., 1980; Finch et al., 1984).

Cumulative exposure to estrogen secreted by the ovary may play a role, as treatment with estrogens during adulthood can accelerate the age-related loss of ovarian function (Brawer & Finch, 1983). In contrast, the principal cause of the loss of ovarian cycling in humans appears to be the depletion of oocytes (Mattison, 1985).

Prenatal or postnatal treatment of females with estrogens or estrogenic pesticides can also cause impaired ovulation and sterility (Gorski, 1979). These observations imply that alterations in ovarian function may not be noticeable immediately after treatment but may become evident at puberty or influence the age at which reproductive senescence occurs.

### **Adverse effects**

Significant effects on measures showing a decrease in the age of onset of reproductive senescence in females should be considered adverse. Included as adverse effects should be cessation of normal cycling measured by vaginal smear cytology, ovarian histopathology, or an endocrine pattern that is consistent with this interpretation.

## **III.C. HUMAN STUDIES**

In principle, human data are preferred for risk assessment. At this time, reproductive data for humans are available for only a limited number of toxicants. As the field develops further, expanding both the endpoints available for study and agents covered, risk assessments will more frequently incorporate human data. The following describes the methods of generation and evaluation of human data and the weight human data should be given in risk assessments.

"Human studies" include 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, because they can best evaluate exposure-response relationships. 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



## **DRAFT--DO NOT QUOTE OR CITE**

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 support 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.

### **III.C.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.

#### **III.C.1.a. 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; Hatch & Kline, 1981; Wilcox, 1983; Sever & Hessol, 1984; Axelson, 1985; Tilley et al., 1985; Kimmel, C.A. et al., 1986). Some of the more prominent factors are discussed below.

##### **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, groups are defined by exposure, and their health outcomes examined. Common outcomes, such as recognized fetal loss, require hundreds of pregnancies to have a high probability of detecting a modest increase in risk (e.g., 133 participants in both exposed and unexposed groups to detect a twofold increase;  $\alpha = 0.05$ , 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 pregnancies in both exposed and unexposed groups)

## **DRAFT--DO NOT QUOTE OR CITE**

(Bloom, 1981; Selevan, 1981, 1985; Sever & Hessol, 1984; Stein, Z. et al., 1985; Kimmel, C.A. et al., 1986). Semen evaluation may require fewer subjects depending on the sperm parameters evaluated, especially when each man is used as his own control (Wyrobek, 1982, 1984). In case-referent studies, groups are defined by health status and prior exposures are examined. Study sizes are dependent upon the frequency of exposure within the source population. The confidence one has in the results of a study with negative findings is related directly to the power of the study to detect meaningful differences in the endpoints.

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 is useful in evaluating negative findings. Negative findings in a study of low power would be given considerably less weight than either a positive study or a negative study with high power. Positive findings from very small studies are open to question because of the instability of the risk estimates and the highly selected nature of the population.

### Potential bias in data collection

Bias may result from the way the study group is selected or information collected (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 under-ascertain events, because women are not always hospitalized for these outcomes. More weight would be given in a risk assessment to a study in which a more complete list of pregnancies is obtained by, for example, collecting biologic data [e.g., human chorionic gonadotropin (hCG) measurements] of pregnancy status from study members. These studies may also be affected by bias. The

## **DRAFT--DO NOT QUOTE OR CITE**

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). 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 because some factors that influence employment status may also affect 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). 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 & Pinney, 1989).

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 reduces the likelihood of such bias. Studies with lower likelihood of such bias 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 more complete and valid than those of the husbands (Selevan, 1980; Selevan et

## **DRAFT--DO NOT QUOTE OR CITE**

al., 1982). Studies based on interview data from the appropriate respondents 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 on male workers' exposures and factors relating to semen quality (e.g., fever within the past 2 to 3 months) should be obtained from the workers themselves.

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, A. & 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).

### **Collection of data on other risk factors, effect modifiers, and confounders**

Risk factors for reproductive toxicity include such characteristics as age, smoking, alcohol consumption, drug use, and past reproductive history. Additionally, occupational and environmental exposures may be risk factors for these effects. Known and potential risk factors should be examined to identify those that may be confounders or effect modifiers. An effect modifier is a factor that produces different exposure-response relationships at different levels of that factor. For example, age would be an effect modifier if the risk associated with a given exposure increased with the individual'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, smoking might be a confounder in a study of the association of socioeconomic status and fertility because smoking may be associated with both.

## **DRAFT--DO NOT QUOTE OR CITE**

Both effect modifiers and confounders 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 for the Interagency Regulatory Liaison Group, 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). Studies that fail to account for these important factors should be given less weight in a risk assessment.

### Statistical factors

As in studies of test animals, pregnancies experienced by the same woman are not fully independent events (Kissling, 1981; Selevan, 1985). Women who have had fetal loss are reported to be more likely to have subsequent losses (Leridon, 1977). In test 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, making analysis considering nonindependence of events difficult (Epidemiology Workgroup for the Interagency Regulatory Liaison Group, 1981; Kissling, 1981; Selevan, 1981). If more than one pregnancy per woman is included, as is often necessary with 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 time, a generally accepted solution to this problem has not been developed.

### III.C.1.b. Selection of Outcomes for Study

As already discussed, a number of endpoints can be considered in the evaluation of adverse reproductive effects. However, some of the outcomes are not easily observed in humans, such as early embryonic loss, reproductive capacity of the offspring, and invasive evaluations of reproductive function (e.g., testicular biopsies). Currently, the most feasible

## **DRAFT--DO NOT QUOTE OR CITE**

endpoints for epidemiologic studies are (1) indirect measures of fertility/infertility; (2) reproductive history studies of some pregnancy outcomes (e.g., embryonic/fetal loss, birth weight, sex ratio, congenital malformations, postnatal function, and neonatal growth and survival); (3) semen evaluations; and (4) menstrual history. 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 reproductive outcomes available for epidemiologic examination are limited by a number of factors, including the relative magnitude of the exposure, the size and demographic characteristics of the population, and the ability to observe the reproductive outcome in humans. Improved methods for identifying some outcomes such as embryonic loss using more sensitive hCG assays may change the spectrum of outcomes available for study (Wilcox et al., 1985; Sweeney et al., 1988). Other endpoints require invasive techniques to obtain samples (e.g., histopathology) or have high intra- or interindividual variability (e.g., serum hormone levels, sperm count).

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 birth control practices would also affect the number of outcomes available for study.

In addition to the above-mentioned factors, reproductive endpoints may be envisioned as effects recognized at various points in a continuum, starting before conception and continuing 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 (Bloom, 1981). A shift in the patterns of outcomes could result from differences in timing or in level of exposure (Selevan & Lemasters, 1987).

The following section discusses various human male and female reproductive endpoints. These are followed by a discussion of reproductive history studies.

### Male Endpoints - Semen Evaluations

## **DRAFT--DO NOT QUOTE OR CITE**

The use of semen analysis was discussed in Section III.B.3.d. Most epidemiologic studies of semen characteristics have been conducted in occupational groups and patients receiving drug therapy. Obtaining specimens with a high level of participation in the workforce has been difficult, because social and cultural attitudes concerning sex and reproduction may affect cooperation of the study groups. Increased participation may occur in men who are planning to have children or who are concerned either about existing reproductive problems or about possible ill effects of their exposures. Unless controlled, such biased participation may yield unrepresentative estimates of risk associated with exposure, resulting in data that are less useful for risk assessment. Response rates are typically less than 70% in such studies and may be even lower in the comparison group (Egnatz et al., 1980; Lipshultz et al., 1980; Milby & Whorton, 1980; Lantz et al., 1981; Meyer, 1981; Milby et al., 1981; Rosenberg et al., 1985). Some of the low response rates may be caused by inclusion of vasectomized men in the total population, although this could vary widely by population (Milby & Whorton, 1980). Participation in the comparison group may be biased toward those with pre-existing reproductive problems. The response rate may be improved substantially with proper education and payment of subjects (Ratcliffe et al., 1986, 1987).

Several factors may influence the semen evaluation, including the period of abstinence preceding collection of the sample, health status, and social habits (e.g., alcohol, drugs, smoking). Data on these factors may be collected by interview, subject to the limitations described for pregnancy outcome studies.

Such studies have also included an evaluation of endocrine status of exposed males. These evaluations include determination of hormone levels in the blood and urine.

### **Female Endpoints**

Reproductive effects may result from a variety of exposures. For example, environmental exposures may result in oocyte toxicity, in which a loss of primary oocytes irreversibly affects the woman's fertility. The exposures of importance may occur during the prenatal period, and beyond. Oocyte depletion is difficult to examine directly in women

## **DRAFT--DO NOT QUOTE OR CITE**

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).

Numerous diagnostic methods have been developed to evaluate female reproductive dysfunction. Although these methods have rarely been used for occupational or environmental toxicologic evaluations, they may be helpful in defining biologic parameters and the mechanisms related to female reproductive toxicity. If clinical observations are able to link exposures to the reproductive effect of concern, these data may aid the assessment of adverse female reproductive toxicity. The following clinical observations include endpoints that may be reported in case reports or epidemiologic research studies.

Reproductive dysfunction can be studied by the evaluation of irregularities of menstrual cycles. However, menstrual cyclicity is affected by many parameters such as age, nutritional status, stress, certain drugs, and the use of contraceptive measures that alter endocrine feedback. Vaginal bleeding at menstruation is a reflection of withdrawal of steroidogenic support, particularly progesterone. Vaginal bleeding can occur at midcycle, in early miscarriage, after withdrawal of contraceptive steroids, or after an inadequate luteal phase. The length of the menstrual cycle, particularly the follicular phase, can vary between individuals and may make it difficult to determine significant effects in populations of women (Burch et al., 1967; Treloar et al., 1967). However, menstrual dysfunction data have been used to examine adverse reproductive effects in women exposed to styrene in the workplace (Lemasters et al., 1985).

Vaginal cytology may provide information on the functional state of reproductive cycles. Cytologic evaluations, along with the evaluation of changes in cervical mucus viscosity, can be used to estimate the occurrence of ovulation and determine different stages of the reproductive cycle.

The endocrine status of a woman can be evaluated by the measurement of hormones in blood and urine. Progesterone can also be measured in saliva. Because the female reproductive endocrine milieu is changing in a cyclic pattern, single sample analysis does



## **DRAFT--DO NOT QUOTE OR CITE**

not provide adequate information for evaluating alterations in the reproductive function. Still, a single sample for progesterone determination some 7 to 9 days after the estimated midcycle surge of gonadotropins in a regularly cycling woman may provide suggestive evidence for the presence of a functioning corpus luteum and prior follicular maturation and ovulation. Clearly clinically abnormal levels of gonadotropins, steroids, or other biochemical parameters may be detected from a single sample. Preferably, multiple samples could be collected and observed in conjunction with events in the menstrual cycle.

Ovulation can be estimated by the biphasic shift in basal body temperature. Ovulation can also be detected by serial measurement of hormones in the blood or urine and analysis of estradiol and gonadotropins at midcycle. After ovulation, luteal phase function can be assessed by analysis of progesterone secretion and by evaluation of endometrial histology. Tubal patency is an important endpoint that can be observed in clinical evaluations of reproductive function (Forsberg, 1981). These latter evaluations are less likely to be present in epidemiologic studies or surveillance programs because of the invasiveness of the procedures.

### **III.C.1.c. Reproductive History Studies**

#### **Measures of fertility**

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, and contraceptive use (Wong et al., 1979; Levine et al., 1980, 1981, 1983; 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

## **DRAFT--DO NOT QUOTE OR CITE**

has similar limitations in interpretation (Gaffey, 1976; McMichael, 1976; Tsai & Wen, 1986).

Analysis of the time between recognized pregnancies or live births has been suggested as another indirect measure of fertility (Dobbins et al., 1978; Baird & Wilcox, 1985; Baird et al., 1986; Weinberg & Gladen, 1986). Because the time 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 or couple.

Fertility may also be affected by alterations in sexual behavior. However, data linking toxic exposures to these alterations in humans are limited and are not easily obtained in epidemiology studies (see Section III.C.1.e.).

### **Pregnancy outcomes**

Pregnancy outcomes examined in human studies of parental exposures may include embryo or fetal loss, congenital malformations, birth weight effects, sex ratio at birth, and possibly 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 Guidelines for Developmental Toxicity Risk Assessment (U.S. Environmental Protection Agency, 1991). As mentioned above, epidemiologic studies that focus on only one type of pregnancy outcome may miss a true effect of exposure. Studies that examine multiple endpoints could yield more information, but results may be more difficult to interpret.

Evidence of a dose-response relationship is usually an important criterion in the assessment of exposure to a potentially toxic agent. However, traditional dose-response relationships may not always be observed for some endpoints (Wilson, 1973). For example, with increasing dose, a pregnancy might end in embryo or 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 & Lemasters, 1987) (for a more detailed description, see Section III.C.1.e.). Therefore, a risk assessment

## **DRAFT--DO NOT QUOTE OR CITE**

should, when possible, attempt to look at the relationship of different reproductive endpoints and patterns of exposure.

In addition to the above effects, genetic damage to germ cells may potentially result from exposures to the reproductive system. Outcomes resulting from germ cell mutations could include reduced probability of fertilization and increased probability of embryo or fetal loss and postnatal developmental effects. Based on studies with test species, critical exposures are to germ cells or early zygotes. Germ cell mutagenicity could be expressed also as genetic diseases in future generations. Unfortunately, these studies are difficult to conduct in human populations because of the long time between exposure and outcome. For more information, refer to the Guidelines for Mutagenicity Risk Assessment (U.S. Environmental Protection Agency, 1986c).

### **III.C.1.d. Community Studies and 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 toxicants in the water system and adverse pregnancy outcome (Deane et al., 1989; Swan 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 with 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 & Fine, 1980; Papier, 1985), and childhood cancer (Fabia & Thuy, 1974; Hakulinen et al., 1976; Kwa & 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 & Fine, 1980). Such studies are most helpful in the identification of

## **DRAFT--DO NOT QUOTE OR CITE**

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 or semen evaluations could be followed to monitor reproductive effects of exposures. Both could yield very useful data for risk assessment; however, a semen evaluation or other clinical program would be costly to maintain, and there are numerous impediments to the collection of reliable and valid information in the workplace. In addition to these concerns, impediments include potentially low employee participation rates, employee sensitivities, and confidentiality requirements.

### **III.C.1.e. Identification of Important Exposures for Reproductive Effects**

For all examinations of the relationship between reproductive 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, low birth weight, and measures of infertility). These exposures, plus postnatal exposure from breast milk, food, and the environment, may also be associated with postnatal developmental effects (e.g., changes in behavioral and cognitive function or growth).

General environmental exposures are typically lower than in industrial or 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 & 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.

Data on exposure from human studies are frequently qualitative, such as employment or residence histories. More quantitative data may be difficult to obtain because of the

## **DRAFT--DO NOT QUOTE OR CITE**

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 outcomes studied, the biologic mechanism affected by exposure, and the biologic half-life of the agent. The 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 (Smith, P.E., 1939; Leridon, 1977; Collins, 1978; Scommegna et al., 1980; Deane et al., 1989; Swan et al., 1989). As more prospective studies are done, better estimates of exposure will be developed.

### **III.C.2. Examination of Clusters, Case Reports, or Series**

The identification of cases or clusters of adverse reproductive effects is generally limited to those identified by the individuals involved or clinically by their physicians. The likelihood of identification varies with the gender of the exposed person. Identification of infertility in either gender is difficult. This might be thought of as identification of a nonevent (e.g., lack of pregnancies or children), and thus is much harder to recognize than are some developmental effects, including malformations, resulting from in utero exposure.

The identification of cases or clusters of adverse male reproductive outcomes may be limited because of cultural norms that may inhibit the reporting of impaired fertility in men. Identification is also limited by the decreased likelihood of recognizing adverse pregnancy outcomes as a result of paternal exposure rather than maternal exposure. Thus far, only one human male reproductive toxicant, dibromochloropropane (DBCP), has been identified after observation of a cluster of male infertility through an atypically high level of communication among the workers' wives (Whorton et al., 1977, 1979; Biava et al., 1978; Whorton & Milby, 1980).

## **DRAFT--DO NOT QUOTE OR CITE**

Adverse effects identified in females have, thus far, been limited to adverse pregnancy outcomes such as fetal loss and congenital malformations. Identification of other effects, such as infertility or menstrual disorders, may be difficult, as noted above.

Case reports may have importance in the recognition of reproductive toxicants. However, they are probably of greatest use in suggesting topics for further investigation. 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 test animals also occur in humans.

### **III.D. PHARMACOKINETIC CONSIDERATIONS**

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 or peak plasma concentrations, placental metabolism and transfer, comparative metabolism, and concentrations of the parent compound and metabolites in target organs may be useful in predicting risk for reproductive 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 (Wilson et al., 1975, 1977), determining dosimetry at target sites, and comparing pharmacokinetic profiles for various dosing regimens or routes of exposure. EPA's Office of Prevention, Pesticides, and Toxic Substances has published protocols for metabolism studies that may be adapted to provide information useful in reproductive toxicity risk assessment for a suspect toxicant. Pharmacokinetic studies in reproductive toxicology are most useful if the data are obtained with animals that are at the same reproductive status and stage of life (e.g., pregnant, nonpregnant, embryo or fetus, neonate, prepubertal, adult) at which reproductive insults are expected to occur in humans.

Specific guidance regarding both the development and application of pharmacokinetic data was agreed on by the participants of the Workshop on Dermal Developmental Toxicity Studies (Kimmel, C.A. & Francis, 1990). This guidance is also

## **DRAFT--DO NOT QUOTE OR CITE**

applicable to non-dermal reproductive toxicity studies. Participants of the Workshop concluded that absorption data are needed both when a dermal study does or does not show effects. The results of a dermal study showing no effects and without blood level data are potentially misleading and are inadequate for risk assessment, especially if interpreted as a "negative" study. In studies where adverse effects are detected, regardless of the route of exposure, pharmacokinetic data can be used to establish the internal dose in maternal and paternal animals for risk extrapolation purposes.

The existence of a Sertoli cell barrier (formerly called the blood-testis barrier) in the seminiferous tubules may influence the pharmacokinetics of a potential testicular toxicant by restricting access of compounds to the adluminal compartment of seminiferous tubules. The Sertoli cell barrier is formed by tight junctions between Sertoli cells and divides the seminiferous epithelium into basal and adluminal compartments (Russell et al., 1990). The basal compartment contains the spermatogonia and primary spermatocytes to the preleptotene stage, whereas more advanced germ cells are located on the adluminal side. This selectively permeable barrier is most effective in limiting the access of large, hydrophilic molecules in the intertubular lymph to cells on the adluminal side. An analogous barrier in the ovary has not been found, although the zona pellucida and granulosa cells may modulate access of chemicals to oocytes.

The reproductive organs appear to have a wide range of metabolic capabilities directed at both steroid and xenobiotic metabolism. However, there are substantial differences between compartments within the organs in types and levels of enzyme activities (Mukhtar et al., 1978). Recognition of these differences can be important in understanding the potential of agents to have specific toxic effects.

Most pharmacokinetic studies have incompletely characterized the distribution of toxic agents and their subsequent metabolic fate within the reproductive organs. Generalizations based on hepatic metabolism are not necessarily adequate to predict the fate of the agent in the testis, ovary, placenta, or conceptus. For example, the metabolic profile for a given agent may differ in the male between the liver and the testis and between the

## **DRAFT--DO NOT QUOTE OR CITE**

maternal liver and placenta. Detailed interspecies comparisons of the metabolic capabilities of the testis, ovary, placenta, and conceptus also have not been conducted. For some xenobiotics, significant differences in metabolism have been identified between males and females. This is, in part, attributable to organizational effects of the gonadal steroids in the developing liver (Gustafsson et al., 1980; Skett, 1988). Also, in adults, the sex steroids have been shown to affect the activity of a number of enzymes involved in the metabolism of administered compounds. Thus, the blood levels of a toxicant, as well as the final concentration in the target tissue, may differ significantly between sexes. If data are to be used effectively in interspecies comparisons and extrapolations for these target systems, more attention should be directed to the pharmacokinetic properties of chemicals in the reproductive organs and in other organs that are affected by reproductive hormones.

### **III.E. COMPARISONS OF MOLECULAR STRUCTURE**

Comparisons of the chemical or physical properties of an agent with those of known reproductive toxicants may provide some indication of a potential for reproductive 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 reproductive toxicology, although data are available that suggest structure-activity relationships for certain classes of chemicals (e.g., glycol ethers, some estrogens, androgens, other steroids, retinoids, phthalate esters, short-chain halogenated hydrocarbon pesticides, metals). The literature has been reviewed and a set of classifications offered relating structure to reported male reproductive activity (Bernstein, 1984). Although limited in scope and in need of validation, such schemes do provide hypotheses that can be tested.

In spite of the limited information available on structure-activity relationships in this field, 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 other data are available.



## **DRAFT--DO NOT QUOTE OR CITE**

### **III.F. EVALUATION OF DOSE-RESPONSE RELATIONSHIPS**

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

The dose-response relationships for individual endpoints, as well as the combination of endpoints, must be examined in data interpretation.

Dose-response evaluations should consider the effects that competing risks between different endpoints may have on outcomes observed at different exposure levels. For example, a toxicant may interfere with cell function in such a manner that, at a low dose level, an increase in abnormal sperm morphology is observed. At higher doses cell death may occur, leading to a decrease in sperm counts and a possible decrease in proportion of abnormal sperm.

When data on several species are available, the selection of the data for the dose-response evaluation is based ideally on the response of the species most relevant to humans (e.g., comparable physiologic, pharmacologic, pharmacokinetic, and pharmacodynamic processes), the adequacy of dosing, the appropriateness of the route of administration, and the endpoints selected. However, availability of information on many of those components is usually very limited. For dose-response assessment, no single laboratory animal species can be considered the best in all situations for predicting reproductive toxicologic risk to humans. However, in some cases, such as in the assessment of physiologic parameters related to menstrual disorders, higher nonhuman primates are considered generally similar to the human. In the absence of a clearly most relevant species, data from the most sensitive species (i.e., the species showing a toxic effect at the lowest administered dose) are used, because humans are assumed to be as sensitive generally as the most sensitive

## **DRAFT--DO NOT QUOTE OR CITE**

animal species tested (Nisbet & Karch, 1983; Kimmel, C.A. et al., 1984, 1990; Hemminki & Vineis, 1985; Meistrich, 1986; Working, 1988).

The evaluation of dose-response relationships includes the identification of effective dose levels as well as doses that are associated with low or no increased incidence of adverse effects compared with controls. Much of the focus is on the identification of the critical effect(s) (i.e., the adverse effect occurring at the lowest dose level) and the LOAEL and NOAEL associated with the effect(s). The NOAEL is the highest dose at which there is no significant increase in the frequency of an adverse effect in any manifestation of reproductive toxicity compared with the appropriate control group in a data base having sufficient evidence for use in a risk assessment (see Section III.H. below). The LOAEL is the lowest dose at which there is a significant increase in the frequency of adverse reproductive effects compared with the appropriate control group in a data base having sufficient evidence. An effect, whose incidence is statistically significant at a higher exposure level, may be considered exposure-related if a biologically consistent trend is seen at a lower level in which the observed difference in incidence from the concurrent control group may not reach statistical significance. Although a threshold is assumed for reproductive effects, the existence of a NOAEL in an experimental animal study does not prove or disprove the existence or level of a biologic threshold; it only defines the highest level of exposure under the conditions of the study that is not associated with a significant increase in an adverse effect.

Several limitations in the use of the NOAEL have been described (Gaylor, 1983, 1989; Crump, 1984; Kimmel, C.A. & Gaylor, 1988; Brown & Erdreich, 1989; Kimmel, C.A., 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) Because 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 are 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

## **DRAFT--DO NOT QUOTE OR CITE**

in a study can influence the dose that is chosen for the NOAEL; and 5) Because the NOAEL is defined as a dose that does not produce an observable change 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). The upper confidence limit on developmental risk at the NOAEL has been estimated for several data sets to be 2 to 6% (Crump, 1984; Gaylor, 1989); similar evaluations have not been conducted on data for other reproductive effects.

Because of the limitations associated with the use of the NOAEL (Kimmel, C.A. & Gaylor, 1988; Gaylor, 1989; Kimmel, C.A., 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). Calculation and use of the benchmark dose are described in the Guidelines for Developmental Toxicity Risk Assessment (U.S. Environmental Protection Agency, 1991). The benchmark dose is based on a model-derived estimate of a particular incidence level, such as a 5 or 10% incidence. More specifically, the benchmark dose 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 benchmark dose that is the lower confidence limit on dose for that level of excess response, in this case, the  $LED_{10}$ .

With the benchmark dose approach, an  $LED_{10}$  should 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 may be closer to a true no-effect dose. A level between the  $ED_{01}$  and the  $ED_{10}$  is usually the lowest level of risk that can be estimated adequately for binomial endpoints from standard developmental toxicity studies.

## **DRAFT--DO NOT QUOTE OR CITE**

Various mathematical approaches have been proposed for deriving a benchmark dose in modeling developmental toxicity data (Crump, 1984; Rai & Van Ryzin, 1985; Kimmel, C.A. & Gaylor, 1988; Chen & Kodell, 1989; Faustman et al., 1989; Kodell et al., 1991). A benchmark dose approach has been applied to male reproductive effects of dibromochloropropane (Pease et al., 1991). Such models may be used to calculate the benchmark dose, and choice of the model may not be critical since estimation is within the observed dose range. Because the model is only used to fit the observed data, the assumptions about the existence of a threshold do not affect choice of model. Thus, any model that fits the empirical data well is likely to provide a reasonable estimate of the benchmark dose, although if there is some biologic reason to incorporate particular factors in the model (e.g., intralitter correlation; sex-specific dosing), these should be included to account as much as possible for variability in the data. The Agency is exploring the application of several models to data sets for calculating the benchmark dose, as well as to determine the minimum data set that can be modeled and how to apply this approach 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 or RfC since the Agency has limited experience with this approach.

Generally, in studies that do not evaluate reproductive toxicity, only adult male and nonpregnant females are examined. Therefore, the possibility that pregnant females may be more sensitive to the agent is not tested. In studies in which reproductive toxicity has been evaluated, the NOAEL, LOAEL, or benchmark dose should be identified for both reproductive and other forms of systemic toxicity. The NOAEL, LOAEL, or benchmark dose for systemic toxicity in the reproductive study should be compared with the corresponding values from other adult toxicity data to determine if the pregnant or lactating female may be more sensitive to an agent based on results from that study or compared with results for adult males or nonpregnant females in other toxicity studies.

In addition to identification of the NOAEL, LOAEL, or benchmark dose, the dose-response evaluation defines the range of doses that is effective in producing reproductive

## **DRAFT--DO NOT QUOTE OR CITE**

and other forms of systemic toxicity for a given agent, the route of exposure, 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 data base (discussed in the next section).

For developmental toxic effects, an assumption is made that a single exposure at a critical time in development may produce an adverse developmental effect (U.S. Environmental Protection Agency, 1991). Therefore, the daily dose is usually not adjusted for duration of exposure with developmental toxicity unless appropriate pharmacokinetic data are available. However, for other reproductive effects, daily dose may be adjusted for duration of exposure. The Agency is planning to review these stances to determine the most appropriate approach for the future.

### **III.G. CHARACTERIZATION OF THE HEALTH-RELATED DATA BASE**

This section describes evaluation of the health-related data base on a particular chemical and provides criteria for judging the potential for that chemical to produce reproductive toxicity under the exposure conditions inherent in the data base. This determination provides the basis for judging whether the available data are sufficient to proceed with a quantitative risk assessment. Characterizing the available evidence in this way clarifies the strengths and uncertainties in a particular data base. It does not address the level of concern, nor does it completely address determining relevancy of available data for estimating human risk. Both level of concern and relevancy are discussed as part of the final characterization of risk because they depend on information concerning potential human exposure.

A complex interrelationship exists among study design, statistical analysis, and biologic significance of the data. Thus, substantial scientific judgment, based on experience with reproductive toxicity data and with the principles of study design and statistical analysis, may be required to adequately evaluate the data base. In some cases, a data base

## **DRAFT--DO NOT QUOTE OR CITE**

may contain conflicting data. In these instances, the risk assessor must consider each study's strengths and weaknesses within the context of the overall data base to characterize the evidence for assessing the potential hazard for reproductive toxicity. Scientific judgment is always necessary, and in many cases, interaction with scientists in specific disciplines (e.g., reproductive toxicology, developmental toxicology, epidemiology, genetic toxicology, statistics) is recommended.

A scheme for judging the available evidence on the reproductive toxicity of a particular agent is presented below (Table 5). The scheme contains two broad categories, "Sufficient" and "Insufficient," which are defined in Table 5. Data from all available studies, whether or not indicative of potential concern, are evaluated and used to judge whether available evidence allows a hazard assessment for reproductive toxicity. The primary considerations are the human data, if available, and the experimental animal data. The judgment of whether data are sufficient or insufficient should consider a variety of parameters that contribute to the overall quality of the data, such as the power of the studies (e.g., number of animals and variation in the data), the number and types of endpoints examined, replication of effects, relevance of route and timing of exposure for both human and experimental animal studies, and the appropriateness of the dose selection in experimental animal studies. In addition, pharmacokinetic data and structure-activity considerations, data from other toxicity studies, as well as other factors that may affect the overall decision about the evidence, should be taken into account.

In general, the characterization is based on criteria defined by these Guidelines as the minimum evidence necessary to complete a hazard identification/dose-response evaluation. Establishing the minimum human evidence to do a hazard identification/dose-response evaluation is often difficult because there are often considerable variations in study designs and study group selection. The body of 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, few agents can be classified in this category. Agents that have been tested in laboratory animals according to

## **DRAFT--DO NOT QUOTE OR CITE**

EPA's current two-generation reproductive effects test guidelines (U.S. Environmental Protection Agency, 1982, 1985b), but not limited to such designs (e.g., a continuous breeding study with two generations), generally would be included in the "Sufficient Experimental Animal Evidence/Limited Human Data" category. There are occasions in which more limited data regarding the potential reproductive toxicity of an agent (e.g., a one-generation reproductive effects study, a standard subchronic or chronic toxicity study in which the reproductive organs were well examined) are available. If reproductive toxicity is observed in these limited studies, the data may be used to the extent possible to reach a decision regarding hazard to the reproductive system. In cases in which only such limited data are available, it would be appropriate to adjust the uncertainty factor to reflect the attendant increased uncertainty regarding the use of these data until more definitive data are developed. Identification of the increased uncertainty and justification for the adjustment of the uncertainty factor should be stated clearly.

Because it is more difficult both biologically and statistically to support a finding of no apparent hazard, more data are generally required to support this conclusion than a finding for a potential hazard. For example, to judge that a hazard for reproductive toxicity could exist for a given agent, the minimum evidence could be data from a single appropriate, well-executed study in a single test species that demonstrates an adverse reproductive effect, or suggestive evidence from adequately conducted clinical or epidemiologic studies. As in all situations, it is important that the results be biologically consistent. On the other hand, to judge that an agent is unlikely to pose a hazard for reproductive toxicity, the minimum evidence would include data on an array of endpoints and from more than one study that showed no reproductive effects at doses that were otherwise minimally toxic to the adult animal. In addition, there may be human data from appropriate studies that are supportive of no apparent hazard. In the event that a substantial data base 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 data base may be viewed as meeting the "Sufficient" criteria.

## DRAFT--DO NOT QUOTE OR CITE

TABLE 5

### CATEGORIZATION OF THE HEALTH-RELATED DATA BASE HAZARD IDENTIFICATION/DOSE-RESPONSE EVALUATION

#### SUFFICIENT EVIDENCE

The Sufficient Evidence category includes data that collectively provide enough information to judge whether or not a reproductive 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 test 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 reproductive toxicity exists. Generally, agents that have been tested according to EPA's current two-generation reproductive effects test guidelines (but not limited to such designs) would be included in this category. The minimum evidence necessary to judge that a potential hazard exists would be data demonstrating an adverse reproductive effect in a single appropriate, well-executed study in a single test species. The minimum evidence needed to judge that a potential hazard does not exist would include data on an adequate array of endpoints and from more than one study that showed no adverse reproductive effects at doses that were minimally toxic to the adult animal.

#### INSUFFICIENT EVIDENCE

This category includes situations for which there is less than the minimum sufficient evidence necessary for assessing the potential for reproductive toxicity, such as when no data are available on reproductive toxicity, as well as for data bases from studies in test animals or humans that have a limited study design (e.g., small numbers of animals or human subjects, inappropriate dose selection or exposure information, other uncontrolled factors), data from studies that examined only a limited number of endpoints and reported no adverse reproductive effects, or data bases that were limited to information on structure-activity relationships, short-term tests, pharmacokinetic data, or metabolic precursors.attendant increased uncertainty regarding the use of these data until more



**DRAFT--DO NOT QUOTE OR CITE**

TABLE 5 (Continued)

definitive data are developed. Identification of the increased uncertainty and justification for the adjustment of the uncertainty factor should be stated clearly.

Because it is more difficult both biologically and statistically to support a finding of no apparent hazard, more data are generally required to support this conclusion than a finding for a potential hazard. For example, to judge that a hazard for reproductive toxicity could exist for a given agent, the minimum evidence could be data from a single appropriate, well-executed study in a single test species that demonstrates an adverse reproductive effect, or suggestive evidence from adequately conducted clinical or epidemiologic studies. As in all situations, it is important that the results be biologically consistent. On the other hand, to judge that an agent is unlikely to pose a hazard for reproductive toxicity, the minimum evidence would include data on an array of endpoints and from more than one study that showed no reproductive effects at doses that were otherwise minimally toxic to the adult animal. In addition, there may be human data from appropriate studies that are supportive of no apparent hazard. In the event that a substantial data base 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 data base may be viewed as meeting the "Sufficient" criteria.

## **DRAFT--DO NOT QUOTE OR CITE**

Some important considerations in determining the confidence in the health data base are as follows:

- \* Data of equivalent quality from human exposures are given more weight than data from exposures of test species.
- \* Although a single study of high quality could be sufficient to achieve a relatively high level of confidence, replication increases the confidence that may be placed in such results.
- \* Data are available from one or more in vivo studies of acceptable quality with humans or other mammalian species that are believed to be predictive of human responses.
- \* Data exhibit a dose-response relationship.
- \* Results are statistically significant and biologically plausible.
- \* When multiple studies are available, results are reproducible.
- \* When multiple studies are available, the lines of evidence from independent study types are reinforcing.
- \* Sufficient information is available to reconcile discordant data.
- \* Route, level, duration, and frequency of exposure are appropriate.
- \* An adequate array of endpoints has been examined.
- \* The power and statistical treatment of the studies are appropriate.

Any statistically significant deviation from baseline levels for an in vivo effect warrants closer examination. To determine whether such a deviation constitutes an adverse effect requires an understanding of its role within a complex system and the determination of whether a "true effect" has been observed. Application of the above criteria, combined with guidance presented in Section III.B. can facilitate such determinations.

The greatest confidence for reproductive hazard identification should be placed on significant adverse effects on sexual behavior, fertility or pregnancy outcomes, or other endpoints that are directly related to reproductive function such as menstrual (estrous) cycle normality, sperm evaluations, reproductive histopathology, reproductive organ weights, and

## **DRAFT--DO NOT QUOTE OR CITE**

possibly reproductive endocrinology (see III.B.3.e. for qualifying statement). Agents producing effects on these endpoints can be assigned to the "Sufficient Evidence" category if study quality is adequate.

Less confidence should be placed in results from other measures such as in vitro tests, data from nonmammals or structure-activity relationships, but positive results may trigger followup studies to determine the likelihood and extent to which function might be affected. Results from these types of studies alone, whether or not they demonstrate an effect, should be assigned to the "Insufficient Evidence" category.

The absence of effects with test species on the endpoints that are evaluated routinely (i.e., fertility, histopathology, and organ weights) may constitute sufficient evidence to place a low priority on the potential reproductive toxicity of a chemical. However, in such cases, careful consideration should be given to the sensitivity of these endpoints and to the quality of the data on these endpoints. Consideration should also be given to the possibility of adverse effects that may not be reflected in these routine measures (e.g., germ-cell mutation, alterations in estrous cyclicity or sperm measures such as motility or morphology).

Judging that the health data base indicates a potential reproductive 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., the type and degree of hazard may vary significantly depending on route and timing of exposure). In the final risk characterization, the characterization of the health-related data base should always be presented with information on the dose-response evaluation (e.g., LOAEL, NOAEL, or benchmark dose), exposure route, timing and duration of exposure, and if available, with the human exposure estimate (for further discussion, see Section V).

## **DRAFT--DO NOT QUOTE OR CITE**

### **III.H. DETERMINATION OF THE REFERENCE DOSE OR REFERENCE CONCENTRATION FOR REPRODUCTIVE TOXICITY**

In quantitative risk assessment, the existence of a threshold is usually assumed for noncarcinogenic health effects. The assumption of a threshold suggests that the application of adequate uncertainty factors to a NOAEL, LOAEL, or benchmark dose will result in an exposure level for all humans that is not attended with significant risk. In the absence of a threshold, it is assumed that some finite risk exists at any level of exposure, with risk decreasing as exposure decreases. In the absence of data on the responses at low levels of exposure and associated mechanistic information, a threshold is assumed for reproductive effects. It is plausible that certain biologic processes (e.g., Sertoli cell barrier selectivity, metabolic and repair capabilities of the germ cells) may impede the attainment or maintenance of concentrations of the agent at the target site following exposure to low dose levels that would be associated with adverse effects.

The RfD or RfC is an estimate of a daily exposure to the human population that is assumed to be without appreciable risk of deleterious reproductive effects over a lifetime of exposure. The RfD or RfC is derived by applying uncertainty factors to the NOAEL (or the LOAEL if a NOAEL is not available) or to the benchmark dose. To date, the Agency has applied uncertainty factors only to the NOAEL or LOAEL to derive an RfD or RfC. The Agency is considering the use of the benchmark dose approach as the basis for derivation of the RfD or RfC and will develop guidance as information is acquired and analyzed from ongoing Agency studies. Because of the short duration of most studies of developmental toxicity, a unique value ( $\text{RfD}_{\text{DT}}$  or  $\text{RfC}_{\text{DT}}$ ) is determined for adverse developmental effects. For adverse reproductive effects on endpoints other than those of developmental toxicity, no special designator is attached.

The effect used for determining the NOAEL, LOAEL, or benchmark dose in deriving the RfD or RfC is the most sensitive adverse reproductive endpoint (i.e., the critical effect) from the most appropriate or, in the absence of such information, the most sensitive mammalian species (see Sections II and III.B.1.). Uncertainty factors for

## **DRAFT--DO NOT QUOTE OR CITE**

reproductive and other forms of systemic toxicity applied to the NOAEL generally include factors of 3 or 10 each for interspecies variation and for intraspecies variation. Additional factors may be applied to account for other uncertainties that may exist in the data base. For example, the standard study design for a reproductive toxicity study calls for a low dose that demonstrates a NOAEL, but in some cases, the lowest dose administered may cause significant adverse effects and thus be identified as the LOAEL. In circumstances where only a LOAEL is available, the use of an additional uncertainty factor of 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 NOAEL), and thus affect the size of the uncertainty factor to be applied. Additional areas of uncertainty may be identified and modifying factors used depending on the characterization of the data base (e.g., if the only data available are from a one-generation reproductive effects study; see Section III.G.), data on pharmacokinetics, or other considerations that may alter the level of confidence in the data (U.S. Environmental Protection Agency, 1987). The total size of the uncertainty factor will vary from agent to agent and requires scientific judgment, taking into account interspecies differences, variability within species, the slope of the dose-response curve, the types of reproductive effects observed, the background incidence of the effects, the route of administration, and pharmacokinetic data.

There is no experience with the application of uncertainty factors to the benchmark dose approach for calculating the RfD or RfC, and there are several issues that must be addressed prior to its use for this purpose; for example, which benchmark dose (e.g., LED<sub>01</sub>, LED<sub>05</sub>, LED<sub>10</sub>) should be used for calculating the RfD or RfC, and what are the appropriate uncertainty factors that should be applied to the benchmark dose for deriving the RfD or RfC? That is, should the uncertainty factor applied to an LED<sub>10</sub> be similar to that applied to a LOAEL, or should the uncertainty factor applied to an LED<sub>01</sub> be equal to or less than that applied to a NOAEL? These questions are being addressed in ongoing

## **DRAFT--DO NOT QUOTE OR CITE**

Agency studies on the calculation of the RfD or RfC 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 or most sensitive mammalian species to determine the RfD or RfC. If the NOAEL or LOAEL (or benchmark dose) for other forms of systemic toxicity is lower than that for reproductive toxicity, this should be noted in the risk characterization, and this value compared with data from other studies in which adult animals are exposed. Thus, the reproductive toxicity data should be discussed in the context of other toxicity data.

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 biologic processes or mechanisms (Crump, 1984; Rai & Van Ryzin, 1985; Kimmel, C.A. & Gaylor, 1988; Chen & Kodell, 1989; Faustman et al., 1989; Kodell et al., 1991). These approaches may also be applicable for modeling reproductive toxicity data and 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 biologic threshold for reproductive toxicity, based on known homeostatic, compensatory, or adaptive mechanisms that must be overcome before a toxic endpoint is manifested and on the rationale that cells and organs of the reproductive system and the embryo are known to have some capacity for repair of damage. However, a threshold for a population 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 of adverse effects for some, but not necessarily all individuals within the population.

Efforts are underway to develop models that are more biologically based. These models should provide a more accurate estimation of low-dose risk to humans. The development of biologically based dose-response models in reproductive toxicology have

## **DRAFT--DO NOT QUOTE OR CITE**

been impeded by a number of factors, including limited understanding of the biologic mechanisms underlying reproductive toxicity, intra- and interspecies differences in the types of reproductive events, lack of appropriate pharmacokinetic data, and inadequate information on the influence of other types of systemic toxicity on the dose-response curve.

### **III.I. SUMMARY**

The hazard identification/dose-response evaluation of reproductive toxicity data is incorporated into the final characterization of risk along with information on estimates of human exposure. The analysis depends on and should describe scientific judgments as to the accuracy and sufficiency of the health-related data in experimental animals and humans (if available), the biologic relevance of significant effects, and other considerations important in the interpretation and application of data to humans. Scientific judgment is always necessary, and in many cases, interaction with scientists in specific disciplines (e.g., reproductive toxicology, developmental toxicology, epidemiology, statistics) is recommended.

### **IV. EXPOSURE ASSESSMENT**

To obtain a quantitative estimate of risk for the human population, an estimate of human exposure is required. The Guidelines for Exposure Assessment have been published separately (U.S. Environmental Protection Agency, 1992) and will not be discussed in detail here. Rather, issues important to reproductive toxicity risk assessment are addressed. In general, the exposure assessment describes the magnitude, duration, schedule, and route of human exposure. This information is usually developed from monitoring data, from estimates based on modeling of environmental exposures, and from application of paradigms to exposure data bases. Often quantitative estimates of exposures may not be available (e.g., workplace or environmental measurements). In such instances, employment or residential histories also may be used in characterizing exposure in a qualitative sense. The potential use of biomarkers as indicators of exposure is an area of active interest.

## **DRAFT--DO NOT QUOTE OR CITE**

Studies of occupational populations may provide valuable information on the potential environmental health risks for certain agents. Exposures among environmentally exposed human populations tend to be lower (but of longer duration) than those in studies of occupationally exposed populations and therefore may require more observations to assure sufficient statistical power. Also, reconstruction of exposures is more difficult in an environmental study than in those done in workplace settings where industrial hygiene monitoring may provide more detailed exposure data.

The nature of the exposure may be defined at a particular point or may reflect cumulative exposure. Each approach makes an assumption about the underlying relationship between exposure and outcome. For example, a cumulative exposure measure assumes that total exposure is important, with a greater probability of effect with greater total exposure or body burden. A dichotomous exposure measure (ever exposed versus never exposed) assumes an irreversible effect of exposure. Models that define exposure only at a specific time may assume that only the present exposure is important (Selevan & Lemasters, 1987). The appropriate exposure model depends on the biologic processes affected. Thus, a cumulative or dichotomous exposure model may be appropriate if injury occurs in cells that cannot be replaced or repaired (e.g., Sertoli cells, oocytes); on the other hand, a concurrent exposure model may be appropriate for cells that are being generated continually (e.g., spermatids).

There are a number of unique considerations regarding the exposure assessment for reproductive toxicity. Exposure at different stages of male and female development can result in different outcomes. Such age-dependent variation has been well documented in both experimental animal and human studies. Prenatal and neonatal treatment can irreversibly alter reproductive function in a manner that may not be predicted from adult-only exposure. Moreover, chemicals that alter sexual differentiation in rodents during these periods may have similar effects in humans, because the mechanisms underlying these developmental processes appear to be similar in all mammalian species (Gray, 1991).



## **DRAFT--DO NOT QUOTE OR CITE**

The susceptibility of elderly males and females to chemical insult has not been well studied. Although procreative competence may not be a major health concern with elderly individuals, other biologic functions maintained by the gonads (e.g., hormone production) are of significance (Walker, 1986). An exposure assessment should characterize the likelihood of exposure of these different subgroups (embryo or fetus, neonate, juvenile, young adult, older adult) and the risk assessment should factor in the susceptibility of different age groups to the extent possible.

The relationship between time or duration of exposure and observation of male reproductive effects has particular significance for short-term exposures. Spermatogenesis is a temporally synchronized process. In humans, germ cells that were spermatozoa, spermatids, spermatocytes, or spermatogonia at the time of an acute exposure require 1 to 2, 3 to 5, 5 to 8, or 8 to 12 weeks, respectively, to appear in an ejaculate. That timing may vary somewhat depending on degree of sexual activity. It is possible that an end point may be examined too early or too late to detect an effect if only a particular cell type was affected during a relatively brief exposure to an agent. The absence of an effect when observations were made too late suggests either a reversible effect or no effect. However, an effect that is reversible at lower exposures might become irreversible with higher or longer exposures or exposure of a more susceptible individual. Thus, the failure to detect transient effects because of improper timing of observations may be important. If information is available on the type of effect expected from a class of agents, it may be possible to evaluate whether the timing of endpoint measurement relative to the timing of the short-term exposure is appropriate. Some information on the appropriateness of the protocol can be obtained if test animal data are available to identify the most sensitive cell type or the putative mechanism of action for a given agent.

Compared with acute exposures, the link between exposure and outcome may be more apparent with relatively constant subchronic or longer exposures that are of sufficient duration to cover all phases of spermatogenesis (Russell et al., 1990). Assessments may be made at any time after this point as long as exposure remains constant. Time required for

## **DRAFT--DO NOT QUOTE OR CITE**

the agent or metabolite to attain steady-state levels should also be considered. Again, application of models of exposure (e.g., dichotomous, concurrent, or cumulative) depends on the suspected target or mechanism of action.

The reversibility of an adverse effect on the male reproductive system can be affected by the degree and duration of exposure. The degree of stem cell loss is inversely related to the degree of restoration of sperm production, because repopulation of the germinal epithelium is dependent on the stem cells (Foote & Berndtson, 1992). For agents that bioaccumulate, increasing duration of exposure may also increase the extent of damage to the stem cell population. Damage to other spermatogenic cell types reduces the number of sperm produced, but recovery should occur when the toxic agent is removed. Less is known about the effects of toxicity on the Sertoli cells. Temporary impairment of Sertoli cell function may produce long-lasting effects on spermatogenesis. Destruction of Sertoli cells or interference with their proliferation before puberty are irreversible effects because replication ceases after puberty. Sertoli cells are essential for support of the spermatogenic process and loss of those cells results in a permanent reduction of spermatogenic capability (Foster, 1992).

When recovery is possible, the duration of the recovery period is determined by the time for regeneration (for stem cells) and repopulation of the affected spermatogenic cell types and appearance of those cells as sperm in the ejaculate. The time required for these events to occur varies with the species, the pharmacokinetic properties of the agent, the extent to which the stem cell population has been destroyed, and the degree of sublethal toxicity inflicted on the stem cells or Sertoli cells. When the stem cell population has been partially destroyed, humans require longer than mice to reach the same degree of recovery (Meistrich & Samuels, 1985).

Unique considerations in the assessment of female reproductive toxicity include the duration and period of exposure as related to the development or stage of reproductive life (e.g., prenatal, prepubescent, reproductive, or postmenopausal) or considerations of different physiologic states (e.g., nonpregnant, pregnant, lactating).

## **DRAFT--DO NOT QUOTE OR CITE**

For infertility, a cumulative exposure measure assumes destruction of increasing numbers of primary oocytes with greater lifetime exposure or increasing body burden. However, humans may be exposed to varying levels of toxicants within the study period. Exposures during certain critical points in the reproductive process may affect the outcomes observed in humans (Lemasters & Selevan, 1984). In test species, perinatal exposure to androgens or estrogens such as zearalenone, methoxychlor, and DDT (Bulger & Kupfer, 1985; Gray et al., 1985), have been shown to advance puberty and masculinize females. Similar effects have been reported in humans (both sexes) exposed neonatally to synthetic estrogens or progestins (Schardein, 1985; Steinberger & Lloyd, 1985). Studies using test species have also shown that exposure to some environmental agents such as ionizing radiation (Dobson & Felton, 1983) and glycol ethers (Heindel et al., 1989) can deplete the pool of primordial follicles and thus significantly shorten the female's reproductive lifespan. Furthermore, exposure to compounds at different stages of the ovarian cycle can disrupt or delay follicular recruitment and development (Armstrong, 1986), ovulation (Everett & Sawyer, 1950; Terranova, 1980), and ovum transport (Cummings & Perreault, 1990). Compounds that delay ovulation can lead to significant alterations in egg viability (Peluso et al., 1979), fertilizability of the egg (Fugo & Butcher, 1966; Butcher & Fugo, 1967; Butcher et al., 1975) and a reduction in litter size (Fugo & Butcher, 1966). After ovulation, single exposures to compounds such as carbendazim also alter the fertilizability of the ova (Perreault et al., 1992). Thus, knowledge of when acute exposures are administered relative to the female's lifespan and reproductive cycle can provide insight into how an agent disrupts reproductive function.

DES is a classic example of an agent causing different effects on the reproductive system in the developing organism compared with those in adults (McLachlan, 1980). DES interferes with the development of the Mullerian and Wolffian duct systems and thereby causes irreversible structural and functional damage to the developing reproductive system. In adults, the reproductive effects that are caused by the estrogenic activity of DES do not necessarily result in permanent damage.

## **DRAFT--DO NOT QUOTE OR CITE**

Unique considerations for outcomes of pregnancy are duration and period of exposure as related to stage of development (i.e., critical periods) and the possibility that even a single exposure may be sufficient to produce adverse developmental effects. 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 long enough to produce an increasing body burden over time. For these reasons, it is assumed that, in most cases, a single exposure at the critical time in development is sufficient to produce an adverse developmental effect. Therefore, the human exposure estimates used to calculate the MOE for an adverse developmental effect or to compare to the RfD or RfC are usually based on a single daily dose that is not adjusted for duration or pattern (e.g., continuous or intermittent) of exposure. For example, it would be inappropriate 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-hour inhalation exposure to account for a 24-hour exposure scenario) unless pharmacokinetic data were available to indicate an accumulation with continuous exposure. In the case of intermittent exposures, examination of the peak exposures as well as the average exposure over the time of exposure would be important.

It should be recognized that, based on the definitions used in these Guidelines for reproductive toxicity, almost any segment of the human population may be at risk for a reproductive effect. Although the reproductive effects of exposures may be manifested while the exposure is occurring (e.g., menstrual disorder, decreased sperm count, spontaneous abortion) some effects may not be detectable until later in life (e.g., premature reproductive senescence due to oocyte depletion), long after exposure has ceased.

## **V. RISK CHARACTERIZATION**

### **V.A. OVERVIEW**

A risk characterization is a necessary part of any Agency report on risk whether the report is a preliminary one prepared to support allocation of resources toward further study

## **DRAFT--DO NOT QUOTE OR CITE**

or a comprehensive one prepared to support regulatory decisions. In this final step of a risk assessment, the risk characterization involves integration of 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 results of the risk assessment to a risk manager. A risk manager can then use the risk assessment, along with other risk management elements, to make public health decisions. The information should also assist others outside the Agency in understanding the scientific basis for regulatory decisions.

Risk characterization is the culmination of the risk assessment process and is intended to summarize key aspects of the following components of the risk assessment:

1. The nature, reliability and consistency of the data used.
2. The reasons for selection of the key study(ies) and the critical effect(s) and their relevance to human outcomes.
3. The qualitative and quantitative descriptors of the results of the risk assessment.
4. The limitations of the available data, the assumptions used to bridge knowledge gaps in working with those data, and implications of using alternative assumptions.
5. Discussion of the strengths and weaknesses of the risk assessment and the level of scientific confidence in the assessment.
6. Identification of the areas of uncertainty, additional data/research needs to improve confidence in the risk assessment, and the potential impacts of the new research.

The risk characterization should be limited to the most significant and relevant data, conclusions and uncertainties. When special circumstances exist that preclude full assessment, those circumstances should be explained and the related limitations identified.

The following sections describe these aspects of the risk characterization in more detail, but do not attempt to provide a full discussion of risk characterization. Rather, these

## **DRAFT--DO NOT QUOTE OR CITE**

Guidelines point out issues that are important to risk characterization for reproductive toxicity. Comprehensive general guidance for risk characterization is provided by Habicht (1992).

### **V.B. INTEGRATION OF HAZARD IDENTIFICATION/DOSE-RESPONSE AND EXPOSURE ASSESSMENTS**

In developing the hazard identification/dose-response and exposure assessment portions of the risk assessment, risk assessors must make judgments concerning human relevance of the toxicity data, including the appropriateness of the various test animal models for which data are available, the route, timing and duration of exposure relative to the expected human exposure. These judgments should be summarized at each stage of the risk assessment process. When data are not available to make such judgments, as is often the case, the background information and assumptions discussed in the Overview (Section I) provide default positions. The rationale behind the use of a default position should be clearly stated. In integrating the parts of the assessment, risk assessors 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 and quality 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 these analyses. The hazard identification/dose response and exposure assessment sections should each have their own characterization, and these characterizations should be summarized and integrated into the overall risk characterization. Interpretation of data should be explained, and risk managers should be given a clear picture of consensus or lack of consensus that exists about significant aspects of the assessment. When more than one interpretation is supported by the data, the alternative plausible approaches should be presented along with the strengths, weaknesses, and impacts of those options. If one interpretation or option has been selected

## **DRAFT--DO NOT QUOTE OR CITE**

over another, the rationale should be given; if not, then both should be presented as plausible alternatives.

The risk characterization should not only examine the judgements, 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 chemical 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 and relevant to use in a risk assessment.

- \* A discussion of the dose-response patterns for the critical effect(s) and their relationships to the occurrence of other toxicity, data such as the shapes and slopes of the dose-response curves for the various other endpoints, the rationale behind the determination of the NOAEL, LOAEL, and benchmark dose, and the assumptions underlying the estimation of the RfD or RfC.

- \* Descriptions of the estimates of the range of human exposure (e.g., central tendency, high end), the route, duration, and pattern of the exposure, relevant pharmacokinetics, and the size and characteristics of the various populations that might be exposed.

- \* The risk characterization of an agent being assessed for reproductive toxicity 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 an adverse reproductive effect, whether in the male, the female (nonpregnant or pregnant), or the developing organism, and should be considered in relation to other forms of toxicity. The relevance of this indicator to human reproductive outcomes should be described.

If data to be used in a risk characterization 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

## **DRAFT--DO NOT QUOTE OR CITE**

certain assumptions concerning the amount of absorption likely or the applicability of the data from one route to another (U.S. Environmental Protection Agency, 1985a, 1986b). Discussion of some of these issues may be found in the Proceedings of the Workshop on Acceptability and Interpretation of Dermal Developmental Toxicity Studies (Kimmel, C.A. & Francis, 1990). and Principles of Route-to-Route Extrapolation for Risk Assessment (Gerrity et al., 1990).

The level of confidence in the hazard identification/dose-response evaluation should be stated to the extent possible, including placement of the agent into the appropriate category regarding the sufficiency of the health-related data (see Section III.G.). A comprehensive risk assessment ideally includes information on a variety of endpoints that provide insight into the full spectrum of potential reproductive responses. A profile that integrates both human and test species data and incorporates both sensitive endpoints (e.g., properly performed and fully evaluated histopathology) and functional correlates (e.g., fertility) allows more confidence in a risk assessment for a given agent.

Descriptions of the nature of potential human exposures are important for prediction of specific outcomes and the likelihood of persistence or reversibility of the effect in different exposure situations with different subpopulations (U.S. Environmental Protection Agency, 1992). Where possible, several descriptors of exposure such as the nature and range of populations and their various exposure conditions, central tendencies, and high end exposure estimates should be presented. Even with similar exposure patterns and levels, different subpopulations may react differently. For example, the consequences of exposure to developing individuals versus adults can differ markedly, including whether the effects are permanent or transient. Other considerations relative to human exposures might include potential for exposures to other agents, concurrent disease, and nutritional status, and the possible consequences.



## **DRAFT--DO NOT QUOTE OR CITE**

### **V.C. DESCRIPTORS OF REPRODUCTIVE RISK**

There are a number of ways to describe risk. Some ways that are relevant to describing reproductive risk are as follows.

#### **V.C.1. Estimation of the Number/Proportion of Individuals Exposed to Levels Above the RfD or RfC**

The RfD or RfC is assumed to be a level below which no significant risk occurs. Therefore, information from the exposure assessment on the populations below the RfD or RfC ("not likely to be at risk") and above the RfD or RfC ("may be at risk") may be useful information for risk managers. Estimating the number of persons potentially removed from the "at risk" category after a contemplated action is taken may be particularly useful to a risk manager considering possible actions to ameliorate risk for a population.

#### **V.C.2. Presenting Situation-Specific Exposure Scenarios**

Presenting situation-specific scenarios for important exposure situations and subpopulations in the form of "what if?" questions may be particularly useful to give perspective to risk managers on possible future events. The question being asked in these cases is, for any given exposure level, what would be the resulting number or proportion of individuals that may be exposed to levels above that value?

#### **V.C.3. Margin of Exposure**

In the risk characterization, dose-response information and the human exposure estimates may be combined either by comparing the RfD or RfC 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. Environmental Protection Agency, 1985a). If a NOAEL is not available, a LOAEL may be used in the calculation of the MOE, but consideration for the acceptability would be different than when a NOAEL is used.

## **DRAFT--DO NOT QUOTE OR CITE**

Alternatively, a benchmark dose may be compared with the estimated human exposure level to obtain an MOE. Considerations for the acceptability of the MOE are similar to those for the selection of uncertainty factors applied to the NOAEL, LOAEL, or the benchmark dose for the derivation of an RfD. The MOE is presented along with the characterization of the data base, 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 or RfC comparison with the human exposure estimate and the calculation of the MOE are conceptually similar, but may be used in different regulatory situations.

The choice of approach is dependent on several factors, including the statute involved, the situation being addressed, the data base used, and the needs of the decision maker. The RfD, RfC or MOE are considered along with other risk assessment and risk management issues in making risk management decisions, but the scientific issues that should be taken into account in establishing them have been addressed here.

### **V.C.4. Risk Characterization for Highly Exposed Individuals**

This measure and the next are examples of specific scenarios. In some situations, it may be appropriate to combine them.

The purpose of this measure is to describe the upper end of the exposure distribution, allowing risk managers to evaluate whether certain individuals are at disproportionately high or unacceptably high risk. The objective is to look at the upper end of the exposure distribution to derive a realistic estimate of relatively highly exposed individual(s). The "high end" of the risk distribution has been defined (Habicht, 1992) as above the 90th percentile of the actual (either measured or estimated) distribution. Whenever possible, it is important to express the number or proportion of individuals who comprise the selected highly exposed group and, if data are available, 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

## **DRAFT--DO NOT QUOTE OR CITE**

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.

### **V.C.5. Risk Characterization for Highly Sensitive or Susceptible Individuals**

The purpose of this measure is to quantify exposure of identified sensitive or susceptible populations to the agent of concern. Sensitive or susceptible individuals are those within the exposed population at increased risk of expressing the adverse effect. Examples might be lactating women, women with reduced oocyte numbers, men with "borderline" sperm counts, or infants.

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

### **V.D. SUMMARY AND RESEARCH NEEDS**

These Guidelines summarize the procedures that the U.S. Environmental Protection Agency will follow in evaluating the potential for agents to cause reproductive toxicity. They discuss the assumptions that must be made in risk assessment for reproductive toxicity because of gaps in our knowledge about underlying biologic processes and how these compare across species. Research to improve the risk assessment process is needed. Further studies that 1) more completely characterize and define female and male reproductive endpoints, 2) evaluate the interrelationships among endpoints, 3) examine quantitative extrapolation between endpoints (e.g., sperm count) and function (e.g., fertility), 4) provide a better understanding of the relationships between reproductive toxicity and other forms of toxicity, 5) explore pharmacokinetic disposition of the target and, 6) examine mechanistic phenomena related to pharmacokinetic disposition, will aid in the interpretation of data and interspecies extrapolation. These types of studies, along with further evaluation

**DRAFT--DO NOT QUOTE OR CITE**

of a threshold for susceptible populations, should provide methods to more precisely assess risk.

## DRAFT--DO NOT QUOTE OR CITE

### VI. REFERENCES

- Aafjes, J.H., Vels, J.M., Schenck, E. (1980) Fertility of rats with artificial oligozoospermia. *J. Reprod. Fertil.* 58:345-351.
- Adler, N.T. & Toner, J.P. (1986) The effect of copulatory behavior on sperm transport and fertility in rats. In: Komisaruk, B.R., Siegel, H.I., Chang, M.F., Feder, H.H. *Reproduction: Behavioral and Neuroendocrine Perspective*. Ann. NY Acad. Sci. , pp. 21-32.
- Amann, R.P. (1981) A critical review of methods for evaluation of spermatogenesis from seminal characteristics. *J. Androl.* 2:37-58.
- Armstrong, D.L. (1986) Environmental stress and ovarian function. *Biol. Reprod.* 34:29-39.
- Atterwill, C.K. & Flack, J.D. (1992) *Endocrine Toxicology*. Cambridge University Press, Cambridge.
- Axelson, O. (1985) Epidemiologic methods in the study of spontaneous abortions: Sources of data, methods, and sources of error. In: Hemminki, K., Sorsa, M., Vaino, H. *Occupational Hazards and Reproduction*. Hemisphere, Washington. pp. 231-236.
- Baird, D.D. & Wilcox, A.J. (1985) Cigarette smoking associated with delayed conception. *J. Am. Med. Assoc.* 253:2979-2983.
- Baird, D.D., Wilcox, A.J., Weinberg, C.R. (1986) Using time to pregnancy to study environmental exposures. *Am. J. Epidemiol.* 124:470-480.
- Barlow, S.M. & Sullivan, F.M. (1982) *Reproductive Hazards of Industrial Chemicals*. Academic Press, London.
- Barsotti, D.A., Abrahamson, L.J., Allen, J.R. (1979) Hormonal alterations in female rhesus monkeys fed a diet containing 2,3,7,8-TCDD. *Bull. Environ. Contam. Toxicol.* 21:463-469.
- Beach, F.A. (1979) Animal models for human sexuality. In: Ciba Foundation Symposium No. 62, *Sex, Hormones and Behavior*. Elsevier-North Holland, London. pp. 113-143.
- Berndtson, W.E. (1977) Methods for quantifying mammalian spermatogenesis: A review. *J. Anim. Sci.* 44:818-833.
- Bernstein, M.E. (1984) Agents affecting the male reproductive system: Effects of structure on activity. *Drug Metab. Rev.* 15:941-996.

## **DRAFT--DO NOT QUOTE OR CITE**

Biava, C.G., Smuckler, E.A., Whorton, D. (1978) The testicular morphology of individuals exposed to dibromochloropropane. *Exp. Molec. Pathol.* 29:448-458.

Blazak, W.F., Ernst, T.L., Stewart, B.E. (1985) Potential indicators of reproductive toxicity, testicular sperm production and epididymal sperm number, transit time and motility in Fischer 344 rats. *Fund. Appl. Toxicol.* 5:1097-1103.

Blazak, W.F., Treinen, K.A., Juniewicz, P.E. (1993) Application of testicular sperm head counts in the assessment of male reproductive toxicity. In: Chapin, R.E. & Heindel, J.J. *Methods in Toxicology: Male Reproductive Toxicology*. Academic Press, San Diego. pp. 86-94.

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

Boyers, S.P., Davis, R.O., Katz, D.F. (1989) Automated semen analysis. *Curr. Prob. Obstet. Gynecol. Fertil.* 12:173-200.

Brawer, J.R. & Finch, C.E. (1983) Normal and experimentally altered aging processes in the rodent hypothalamus and pituitary. In: Walker, R.F. & Cooper, R.L. *Experimental and Clinical Interventions in Aging*. Marcel Dekker, New York. pp. 45-65.

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

Brown-Grant, K., Davidson, J.M., Grieg, F. (1973) Induced ovulation in albino rats exposed to constant light. *J. Endocrinol.* 57:7-22.

Bulger, W.H. & Kupfer, D. (1985) Estrogenic activity of pesticides and other xenobiotics on the uterus and male reproductive tract. In: Thomas, J.A., Korach, K.S., McLachlan, J.A. *Endocrine Toxicology*. Raven Press, New York. pp. 1-33.

Burch, T.K., Macisco, J.J., Parker, M.P. (1967) Some methodologic problems in the analysis of menstrual data. *Int. J. Fertil.* 12:67-76.

Burger, E.J., Tardiff, R.G., Scialli, A.R., Zenick, H. (1989) *Sperm Measures and Reproductive Success*. Alan R. Liss, New York.

Butcher, R.L. & Fugo, N.W. (1967) Overripeness and the mammalian ova. II. Delayed ovulation and chromosome anomalies. *Fertil. Steril.* 18:297-302.

## **DRAFT--DO NOT QUOTE OR CITE**

Butcher, R.L., Blue, J.D., Fugo, N.W. (1969) Overripeness and the mammalian ova. III. Fetal development at midgestation and at term. *Fertil. Steril.* 20:223-231.

Butcher, R.L., Collins, W.E., Fugo, N.W. (1975) Altered secretion of gonadotropins and steroids resulting from delayed ovulation in the rat. *Endocrinol.* 96:576-586.

Carlsen, E., Giwercman, A., Keiding, N., Skakkebaek, N.E. (1992) Evidence for decreasing quality of semen during past 50 years. *Br. Med. J.* 305:609-613.

Cassidy, S.L., Dix, K.M., Jenkins, T. (1983) Evaluation of a testicular sperm head counting technique using rats exposed to dimethoxyethyl phthalate (DMEP), glycerol alpha-monochlorohydrin (GMCH), epichlorohydrin (ECH), formaldehyde (FA), or methyl methanesulphonate (MMS). *Arch. Toxicol.* 53:71-78.

Chapin, R.E. (1988) Morphologic evaluation of seminiferous epithelium of the testis. In: Lamb, J.C. & Foster, P.M.D. *Physiology and Toxicology of Male Reproduction*. Academic Press, New York. pp. 155-177.

Chapin, R.E. & Heindel, J.J. (1993) *Methods in Toxicology: Male Reproductive Toxicology*. Academic Press, San Diego.

Chapin, R.E., Gulati, D.K., Barnes, L.H., Teague, J.L. (1993a) The effects of feed restriction on reproductive function in Sprague-Dawley rats. *Fund. Appl. Toxicol.* 20:23-29.

Chapin, R.E., Gulati, D.K., Fail, P.A., Hope, E., Russell, S.R., Heindel, J.J., George, J.D., Grizzle, T.B., Teague, J.L. (1993b) The effects of feed restriction on reproductive function in Swiss CD-1 mice. *Fund. Appl. Toxicol.* 20:15-22.

Chapman, R.M. (1983) Gonadal injury resulting from chemotherapy. In: Mattison, D.R. *Reproductive Toxicology*. Alan R. Liss, New York. pp. 149-161.

Chen, J.J. & Kodell, R.L. (1989) Quantitative risk assessment for teratological effects. *J. Am. Stat. Assoc.* 84:966-971.

Colborn, T., vom Saal, F.S., Soto, A.M. (1993) Developmental effects of endocrine-disrupting chemicals in wildlife and humans. *J. Nat. Inst. Environ. Health Sci.* 101:378-384.

Collins, T.F.X. (1978) Multigeneration reproduction studies. In: Wilson, J.G. & Fraser, F.C. *Handbook of Teratology*. Plenum Press, New York. pp. 191-214.

Cooper, R.L. & Walker, R.F. (1979) Potential therapeutic consequences of age-dependent changes in brain physiology. *Interdis. Topics Gerontol.* 15:54-76.

## **DRAFT--DO NOT QUOTE OR CITE**

Cooper, R.L., Conn, P.M., Walker, R.F. (1980) Characterization of the LH surge in middle-aged female rats. *Biol. Reprod.* 23:611-615.

Cooper, R.L., Chadwick, R.W., Rehnberg, G.L., Goldman, J.M., Booth, K.C., Hein, J.F., McElroy, W.K. (1989) Effect of lindane on hormonal control of reproductive function in the female rat. *Toxicol. Appl. Pharmacol.* 99:384-394.

Cooper, R.L., Goldman, J.M., Vandenberg, J.G. (1993) Monitoring of the estrous cycle in the laboratory rodent by vaginal lavage. In: Heindel, J.J. & Chapin, R.E. *Methods in Toxicology: Female Reproductive Toxicology*. Academic Press, San Diego. pp. 45-56.

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

Csapo, A.I. & Pulkkinen, M. (1978) Indispensability of the human corpus luteum in the maintenance of early pregnancy: luteectomy evidence. *Obstet. Gynecol. Surv.* 33:69.

Cummings, A.M. & Gray, L.E. (1987) Methoxychlor affects the decidual cell response of the uterus but not other progestational parameters in female rats. *Toxicol. Appl. Pharmacol.* 90:330-336.

Cummings, A.M. & Perreault, S.D. (1990) Methoxychlor accelerates embryo transport through the rat reproductive tract. *Toxicol. Appl. Pharmacol.* 102:110-116.

Davis, D.L., Friedler, G., Mattison, D., Morris, R. (1992) Male-mediated teratogenesis and other reproductive effects: Biologic and epidemiologic findings and a plea for clinical research. *Reprod. Toxicol.* 6:289-292.

Deane, M., Swan, S.H., Harris, J.A., Epstein, D.M., Neutra, R.R. (1989) Adverse pregnancy outcomes in relation to water contamination, Santa Clara County, California, 1981-1983. *Am. J. Epidemiol.* 129:894-904.

de Boer, P., van der Hoeven, F.A., Chardon, J.A.P. (1976) The production, morphology, karyotypes and transport of spermatozoa from tertiary trisomic mice and the consequences for egg fertilization. *J. Reprod. Fertil.* 48:249-256.

Dixon, R.L. & Hall, J.L. (1984) Reproductive toxicology. In: Hayes, A.W. *Principles and Methods of Toxicology*. Raven Press, New York. pp. 107-140.

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, Seattle, WA: June, 1978.



## **DRAFT--DO NOT QUOTE OR CITE**

Dobson, R.L. & Felton, J.S. (1983) Female germ cell loss from radiation and chemical exposure. *Am. J. Ind. Med.* 4:175-190.

Drouva, S.V., Laplante, E., Kordon, C. (1982) Alpha 1-adrenergic receptor involvement in the LH surge in ovariectomized estrogen-primed rats. *Eur. J. Pharmacol.* 81:341-344.

Egnatz, D.G., Ott, M.G., Townsend, J.C., Olson, R.D., Johns, D.B. (1980) DBCP and testicular effects in chemical workers; an epidemiological survey in Midland. *J. Occup. Med.* 22:727-732.

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

Everett, J.W. & Sawyer, C.H. (1950) A 24-hour periodicity in the "LH-release apparatus" of female rats disclosed by barbiturate sedation. *Endocrinol.* 47:198-218.

Everson, R.B., Sandler, D.P., Wilcox, A.J., Schreinemachers, D., Shore, D.L., Weinberg, C. (1986) Effect of passive exposure to smoking on age at natural menopause. *Br. Med. J.* 293:792.

Fabia, J. & Thuy, T.D. (1974) Occupation of father at time of children dying of malignant disease. *Br. J. Prev. Soc. Med.* 28:98-100.

Faustman, E.M., Wellington, D.G., Smith, W.P., Kimmel, C.A. (1989) Characterization of a developmental toxicity dose-response model. *Environ. Health* 79:229-241.

Fawcett, D.W. (1986) Bloom and Fawcett: A Textbook of Histology. W. B. Saunders, Philadelphia, PA.

Filler, R. (1993) Methods for evaluation of rat epididymal sperm morphology. In: Chapin, R.E. & Heindel, J.J. *Methods in Toxicology: Male Reproductive Toxicology*. Academic Press, San Diego. pp. 334-343.

Finch, C.E., Felicio, L.S., Mobbs, C.V. (1984) Ovarian and steroidal influences on neuroendocrine aging processes in female rodents. *Endocrinol. Rev.* 5:467-497.

Fink, G. (1988) Gonadotropin secretion and its control. In: Knobil, E. & Neill, J.D. *The Physiology of Reproduction*. Raven Press, New York. pp. 1349-1377.

Foote, R.H. & Berndtson, W.E. (1992) The Germinal Cells. In: Scialli, A.R. & Clegg, E.D. *Reversibility in Testicular Toxicity Assessment*. CRC Press, Boca Raton. pp. 1-55.

## **DRAFT--DO NOT QUOTE OR CITE**

- Foote, R.H., Schermerhorn, E.C., Simkin, M.E. (1986) Measurement of semen quality, fertility, and reproductive hormones to assess dibromochloropropane (DBCP) effects in live rabbits. *Fund. Appl. Toxicol.* 6:628-637.
- Forsberg, J.G. (1981) Permanent changes induced by DES at critical stages in human and model systems. *Biol. Res. Pregnancy* 2:168-175.
- Foster, P.M.D. (1992) The Sertoli cell. In: Scialli, A.R. & Clegg, E.D. *Reversibility in Testicular Toxicity Assessment*. CRC Press, Boca Raton. pp. 57-86.
- Francis, E.Z. & Kimmel, G.L. (1988) Proceedings of the workshop on one- vs two-generation reproductive effects studies. *J. Amer. Coll. Toxicol.* 7:911-925.
- Franken, D.R., Burkman, L.J., Coddington, C.C., Oehninger, S., Hodgen, G.D. (1990) Human hemizona attachment assay. In: Acosta, A.A., Swanson, R.J., Ackerman, S.B., Kruger, T.F., VanZyl, J.A., Menkveld, R. *Human Spermatozoa in Assisted Reproduction*. Williams & Wilkins, Baltimore. pp. 355-371.
- Fugo, N.W. & Butcher, R.L. (1966) Overripeness and the mammalian ova. I. Overripeness and early embryonic development. *Fertil. Steril.* 17:804-814.
- Gaffey, W.R. (1976) A critique of the standard mortality ratio. *J. Occup. Med.* 18:157-160.
- Galbraith, W.M., Voytek, P., Ryon, M.S. (1983) Assessment of risks to human reproduction and development of the human conceptus from exposure to environmental substances. In: Christian, M.S., Galbraith, W.M., Voytek, P., Mehlman, M.A. *Advances in Modern Environmental Toxicology*. Princeton Scientific Publ., Princeton. pp. 41-153.
- 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* 79:243-246.
- Gelletti, F. & Kloppe, A. (1964) The effect of progesterone on the quantity and distribution of body fat in the female rat. *Acta Endocrinol.* 46:379-386.
- Generoso, W.M., Rutledge, J.C., Cain, K.T., Hughes, L.A., Braden, P.W. (1987) Exposure of female mice to ethylene oxide within hours after mating leads to fetal malformation and death. *Mutat. Res.* 176:269-274.

## **DRAFT--DO NOT QUOTE OR CITE**

Gerrity, T.R., Henry, C.J., Bronaugh, R., et al. (1990) Summary report of the workshops on principles of route-to-route extrapolation for risk assessment. In: Gerrity, T.R. & Henry, C.J. Principles of Route-To-Route Extrapolation for Risk Assessment. Elsevier Science Publ. Co., New York. pp. 1-12.

Giwercman, A., Carlsen, E., Keiding, N., Skakkebaek, N.E. (1993) Evidence for increasing incidence of abnormalities of the human testis: A review. *Envir. Health Perspect.* 101:65-71.

Goldman, J.M., Cooper, R.L., Laws, S.C., Rehnberg, G.L., Edwards, T.L., McElroy, W.K., Hein, J.F. (1990) Chlordimeform-induced alterations in endocrine regulation within the male rat reproductive system. *Toxicol. Appl. Pharmacol.* 104:25-35.

Goldman, J.M., Cooper, R.L., Edwards, T.L., Rehnberg, G.L., McElroy, W.K., Hein, J.F. (1991) Suppression of the luteinizing hormone surge by chlordimeform in ovariectomized, steroid-primed female rats. *Pharmacol. Toxicol.* 68:131-136.

Gorski, R.A. (1979) The neuroendocrinology of reproduction: An overview. *Biol. Reprod.* 20:111-127.

Gray, L.E. (1991) Delayed effects on reproduction following exposure to toxic chemicals during critical periods of development. In: Cooper, R.L., Goldman, J.M., Harbin, T.J. Aging and Environmental Toxicology: Biological and Behavioral Perspectives. Johns Hopkins University Press, Baltimore. pp. 183-210.

Gray, L.E., Ferrell, J.M., Ostby, J.S. (1985) Alteration of behavioral sex differentiation by exposure to estrogenic compounds during a critical neonatal period: Effects of zearalenone, methoxychlor, and estradiol in hamster. *Toxicol. Appl. Pharmacol.* 80:127-136.

Gray, L.E., Ostby, J., Sigmon, R., Ferrell, J., Linder, R., Cooper, R., Goldman, J., Laskey, J. (1988) The development of a protocol to assess reproductive effects of toxicants in the rat. *Reprod. Toxicol.* 2:281-287.

Gray, L.E., Ostby, J., Ferrell, J., Rehnberg, G., Linder, R., Cooper, R., Goldman, J., Slott, V., Laskey, J. (1989) A dose-response analysis of methoxychlor-induced alterations of reproductive development and function in the rat. *Fund. Appl. Toxicol.* 12:92-108.

Gray, L.E., Ostby, J., Linder, R., Goldman, J., Rehnberg, G., Cooper, R. (1990) Carbendazim-induced alterations of reproductive development and function in the rat and hamster. *Fund. Appl. Toxicol.* 15:281-297.

## **DRAFT--DO NOT QUOTE OR CITE**

Green, S., Auletta, A., Fabricant, R., Kapp, M., Sheu, C., Springer, J., Whitfield, B. (1985) Current status of bioassays in genetic toxicology: The dominant lethal test. *Mutat. Res.* 154:49-67.

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

Gulati, D.K., Hope, E., Teague, J., Chapin, R.E. (1991) Reproductive toxicity assessment by continuous breeding in Sprague-Dawley rats: A comparison of two study designs. *Fund. Appl. Toxicol.* 17:270-279.

Gustafsson, J.-A., Mode, A., Norstedt, G., Hokfelt, T., Sonnenschein, C., Eneroth, P., Skett, P. (1980) The hypothalamo-pituitary-liver axis: A new hormonal system in control of hepatic steroid and drug metabolism. *Biochem. Act. Hormones* 14:47-89.

Habicht, F.H. (1992). Guidance on Risk Characterization for Risk Managers and Risk Assessors. U.S. EPA, Memorandum to Assistant Administrators and Regional Administrators, February 26, 1992.

Hakulinen, T., Salonen, T., Teppo, L. (1976) Cancer in the offspring of fathers in hydrocarbon-related occupations. *Br. J. Prev. Soc. Med.* 30:130-140.

Harris, M.W., Chapin, R.E., Lockhart, A.C., Jokinen, M.P., Allen, J.D., Haskins, E.A. (1992) Assessment of a short-term reproductive and developmental toxicity screen. *Fund. Appl. Toxicol.* 19:186-196.

Haschek, W.M. & Rousseaux, C.G. (1991) *Handbook of Toxicologic Pathology*. Academic Press, New York.

Hatch, M. & Kline, J. (1981). Spontaneous abortion and exposure to the herbicide 2,4,5-T: A pilot study. U.S. Environmental Protection Agency, Washington, D.C. EPA-560/6-81-006.

Heindel, J.J. & Chapin, R.E. (1993) *Methods in Toxicology: Female Reproductive Toxicology*. Academic Press, San Diego.

Heindel, J.J., Thomford, P.J., Mattison, D.R. (1989) Histological assessment of ovarian follicle number in mice as a screen of ovarian toxicity. In: Hirshfield, A.N. *Growth Factors and the Ovary*. Plenum Press, New York. pp. 421-426.

**DRAFT--DO NOT QUOTE OR CITE**

- Hemminki, K. & Vineis, P. (1985) Extrapolation of the evidence on teratogenicity of chemicals between humans and experimental animals: Chemicals other than drugs. *Terat. Carcin. Mutagen.* 5:251-318.
- Hemminki, K., Mutanen, P., Luoma, K., Saloniemi, I. (1980) Congenital malformations by the parental occupation in Finland. *Int. Arch. Occup. Environ. Health* 46:93-98.
- Hemminki, K., Saloniemi, I., Salonen, T. (1981) Childhood cancer and paternal occupation in Finland. *J. Epidemiol. Community Health* 35:11-15.
- Hertig, A.T. (1967) The overall problem in man. In: Benirschke, K. *Comparative Aspects of Reproductive Failure*. Springer-Verlag, New York. pp. 11-41.
- Hervey, E. & Hervey, G.R. (1967) The effects of progesterone on body weight and composition in the rat. *J. Endocrinol.* 37:361-384.
- Hess, R.A. (1990) Quantitative and qualitative characteristics of the stages and transitions in the cycle of the rat seminiferous epithelium: Light microscopic observations of perfusion-fixed and plastic-embedded testes. *Biol. Reprod.* 43:525-542.
- Hess, R.A. & Moore, B.J. (1993) Histological methods for evaluation of the testis. In: Chapin, R.E. & Heindel, J.J. *Methods in Toxicology: Male Reproductive Toxicology*. Academic Press, San Diego. pp. 52-85.
- Hess, R.A., Moore, B.J., Forrer, J., Linder, R.E., Abuel-Atta, A.A. (1991) The fungicide Benomyl (methyl 1-(butylcarbamoyl)-2-benzimidazolecarbamate) causes testicular dysfunction by inducing the sloughing of germ cells and occlusion of efferent ductules. *Fund. Appl. Toxicol.* 17:733-745.
- Heywood, R. & James, R.W. (1985) Current laboratory approaches for assessing male reproductive toxicity. In: Dixon, R.L. *Reproductive Toxicology*. Raven Press, New York. pp. 147-160.
- Holloway, A.J., Moore, H.D.M., Foster, P.M.D. (1990a) The use of in vitro fertilization to detect reductions in the fertility of male rats exposed to 1,3-dinitrobenzene. *Fund. Appl. Toxicol.* 14:113-122.
- Holloway, A.J., Moore, H.D.M., Foster, P.M.D. (1990b) The use of rat in vitro fertilization to detect reductions in the fertility of spermatozoa from males exposed to ethylene glycol monomethyl ether. *Reprod. Toxicol.* 4:21-27.

## **DRAFT--DO NOT QUOTE OR CITE**

Holmes, R.L. & Ball, J.N. (1974) *The Pituitary Gland: A Comparative Account*. Cambridge University Press, Cambridge.

Hood, R.D. (1989) Paternally mediated effects. In: Hood, R.D. *Developmental Toxicity: Risk Assessment and the Future*. Van Nostrand Reinhold, New York. pp. 77-79.

Huang, H.H. & Meites, J. (1975) Reproductive capacity of aging female rats. *Neuroendocrinol.* 17:289-295.

Hughenoltz, A.P. & Bruce, W.R. (1983) Radiation induction of mutations affecting sperm morphology in mice. *Mutat. Res.* 107:177-185.

Hughes, C.L. (1988) Phytochemical mimicry of reproductive hormones and modulation of herbivore fertility by phytoestrogens. *Environ. Health* 78:171-175.

Hurt, M.E. & Zenick, H. (1986) Decreasing epididymal sperm reserves enhances the detection of ethoxyethanol-induced spermatotoxicity. *Fund. Appl. Toxicol.* 7:348-353.

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

Jones, T.C., Mohr, U., Hunt, R.D. (1987) *Genital System*. Springer-Verlag, New York.

Katz, D.F. & Overstreet, J.W. (1981) Sperm motility assessment by videomicrography. *Fertil. Steril.* 35:188-193.

Katz, D.F., Diel, L., Overstreet, J.W. (1982) Differences in the movement of morphologically normal and abnormal human seminal spermatozoa. *Biol. Reprod.* 26:566-570.

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

Kimmel, C.A. & Francis, E.Z. (1990) Proceedings of the Workshop on the Acceptability and Interpretation of Dermal Developmental Toxicity Studies. *Fund. 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.

**DRAFT--DO NOT QUOTE OR CITE**

Kimmel, C.A., Holson, J.F., Hogue, C.J., Carlo, G.L. (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. (1986) Interagency Regulatory Liason Group workshop on reproductive toxicity risk assessment. *Environ. Health* 66:193-221.

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

Kimmel, G.L., Clegg, E.D., Crisp, T.M. (In press) Reproductive toxicity testing: a risk assessment perspective. In: Witorsch, R.J. Reproductive Toxicology. Raven Press, New York.

Kissling, G. (1981) A generalized model for analysis of non-independent observations. Thesis. University of North Carolina.

Kleinbaum, D.G., Kupper, L.L., Morgenstern, H. (1982) Epidemiologic Research: Principle and Quantitative Methods. Lifetime Learning Publications, London.

Knobil, E. & Neill, J.D. (1988) The Physiology of Reproduction. Raven Press, New York.

Kodell, R.L., Howe, R.B., Chen, J.J., Gaylor, D.W. (1991) Mathematical modeling of reproductive and developmental toxic effects for quantitative risk assessment. *Risk Anal.* 11:583-590.

Kupfer, D. (1987) Critical evaluation of methods for detection and assessment of estrogenic compounds in mammals: Strengths and limitations for application to risk assessment. *Reprod. Toxicol.* 2:147-153.

Kurman, R. & Norris, H.J. (1978) Germ cell tumors of the ovary. *Pathol. Annu.* 13:291.

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

La Bella, F.S., Dular, R., Lemons, P., Vivian, S., Queen, M. (1973a) Prolactin secretion is specifically inhibited by nickel. *Nature* 245:330-332.

La Bella, F.S., Dular, R., Vivian, S., Queen, G. (1973b) Pituitary hormone releasing activity of metal ions present in hypothalamic extracts. *Biochem. Biophys. Res. Commun.* 52:786-791.

## **DRAFT--DO NOT QUOTE OR CITE**

Lamb, J.C. (1985) Reproductive toxicity testing: Evaluating and developing new testing systems. *J. Amer. Coll. Toxicol.* 4:163-171.

Lamb, J.C. & Chapin, R.E. (1985) Experimental models of male reproductive toxicology. In: Thomas, J.A., Korach, K.S., McLachlan, J.A. *Endocrine Toxicology*. Raven Press, New York. pp. 85-115.

Lamb, J.C. & Foster, P.M.D. (1988) *Physiology and Toxicology of Male Reproduction*. Academic Press, New York.

Lamb, J.C., Jameson, C.W., Choudhury, H., Gulati, D.K. (1985) Fertility assessment by continuous breeding: Evaluation of diethylstilbestrol and a comparison of results from two laboratories. *J. Amer. Coll. Toxicol.* 4:173-183.

Langley, F.A. & Fox, H. (1987) Ovarian tumors. Classification, histogenesis, etiology. In: Fox, H. Haines and Taylor's *Obstetrical and Gynaecologic Pathology*. Churchill Livingstone, Edinburgh. pp. 542-555.

Lantz, G.D., Cunningham, G.R., Huckins, C., Lipshultz, L.I. (1981) Recovery from severe oligospermia after exposure to dibromochloropropane. *Fertil. Steril.* 35:46-53.

LeFevre, J. & McClintock, M.K. (1988) Reproductive senescence in female rats: A longitudinal study of individual differences in estrous cycles and behavior. *Biol. Reprod.* 38:780-789.

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.

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

Lemasters, G.K., Hagen, A., Samuels, S. (1985) Reproductive outcomes in women exposed to solvents in 36 reinforced plastic companies. 1. Menstrual dysfunction. *J. Occup. Med.* 27:490-494.

Leridon, H. (1977) *Human Fertility: The Basic Components*. The University of Chicago Press, Chicago.

Le Vier, R.R. & Jankowiak, M.E. (1972) The hormonal and antifertility activity of 2,6-cis-diphenylhexamethylcyclotetra-siloxane in the female rat. *Biol. Reprod.* 7:260-266.



## **DRAFT--DO NOT QUOTE OR CITE**

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, R.J., Symons, M.J., Balogh, S.A., Arndt, D.M., Kaswandik, N.R., Gentile, J.W. (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., Milby, T.H., Whorton, M.D. (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.

Levine, R.J., Blunden, P.B., DalCorso, R.D., Starr, T.B., Ross, C.E. (1983) Superiority of reproductive histories to sperm counts in detecting infertility at a dibromochloropropane manufacturing plant. *J. Occup. Med.* 25:591-597.

Lewis, R.J. (1991) *Reproductively Active Chemicals: A Reference Guide*. Van Nostrand Reinhold, New York.

Linder, R.E., Hess, R.A., Strader, L.F. (1986) Testicular toxicity and infertility in male rats treated with 1,3-dinitrobenzene. *J. Toxicol. Environ. Health* 19:477-489.

Linder, R.E., Strader, L.F., Barbee, R.R., Rehnberg, G.L., Perreault, S.D. (1990) Reproductive Toxicity of a Single Dose of 1,3-Dinitrobenzene in 2 Ages of Young Adult Male Rats. *Fund. Appl. Toxicol.* 14:284-298.

Lipshultz, L.I., Ross, C.E., Whorton, D., Thomas, M., Smith, R., Joyner, R.E. (1980) Dibromochloropropane and its effect on testicular function in man. *J. Urol.* 124:464-468.

Long, J.A. & Evans, H.M. (1922) The oestrous cycle in the rat and its associated phenomena. *Mem. Univ. Calif.* 6:1-111.

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

Manson, J.M. (1994) Testing of pharmaceutical agents for reproductive toxicity. In: Kimmel, C.A. & Buelke-Sam, J. *Developmental Toxicology*. Raven Press, New York. p. 379.

Manson, J.M. & Kang, Y.J. (In press) Test methods for assessing female reproductive and developmental toxicology. In: Hayes, A.W. *Principles and Methods of Toxicology*. Raven Press, New York.

## **DRAFT--DO NOT QUOTE OR CITE**

Mattison, D.R. (1985) Clinical manifestations of ovarian toxicity. In: Dixon, R.L. Reproductive Toxicology. Raven Press, New York. pp. 109-130.

Mattison, D.R. & Nightingale, M.R. (1980) The biochemical and genetic characteristics of murine ovarian aryl hydrocarbon (benzo(a)pyrene) hydroxylase activity and its relationship to primary oocyte destruction by polycyclic aromatic hydrocarbons. *Toxicol. Appl. Pharmacol.* 56:399-408.

Mattison, D.R. & Thomford, P.J. (1989) The mechanisms of action of reproductive toxicants. *Toxicol. Pathol.* 17:364-376.

Mattison, D.R. & Thorgeirsson, S.S. (1978) Gonadal aryl hydrocarbon hydroxylase in rats and mice. *Cancer Res.* 38:1368-1373.

McLachlan, J.A. (1980) Estrogens in the Environment. Elsevier North Holland, New York.

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

McNatty, K.P. (1979) Follicular determinants of corpus luteum function in the human ovary. *Adv. Exp. Med. Biol.* 112:465-481.

Meistrich, M.L. (1982) Quantitative correlation between testicular stem cell survival, sperm production, and fertility in the mouse after treatment with different cytotoxic agents. *J. Androl.* 3:58-68.

Meistrich, M.L. (1986) Critical components of testicular function and sensitivity to disruption. *Biol. Reprod.* 34:17-28.

Meistrich, M.L. & Brown, C.C. (1983) Estimation of the increased risk of human infertility from alterations in semen characteristics. *Fertil. Steril.* 40:220-230.

Meistrich, M.L. & Samuels, R.C. (1985) Reduction in sperm levels after testicular irradiation of the mouse: A comparison with man. *Rad. Res.* 102:138-147.

Meistrich, M.L. & van Beek, M.E.A.B. (1993) Spermatogonial stem cells: assessing their survival and ability to produce differentiated cells. In: Chapin, R.E. & Heindel, J.J. *Methods in Toxicology: Male Reproductive Toxicology*. Academic Press, San Diego. pp. 106-123.

Meyer, C.R. (1981) Semen quality in workers exposed to carbon disulfide compared to a control group from the same plant. *J. Occup. Med.* 23:435-439.

## **DRAFT--DO NOT QUOTE OR CITE**

- Milby, T.H. & Whorton, D. (1980) Epidemiological assessment of occupationally related chemically induced sperm count suppression. *J. Occup. Med.* 22:77-82.
- Milby, T.H., Whorton, M.D., Stubbs, H.A., Ross, C.E., Joyner, R.E., Lipshultz, L.I. (1981) Testicular function among epichlorohydrin workers. *Br. J. Ind. Med.* 38:372-377.
- Morris, I.D., Bardin, C.W., Gunsalus, G., Ward, J.A. (1990) Prolonged suppression of spermatogenesis by oestrogen does not preserve the seminiferous epithelium in Procarbazine-treated rats. *Int. J. Androl.* 13:180-189.
- Morrissey, R.E., Lamb, J.C., Morris, R.W., Chapin, R.E., Gulati, D.K., Heindel, J.J. (1989) Results and evaluations of 48 continuous breeding reproduction studies conducted in mice. *Fund. Appl. Toxicol.* 13:747-777.
- Mosher, W.D. & Pratt, W.F. (1990). Fecundity and infertility in the United States, 1965-88. Report 192, National Center for Health Statistics, Hyattsville, MD.
- Mukhtar, H., Philpot, R.M., Lee, I.P., Bend, J.R. (1978) Developmental aspects of epoxide-metabolizing enzyme activities in adrenals, ovaries, and testes of the rat. In: Mahlum, D.D., Sikov, M.R., Hackett, P.L., Andrew, F.D. *Developmental Toxicology of Energy Related Pollutants*. Technical Information Center, U.S. Department of Energy, Springfield, VA. pp. 89-104.
- Na, J.Y., Garza, F., Terranova, P.F. (1985) Alterations in follicular fluid steroids and follicular hCG and FSH binding during atresia in hamsters. *Proc. Soc. Exp. Biol. Med.* 179:123-127.
- Nagao, T. & Fujikawa, K. (1990) Genotoxic potency in mouse spermatogonial stem cells of triethylenemelamine, mitomycin-C, ethylnitrosourea, procarbazine, and propyl methanesulfonate as measured by F1 congenital defects. *Mutat. Res.* 229:123-128.
- Nakai, M., Moore, B.J., Hess, R.A. (1993) Epithelial reorganization and irregular growth following carbendazim-induced injury of the efferent ductules of the rat testis. *Anat. Rec.* 235:51-60.
- National Research Council. (1977) Reproduction and teratogenicity tests. In: *Principles and Procedures for Evaluating the Toxicity of Household Substances*. National Academy Press, Washington, DC.
- National Research Council. (1983) *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press, Washington.

## **DRAFT--DO NOT QUOTE OR CITE**

National Research Council. (1989) *Biologic Markers in Reproductive Toxicity*. National Academy Press, Washington, DC.

Nestor, A. & Handel, M.A. (1984) The transport of morphologically abnormal sperm in the female reproductive tract of mice. *Gamete Res.* 10:119-125.

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

Organization for Economic Cooperation and Development (1983). First addendum to OECD guideline 415 for testing of chemicals, "One-Generation Rreproduction Toxicity".

Organization for Economic Cooperation and Development (1993a). Draft guidelines for testing chemicals: Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test. #422.

Organization for Economic Cooperation and Development (1993b). First amendment to OECD guidelines 416 "Two Generation Reproduction Toxicity".

Overstreet, J.W. (1984) Laboratory tests for human male reproductive risk assessment. In: Legator, M.S., Rosenberg, M., Zenick, H. *Environmental Influences on Fertility, Pregnancy and Development. Strategies for Measurement and Evaluation*. Alan R. Liss, New York. pp. 67-82.

Pang, C.N., Zimmerman, E., Sawyer, C.H. (1977) Morphine inhibition of preovulatory surges of plasma luteinizing hormone and follicle stimulating hormone in the rat. *Endocrinol.* 101:1726-1732.

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.

Pease, W., Vandenberg, J., Hooper, K. (1991) Comparing alternative approaches to establishing regulatory levels for reproductive toxicants: DBCP as a case study. *Environ. Health* 91:141-155.

Peluso, J.J., Bolender, D.L., Perri, A. (1979) Temporal changes associated with the degeneration of the rat oocyte. *Biol. Reprod.* 20:423-430.

Perreault, S.D. (1989) Impaired gamete function: Implications for reproductive toxicology. In: Working, P.K. *Toxicology of the Male and Female Reproductive Systems*. Hemisphere, New York. pp. 217-229.

## **DRAFT--DO NOT QUOTE OR CITE**

- Perreault, S.D., Jeffay, S., Poss, P., Laskey, J.W. (1992) Use of the fungicide carbendazim as a model compound to determine the impact of acute chemical exposure during oocyte maturation and fertilization on pregnancy outcome in the hamster. *Toxicol. Appl. Pharmacol.* 114:225-231.
- Peters, J.M., Preston-Martin, S., Yu, M.C. (1981) Brain tumors in children and occupational exposure of the parents. *Science* 213:235-237.
- Plowchalk, D.R., Smith, B.J., Mattison, D.R. (1993) Assessment of toxicity to the ovary using follicle quantitation and morphometrics. In: Heindel, J.J. & Chapin, R.E. *Methods in Toxicology: Female Reproductive Toxicology*. Academic Press, San Diego. pp. 57-68.
- Rai, K. & Van Ryzin, J. (1985) A dose response model for teratological experiments involving quantal responses. *Biometrics* 41:1-10.
- Ratcliffe, J.M., Clapp, D.E., Schrader, S.M., Turner, T.W., Oser, J., Tanaka, S., Hornung, R.W., Halperin, W.E. (1986). Semen quality in 2-ethoxyethanol-exposed workers. Health Hazard evaluation report, HETA 84-415-1688. Department of Health and Human Services, National Institute for Occupational Safety and Health, Cincinnati, Ohio.
- Ratcliffe, J.M., Schrader, S.M., Steenland, K., Clapp, D.E., Turner, T., Hornung, R.W. (1987) Semen quality in papaya workers with long term exposure to ethylene dibromide. *Br. J. Ind. Med.* 44:317-326.
- Redi, C.A., Garagna, S., Pellicciari, C., Manfredi-Romanini, M.G., Capanna, E., Winking, H., Gropp, A. (1984) Spermatozoa of chromosomally heterozygous mice and their fate in male and female genital tracts. *Gamete Res.* 9:273-286.
- Robaire, B., Smith, S., Hales, B.F. (1984) Suppression of spermatogenesis by testosterone in adult male rats: Effect on fertility, pregnancy outcome and progeny. *Biol. Reprod.* 31:221-230.
- Rosenberg, M.J., Wyrobeck, A.J., Ratcliffe, J., Gordon, L.A., Watchmaker, G., Fox, S.H., Moore, D.H. (1985) Sperm as an indicator of reproductive risk among petroleum refinery workers. *Br. J. Ind. Med.* 42:123-127.
- Rothman, K.J. (1986) *Modern epidemiology*. Little, Brown, Boston.
- Rubin, H.B. & Henson, D.E. (1979) Effects of drugs on male sexual function. In: *Advances in Behavioral Pharmacology*. Academic Press, New York. pp. 65-86.

## **DRAFT--DO NOT QUOTE OR CITE**

- Russell, L.D. (1983) Normal testicular structure and methods of evaluation under experimental and disruptive conditions. In: Clarkson, T.W., Nordberg, G.F., Sager, P.R. Reproductive and Developmental Toxicity of Metals. Plenum Publishing Co., New York. pp. 227-252.
- Russell, L.D., Malone, J.P., McCurdy, D.S. (1981) Effect of microtubule disrupting agents, colchicine and vinblastine, on seminiferous tubule structure in the rat. *Tiss. Cell* 13:349-367.
- Russell, L.D., Ettlin, R., Sinha Hikim, A.P., Clegg, E.D. (1990) Histological and Histopathological Evaluation of the Testis. Cache River Press, Clearwater, FL.
- Sakai, C.N. & Hodgen, G.D. (1987) Use of primate folliculogenesis models in understanding human reproductive biology and applicability to toxicology. *Reprod. Toxicol.* 1:207-222.
- Scala, R.A., Bevan, C., Beyer, B.K. (1992) An Abbreviated Repeat Dose and Reproductive/Developmental Toxicity Test for High Production Volume Chemicals. *Regul. Toxicol. Pharmacol.* 16:73-80.
- Schardein, J.L., Schwetz, B.B., Kenel, M.F. (1985) Species sensitivities and prediction of teratogenic potential. *Environ. Health Perspect.* 61:55-62.
- Schrag, S.D. & Dixon, R.L. (1985a) Occupational exposures associated with male reproductive dysfunction. *Ann. Rev. Pharmacol. Toxicol.* 25:567-592.
- Schrag, S.D. & Dixon, R.L. (1985b) Reproductive effects of chemical agents. In: Dixon, R.L. Reproductive Toxicology. Raven Press, New York. pp. 301-319.
- Schwetz, B.A., Rao, K.S., Park, C.N. (1980) Insensitivity of tests for reproductive problems. *J. Environ. Pathol. Toxicol.* 3:81-98.
- Scialli, A.R. & Clegg, E.D. (1992) Reversibility in Testicular Toxicity Assessment. CRC Press, Boca Raton.
- Scommegna, A., Vorys, N., Givens, J.R. (1980) Chapter 15: Menstrual dysfunction. In: Gold, J.J. & Josimovich, J.B. Gynecologic Endocrinology. Harper and Row, Hagerstown, Maryland.
- Selevan, S.G. (1980) Evaluation of data sources for occupational pregnancy outcome studies. University of Cincinnati from University Microfilms, Ann Arbor.

## **DRAFT--DO NOT QUOTE OR CITE**

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. *Occupational Hazards and Reproduction*. Hemisphere, Washington, DC. pp. 219-229.

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

Selevan, S.G., Edwards, B., Samuels, S. (1982) Interview data from both parents on pregnancies and occupational exposures. How do they compare? *Am. J. Epidemiol.* 116:583.

Sever, L.E. & Hessol, N.A. (1984) Overall design considerations in male and female occupational reproductive studies. In: Lockey, J.E., Lemasters, G.K., Key, W.R. *Reproduction: The New Frontier in Occupational and Environmental Research*. Alan R. Liss, Inc., New York. pp. 15-48.

Sheehan, D.M., Young, J.F., Slikker, W., Gaylor, D.W., Mattison, D.R. (1989) Workshop on risk assessment in reproductive and developmental toxicology: Addressing the assumptions and identifying the research needs. *Regul. Toxicol. Pharmacol.* 10:110-122.

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

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

Skett, P. (1988) Biochemical basis of sex differences in drug metabolism. *Pharm. Thera.* 38:269-304.

Slott, V.L., Suarez, J.D., Simmons, J.E., Perreault, S.D. (1990) Acute inhalation exposure to epichlorohydrin transiently decreases rat sperm velocity. *Fund. Appl. Toxicol.* 15:597-606.

Slott, V.L., Suarez, J.D., Perreault, S.D. (1991) Rat sperm motility analysis: Methodologic considerations. *Reprod. Toxicol.* 5:449-458.

Smith, B.J., Plowchalk, D.R., Sipes, I.G., Mattison, D.R. (1991) Comparison of random and serial sections in assessment of ovarian toxicity. *Reprod. Toxicol.* 5:379-383.

## **DRAFT--DO NOT QUOTE OR CITE**

Smith, C.G. (1983) Reproductive toxicity: hypothalamic-pituitary mechanisms. *Am. J. Ind. Med.* 4:107-112.

Smith, C.G. & Gilbeau, P.M. (1985) Drug abuse effects on reproductive hormones. In: Thomas, J.A., Korach, K.S., McLachlan, J.A. *Endocrine Toxicology*. Raven Press, New York. pp. 249-267.

Smith, E.R. & Davidson, J.M. (1974) Luteinizing hormone releasing factor in rats exposed to constant light: Effects of mating. *Neuroendocrinol.* 14:129-138.

Smith, P.E. (1939) The effect on the gonads of the ablation and implantation of the hypophysis and the potency of the hypophysis under various conditions. In: Allen, E. *Sex and Internal Secretions*. Williams and Wilkins., Baltimore, Maryland.

Smith, S.K., Lenton, E.A., Landgren, B.M., Cooke, I.D. (1984) The short luteal phase and infertility. *Br. J. Obstet. Gynaecol.* 91:1120-1122.

Sonawane, B.R. & Yaffe, S.J. (1983) Delayed effects of drug exposure during pregnancy: Reproductive function. *Biol. Res. Pregnancy* 4:48-55.

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

Stein, A. & Hatch, M. (1987) Biological markers in reproductive epidemiology: Prospects and precautions. *Environ. Health* 74:67-75.

Stein, Z., Kline, J., Shrout, P. (1985) Power in surveillance. In: Hemminki, K., Sorsa, M., Vainio, H. *Occupational Hazards and Reproduction*. Hemisphere, Washington, DC. pp. 203-208.

Steinberger, E. & Lloyd, J.A. (1985) Chemicals affecting the development of reproductive capacity. In: Dixon, R.L. *Reproductive Toxicology*. Raven Press, New York, New York. .

Stevens, K.R. & Gallo, M.A. (1989) Practical considerations in the conduct of chronic toxicity studies. In: Hayes, A.W. *Principles and Methods of Toxicology*. Raven Press, New York. pp. 237-250.

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



**DRAFT--DO NOT QUOTE OR CITE**

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

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

Tanaka, S., Kawashima, K., Naito, K., Usami, M., Nakadate, M., Imaida, K., Takahashi, M., Hayashi, Y., Kurokawa, Y., Tobe, M. (1992) Combined Repeat Dose and Reproductive/Developmental Toxicity Screening Test (OECD): Familiarization using cyclophosphamide. *Fund. Appl. Toxicol.* 18:89-95.

Terranova, P.F. (1980) Effects of phenobarbital-induced ovulatory delay on the follicular population and serum levels of steroids and gonadotropins in the hamster: A model for atresia. *Biol. Reprod.* 23:92-99.

Thomas, J.A. (1981) Reproductive hazards and environmental chemicals: A review. *Toxic Subst. J.* 2:318-348.

Thomas, J.A. (1991) Toxic responses of the reproductive system. In: Amdur, M.O., Doull, J., Klaassen, C.D. Casarett and Doull's Toxicology. Pergamon Press, New York. pp. 484-520.

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

Toth, G.P., Stober, J.A., Read, E.J., Zenick, H., Smith, M.K. (1989a) The automated analysis of rat sperm motility following subchronic epichlorohydrin administration: Methodologic and statistical considerations. *J. Androl.* 10:401-415.

Toth, G.P., Zenick, H., Smith, M.K. (1989b) Effects of epichlorohydrin on male and female reproduction in Long-Evans rats. *Fund. Appl. Toxicol.* 13:16-25.

Toth, G.P., Stober, J.A., Zenick, H., Read, E.J., Christ, S.A., Smith, M.K. (1991) Correlation of sperm motion parameters with fertility in rats treated subchronically with epichlorohydrin. *J. Androl.* 12:54-61.

Treloar, A.E., Boynton, R.E., Borghild, G.B., Brown, B.W. (1967) Variation in the human menstrual cycle through reproductive life. *Int. J. Fertil.* 12:77-126.

**DRAFT--DO NOT QUOTE OR CITE**

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. Congress (1985). *Reproductive Health Hazards in the Workplace*. Office of Technology Assessment, OTA-BA-266, U.S. Government Printing Office, Washington, DC.

U. S. Congress (1988). *Infertility: Medical and Social Choices*. Office of Technology Assessment, OTA-BA-358, U.S. Government Printing Office, Washington, DC.

U.S. Environmental Protection Agency (1982). *Reproductive and Fertility Effects. Pesticide Assessment Guidelines, Subdivision F. Hazard Evaluation: Human and Domestic Animals*. Office of Pesticides and Toxic substances, Washington, D.C. EPA-540/9-82-025.

U.S. Environmental Protection Agency (1985a). *Hazard Evaluation Division Standard Evaluation Procedure. Teratology Studies*. Office of Pesticide Programs, Washington, DC. pp. 22-23.

U.S. Environmental Protection Agency (1985b). *Toxic Substances Control Act Test Guidelines: Final Rules*. *Federal Register* 50 (188):39426-39436.

U.S. Environmental Protection Agency (1986a). *Guidelines for Carcinogen Risk Assessment*. *Federal Register*. 51(185):33992-34003.

U.S. Environmental Protection Agency (1986b). *Guidelines for Estimating Exposures*. *Federal Register* 51(185):34042-34054.

U.S. Environmental Protection Agency (1986c). *Guidelines for Mutagenicity Risk Assessment*. *Federal Register* 51(185):34006-34012.

U.S. Environmental Protection Agency (1987). *Reference Dose (RfD): Description and Use in Health Risk Assessments. Intergrated Risk Information System (IRIS): Appendix A. Integrated Risk Information System Documentation, Vol. 1*. EPA/600/8-66/032a.

U.S. Environmental Protection Agency. (1991) *Guidelines for Developmental Toxicity Risk Assessment*. *Fed. Reg.* 56(234):63798-63826.

U.S. Environmental Protection Agency (1992). *Guidelines for Exposure Assessment*. *Federal Register* 57(104):22888-22938.

Wade, G.N. (1972) Gonadal hormones and behavioral regulation of body weight. *Physiol. Behav.* 8:523-534.

## **DRAFT--DO NOT QUOTE OR CITE**

Walker, R.F. (1986) Age factors potentiating drug toxicity in the reproductive axis. *Environ. Health* 70:185-191.

Walker, R.F., Schwartz, L.W., Manson, J.M. (1988) Ovarian effects of an anti-inflammatory-immunomodulatory drug in the rat. *Toxicol. Appl. Pharmacol.* 94:266-275.

Waller, D.P., Killinger, J.M., Zaneveld, L.J.D. (1985) Physiology and toxicology of the male reproductive tract. In: Thomas, J.A., Korach, K.S., McLachlan, J.A. *Endocrine Toxicology*. Raven Press, New York. pp. 269-333.

Wang, G.H. (1923) The relation between the "spontaneous" activity and the oestrous cycle in the white rat. *Comp. Psychol. Monographs* 2:1-27.

Warren, J.C., Cheatum, S.G., Greenwald, G.S., Barker, K.L. (1967) Cyclic variation of uterine metabolic activity in the golden hamster. *Endocrinol.* 80:714-718.

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

Whorton, D. & Milby, T.H. (1980) Recovery of testicular function among DBCP workers. *J. Occup. Med.* 22:177-179.

Whorton, D., Krauss, R.M., Marshall, S., Milby, T.H. (1977) Infertility in male pesticide workers. Preliminary communication. *Lancet* 2(8051):1259-1261.

Whorton, D., Milby, T.H., Krauss, R.M., Stubbs, H.A. (1979) Testicular function in DBCP exposed pesticide workers. *J. Occup. Med.* 21:161-166.

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

Wilcox, A.J., Weinburg, C.R., Wehmann, R.E., Armstrong, E.G., Canfield, R.E., Nisula, B.C. (1985) Measuring early pregnancy loss: laboratory and field methods. *Fertil. Steril.* 44:366-374.

Williams, J., Gladen, B.C., Schrader, S.M., Turner, T.W., Phelps, J.L., Chapin, R.E. (1990) Semen analysis and fertility assessment in rabbits: Statistical power and design considerations for toxicology studies. *Fund. Appl. Toxicol.* 15:651-665.

Wilson, J.G. (1973) *Environment and birth defects*. Academic Press, New York.

**DRAFT--DO NOT QUOTE OR CITE**

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

Wilson, J.G., Scott, W.J., Ritter, E.J., Fradkin, R. (1975) Comparative distribution and embryotoxicity of hydroxyurea in pregnant rats and rhesus monkeys. *Teratol.* 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.

Working, P.K. (1988) Male reproductive toxicity: Comparison of the human to animal models. *Environ. Health* 77:37-44.

Working, P.K. (1989) Toxicology of the Male and Female Reproductive Systems. Hemisphere, New York.

Working, P.K. & Hurtt, M. (1987) Computerized videomicrographic analysis of rat sperm motility. *J. Androl.* 8:330-337.

Wyrobek, A.J. (1982) Sperm assays as indicators of chemically-induced germ cell damage in man. In: . Mutagenicity: New Horizons in Genetic Toxicology. Academic Press, New York. pp. 337-349.

Wyrobek, A.J. (1984). Identifying agents that damage human spermatogenesis: Abnormalities in sperm concentration and morphology. In: Monitoring Human Exposure to Carcinogenic and Mutagenic agents. Proceedings of a joint symposium held in Espoo, Finland. Dec. 12-15, 1983. International Agency for Research on Cancer, Lyon, France.

Wyrobek, A.J. & Bruce, W.R. (1978) The induction of sperm-shape abnormalities in mice and humans. In: Hollander, A. & de Serres, F.J. Chemical Mutagens: Principles and Methods for Their Detection. Plenum Press, New York.

Wyrobek, A.J., Gordon, L.A., Burkhart, J.G., Francis, M.W., Kapp, R.W., Letz, G., Malling, H.V., Topham, J.C., Whorton, D.M. (1983a) An evaluation of the mouse sperm morphology test and other sperm tests in nonhuman mammals. *Mutat. Res.* 115:1-72.

Wyrobek, A.J., Gordon, L.A., Burkhart, J.G., Francis, M.W., Kapp, R.W., Jr., Letz, G., Malling, H., V, Topham, J.C., Whorton, D.M. (1983b) An evaluation of human sperm as

**DRAFT--DO NOT QUOTE OR CITE**

indicators of chemically induced alterations of spermatogenic function. *Mutat. Res.* 115:73-148.

Wyrobek, A.J., Watchmaker, G., Gordon, L. (1984) An evaluation of sperm tests as indicators of germ-cell damage in men exposed to chemical or physical agents. In: Lockey, J.E., Lemasters, G.K., Keye, W.R. *Reproduction: The New Frontier in Occupational and Environmental Health Research*. Alan R. Liss, New York. pp. 385-407.

Zack, M., Cannon, S., Lloyd, D., Heath, C.W., Falletta, J.M., Jones, B., Housworth, J., Crowley, S. (1980) Cancer in children of parents exposed to hydrocarbon-related industries and occupations. *Am. J. Epidemiol.* 3:329-336.

Zenick, H. & Clegg, E.D. (1986) Issues in risk assessment in male reproductive toxicology. *J. Amer. Coll. Toxicol.* 5:249-259.

Zenick, H., Blackburn, K., Hope, E., Baldwin, D.J. (1984) Evaluating male reproductive toxicity in rodents: A new animal model. *Terat. Carcin. Mutagen.* 4:109-128.

Zenick, H., Clegg, E.D., Perreault, S.D., Klinefelter, G.R., Gray, L.E. (In press) Assessment of male reproductive toxicity: A risk assessment approach. In: Hayes, A.W. *Principles and Methods of Toxicology*. Raven Press, New York.