EPA/630/R-96/009
October 1996

Guidelines for Reproductive Toxicity Risk Assessment

Published on October 31, 1996, Federal Register 61(212):56274-56322

These guidelines replace two proposed guidelines: Proposed Guidelines for Female Reproductive Risk and Proposed Guidelines for Male Reproductive Risk, both dated June 30, 1988.

Risk Assessment Forum
U.S. Environmental Protection Agency
Washington, DC

DISCLAIMER

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

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

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

Part B: Response to Science Advisory Board and Public Comments

LIST OF TABLES

GUIDELINES FOR REPRODUCTIVE TOXICITY RISK ASSESSMENT

[FRL-5630-6]

AGENCY: U.S. Environmental Protection Agency

ACTION: Notice of availability of final Guidelines for Reproductive Toxicity Risk Assessment

SUMMARY: The U.S. Environmental Protection Agency (EPA) is today publishing in final form a document entitled *Guidelines for Reproductive Toxicity Risk Assessment* (hereafter "Guidelines"). These Guidelines were developed as part of an interoffice guidelines development program by a Technical Panel of the Risk Assessment Forum. They were proposed initially in 1988 as separate guidelines for the female and male reproductive systems. Subsequently, based upon the public comments and Science Advisory Board (SAB) recommendations, changes made included combining those two guidelines, integrating the hazard identification and dose-response sections, assuming as a default that an agent for which sufficient data were available on only one sex may also affect reproductive function in the other sex, expansion of the section on interpretation of female endpoints, and consideration of the benchmark dose approach for quantitative risk assessment. These Guidelines were made available again for public comment and SAB review in 1994. This notice describes the scientific basis for concern about exposure to agents that cause reproductive toxicity, outlines the general process for assessing potential risk to humans from exposure to environmental agents, and addresses Science Advisory Board and public comments on the 1994 *Proposed Guidelines for Reproductive Toxicity Risk Assessment*. Subsequent reviews have included the Agency's Risk Assessment Forum and interagency comment by members of subcommittees of the Committee on the Environment and Natural Resources of the Office of Science and Technology Policy. The EPA appreciates the efforts of all participants in the process and has tried to address their recommendations in these Guidelines.

EFFECTIVE DATE: The Guidelines will be effective October 31, 1996.

ADDRESSES: The Guidelines will be made available in the following ways:

(1) The electronic version will be accessible on EPA's Office of Research and Development home page on the Internet at http://www.epa.gov/ORD/WebPubs/repro/.

(2) 3 ½-inch high-density computer diskettes in WordPerfect 5.1 will be available from ORD Publications, Technology Transfer and Support Division, National Risk Management Research Laboratory, Cincinnati, OH; telephone: 513-569-7562; fax: 513-569-7566. Please provide the EPA No. (EPA/630/R-96/009a) when ordering.

(3) This notice contains the full document. In addition, copies of the Guidelines will be available for inspection at EPA headquarters in the Air and Radiation Docket and Information Center and in EPA headquarters and regional libraries. The Guidelines also will be made available through the U.S. Government Depository Library program and for purchase from the National Technical Information Service (NTIS), Springfield, VA; telephone: 703-487-4650; fax: 703-321-8547. Please provide the NTIS PB No. (PB97-100093) when ordering.

FOR FURTHER INFORMATION, CONTACT: Dr. Eric D. Clegg, Effects Identification and Characterization Group, National Center for Environmental Assessment-Washington Division (8623D), U.S. Environmental Protection Agency, 401 M Street, S.W., Washington, DC 20460; telephone: 202-564-3297; e-mail: clegg.eric@epamail.epa.gov.

SUPPLEMENTARY INFORMATION:

A. APPLICATION OF THE GUIDELINES

The EPA 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. EPA, 1982, 1985b, 1996a) that provide protocols designed to determine the potential of a test substance to produce reproductive (including developmental) toxicity in laboratory animals. Proposed revisions to these

testing guidelines are in the final stages of completion (U.S. EPA, 1996a). The Organization for Economic Cooperation and Development (OECD) also has issued testing guidelines (which are under revision) for reproduction studies (OECD, 1993b).

These Guidelines apply within the framework of policies provided by applicable EPA statutes and do not alter such policies. They do not imply that one kind of data or another is prerequisite for action concerning any agent. The Guidelines are not intended, nor can they be relied upon, to create any rights enforceable by any party in litigation with the United States. This document is not a regulation and is not intended to substitute for EPA regulations. These Guidelines set forth current scientific thinking and approaches for conducting reproductive toxicity risk assessments. EPA will revisit these Guidelines as experience and scientific consensus evolve.

The procedures outlined here in the Guidelines provide guidance for interpreting, analyzing, and using the data from studies that follow the above testing guidelines (U.S. EPA 1982, 1985b, 1996a). 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 in the future 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. More specific guidance on developmental effects is provided by the *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991). Other health effects guidance is provided by the *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 1986a, 1996b), the *Guidelines for Mutagenicity Risk Assessment* (U.S. EPA, 1986c), and the *Proposed Guidelines for Neurotoxicity Risk Assessment* (U.S. EPA, 1995a). These Guidelines and the four cited above are complementary.

The Agency has sponsored or participated in several conferences that addressed issues related to evaluations of reproductive toxicity data which provide some of the scientific bases for these risk assessment guidelines. Numerous publications from these and other efforts are available which provide background for these Guidelines (U.S. EPA, 1982, 1985b, 1995b; Galbraith et al., 1983; OECD, 1983; U.S. Congress, 1985, 1988; Kimmel, C.A. et al., 1986; Francis and Kimmel, 1988; Burger et

al., 1989; Sheehan et al., 1989; Seed et al., 1996). Also, numerous resources provide background information on the physiology, biochemistry, and toxicology of the male and female reproductive systems (Lamb and Foster, 1988; Working, 1989; Russell et al., 1990; Atterwill and Flack, 1992; Scialli and Clegg, 1992; Chapin and Heindel, 1993; Heindel and Chapin, 1993; Paul, 1993; Manson and Kang, 1994; Zenick et al., 1994; Kimmel, G.L. et al., 1995; Witorsch, 1995). A comprehensive text on reproductive biology also has been published (Knobil et al., 1994).

B. ENVIRONMENTAL AGENTS AND REPRODUCTIVE TOXICITY

Disorders of reproduction and hazards to reproductive health have become prominent public health issues. A variety of factors are associated with reproductive system disorders, including nutrition, environment, socioeconomic status, lifestyle, and stress. 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.

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 and Pratt, 1990). In their 1988 survey, an estimated 4.9 million women ages 15-44 (8.4%) had impaired fertility. The proportion of married couples that was infertile, from all causes, was 7.9%.

Carlsen et al. (1992) have reported from a meta analysis that human sperm concentration has declined from 113 x 10^6 per mL of semen prior to 1960 to 66 x 10^6 per mL subsequently. When combined with a reported decline in semen volume from 3.4 mL to 2.75 mL, that suggests a decline in total number of sperm of approximately 50%. Increased incidence of human male hypospadias, cryptorchidism, and testicular cancer have also been reported over the last 50 years (Giwercman et al., 1993). Several other retrospective studies that examined semen characteristics from semen donors have obtained conflicting results (Auger et al., 1995; Bujan et al., 1996; Fisch et al., 1996; Ginsburg et al., 1994; Irvine et al., 1996; Paulsen et al., 1996; Van Waeleghem et al., 1996; Vierula et al., 1996). While concerns exist about the validity of some of those conclusions, the data indicating an increase in human testicular cancer, as well as possible occurrence of other plausibly related effects such as

reduced sperm production, hypospadias, and cryptorchidism, suggest that an adverse effect may have occurred. However, there is no definitive evidence that such adverse human health effects have been caused by environmental chemicals.

Endometriosis is a painful reproductive and immunologic disease in women that is characterized by aberrant location of uterine endometrial cells, often leading to infertility. It affects approximately five million women in the United States between 15 and 45 years of age. Very limited research has suggested a link between dioxin exposure and development of endometriosis in rhesus monkeys (Rier et al., 1993). Gerhard and Runnebaum (1992) reported an association in women between occurrence of endometriosis and elevated blood PCB levels, while a subsequent small clinical study found no significant correlations between disease severity in women and serum levels of halogenated aromatic hydrocarbons (Boyd et al., 1995).

Even though not all infertile couples seek treatment, and infertility is not the only adverse reproductive effect, it is estimated that in 1986, Americans spent about $1 billion on medical care to treat infertility alone (U.S. Congress, 1988). With the increased use of assisted reproduction techniques in the last 10 years, that amount has increased substantially.

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; Kline et al., 1989). Methods that detect pregnancy as early as eight days after conception have shown that 32%-34% of postimplantation pregnancies end in embryonic or fetal loss (Wilcox et al., 1988; Zinaman et al., 1996). 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).

A variety of developmental alterations may be detected after either pre- or postnatal exposure. Several of these are discussed in the *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991), and developmental neurotoxicity is discussed in the *Proposed Guidelines for Neurotoxicity Risk Assessment* (U.S. EPA, 1996a). Relative to developmental reproductive

alterations, 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. Examples include the actions of estrogens, anti-androgens or dioxin in interfering with male sexual differentiation (Gill et al., 1979; Gray et al., 1994, 1995; Giusti et al., 1995; Gray and Ostby, 1995). 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 and Hall, 1984; Schrag and Dixon, 1985b; U.S. Congress, 1985). For example, developmental exposure to environmental compounds that possess steroidogenic (Mattison, 1985) or antisteroidogenic (Schardein, 1993) activity affect the onset of puberty and reproductive function in adulthood.

Numerous agents have been shown to cause reproductive toxicity in adult male and female laboratory animals and in humans (Mattison, 1985; Schrag and Dixon, 1985a,b; Waller et al., 1985; Lewis, 1991). In adult males and females, exposure to agents of abuse, e.g., cocaine, disrupts normal reproductive function in both test species and humans (Smith, C.G. and Gilbeau, 1985). Numerous chemicals disrupt the ovarian cycle, alter ovulation, and impair fertility in experimental animals and humans. These include agents with steroidogenic activity, certain pesticides, and some metals (Thomas, 1981; Mattison, 1985). In males, estrogenic compounds can be testicular toxicants in rodents and humans (Colborn et al., 1993; Toppari et al., 1995). 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 and 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 (McNatty, 1979; Scommegna et al., 1980; Smith, S.K. et al., 1984; Sakai and 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).

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 20×10^6 sperm per mL 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. However, some men with low sperm concentrations are able to achieve conception and many subfertile men have concentrations greater than 20×10^6, illustrating the importance of sperm quality. Toxic agents may further decrease production of sperm and increase risk of impaired fertility.

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

Risk assessment is the process by which scientific judgments are made concerning the potential for toxicity to occur in humans. In 1983, the National Research Council (NRC) defined risk assessment as comprising some or all of the following components: hazard identification, dose-response assessment, exposure assessment, and risk characterization (NRC, 1983). In its 1994 report, *Science and Judgment in Risk Assessment*, the NRC extended its view of the paradigm to include characterization of each component (NRC, 1994). In addition, it noted the importance of an interactive approach that deals with recurring conceptual issues that cut across all stages of risk assessment.

These Guidelines adopt an interactive approach by organizing the process around the components of hazard characterization, the quantitative dose-response analysis, the exposure assessment, and the risk characterization where hazard characterization combines hazard identification with qualitative consideration of dose-response relationships, route, timing, and duration of exposure. This is done because, in practice, hazard identification for reproductive toxicity and other noncancer health effects includes an evaluation of dose-response relationships, route, timing, and duration of exposure in the studies used to identify the hazard. Determining a hazard often depends on whether a dose-response relationship is present (Kimmel, C.A. et al., 1990). This approach combines the information important in comparing the toxicity of a chemical to potential human exposure scenarios identified as part of the exposure assessment. Also, it minimizes the potential for labeling chemicals inappropriately as "reproductive toxicants" on a purely qualitative basis.

In *hazard characterization*, all available experimental animal and human data, including observed effects, associated doses, routes, timing, and duration of exposure, are examined to determine if an agent causes reproductive toxicity in that species and, if so, under what conditions. From the hazard characterization and criteria provided in these Guidelines, the health-related database can be characterized as sufficient or insufficient for use in risk assessment (Section 3.7). This approach does not preclude the evaluation and use of the data for other purposes when adequate quantitative information for setting reference doses (RfDs) and reference concentrations (RfCs) is not available.

The next step, the *quantitative dose-response analysis* (Section 4), includes determining the no-observed-adverse-effect-level (NOAEL) and/or the lowest-observed-adverse-effect-level (LOAEL) for each study and type of effect. Because of the limitations associated with the use of the NOAEL, the Agency is beginning to use an additional approach, the benchmark dose approach (Crump, 1984; U.S. EPA. 1995b), for a more quantitative dose-response evaluation when allowed by the data. The benchmark dose approach takes into account the variability in the data and the slope of the dose-response curve, and thus, provides more complete use of the data for calculation of the RfD or RfC. If the data are considered sufficient for risk assessment, and if reproductive toxicity occurs at the lowest toxic dose level (i.e., the critical effect), an RfD or RfC, based on adverse reproductive effects, could be derived. This RfD or RfC is derived using the NOAEL or benchmark dose divided

by uncertainty factors to account for interspecies differences in response, intraspecies variability and deficiencies in the database.

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. EPA, 1992). However, unique considerations for reproductive toxicity exposure assessments are detailed in Section 5.

A statement of the potential for human risk and the consequences of exposure can come only from integrating the hazard characterization and quantitative dose-response analysis with human exposure estimates in the *risk characterization*. 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.

In 1992, EPA issued a policy memorandum (Habicht, 1992) and guidance package on risk characterization to encourage more comprehensive risk characterizations, to promote greater consistency and comparability among risk characterizations, and to clarify the role of professional judgment in characterizing risk. In 1995, the Agency issued a new risk characterization policy and guidance (Browner, 1995) that refines and reaffirms the principles found in the 1992 policy and outlines a process within the Agency for implementation. Although specific program policies and procedures are still evolving, these Guidelines discuss attributes of the Agency's risk characterization policy as it applies to reproductive toxicity.

Risk assessment is just one component of the regulatory process. The other component, *risk management*, uses risk characterization along with directives of the enabling regulatory legislation and other factors to decide whether to control exposure to the suspected agent and the level of control. Risk management decisions also consider socioeconomic, technical, and political factors. Risk management is not discussed directly in these guidelines because the basis for decisionmaking goes beyond scientific considerations alone. However, the use of scientific information in this process is

discussed. For example, the acceptability of the margin of exposure (MOE) is a risk management decision, but the scientific bases for generating this value are discussed here.

Dated: October 15, 1996

Signed by EPA Administrator
Carol M. Browner

PART A: GUIDELINES FOR REPRODUCTIVE TOXICITY RISK ASSESSMENT

1. OVERVIEW

These Guidelines describe the procedures that the EPA follows in using existing data to evaluate the potential toxicity of environmental agents to the human male and female reproductive systems and to developing offspring. These Guidelines focus on reproductive system 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. EPA, 1991) focus specifically on effects of agents on development 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 mutagenicity or carcinogenicity as well as other forms of general systemic toxicity. 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. EPA, 1986c) should be consulted in conjunction with these Guidelines. The *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 1986a, 1996b) 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 having adequate study design and power. 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 default assumptions are

made. These default assumptions, which are summarized in Table 1, should not preclude inquiry into the relevance of the data to potential human risk and should be invoked only after examination of the available information indicates that necessity. 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 judgment as to human risk for each agent, and that information summarized in the risk characterization.

An agent that produces an adverse reproductive effect in experimental animal studies is assumed to pose a potential reproductive threat to humans. This assumption is based on comparisons of data for agents that are known to cause human reproductive toxicity (Thomas, 1981; Nisbet and Karch, 1983; Kimmel, C.A. et al., 1984, 1990; Hemminki and 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 developmental outcomes, it is assumed that the specific outcomes seen in experimental animal studies are not necessarily the same as those produced in humans. This latter assumption is made because of the possibility of species-specific differences in timing of exposure relative to critical periods of development, pharmacokinetics (including metabolism), developmental patterns, placentation, or modes of action. However, adverse developmental outcomes in laboratory mammalian studies are presumed to predict a hazard for adverse developmental outcome in humans.

When sufficient data are available (e.g., pharmacokinetic) to allow a decision, 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 agents known to cause human reproductive toxicity, humans appear to be as or more sensitive than the most sensitive animal species tested (Nisbet and Karch, 1983; Kimmel, C.A. et al., 1984, 1990; Hemminki and 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 affects reproductive function 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; and (3) Many of the mechanisms controlling important aspects of reproductive system function are

Table 1. Default assumptions in reproductive toxicity risk assessment

1. An agent that produces an adverse reproductive effect in experimental animals is assumed to pose a potential threat to humans.

2. Effects of xenobiotics on male and female reproductive processes are assumed generally to be similar unless demonstrated otherwise. For developmental outcomes, the specific effects in humans are not necessarily the same as those seen in the experimental species.

3. In the absence of information to determine the most appropriate experimental species, data from the most sensitive species should be used.

4. In the absence of information to the contrary, an agent that affects reproductive function in one sex is assumed to adversely affect reproductive function in the other sex.

5. A nonlinear dose-response curve is assumed for reproductive toxicity.

similar in females and males, and therefore could be susceptible to the same agents. Information that would negate this assumption would demonstrate that either a mechanistic difference existed between the sexes that would preclude toxic action on the other sex or, on the basis of sufficient testing, an agent did not produce an adverse reproductive effect when administered to the other sex. Mechanistic differences could include functions that do not exist in the other sex (e.g., lactation), differences in endocrine control of affected organ development or function, or pharmacokinetic and metabolic differences between sexes.

In a quantitative dose-response analysis, mode of action, pharmacokinetic, and pharmacodynamic information should be used to predict the shape of the dose-response curve when sufficient information of that nature is available. When that information is insufficient, it has generally been assumed that there is a nonlinear dose-response for reproductive toxicity. This is 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 developing organism are known to have some capacity for repair of damage. However, in a population, background levels of toxic agents and preexisting conditions may increase the sensitivity of some individuals in the population. Thus, exposure to a toxic agent may result in an increased risk of adverse effects for some, but not necessarily all, individuals within the population. Although a threshold may exist for endpoints of reproductive toxicity, it usually is not feasible to distinguish empirically between a true threshold and a nonlinear low-dose relationship. The shift to the term nonlinear does not change the RfD/RfC methodology for reproductive system health effects, including the use of uncertainty factors.

2. DEFINITIONS AND TERMINOLOGY

For the purposes of these Guidelines, the following definitions will be used:

Reproductive toxicity - The occurrence of biologically adverse effects on the reproductive systems of females or males 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, developmental toxicity, premature reproductive senescence, or modifications in other functions that are dependent on the integrity of the reproductive systems.

Fertility - The capacity to conceive or induce conception.

Fecundity - The ability to produce offspring within a given period of time. For litter-bearing species, the ability to produce large litters is also a component of fecundity.

Fertile - 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 - 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 lifespan of the organism. The major manifestations of developmental toxicity include (1) death of the developing organism, (2) structural abnormality, (3) altered growth, and (4) functional deficiency (U.S. EPA, 1991).

3. HAZARD CHARACTERIZATION FOR REPRODUCTIVE TOXICANTS

Identification and characterization of reproductive hazards can be based on data from either human or experimental animal studies. Such data can result from routine or accidental environmental or occupational exposures or, for experimental animals, controlled experimental exposures. A hazard characterization should evaluate all of the information available and should:

- Identify the strengths and limitations of the database, including all available epidemiologic and experimental animal studies as well as pharmacokinetic and mechanistic information.
- Identify and describe key toxicological studies.
- Describe the type(s) of effects.
- Describe the nature of the effects (irreversible, reversible, transient, progressive, delayed, residual, or latent effects).
- Describe how much is known about how (through what biological mechanism) the agent produces adverse effects.
- Discuss the other health endpoints of concern.
- Discuss any nonpositive data in humans or experimental animals.
- Discuss the dose-response data (epidemiologic or experimental animal) available for further dose-response analysis.
- Discuss the route, level, timing, and duration of exposure in studies as compared to expected human exposures.
- Summarize the hazard characterization, including:
  - Major assumptions used,
  - Confidence in the conclusions,
  - Alternative conclusions also supported by the data,
  - Major uncertainties identified, and
  - Significant data gaps.

Conduct of a hazard characterization requires knowledge of the protocols in which data were produced and the endpoints that were evaluated. Sections 3.1 and 3.2 present 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. These are followed by presentation of many of the specific characteristics of human studies (Section 3.3) and limited discussions of pharmacokinetic and structure-activity factors (Sections 3.4 and 3.5).

3.1. LABORATORY TESTING PROTOCOLS

3.1.1. 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.

Integral parts of the hazard characterization and quantitative dose-response processes are the 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.

3.1.2. 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 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.

3.1.3. 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 impact to sperm production is necessary before an adverse effect on fertility can be detected.

3.1.4. Number of Females Mated to Each Male

The 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. EPA, 1982, 1985b, 1996a). However, in some tests that were not designed to conform to EPA test guidelines (OECD, 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.

3.1.5. 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. EPA, 1982, 1985b, 1996a; Galbraith et al., 1983; OECD, 1983).

The single-generation reproduction test evaluates effects of subchronic exposure of peripubertal and adult animals. In the multigeneration reproduction protocol, F_1 and F_2 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 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 and increase the likelihood of detecting histologic lesions. 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_1, F_2). Only the P generation is mated in a single-generation test, while both the P and F_1 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 can be allowed for up to 3 weeks (U.S. EPA, 1982, 1985b), 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_1 male and female offspring continue to be exposed after weaning (day 21) and through the mating period. Treatment of mated F_1 females is continued throughout gestation and lactation. More than one litter may be produced from either P or F_1 animals. Depending on the route of exposure of lactating females, it is important to consider that offspring may be exposed 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, as well as via the milk.

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 from males, and histopathology of the vagina, uterus, cervix, ovaries, and mammary glands from females. Uterine and ovarian weights also are often available. Male and female mating and fertility indices (Section 3.2.2.1) 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 and 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. For example, adverse transgenerational effects on reproductive system development by agents that disrupt endocrine control of sexual differentiation would be missed. 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. These limitations are shared with the shorter-term "screening" protocols described below. Because of these limitations, a comprehensive reproductive risk assessment should include results from a two-generation test or its equivalent. A further recommendation from the workshop was to include sperm analyses and estrous cycle normality as endpoints in reproductive effects studies. These endpoints have been included in the proposed revisions to the EPA test guideline (U.S. EPA, 1996a).

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 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 decreased with succeeding generations. When a decrease occurs, one explanation may be that the animals in the F_1 and F_2 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.

3.1.6. Alternative Reproductive Tests

A number of alternative test designs have appeared in the literature (Lamb, 1985; Lamb and 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 and Kimmel, 1988).

A continuous breeding protocol, Fertility (or Reproductive) Assessment by Continuous Breeding (FACB or RACB), has been developed by the National Toxicology Program (NTP) (Lamb and Chapin, 1985; Morrissey et al., 1989; Gulati et al., 1991). As originally described, this protocol (FACB) was a one-generation test. However, in the current design (RACB), dosing is extended into the F_1 generation to make it compatible with the EPA workshop recommendations for a two-generation design (Francis and 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 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 and then discarded. 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 improves the 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 were allowed to grow and reproduce (as they are routinely 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 and their utility has been examined using model compounds in the mouse (Morrissey et al., 1989).

Another test method 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_1 generation is mated in a continuous breeding 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_1 generations at all doses, and terminal studies on both generations include comprehensive assessment of sperm measures (number, 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 or antiandrogens.

3.1.7. 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 protocol (Scala et al., 1992; Tanaka et al., 1992; OECD, 1993a), a tripartite protocol developed by the International Conference on Harmonization (International Conference on Harmonization of Technical Requirements of Pharmaceuticals for Human Use, 1994; Manson, 1994), and the NTP's Short-Term Reproductive and Developmental Toxicity Screen (Harris, M.W. 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 multigeneration reproduction protocols.

The male *dominant lethal* test was designed to detect mutagenic effects in the male spermatogenic process that are lethal to the offspring. A female dominant lethal protocol has also been used to detect equivalent effects on oogenesis (Generoso and Piegorsch, 1993).

A review of the male dominant lethal 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 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. Dose levels used with the acute protocol are usually higher than those used with the subchronic protocol. Females are monitored for evidence of mating, killed at approximately midgestation, and examined for incidence of pre- and postimplantation loss (see Section 3.2.2 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. EPA, 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 6-8 weeks of age and is continued for 90 days (U.S. EPA, 1982, 1985b). Initiation of exposure at 8 weeks of age (compared with 6) 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 females, 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, data from 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 pathologic conditions is often increased in the reproductive tracts of aged control animals. Therefore, findings should be interpreted carefully.

3.2. ENDPOINTS FOR EVALUATING MALE AND FEMALE REPRODUCTIVE TOXICITY IN TEST SPECIES

3.2.1. 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, secondary 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 can have a contributing role if both partners are exposed. Thus, exposure of 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, related endpoints that can be used with humans are identified. Information that is specific for evaluation of effects on humans is presented in Section 3.3.

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

3.2.2. 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 only 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.

3.2.2.1. *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 2. Many endpoints that are pertinent for developmental toxicity are also listed and discussed in the Agency's *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991). Also included in Table 2 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.

Some of the endpoints identified above are used to calculate ratios or indices (NRCl, 1977; Collins, 1978; Schwetz et al., 1980; U.S. EPA, 1982, 1985b; Dixon and 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 3.

Table 2. Couple-mediated endpoints of reproductive toxicity

Multigeneration studies	**Other reproductive endpoints**
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
Offspring gender* (sex ratio)	variations*
Birth weight*	Postnatal structural and
Postnatal weights*	functional development*
Offspring survival*	
External malformations and variations*	
Offspring reproduction*	

*Endpoints that can be obtained with humans.

Table 3. 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 often is 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 (PREGNANCY) 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)

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

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

Table 3. Selected indices that may be calculated from endpoints of reproductive toxicity in test species (continued)

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

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

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.

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. Therefore, mating rate and fertility data from the first estrous cycle after initiation of cohabitation should be more discriminating than measurements involving multiple cycles. 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 after the start of cohabitation (*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 the start of cohabitation. If the stage of the estrous cycle at the time of cohabitation is known, the component of the variance due to variation in stage at cohabitation can be removed in the data 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 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 3.1.3 and 3.1.4).

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 increased 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 and Perreault, 1990). Preimplantation loss (described below) occurs in untreated as well as treated rodents and contributes to the normal 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 can 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 oolemma, 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 oocyte maturation and normal zygote formation after sperm-egg fusion by affecting meiosis (Perreault et al., 1992; Zuelke and Perreault, 1995). 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 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 Supplementary 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 4) 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 4).

Both the blastocyst and the uterus must be ready for implantation, and their synchronous development is critical (Cummings and 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 or inhibit protein synthesis, mitosis, or cell differentiation can block implantation and cause embryo death.

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 constitutes a maternal health threat for humans as well as test species. 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. Postimplantation loss can be determined also by examining uteri from postparturient females. *Preimplantation loss* is the (number of corpora lutea minus number of implantation sites)/number of corpora lutea. *Postimplantation loss*, determined following delivery of a litter, is the (total number of implantation sites minus number of full-term pups)/number of implantation sites.

Offspring gender in mammals 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. Further, alteration of the external sexual characteristics of offspring by agents that disrupt sexual development may produce apparent effects on sex ratios. Although not examined routinely, these factors provide the most likely explanations for alterations in the sex ratio.

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 (Tyl, 1987). Individual pups in large litters tend to be smaller than 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. Multivariate analyses may be used to adjust pup weights for litter size (e.g., analysis of covariance, multiple regression). 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* and the *Proposed Guidelines for Neurotoxicity Risk Assessment* (U.S. EPA, 1991, 1995a). These documents should be consulted for additional information on those parameters.

3.2.2.1.1. ***Adverse effects.*** Table 2 lists couple-mediated endpoints that may be measured in reproduction studies. Table 3 presents examples of indices that may be calculated from couple-mediated reproductive toxicity data. Significant detrimental effects on any of those endpoints or on indices derived from those data 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.

3.2.2.2. *Sexual Behavior*

Sexual behavior reflects complex neural, endocrine, and reproductive organ interactions and is therefore susceptible to disruption by a variety of toxic agents 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 studies on effects of drugs on sexual behavior or 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 suggests 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 and 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 and 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 that can be reflected in reduced numbers of females with vaginal plugs or vaginal sperm, alterations in lordosis behavior, and increased time to mating after start of cohabitation. In the male, measures include latency periods to first mount, mount with intromission, and first ejaculation, number of mounts with intromission to ejaculation, and the postejaculatory 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 central or peripheral neurotoxicity. 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.

3.2.2.2.1. ***Adverse effects.*** Effects on sexual behavior (within the limited definition of these Guidelines) should be considered as adverse reproductive effects. Included is evidence of impaired sexual receptivity and copulatory behavior. Impairment that is secondary to more generalized physical debilitation (e.g., impaired rear leg motor activity or general lethargy) should not be considered an adverse reproductive effect, although such conditions represent adverse systemic effects.

3.2.3. Male-Specific Endpoints

3.2.3.1. *Introduction*

The following sections (3.2.3 and 3.2.4) 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.

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 4.

3.2.3.2. *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 3.2.2). Depression in body weight or reduction in weight gain may reflect a variety of responses, including rejection of chemical-containing food or water because of reduced palatability, treatment-induced anorexia, or systemic toxicity. Less than severe reductions in adult body weight induced by restricted nutrition have shown 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 nonreproductive 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 or other severe systemic debilitation, it

should be noted that an adverse effect on a reproductive endpoint occurred, but the effect may have resulted from a more generalized toxic effect. Regardless, adverse effects would have been 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

Table 4. Male-specific endpoints of reproductive toxicity

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

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

reported relative to brain weight since, subsequent to development, the weight of the brain usually remains quite stable (Stevens and 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 absolute 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; Ku et al., 1993). This contradiction may arise from several factors, including a delay before cell deaths are reflected in a weight decrease (due to preceding 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, while testis weight measurements may not reflect certain adverse testicular effects and do not indicate the nature of an effect, 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. 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, with or without their secretory fluids. Differential loss of secretory fluids prior to weighing could produce artifactual weights. 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.

3.2.3.2.1. ***Adverse effects.*** 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.

3.2.3.3. *Histopathologic Evaluations*

Histopathologic evaluations of test animal tissues 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 from a variety of protocols; and (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 and 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 or degenerating, if multinucleated giant cells are obvious, or if sloughed cells are present in the tubule lumen. More subtle lesions, such as retained spermatids or missing germ cell types, 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 book is available (Russell et al., 1990) which 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 and Rousseaux, 1991). Compared with the testes, less is known about structural 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. Evaluation should include information on the caput, corpus, and cauda segments of the epididymis. Presence of debris and sloughed cells in the epididymal lumen are valuable indicators of damage to the germinal epithelium or the excurrent ducts. The presence of lesions such as sperm granulomas, leucocyte infiltration (inflammation) or absence of clear cells in the cauda epididymal epithelium should be noted. 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 standardized tissue preparation techniques and a standardized quantification system, the evaluation of histopathologic data would be facilitated by the presentation of the evaluation criteria and procedure by 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).

3.2.3.3.1. ***Adverse effects.*** Significant and biologically meaningful 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 evaluations 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.

3.2.3.4. *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 testicular 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 in humans. Brief descriptions of these measures are provided below, followed by a discussion of the use of various sperm measures in male reproductive risk assessment.

3.2.3.4.1. ***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 (Seed et al., 1996). 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.

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 variation 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 when allowed by protocol), 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, but the samples for sperm motility and morphology may be derived also from the vas deferens. 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 and 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 and van Beek, 1993). If evaluation was conducted when the effect of a lesion would be reflected adequately in the spermatid count, then 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 the spermatid count. Also, spermatid counts reported from some laboratories have large coefficients of variation that may reduce the statistical power and thus the usefulness of that measure.

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, vas deferens, or ejaculated samples.

3.2.3.4.2. ***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. Thus, the link between sperm morphology and mutagenicity is not necessarily sensitive or specific.

An increase in abnormal sperm morphology has been considered evidence that the agent has gained access to the germ cells (U.S. EPA, 1986c). Exposure of males to toxic agents may lead to sperm abnormalities in their progeny (Wyrobek and Bruce, 1978; Hugenholtz and Bruce, 1983; Morrissey et al., 1988a,b). 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, midpiece, and tail defects in either stained preparations by bright field microscopy (Filler, 1993) or fixed, unstained preparations by phase contrast microscopy (Linder et al., 1992; Seed et al., 1996). Such an approach may be adequate for mice 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 categorize the types of abnormalities observed and quantify 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 in 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 et al., 1994). 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 and 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.

3.2.3.4.3. ***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 (Morrissey et al., 1988a,b; Toth et al., 1989b, 1991b), and sperm motility assessments are an integral part of some reproductive toxicity tests (Gray et al., 1988; Morrissey et al., 1989; U.S. EPA, 1996a).

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 experimental variables, including abstinence interval, method of sample collection and handling, elapsed time between sampling and observation, the temperature at which the sample is stored and analyzed, the extent of sperm dilution, the nature of the dilution medium, and the microscopic chamber employed for the observations (Slott et al., 1991; Toth et al., 1991a; Chapin et al., 1992; Schrader et al., 1992; Weir and Rumberger, 1995; Seed et al., 1996).

Sperm motility can be evaluated in fresh samples under phase contrast microscopy, or sperm images can be recorded and stored in video or digital format and analyzed later, either manually or by computer-aided semen analysis (Linder et al., 1986; Boyers et al., 1989; Toth et al., 1989a; Yeung et al., 1992; Slott and Perreault, 1993). For manual assessments, the percentage of motile and progressively motile sperm can be estimated and a simple scale used to describe the vigor of the sperm motion.

The recent application of video and/or digital technology to sperm analysis allows a more detailed evaluation of sperm motion including information about the individual sperm tracks. It also provides permanent storage of the sperm tracks which can be reanalyzed as necessary (manually or computer-assisted). With computer-assisted technology, information about sperm velocity (straight-line and curvilinear) as well as the amplitude and frequency of the track are obtained rapidly and efficiently on large numbers of sperm. Using this technology, chemically induced alterations in sperm motion have been detected (Toth et al., 1989a, 1992; Slott et al., 1990; Klinefelter et al., 1994a), and such changes have been related to the fertility of the exposed animals (Toth et al., 1991a; Oberlander et al., 1994; Slott et al., 1995). These preliminary studies indicate that significant reductions in sperm velocity are associated with infertility, even when the percentage of motile sperm is not affected. The ability to distinguish between the proportion of sperm showing any type of motion and those with progressive motility is important (Seed et al., 1996).

Changes in endpoints that measure effects on spermatogenesis and sperm maturation have been related to fertility in several test species, but the ability to predict infertility from these data (in the

absence of fertility data) is not reliable. This is in part due to the observation, in both test species and humans, that fertility is dependent not only on having adequate numbers of sperm, but also on the degree to which those sperm are normal. If sperm quality is high, then sperm number must be substantially reduced before fertility is affected. For example, in a rat model that employs artificial insemination of differing numbers of good quality sperm, sperm numbers can be reduced substantially before fertility is affected (Klinefelter et al., 1994b). In humans, the distribution of sperm counts for fertile and infertile men overlap, with the mean for fertile men being higher (Meistrich and Brown, 1983), but fertility is likely to be impaired when counts drop below 20 million/mL (WHO, 1992). Similarly, if sperm numbers are normal in rodents, a relatively large effect on sperm motility is required before fertility is affected. For example, rodent sperm velocity must be substantially reduced, in the presence of adequate numbers of sperm, before fertility is affected (Toth et al., 1991a; Slott et al., 1995). These models also show that relatively modest changes in sperm numbers or quality may not cause infertility, but can nevertheless be predictive of infertility. On the other hand, fertility may be impaired by smaller decrements in both number and motility (or other qualitative characteristics).

Thus, the process of reproductive risk assessment is facilitated by having information on a variety of sperm measures and reproductive organ histopathology in addition to fertility. Specific information about reproductive organ and gamete function can then be used to evaluate the occurrence and extent of injury, and the probable site of toxicity in the reproductive system. The more information that is available from supplementary endpoints, the more the risk assessment can be based on science rather than uncertainty.

3.2.3.4.4. ***Adverse effects.*** Human male fertility is generally lower than that of test species and may be more susceptible to damage from toxic agents (see Supplementary Information). Therefore, the conservative approach should be taken that, within the limits indicated in the sections on those parameters, statistically significant changes in measures of sperm count, morphology, or motility as well as number of normal sperm should be considered adverse effects.

3.2.3.5. *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 (Davis et al., 1992; Colie, 1993; Savitz et al., 1994; Qiu et al., 1995). Paternally mediated effects include pre-and postimplantation loss, growth and

behavioral deficits, and malformations. A large proportion of the chemicals reported to cause paternally mediated effects have genotoxic activity, and are considered to exert this effect via transmissible genetic alterations. Low doses of cyclophosphamide have resulted in induction of single strand DNA breaks during rat spermatogenesis which, due in part to absence of subsequent DNA repair capability, remain at fertilization (Qiu et al., 1995). The results of such damage have been observed in the F_2 generation offspring (Hales et al., 1992). Other mechanisms of induction of paternally mediated effects are also possible. 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 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 or in humans as causing a paternally mediated adverse effect on offspring, the effect should be considered an adverse reproductive effect.

3.2.4. Female-Specific Endpoints

3.2.4.1. *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 et al., 1994). 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 amount of information obtained routinely to detect toxic effects on the female reproductive system has been 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.

Adverse 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 and Jankowiak, 1972; Barsotti et al., 1979; Sonawane and Yaffe, 1983; Cummings and 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 agent's effect(s). Areas that may be examined in evaluations of the female reproductive system are listed in Table 5.

Reproductive function in the female is controlled through complex interactions involving the central nervous system (particularly the hypothalamus), pituitary, ovaries, the reproductive tract, and the secondary sexual organs. Other nongonadotrophic 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 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 used below 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.

3.2.4.2. *Body Weight, Organ Weight, Organ Morphology, and Histology*

3.2.4.2.1. ***Body weight***. Toxicologists are often concerned about how a change in body weight may affect reproductive function. In females, an important consideration is that body weight fluctuates 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 (Galletti and Klopper, 1964; Hervey and Hervey, 1967). Food consumption is elevated during pregnancy, in part

Table 5. Female-specific endpoints of reproductive toxicity

Organ weights	Ovary, uterus, vagina, pituitary
Visual examination and histopathology	Ovary, uterus, vagina, pituitary, oviduct, mammary gland
Estrous (menstrual*) cycle normality	Vaginal smear cytology
Sexual behavior	Lordosis, time to mating, vaginal plugs, or sperm
Hormone levels*	LH, FSH, estrogen, progesterone, prolactin
Lactation*	Offspring growth, milk quantity and quality
Development	Normality of external genitalia*, vaginal opening, vaginal smear cytology, onset of estrous behavior (menstruation*)
Senescence	Vaginal smear cytology, ovarian histology (menopause*)

*Endpoints that can be obtained relatively noninvasively with humans.

because of the elevated serum progesterone level. One of the most sensitive noninvasive 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 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% and a marginal effect at 90%. 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.

3.2.4.2.2. ***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.

3.2.4.2.2.1. ***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.

3.2.4.2.2.2. ***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 toxicity. A number of pathologic conditions can be detected by ovarian histology (Kurman and Norris, 1978; Langley and 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.

3.2.4.2.2.3. ***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

3.2.4.2.3. ***Uterus***

3.2.4.2.3.1. ***Uterine weight.*** An alteration in the weight of the uterus may be considered an indication of female reproductive organ toxicity. Compounds that inhibit steroidogenesis and cyclicity can dramatically reduce the weight of the uterus so that it appears atrophic and small. However, uterine weight fluctuates three- to fourfold 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 fluctuations in weight, uterine weights taken from cycling animals have a high variance, and large compound-related effects are required to demonstrate a significant effect 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, alkylphenols, and phytoestrogens) possess varying degrees of estrogenic activity and have the potential to stimulate the female reproductive tract (Barlow and Sullivan, 1982; Bulger and 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.

3.2.4.2.3.2. ***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), and extended treatment with these compounds leads to hypertrophy and hyperplasia. Conversely, inhibition of ovarian activity and reduced steroid secretion results in endometrial hypoplasia and atrophy, as well as altered vaginal smear cytology. Effects induced during development may delay or prevent puberty, resulting in persistence of infantile genitalia.

3.2.4.2.3.3. ***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

3.2.4.2.4. ***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 in detecting morphologic anomalies. Descriptions of pathologic effects within the oviducts of animals other than humans are not common. Hypoplasia of otherwise well-formed oviducts and loss of cilia result 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.

3.2.4.2.5. ***Vagina and external genitalia***

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

3.2.4.2.5.2. ***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 3.2.4.4). 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.

3.2.4.2.5.3. ***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.

The sex ratio observed at birth may be affected by exposure of genotypic females in utero to agents that disrupt reproductive tract development. In cases of incomplete sex reversal because of such exposures, female rodents may appear more male-like and have an increased ano-genital distance (Gray and Ostby, 1995).

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 occurs without truly affecting puberty.

3.2.4.2.5.4. ***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 ano-genital distance
- Vaginal hypoplasia or aplasia
- Altered timing of vaginal opening
- Abnormal vaginal smear cytology pattern

3.2.4.2.6. ***Pituitary***

3.2.4.2.6.1. ***Pituitary weight.*** Alterations in weight of the pituitary gland should be considered an adverse effect. The discussion on pituitary weight and histology for males (see Section 3.2.3.2) 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).

3.2.4.2.6.2. ***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 and 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.

3.2.4.2.6.3. ***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.

3.2.4.3. ***Oocyte Production***

3.2.4.3.1. ***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. If the population of follicles is depleted, it cannot be replaced and the female will be rendered infertile. In humans, depletion of oocytes leads to premature menopause. Ovarian follicle biology and toxicology have been reviewed by Crisp (1992).

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

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

3.2.4.3.2. ***Ovulation.*** Chemicals can delay or block ovulation by disrupting the ovulatory surge of luteinizing hormone (LH) or by interfering with the ability of the maturing follicle to respond to that gonadotropic signal. Examples for rats include compounds that interfere with normal central nervous

system (CNS) norepinephrine receptor stimulation such as the pesticides chlordimeform and amitraz (Goldman et al., 1990, 1991) and compounds that interfere with norepinephrine synthesis such as the fungicide thiram (Stoker et al., 1993). 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 and Butcher, 1966; Butcher and Fugo, 1967; Butcher et al., 1969, 1975; Na et al., 1985). Delayed ovulation induced by exposure to the pesticide chlordimeform has also been shown to alter fetal development and pregnancy outcome in rats (Cooper et al., 1994).

3.2.4.3.3. ***Corpus luteum.*** The corpus luteum arises from the ruptured follicle and secretes progesterone, which has an important role in the estrous or menstrual cycle. Luteal progesterone is also required for the maintenance of early pregnancy in most mammalian species, including humans (Csapo and 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.

3.2.4.3.3.1. ***Adverse effects.*** Increased rates of follicular atresia and oocyte toxicity leads to premature menopause in humans. Altered follicular development, ovulation failure, or altered corpus luteum formation and function can result in disruption of cyclicity and reduced fertility, and, in nonprimates, interference with normal sexual behavior. Therefore, significant increases in the rate of follicular atresia, evidence of oocyte toxicity, interference with ovulation, or altered corpus luteum formation or function should be considered adverse effects.

3.2.4.4. *Alterations in the Female Reproductive Cycle*

The pattern of events in the estrous cycle may provide 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 and 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.A. et al., 1995). However, practical limitations in testing may limit the examination to the period before mating or necropsy.

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 the 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 and 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. Even normal, control animals can show irregular cycles. However, a significant alteration compared with controls in the interval between occurrence of estrus for a treatment group is cause for concern. Generally, the cycle will be lengthened or the animals will become acyclic. Lengthening of the cycle may be a result of increased duration of either estrus or diestrus. 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 and 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). Persistent or constant vaginal cornification (or vaginal estrus) may result from one or several effects. Typically, in the adult, if the vaginal epithelium becomes cornified and remains so in response to toxicant exposure, it is the result of the agent's estrogenic properties (i.e., DES or methoxychlor), or the ability of the agent to block ovulation. In the latter case, the follicle persists and endogenous estrogen levels bring about the persistent vaginal cornification. Histologically, the ovaries in persistent estrus will be atrophied following exposure to estrogenic substances. In contrast, the ovaries of females in which ovulation has been blocked because of altered gonadotropin secretion will contain several large follicles and no corpora

lutea. Females in constant estrus may be sexually receptive regardless of the mechanism responsible for this altered ovarian condition. However, if ovulation has been blocked by the treatment, an LH surge may be induced by mating (Brown-Grant et al., 1973; Smith, E.R. and Davidson, 1974) and a pregnancy or pseudopregnancy may ensue. The fertility of such matings is reduced (Cooper et al., 1994). Significant delays in ovulation can result in increased embryonic abnormalities and pregnancy loss (Fugo and Butcher, 1966; Cooper et al., 1994).

Persistent diestrus indicates temporary or permanent cessation of follicular development and ovulation, and thus at least temporary 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 and Nightingale, 1980) or agents such as atrazine that interrupt gonadotropin support of the ovary (Cooper et al., 1996). Pseudopregnancy is another altered endocrine state reflected by persistent diestrus. A pseudopregnant condition also has been shown to result in rats following single or multiple doses of atrazine (Cooper et al., 1996). The ovaries of anestrous females are atrophic, with few primary follicles and an unstimulated uterus (Huang and Meites, 1975). Serum estradiol and progesterone are abnormally low.

3.2.4.4.1. ***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.

3.2.4.5. *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 (Wolff, 1993; Imagawa et al., 1994; Tucker, 1994).

Mammary gland size, milk production and release, and histology can be affected adversely by toxic agents, and many exogenous chemicals and drugs are transferred into milk (American Academy of Pediatrics Committee on Drugs, 1994; Oskarsson et al., 1995; Sonawane, 1995). Reduced growth of young could be caused by reduced milk availability, palatability or quality, by ingestion of a toxic agent secreted into the milk, or by other factors unrelated to lactational ability (e.g., deficient suckling ability or deficient maternal behavior). 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 agents.

Techniques for measuring mammary tissue development, nucleic acid content, milk production and milk composition in rodents are discussed by Tucker (1994). During lactation, the mammary glands can be dissected and weighed only with difficulty. RNA content of the mammary glands may be measured as an index of lactational potential. More direct estimates of milk production may be obtained by measuring litter weights of milk-deprived pups taken before and after nursing. Milk from the stomachs of pups treated similarly can also be weighed at necropsy. Cleared and stained whole mounts of the mammary gland can be prepared at necropsy for histologic examination. 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.

3.2.4.6. *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 and Walker, 1979). Although the mechanisms responsible for this loss of cycling are not thoroughly understood, age-dependent changes occur within the hypothalamic-pituitary control of ovulation (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 and 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.

3.2.4.6.1. ***Adverse effects.*** Significant effects on measures showing a decrease in the age of onset of reproductive senescence in females should be considered adverse. Cessation of normal cycling, which

is measured by vaginal smear cytology, ovarian histopathology, or an endocrine profile that is consistent with this interpretation, should be included as an adverse effect.

3.2.5. Developmental and Pubertal Alterations

3.2.5.1. *Developmental Effects*

Alterations of reproductive differentiation and development, including those produced by endocrine system disruption, can result in infertility, functional and morphologic alterations of the reproductive system, and cancer (Steinberger and 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 toxic agents in utero, through contact with exposed dams, or in milk. Dosing of dams during lactation also can result in developmental effects through impaired nursing capability of the dams.

Effects observed in rodents following developmental exposure to agents can include alterations in the genitalia (including ano-genital distance), inhibited (female) or retained (male) nipple development, impaired sexual behavior, delay or acceleration of the onset of puberty, and reduced fertility (Gray et al., 1985, 1994, 1995; Gray and Ostby, 1995; Kelce et al., 1995). Effects may include altered sexual behavior or ability to produce gametes normally 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. Generally, the type of effect seen may differ depending on the stage of development at which the exposure occurred.

Many of these effects have been detected in human females and males exposed prenatally to diethylstilbestrol (DES), other estrogens, progestins, androgens, and anti-androgens (Giusti et al., 1995; Harrison et al., 1995). Accelerated reproductive aging and tumors of the reproductive tract have been observed in laboratory animal and human females after pre- or perinatal exposure to hormonally active agents. However, capability to alter sexual differentiation is not limited to agents with known direct hormonal activity. Other agents, for which the mode of action is not known (e.g., busulfan, nitrofen), or which affect the endocrine system indirectly (e.g., PCBs, dioxin), may act via different mechanisms during critical periods of development to alter sexual differentiation and reproductive system development.

3.2.5.2. *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 an increase in the blood level of estradiol. The ages and weights of females at the first cornified (estrous) vaginal smear, the first diestrous smear, and the onset of vaginal cycles have also been used as endpoints for onset 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 (Gray et al., 1989, 1995; Gray and Ostby, 1995; Kelce et al., 1995). 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).

3.2.5.3. *Adverse Effects*

Effects induced or observed during the pre- or perinatal period should be judged using guidance from the *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991) as well as from these Guidelines. Significant effects on ano-genital distance or age at puberty, either early or delayed, should be considered adverse as should malformations of the internal or external genitalia. Included as adverse effects for females should be effects on nipple development, age at vaginal opening, onset of cyclic vaginal smears, onset of estrus or 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.

3.2.6. Endocrine Evaluations

Toxic agents can alter endocrine system function by affecting any part of the hypothalamic-pituitary-gonadal-reproductive tract axis. Effects may be induced in either sex by altering hormone synthesis, storage, release, transport, or clearance, as well as by altering hormone receptor recognition or postreceptor responses. The involvement of the endocrine system in female reproductive physiology

and toxicology has been presented to a substantial degree as a necessary component in Section 3.2.4 (Female-Specific Endpoints). The information in that section should be considered together with the following material.

The male reproductive system can be affected adversely by disruption of the normal endocrine balance. In adults, effects that result in interference with normal concentrations or action of LH and/or follicle stimulating hormone (FSH) can decrease or abolish spermatogenesis, affect secondary sex organ (e.g., epididymis) and accessory sex gland (e.g., prostate, seminal vesicle) function, and impair sexual behavior (Sharpe, 1994). In mammals, a female reproductive tract develops unless androgen is produced and utilized normally by the fetus (Byskov and Hoyer, 1994; George and Wilson, 1994). Therefore, the consequences of disruption of the normal endocrine pattern during development of the male reproductive system pre- and postnatally are of particular concern. Differentiation and development of the male reproductive system are especially sensitive to substances that interfere with the production or action of androgens (testosterone and dihydrotestosterone). Sexual differentiation of the CNS can be affected also. Therefore, interference with normal production or response to androgens can result in a range of abnormal effects in genotypic males ranging from a pseudohermaphrodite condition to reduction in sperm production or altered sexual behavior. Chemicals with estrogenic or anti-androgenic activity have been identified that are capable, with sufficient exposure levels, of causing effects of these types in males (Gray et al., 1994; Harrison et al., 1995; Kelce et al., 1995). While sensitivity may differ, it is likely that mechanisms of action for these endocrine disrupting agents will be consistent across mammalian species. Chemicals with the ability to interact with the Ah receptor (e.g., dioxin or PCBs) may also disrupt reproductive system development or function (Brouwer et al., 1995; Safe, 1995). Several of the effects seen with exposure of male and female rats and hamsters differ from those caused by estrogens, indicating a different mechanism of action.

The developing nervous system can be a target of chemicals. In rats, sexual differentiation of the CNS can be modified by hormonal treatments or exposure to environmental agents that mimic or interfere with the action of certain hormones. Prior to gender differentiation, the brain is inherently female or at least bipotential (Gorski, 1986). Thus, the functional and structural sex differences in the CNS are not due directly to sex differences in neuronal genomic expression, but rather are imprinted by the gonadal steroid environment during development.

Chemicals with endocrine activity have been shown to masculinize the CNS of female rats. Examples include chlordecone (Gellert, 1978), DDT (Bulger and Kupfer, 1985), and methoxychlor (Gray et al., 1989). Exposure of newborn female rats to these agents during the critical period of

sexual differentiation can alter the timing of puberty and perturb subsequent reproductive function, presumably by altering the development of the neural mechanisms that regulate gonadotropin secretion.

In females, the situation is more complex than in males due to the female cycle, the fertilization process, gestation and lactation. All of the functions of the female reproductive system are under endocrine control, and therefore can be susceptible to disruption by effects on the reproductive endocrine system.

As with males, disturbance of the normal endocrine patterns during development can result in abnormal development of the female reproductive tract at exposure levels that tend to be lower than those affecting adult females (Gellert, 1978; Brouwer et al., 1995). Consistent with the differentiation mechanism described above, exposure of genotypic females to androgens causes formation of pseudohermaphrodite reproductive tracts with varying degrees of severity as well as alteration of brain imprinting. However, exposure to estrogenic substances during development also results in adverse effects on anatomy and function including, in rats, malformations of the genitalia. Exposure of human females to diethylstilbestrol in utero has been shown to cause an increased incidence of vaginal clear cell adenoma (Giusti et al., 1995). Dioxin, presumably acting through the Ah receptor, also disrupts development of the female reproductive system (Gray and Ostby, 1995).

Endpoints can be included in standardized toxicity testing that are capable of detecting, but are not specific for, effects of reproductive endocrine system disruption. For effects of exposure on adults, endpoints can be incorporated into the subchronic toxicity protocol or into reproductive toxicity protocols. For effects that are induced during development, protocols that include exposure throughout the development process and allow evaluation of the offspring postpubertally are needed. Data from specialized testing, including in vitro screening tests, may be useful to evaluate further the site, timing, and mechanism of action.

Endpoints that can detect endocrine-related effects with adult-only exposure in standardized testing include evaluation of fertility, reproductive organ appearance, weights, and histopathology, oocyte number, cycle normality and mating behavior. Endpoints that can detect effects induced by endocrine system disruption during development include, in addition to those identified for adult-exposed animals, the reproductive developmental endpoints identified in Section 3.2.5. Significant effects on any of these measures may be considered to be adverse if the results are consistent and biologically plausible.

Levels of the reproductive hormones are not available routinely from toxicity testing. However, measurements of the reproductive hormones in males offer useful supplemental information in assessing potential reproductive toxicity for test species (Sever and Hessol, 1984; Heywood and James, 1985;

NRC, 1989). Such measurements have increased importance with humans where invasiveness of approaches must be limited. The reproductive hormones measured often are circulating levels of LH, 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.

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 a 24-hour period (Fink, 1988). Thus, the time of sampling should be controlled rigorously to avoid excessive variability (Nett, 1989). Sequential sampling can allow detection of treatment-related changes in circadian and pulsatile rhythms.

The pattern seen in levels of reproductive system hormones can provide useful information about the possible site and type of effect on reproductive system function. For example, if a compound acts at the level of 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. An agent having antiandrogenic activity in adults might elevate serum LH and testosterone. Testis weight might be unaffected, while the weight and size of the accessory sex glands may be reduced. The endocrine profile presented by exposure to specific antiandrogens can differ markedly because of differences in tissue specificity and receptor kinetics, as well as age at which exposure occurred.

3.2.6.1. *Adverse Effects*

In the absence of endocrine data, significant effects on reproductive system anatomy, sexual behavior, pituitary, uterine or accessory sex gland weights or histopathology, female cycle normality, or 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 estrogen, progesterone, testosterone, prolactin, 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, occur in concert with, or result from alterations in gametogenesis, gamete 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, LOAEL, or benchmark dose. Furthermore, endocrine data may facilitate

identification of sites or mechanisms of toxicant action, especially when obtained after short-term exposures.

3.2.7. In Vitro Tests of Reproductive Function

Numerous in vitro tests are available and under development to measure or detect chemically induced changes in various aspects of both male and female reproductive systems (Kimmel, G.L. et al., 1995). These include in vitro fertilization using isolated gametes, whole organ (e.g., testis, ovary) perfusion, culture of isolated cells from the reproductive organs (e.g., Leydig cells, Sertoli cells, granulosa cells, oviductal or epididymal epithelium), co-culture of several populations of isolated cells, ovaries, quarter testes, seminiferous tubule segments, various receptor binding assays on reproductive cells and transfected cell lines, and others.

Tests of sperm properties and function that have been applied to reproductive toxicology include penetration of sperm through viscous medium (Yeung et al., 1992), in vitro capacitation and fertilization assays (Holloway et al., 1990a,b; Perreault and Jeffay, 1993; Slott et al., 1995), and evaluation of sperm nuclear integrity (Darney, 1991). In addition, evaluation of human sperm function may include sperm penetration of cervical mucus, ability of sperm to undergo an acrosome reaction, and ability to penetrate zona pellucida-free hamster oocytes or bind to human hemi-zona pellucidae (Franken et al., 1990; Liu and Baker, 1992).

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).

3.3. HUMAN STUDIES

In principle, human data are scientifically preferable for risk assessment since test animal to human extrapolation is not required. At this time, reproductive data for humans are available for only a limited number of toxicants. Many of these are from occupational settings in which exposures tend to be higher than in environmental settings. As more data become available, expanding the number of agents and endpoints studied and improving exposure assessment, more risk assessments will include these data. The following describes the methods of generation and evaluation of human data and the relative weight the various types of human data should be given in risk assessments.

"Human studies" include both epidemiologic studies and other reports of individual cases or clusters of events. Typical epidemiologic studies include (1) cohort studies in which groups are defined by exposure and health outcomes are examined; (2) case-referent studies in which groups are defined by health status and prior exposures are examined; (3) cross-sectional studies in which exposure and outcome are determined at the same time; and 4) ecologic studies in which exposure is presumed based typically on residence. Greatest weight should be given to carefully designed epidemiologic studies with more precise measures of exposure, because they can best evaluate exposure-response relationships. This assumes that human exposures occur in broad enough ranges for observable differences in response to occur. 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 for hazard characterization, but are of limited use for quantitative risk determination because of the generally broad categorical groupings of exposure. 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 test animal data, but cannot stand by themselves in risk assessments.

3.3.1. Epidemiologic Studies

Good epidemiologic studies provide valuable data for assessment of human risk. As there are many different designs for epidemiologic studies, simple rules for their evaluation do not exist. Risk assessors should seek the assistance of professionals trained in epidemiology when conducting a detailed analysis. The following is an overview of key issues to consider in evaluation for risk assessment of reproductive effects.

3.3.1.1. *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 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; (4) menstrual history; and (5) blood or urinary hormone measures. Factors requiring control in the design or analysis (such as effect modifiers and confounders, described below) 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 outcome in humans. Use of improved methods for identifying some outcomes, such as embryonic loss detected by more sensitive urinary hCG (human chorionic gonadotropin) assays, change the spectrum of outcomes available for study (Wilcox et al., 1985; Sweeney et al., 1988; Zinaman et al., 1996). Other, less accessible, endpoints may require invasive techniques to obtain samples (e.g., histopathology) or may have high intra- or interindividual variability (e.g., serum hormone levels, sperm count).

Demographic characteristics of the population, such as marital status, age, 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 progeny. Many studies, however, are limited to evaluating endpoints at a particular time in this continuum. For example, in a study of defects observed at live birth, a malformed stillbirth would not be included, even though the etiology could be identical (Bloom, 1981). Also, a different spectrum of outcomes could result from differences in timing or in level of exposure (Selevan and Lemasters, 1987).

3.3.1.1.1. ***Human reproductive endpoints.*** The following section discusses various human male and female reproductive endpoints. These outcomes may be an indicator of sub- or infertility. These are followed by a discussion of reproductive history studies.

3.3.1.1.1.1. ***Male endpoints - semen evaluations.*** The use of semen analysis was discussed in Section 3.2.3.4. Most epidemiologic studies of potential effects of agents on semen characteristics have been conducted in occupational groups and patients receiving drug therapy. Obtaining 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 about existing reproductive problems or 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. While some studies have response rates greater than 70% (Ratcliffe et al., 1987; Welch et

al., 1988), response rates are often less than 70% in such studies and may be even lower in the comparison group (Egnatz et al., 1980; Lipshultz et al., 1980; Milby and Whorton, 1980; Lantz et al., 1981; Meyer, 1981; Milby et al., 1981; Rosenberg et al., 1985; Ratcliffe et al., 1989). 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 and Whorton, 1980). Participation in the comparison group may be biased toward those with preexisting 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, recreational drugs, smoking). Data on these factors may be collected by interview, subject to the limitations described for pregnancy outcome studies.

Reports of studies with semen analyses have rarely included an evaluation of endocrine status (hormone levels in blood or urine) of exposed males (Lantz et al., 1981; Ratcliffe et al., 1989). Conversely, studies that have examined endocrine status typically do not have data on semen quality (Mason, 1990; McGregor and Mason, 1991; Egeland et al., 1994).

3.3.1.1.1.2. ***Female endpoints.*** Reproductive effects may result from a variety of exposures. For example, environmental exposures may be toxic to the oocyte, producing a loss of primary oocytes that irreversibly affects the woman's fecundity. The exposures of importance may occur during the prenatal period, and beyond. Oocyte depletion is difficult to examine directly in women because of the invasiveness of the tests required; however, it can be studied indirectly through evaluation of the age at reproductive senescence (menopause) (Everson et al., 1986).

Numerous diagnostic methods have been developed to evaluate female reproductive dysfunction. Although these methods have been used rarely 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 will 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 also 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, exercise level, 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 (before ovulation), can vary between individuals and may make it difficult to determine significant effects on length in populations of women (Burch et al., 1967; Treloar et al., 1967). Human 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 (Kesner et al., 1992). Menstrual dysfunction data have been used to examine adverse reproductive effects in women exposed to potentially toxic agents occupationally (Lemasters, 1992),

Reports of prospective clinical evaluations of menstrual function (Kesner et al., 1992; Wright et al., 1992), have shown urinary endocrine measures to be practical and useful. 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 changes in a cyclic pattern, single sample analysis does not provide adequate information for evaluating alterations in 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. Clinically abnormal levels of gonadotropins, steroids, or other biochemical parameters may be detected from a single sample. However, a much stronger design involves collection of multiple samples and their observation in conjunction with events in the menstrual cycle.

The day of 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 analyses of estradiol and gonadotropin status at midcycle. After ovulation, luteal phase function can be assessed by analysis of progesterone secretion and by evaluation of endometrial histology. Tubal patency, which could be affected by abnormal development, endometriosis or infection, is an endpoint that can be observed in clinical evaluations of reproductive function (Forsberg, 1981). These latter evaluations of endometrial histology and tubal patency are less likely to be present in epidemiologic studies or surveillance programs because of the invasiveness of the procedures.

3.3.1.2. *Reproductive History Studies*

3.3.1.2.1. ***Measures of fertility.*** Subfertility may be thought of as nonevents: a couple is unable to have children within a specific time frame. Therefore, the epidemiologic measurement of reduced fertility or fecundity is typically indirect and is accomplished by comparing birth rates or time intervals

between births or pregnancies. These outcomes have been examined using several methods: the Standardized Birth Ratio (SBR; also referred to as the Standardized Fertility Ratio) and the length of time to pregnancy or birth. In these evaluations, the couple's joint ability to procreate is estimated. The SBR compares the number of births observed to those expected based on the person-years of observation preferably stratified by factors such as time period, age, race, marital status, parity, and (if possible) 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 has similar limitations in interpretation (Gaffey, 1976; McMichael, 1976; Tsai and Wen, 1986). The SBR was found to be less sensitive in identifying an effect when compared to semen analyses (Welch et al., 1991). These data can also be analyzed using Poisson regression.

Analysis of the time between recognized pregnancies or live births is a more recent approach to indirect measurement of fertility (Dobbins et al., 1978; Baird and Wilcox, 1985; Baird et al., 1986; Weinberg and Gladen, 1986; Rowland et al., 1992). 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 obtained easily in epidemiology studies (see Section 3.3.1.4).

3.3.1.2.2. ***Developmental outcomes.*** Developmental 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). Developmental effects are discussed in more detail in the *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991). As mentioned above, epidemiologic studies that focus on only one type of developmental outcome or exposures to only one parent may miss a true effect of exposure.

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; Selevan and Lemasters, 1987). 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 and Lemasters, 1987) (for a more detailed description, see Section 3.3.1.4). Therefore, a risk assessment should, when possible, attempt to look at the relationship of different reproductive endpoints and patterns of exposure.

In addition to the above effects, exposure may produce genetic damage to germ cells. 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, germ cells or early zygotes are critical targets of potentially toxic agents. 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 and the large study groups needed. For more information and guidance on the evaluation of these data, refer to the *Guidelines for Mutagenicity Risk Assessment* (U.S. EPA, 1986c).

3.3.1.3. *Community Studies and Surveillance Programs*

Epidemiologic studies may be based on broad populations such as a community, a nationwide probability sample, or surveillance programs (such as birth defects registries). Some studies have examined the effects of environmental exposures such as potential toxic agents in outdoor air, food, water, and soil. These studies may assume certain exposures through these routes due to residence (ecologic studies). The link between environmental measurements and critical periods of exposure for a given reproductive effect may be difficult to make. Other studies may go into more detail, evaluating the above routes and also indoor air, house dust, and occupational exposures on an individual basis (Selevan, 1991). Such environmental studies, relating individual exposures to health outcomes should have less misclassification of exposure.

Exposure definition in community studies has some limitations in the assessment of exposure-effect relationships. For example, in many community-based studies, it may not be possible to distinguish maternally mediated effects from paternally mediated effects since both parents spend time in the same home environment. In addition, the presumably lower exposure levels (compared with industrial settings) may require very large groups for the 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; McDonald et al., 1989; Lindbohm et al., 1991), birth defects (Hemminki et al., 1980; Kwa and Fine, 1980; Papier, 1985), and childhood cancer (Fabia and Thuy, 1974; Hemminki et al., 1981; Peters et al., 1981; Gardner et al., 1990a,b). In these reports, jobs are classified typically into broad categories based on the probability of exposure to certain classes or levels of exposure. Such studies are most helpful in the identification of topics for

additional study. However, because of the broad groupings of types or levels of exposure, these studies are not typically useful for risk assessment of any one particular agent.

Surveillance programs may also exist in occupational settings. In this case, reproductive histories (including menstrual cycles) or semen evaluations could be followed to monitor reproductive effects of exposures. With adequate exposure information, these could yield very useful data for risk assessment. Reproductive histories tend to be easier and less costly to collect, whereas, a semen evaluation program would be rather costly. Success with such programs in the workplace will be determined by the confidence the worker has that reproductive data are kept confidential and will not affect employment status (Samuels, 1988; Lemasters and Selevan, 1993).

3.3.1.4. *Identification of Important Exposures for Reproductive Effects*

For all examinations of the relationship between reproductive effects and potentially toxic exposures, defining the exposure that produces the effect is crucial. Preconceptional exposures of 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 in- or subfertility). These exposures, plus postnatal exposure via breast milk, food, and the environment, may also be associated with postnatal developmental effects (e.g., changes in growth or in behavioral and cognitive function).

A number of factors affect the intensity and duration of exposure. General environmental exposures are typically lower than those found 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, home pesticide usage). Larger populations are necessary to achieve sufficient power in settings with lower exposures which are likely to have lower measures of risk (Lemasters and Selevan, 1984). In addition, exposure to individuals may change as they move in and out of areas with differing levels and types of exposures, thus 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 nature of certain study designs (e.g., retrospective studies) and limitations in estimates of historic exposures. Many reproductive effects 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) affects the individual's body burden and consequently the actual dose during the critical period. The probability of misclassification of exposure status may affect the ability to

recognize a true effect in a study (Selevan, 1981; Hogue, 1984; Lemasters and Selevan, 1984; Sever and Hessol, 1984; Kimmel, C.A. et al., 1986). As more prospective studies are done, better estimates of exposure should be developed.

3.3.1.5. *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 and Kline, 1981; Wilcox, 1983; Sever and Hessol, 1984; Axelson, 1985; Tilley et al., 1985; Kimmel, C.A. et al., 1986; Savitz and Harlow, 1991). Some of the more prominent factors are discussed below.

3.3.1.5.1. ***The power of the study.*** The power, or ability of a study to detect a true effect, is dependent on the size of the study group, the frequency of the outcome in the general population, and the level of excess risk to be identified. In a cohort study, common outcomes, such as recognized fetal loss, require hundreds of pregnancies to have a high probability of detecting a modest increase in risk (e.g., 133 pregnancies 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) (Bloom, 1981; Selevan, 1981, 1985; Sever and 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, study sizes are dependent upon the frequency of exposure within the source population. The confidence one has in the results of a study showing no effect 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 could increase confidence in the absence of risk for agents showing no effect. However, caution must be exercised in the combination of potentially dissimilar study groups.

Results of a negative study should be carefully evaluated, examining the power of the study and the degree of concordance or discordance between that study and other studies (including careful examination of comparability in the details such as similarity of adverse endpoints and study design). The consistency among results of different studies could be evaluated by comparing statistical confidence intervals for the effects found in different studies. Studies with lower power will tend to yield

wider confidence intervals. If the confidence intervals from a negative study and a positive study overlap, then there may be no conflict between the results of the two studies.

3.3.1.5.2. ***Potential bias in data collection.*** Bias may result from the way the study group is selected or information is collected (Rothman, 1986). Selection bias may occur when an individual's willingness to participate varies with certain characteristics relating to exposure or health status. In addition, selection bias may operate in the identification of subjects for study. For example, in studies of very early pregnancy loss, use of hospital records to identify the study group 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. The representativeness of these data may be affected by selection factors related to the willingness of different groups of women to continue participation over the total length of the study. Interview data result in more complete ascertainment than hospital records; however this strategy carries with it the potential for recall bias, discussed in further detail below. Other examples of different levels of ascertainment of events include: (1) use of hospital records to study congenital malformations since hospital records contain more complete data on malformations than do birth certificates (Mackeprang et al., 1972; Snell et al., 1992) and (2) use of sperm bank or fertility clinic data for semen studies. Semen data from either source are selected data because semen donors are typically of proven fertility, and men in fertility clinics are part of a subfertile couple who are actively trying to conceive. Thus, studies using the different record sources to identify reproductive outcomes need to be evaluated for ascertainment patterns prior to use in risk assessment.

Studies of women who work outside the home 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; Lemasters and Pinney, 1989). Thus, retrospective studies of female exposure that do not include terminated women workers may be of limited use in risk assessment because the level of risk for these outcomes is likely to be overestimated (Lemasters and Pinney, 1989).

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, and the individual's self-report of his/her occupational exposures and health characteristics more reliable than his/her mate's report (Selevan, 1980; Selevan et al., 1982). Studies based on interview data from the appropriate respondents would carry more weight than those from proxy respondents.

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 suggest 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. and Hatch, 1987; Weinberg et al., 1994).

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).

3.3.1.5.3. ***Collection of data on other risk factors, effect modifiers, and confounders.*** Risk factors for reproductive toxicity include such characteristics as age, smoking, alcohol or caffeine consumption, drug use, and past reproductive history. Groups of individuals may represent susceptible subpopulations based on genetic, acquired (e.g., behavioral), or developmental characteristics (e.g., greater effect of childhood exposures). 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 changed with age (e.g., if older men had semen changes with exposure while younger ones did not). A confounder is a variable that is a risk factor for the outcome 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.

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.

3.3.1.5.4. ***Statistical factors.*** As in studies of test animals, pregnancies experienced by the same woman are not fully independent events. For example, women who have had fetal loss are reported to be more likely to have subsequent losses (Leridon, 1977). In test animal studies, the litter can be used as the unit of measure to deal with nonindependence of response within the litter. In studies of humans, pregnancies are sequential, requiring analyses which consider nonindependence of events (Epidemiology Workgroup for the Interagency Regulatory Liaison Group, 1981; Kissling, 1981; Selevan, 1981; Zeger and Liang, 1986). 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). Analysis problems may occur when (1) prior adverse outcomes are due to the same exposures or (2) when prior adverse outcomes could result in changes in behaviors that could reduce exposures. Some approaches to deal with these issues have been suggested (Kissling, 1981; Stiratelli et al., 1984; Selevan, 1985; Zeger and Liang, 1986). These approaches include selecting one pregnancy per family (Selevan, 1985) or using generalized estimating equations (Zeger and Liang, 1986).

3.3.2. 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 subfecundity 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 fecundity in men. Identification is

also limited by the decreased likelihood of recognizing adverse developmental effects in their offspring as resulting from paternal exposure rather than maternal exposure. Thus far, only one agent causing human male reproductive toxicity, dibromochloropropane (DBCP), has been identified after observation of a cluster of infertility that resulted from male subfecundity. This cluster was identified because of an atypically high level of communication among the workers' wives (Whorton et al., 1977, 1979; Biava et al., 1978; Whorton and Milby, 1980).

Adverse effects identified in females through clusters and case reports have, thus far, been limited to adverse pregnancy outcomes such as fetal loss and congenital malformations. Identification of other effects, such as subfertility/subfecundity or menstrual cycle disorders, may be more difficult, as noted above.

Case reports may have importance in the recognition of agents that cause reproductive toxicity. 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.

3.4. 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. Information on the variability between humans and test species also may be useful in evaluating factors such as age-related differences in the balance between activation and deactivation of a toxic agent. These types of data may be helpful in defining the sequence of events leading to an adverse effect and 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 agent. 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. and Francis, 1990). This guidance is also applicable to nondermal 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 an agent with potential to cause testicular toxicity 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 (Crisp, 1992).

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 in the female between the maternal liver, ovary, 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 (Harris, R.Z. et al., 1995). 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 toxic agent, 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.

3.5. COMPARISONS OF MOLECULAR STRUCTURE

Comparisons of the chemical or physical properties of an agent with those of agents known to cause reproductive toxicity 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 (SAR) have not been well studied in reproductive toxicology, and have had limited success in predicting reproductive toxicity. The early literature has been reviewed and a set of classifications offered relating structure to reported male reproductive system activity (Bernstein, 1984). Data are available that suggest structure-activity relationships with limited utility in risk assessment for certain classes of chemicals (e.g., glycol ethers, some estrogens, androgens, other steroids, substituted phenols, retinoids, phthalate esters, short-chain halogenated hydrocarbon pesticides, alkyl-substituted polychlorinated dibenzofurans, PCBs, vinylcyclohexene and related olefins, halogenated propanes, metals, and azo dyes). McKinney and Waller (1994) have studied the qualitative SAR properties of PCBs with respect to their recognition by thyroxine, Ah and estrogen receptors. Although generally limited in scope and in need of validation, such relationships provide hypotheses that can be tested.

In spite of the limited information available on SAR in reproductive toxicology, under certain circumstances (e.g., in the case of new chemicals), this procedure can be used to evaluate the potential for toxicity when little or no other data are available.

3.6. EVALUATION OF DOSE-RESPONSE RELATIONSHIPS

The description and evaluation of dose-response relationships is a critical component of the hazard characterization. Evidence for a dose-response relationship is an important criterion in establishing a toxic reproductive effect. It includes the evaluation of data from both human and laboratory animal studies. When possible, pharmacokinetic data should be used to determine the effective dose at the target organ(s). 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

quantitative 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, an agent 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 risk of reproductive toxicity 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 at least as sensitive generally as the most sensitive animal species tested (Nisbet and Karch, 1983; Kimmel, C.A. et al., 1984, 1990; Hemminki and 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 or benchmark dose associated with the effect(s) (see Section 4).

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 effective dose range should be identified for both reproductive and other forms of systemic toxicity, and 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.

In addition to identification of the range of doses that is effective in producing reproductive 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 should be identified and evaluated. This information should always accompany the characterization of the health-related database (discussed in the next section).

Because the developing organism is changing rapidly and is vulnerable at a number of stages, an assumption is made with developmental effects that a single exposure at a critical time in development may produce an adverse effect (U.S. EPA, 1991). Therefore, with inhalation exposures, the daily dose is usually not adjusted to a 24-hour equivalent duration with developmental toxicity unless appropriate pharmacokinetic data are available. However, for other reproductive effects, daily doses by the inhalation route may be adjusted for duration of exposure. The Agency is planning to review these stances to determine the most appropriate approach for the future.

3.7. CHARACTERIZATION OF THE HEALTH-RELATED DATABASE

This section describes evaluation of the health-related database 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 database. This determination provides the basis for judging whether the available data are sufficient to characterize a hazard and to conduct quantitative dose-response analyses. It also should provide a summary and evaluation of the existing data and identify data gaps for an agent that is judged to have insufficient information to proceed with a quantitative dose-response analysis. Characterizing the available evidence in this way clarifies the strengths and uncertainties in a particular database. It does not address the level of concern, nor does it completely address determining relevance of available data for estimating human risk. Issues concerning relevance of mechanisms of action and types of effects observed should be included in the hazard characterization. Both level of concern and relevance are discussed further as part of the final characterization of risk, taking into account the information concerning potential human exposure. Data from all potentially relevant studies, whether indicative of potential hazard or not, should be included in the hazard characterization.

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 evaluate the database adequately. In some cases, a database may contain conflicting data. In these instances,

the risk assessor must consider each study's strengths and weaknesses within the context of the overall database 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, 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 6). The scheme contains two broad categories, "Sufficient" and "Insufficient," which are defined in Table 6. Data from all available studies, whether or not indicative of potential concern, are evaluated and used in the hazard characterization 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., sample size 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 test species and 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 characterize a hazard and conduct dose-response analyses. Establishing the minimum human evidence to proceed with quantitative analyses based on the human data is often difficult because there may be considerable variations in study designs and study group selection. The body of human data should contain convincing evidence as described in the "Sufficient Human Evidence" category. Because the human data necessary to judge whether or not a causal relationship exists are generally limited, few agents can be classified in this category. Agents that have been tested in laboratory animals according to EPA's two-generation reproductive effects test guidelines (U.S. EPA, 1982, 1985b, 1996a), 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, including determination of an RfD or RfC. In cases in which 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.

Table 6. Categorization of the health-related database

SUFFICIENT EVIDENCE

The Sufficient Evidence category includes data that collectively provide enough information to judge whether or not a reproductive hazard exists within the context of effect as well as dose, duration, timing, and route of exposure. This category may include both human and experimental animal evidence.

Sufficient Human Evidence

This category includes agents for which there is convincing evidence from epidemiologic studies (e.g., case control and cohort) to judge whether exposure is causally related to reproductive toxicity. A case series in conjunction with other supporting evidence also may be judged as Sufficient Evidence. An evaluation of epidemiologic and clinical case studies should discuss whether the observed effects can be considered biologically plausible in relation to chemical exposure.

Sufficient Experimental Animal Evidence/Limited Human Data

This category includes agents for which there is sufficient evidence from experimental animal studies and/or limited human data to judge if a potential reproductive hazard exists. Generally, agents that have been tested according to EPA's two-generation reproductive effects test guidelines (but not limited to such designs) would be included in this category. The minimum evidence necessary to determine if 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 determine that a potential hazard does not exist would include data on an adequate array of endpoints from more than one study with two species that showed no adverse reproductive effects at doses that were minimally toxic in terms of inducing an adverse effect. Information on pharmacokinetics, mechanisms, or known properties of the chemical class may also strengthen the evidence.

INSUFFICIENT EVIDENCE

This category includes agents for which there is less than the minimum sufficient evidence necessary for assessing the potential for reproductive toxicity. Included are situations such as when no data are available on reproductive toxicity; as well as for data bases from studies on test animals or humans that have a limited study design or conduct (e.g., small numbers of test 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 or in vitro tests, pharmacokinetic data, or metabolic precursors.

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 whether 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 plausible and consistent. On the other hand, to judge whether an agent is unlikely to pose a hazard for reproductive toxicity, the minimum evidence would include data on an array of endpoints and from studies with more than one species 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 database exists for a given chemical, but no single study meets current test guidelines, the risk assessor should use scientific judgment to determine whether the composite database may be viewed as meeting the "Sufficient" criteria.

Some important considerations in determining the confidence in the health database 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, the results are consistent.
- 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 3.2, can facilitate such determinations.

The greatest confidence for identification of a reproductive hazard should be placed on significant adverse effects on sexual behavior, fertility or development, or other endpoints that are directly related to reproductive function such as menstrual (estrous) cycle normality, sperm evaluations, reproductive histopathology, reproductive organ weights, and reproductive endocrinology. Agents producing adverse effects on these endpoints can be assigned to the "Sufficient Evidence" category if study quality is adequate.

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

The absence of effects in test species on the endpoints that are evaluated routinely (i.e., fertility, histopathology, prenatal development, 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 also should 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, functional effects from developmental exposures).

Judging that the health database 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 nonlinear dose-response) 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 summary of the hazard characterization should always be presented with information on the quantitative dose-response analysis and, if available, with the human exposure estimates.

4. QUANTITATIVE DOSE-RESPONSE ANALYSIS

In quantitative dose-response assessment, a nonlinear dose-response is assumed for noncancer health effects unless mode of action or pharmacodynamic information indicate otherwise. If sufficient data are available, a biologically based approach should be used on a chemical-specific basis to predict the shape of the dose-response curve below the observable range. 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 assumption of a nonlinear dose-response 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 above background. With a linear dose-response, it is assumed that some risk exists at any level of exposure, with risk decreasing as exposure decreases.

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 database having sufficient evidence for use in a risk assessment. 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 database having sufficient evidence. A significant increase may be based on statistical significance or on a biologically significant trend. Evidence for biological significance may be strengthened by mode of action or other biochemical evidence at lower exposure levels that supports the causation of such an effect. The existence of a NOAEL in an experimental animal study does not show the shape of the dose-response below the observable range; 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. Alternatively, mathematical modeling of the dose-response relationship may be performed in the experimental range. This approach can be used to determine a benchmark dose, which may be used in place of the NOAEL as a point of departure for calculating an RfD, RfC, MOE, or other exposure estimates.

Several limitations in the use of the NOAEL have been described (Kimmel, C.A. and Gaylor, 1988; U.S. EPA, 1995b): (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 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 95% upper confidence limit on developmental toxicity 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, the Agency is beginning to use the benchmark dose approach for quantitative dose-response evaluation when sufficient data are available.

Calculation and use of the benchmark dose are described in the EPA document *The Use of the Benchmark Dose Approach in Health Risk Assessment* (U.S. EPA, 1995b). The Agency is currently developing guidance for application of the benchmark dose, including a decision process to use for the various steps in the analysis (U.S. EPA, 1996c). The benchmark dose is based on a model-derived estimate of a particular incidence level, such as a 5% or 10% incidence. The BMD/C for a given endpoint serves as a consistent point of departure for low-dose extrapolation. In some cases, mode of action data may be sufficient to estimate a BMD/C at levels below the observable range for the health effect of concern. A benchmark response (BMR) of 5% is usually the lowest level of risk that can be estimated adequately for binomial endpoints from standard developmental toxicity studies (Allen et al., 1994a,b). For fetal weight, a continuous endpoint, a 5% change from the control mean was near the limit of detection for standard prenatal toxicity studies (Kavlock et al., 1995). The modeling approaches that have been proposed for developmental toxicity (U.S. EPA, 1995b) are, for the most part, curve-fitting models that have biological plausibility, but do not incorporate mode of action. Similar approaches can be applied to other reproductive toxicity data 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. Further guidance on the use of the BMD/C is being developed by the Agency (U.S. EPA, 1996c).

The RfD or RfC for reproductive toxicity 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. Because of the short duration of most studies of developmental toxicity, a unique value (RfD_{DT} or RfC_{DT}) is determined for adverse developmental effects. For adverse reproductive effects on endpoints other than those of

developmental toxicity, no special designator is attached. Data on reproductive toxicity (including developmental toxicity) are considered along with other data on a particular chemical in deriving an RfD or RfC.

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 2 and 3.2.1). Uncertainty factors for reproductive and other forms of systemic toxicity applied to the NOAEL or benchmark dose 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 database. In circumstances where only a LOAEL is available, the use of an additional uncertainty factor of up to 10 may be required, depending on the sensitivity of the endpoints evaluated, adequacy of dose levels tested, or general confidence in the LOAEL.

Other areas of uncertainty may be identified and modifying factors used depending on the characterization of the database (e.g., if the only data available are from a one-generation reproductive effects study; see Section 3.7), data on pharmacokinetics, or other considerations that may alter the level of confidence in the data (U.S. EPA, 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.

The NOAEL, LOAEL, or the benchmark dose is divided by the total uncertainty factor selected for the critical effect in the most appropriate or most sensitive mammalian species to determine the RfD or RfC. If the NOAEL, 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 should be compared with data from other studies in which adult animals are exposed. Thus, reproductive toxicity data should be discussed in the context of other toxicity data.

It has generally been assumed that there is a nonlinear dose-response for reproductive toxicity. This is 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 developing organism are known to have some capacity for repair of damage. However, in a population, background levels of toxic agents and preexisting conditions may increase the sensitivity of some individuals in the population. Thus, exposure to a toxic agent 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 has 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. Current research on modes of action in reproductive toxicology is promising and may provide data that are useful for appropriate modeling in the near future.

4.1. UTILIZATION OF INFORMATION IN RISK CHARACTERIZATION

The hazard characterization and quantitative dose-response evaluations are 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, epidemiology, statistics) is recommended.

5. 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* (U.S. EPA, 1992) have been published separately 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. Ideally, existing body burden as well as internal circulating and target organ exposure information for the agent of concern and other synergistic or antagonistic agents should be described. It should include information on the purpose, scope, level of detail and approach used, including estimates of exposure and dose by pathway and route for populations, subpopulations, and individuals in a manner that is appropriate for the intended risk characterization. It also should provide an evaluation of the overall level of confidence in the estimate(s) of exposure and dose and the conclusions drawn. 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.

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 in time 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 and Lemasters, 1987). The appropriate exposure model depends on the biologic processes affected and the nature of the chemical under study. Thus, a cumulative or dichotomous exposure model may be appropriate if injury occurs in cells that cannot be replaced or

repaired (e.g., 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 and other aspects of development in a manner or to an extent 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).

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 endpoint 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 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 and chemical mechanism of action.

The reversibility of an adverse effect on the reproductive system can be affected by the degree and duration of exposure (Clegg, 1995). 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 (Meistrich, 1982; Foote and 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 more time than mice to reach the same degree of recovery (Meistrich and 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). 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 an agent within the study period. Exposures during certain critical points in the reproductive process may affect the outcomes observed in humans (Lemasters and Selevan, 1984). In test species, perinatal exposure to androgens or estrogens such as zearalenone, methoxychlor, and DDT (Bulger and 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 (Steinberger and Lloyd, 1985; Schardein, 1993). Studies using test species also have shown that exposure to some environmental agents such as ionizing radiation (Dobson and 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 and Sawyer, 1950; Terranova, 1980), and ovum transport (Cummings and Perreault, 1990). Compounds that delay ovulation can lead to significant alterations in egg viability (Peluso et al., 1979), fertilizability of the egg (Fugo and Butcher, 1966; Butcher and Fugo, 1967; Butcher et al., 1975), and a reduction in litter size (Fugo and Butcher, 1966). After ovulation, single exposures to microtubule poisons such as carbendazim may impair the completion of meiosis in the fertilized oocyte with adverse developmental consequences (Perreault et al., 1992; Zuelke and Perreault, 1995). Thus, knowledge of when acute exposures occur 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, as well as other agents with estrogenic or anti-androgenic activity, 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.

Unique considerations for developmental effects 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, 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., endocrine disruption of reproductive tract development, premature reproductive senescence due to oocyte depletion), long after exposure has ceased.

6. RISK CHARACTERIZATION

6.1. OVERVIEW

A risk characterization is an essential part of any Agency report on risk whether the report is a preliminary one prepared to support allocation of resources toward further study, a site-specific assessment, or a comprehensive one prepared to support regulatory decisions. A risk characterization should be prepared in a manner that is clear, reasonable, and consistent with other risk characterizations of similar scope prepared across programs in the Agency. It should identify and discuss all the major issues associated with determining the nature and extent of the risk and provide commentary on any constraints limiting more complete exposition. The key aspects of risk characterization are: (1) bridging risk assessment and risk management, (2) discussing confidence and uncertainties, and (3) presenting several types of risk information. In this final step of a risk assessment, the risk characterization involves integration of toxicity information from the hazard characterization and quantitative dose-response analysis 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 intended to summarize key aspects of the following components of the risk assessment:

- The nature, reliability, and consistency of the data used.
- The reasons for selection of the key study(ies) and the critical effect(s) and their relevance to human outcomes.
- The qualitative and quantitative descriptors of the results of the risk assessment.
- The limitations of the available data, the assumptions used to bridge knowledge gaps in working with those data, and implications of using alternative assumptions.
- The strengths and weaknesses of the risk assessment and the level of scientific confidence in the assessment.
- 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 Guidelines point out issues that are important to risk characterization for reproductive toxicity. Comprehensive general guidance for risk characterization is provided by Habicht (1992) and Browner (1995).

6.2. INTEGRATION OF HAZARD CHARACTERIZATION, QUANTITATIVE DOSE-RESPONSE, AND EXPOSURE ASSESSMENTS

In developing each component 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, and 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 1) provide default positions. The default positions used and the rationale behind the use of each 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. Each section (hazard characterization, quantitative dose-response analysis, and exposure assessment) should have its own summary, and these summaries should be 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 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 judgments, but also should 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/or benchmark dose; and the assumptions underlying the estimation of the RfD, RfC, or other exposure estimate.
- 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 also should 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. The rationale for those decisions should be presented.

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 certain assumptions concerning the amount of absorption likely or the applicability of the data from one route to another (U.S. EPA, 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. and Francis, 1990) and *Principles of Route-to-Route Extrapolation for Risk Assessment* (Gerrity et al., 1990). The risk characterization should identify the methods used to extrapolate across exposure routes and discuss the strengths and limitations of the approach.

The level of confidence in the hazard characterization and quantitative 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 3.7). 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. EPA, 1992; Clegg, 1995).

In the risk assessment process, risk is estimated as a function of exposure, with the risk of adverse effects increasing as exposure increases. Information on the levels of exposure experienced by different members of the population is key to understanding the range of risks that may occur. 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. Differences among individuals in absorption rates, metabolism, or other factors mean that individuals or subpopulations with the same level and pattern of exposure may have differing susceptibility. For example, the consequences of exposure can differ markedly between developing individuals, young adults and aged adults, including whether the effects are permanent or transient. Other considerations relative to human exposures might include pregnancy or lactation, potential for exposures to other agents, concurrent disease, nutritional status, lifestyle, ethnic background and genetic polymorphism, and the possible consequences. Knowledge of the molecular events leading to induction of adverse effects may be of use in determining the range of susceptibility in sensitive populations.

An outline to serve as a guide and formatting aid for developing reproductive risk characterizations for chemical-specific risk assessments can be found in Table 7. A common format will assist risk managers in evaluating and using reproductive risk characterization. The outline has two parts. The first part tracks the reproductive risk assessment to bring forward its major conclusions. The second part pulls the information together to characterize the reproductive risk.

6.3. DESCRIPTORS OF REPRODUCTIVE RISK

Descriptors of reproductive risk convey information and answer questions about risk, with each descriptor providing different information and insights. There are a number of ways to describe risk. Details on how to use these descriptors can be obtained from the guidance on risk characterization (Browner, 1995) from which some of the information below has been extracted.

In most cases, the state of the science is not yet adequate to define distributions of factors such as population susceptibility. The guidance principles below discuss a variety of risk descriptors that primarily reflect differences in estimated exposure. If a full description of the range of susceptibility in the population cannot be presented, an effort should be made to identify subgroups that, for various reasons, may be particularly susceptible.

Table 7. Guide for developing chemical-specific risk characterizations for reproductive effects

PART ONE
Summarizing Major Conclusions in Risk Characterization

I. Hazard Characterization

A. What is (are) the key toxicological study (or studies) that provides the basis for health concerns for reproductive effects?

- How good is the key study?
- Are the data from laboratory or field studies? In a single or multiple species?
- What adverse reproductive endpoints were observed, and what is the basis for the critical effect?
- Describe other studies that support this finding.
- Discuss any valid studies which conflict with this finding.

B. Besides the reproductive effect observed in the key study, are there other health endpoints of concern? What are the significant data gaps?

C. Discuss available epidemiological or clinical data. For epidemiological studies:

- What types of data were used (e.g., human ecologic, case-control or cohort studies, or case reports or series)?
- Describe the degree to which exposures were described.
- Describe the degree to which confounding factors were considered.
- Describe the degree to which other causal factors were excluded.

D. How much is known about how (through what biological mechanism) the chemical produces adverse reproductive effects?

- Discuss relevant studies of mechanisms of action or metabolism.
- Does this information aid in the interpretation of the toxicity data?
- What are the implications for potential adverse reproductive effects?

E. Comment on any nonpositive data in animals or people, and whether these data were considered in the hazard characterization.

F. If adverse health effects have been observed in wildlife species, characterize such effects by discussing the relevant issues as in A through E above.

G. Summarize the hazard characterization and discuss the significance of each of the following:

- Confidence in conclusions
- Alternative conclusions that are also supported by the data
- Significant data gaps

- Highlights of major assumptions

Table 7. Guide for developing chemical-specific risk characterizations for reproductive effects (continued)

II. Characterization of Dose-Response

A. What data were used to develop the dose-response curve? Would the result have been significantly different if based on a different data set?

- If laboratory animal data were used:
  Which species were used?
  Most sensitive, average of all species, or other?
  Were any studies excluded? Why?
- If epidemiological data were used:
  Which studies were used?
  Only positive studies, all studies, or some other combination?
  Were any studies excluded? Why?
  Was a meta-analysis performed to combine the epidemiological studies? What approach was used?
  Were studies excluded? Why?

B. Was a model used to develop the dose-response curve and, if so, which one? What rationale supports this choice? Is chemical-specific information available to support this approach?

- How was the RfD/RfC (or the acceptable range) calculated?
- What assumptions and uncertainty factors were used?
- What is the confidence in the estimates?

C. Discuss the route, level, and duration of exposure observed, as compared to expected human exposures.

- Are the available data from the same route of exposure as the expected human exposures? If not, are pharmacokinetic data available to extrapolate across route of exposure?
- How far does one need to extrapolate from the observed data to environmental exposures? One to two orders of magnitude? Multiple orders of magnitude? What is the impact of such an extrapolation?

D. If adverse health effects have been observed in wildlife species, characterize dose-response information using the process outlined in A through C above.

III. Characterization of Exposure

A. What are the most significant sources of environmental exposure?
Are there data on sources of exposure from different media?
What is the relative contribution of different sources of exposure?
What are the most significant environmental pathways for exposure?

Table 7. Guide for developing chemical-specific risk characterizations for reproductive effects (continued)

B. Describe the populations that were assessed, including the general population, highly exposed groups, and highly susceptible groups.

C. Describe the basis for the exposure assessment, including any monitoring, modeling, or other analyses of exposure distributions such as Monte Carlo or krieging.

D. What are the key descriptors of exposure?
Describe the (range of) exposures to: "average" individuals, "high-end" individuals, general population, high exposure group(s), children, susceptible populations, males, females (nonpregnant, pregnant, lactating).
How was the central tendency estimate developed?
What factors and/or methods were used in developing this estimate?
How was the high-end estimate developed?
Is there information on highly exposed subgroups?
Who are they?
What are their levels of exposure?
How are they accounted for in the assessment?

E. Is there reason to be concerned about cumulative or multiple exposures because of biological, ethnic, racial, or socioeconomic reasons?

F. If adverse reproductive effects have been observed in wildlife species, characterize wildlife exposure by discussing the relevant issues as in A through E above.

G. Summarize exposure conclusions and discuss the following:
- Results of different approaches, i.e., modeling, monitoring, probability distributions;
- Limitations of each, and the range of most reasonable values;
- Confidence in the results obtained, and the limitations to the results

PART TWO
Risk Conclusions and Comparisons

IV. Risk Conclusions

A. What is the overall picture of risk, based on the hazard, quantitative dose-response, and exposure characterizations?

B. What are the major conclusions and strengths of the assessment in each of the three main analyses (i.e., hazard characterization, quantitative dose-response, and exposure assessment)?

Table 7. Guide for developing chemical-specific risk characterizations for reproductive effects (continued)

C. What are the major limitations and uncertainties in the three main analyses?

D. What are the science policy options in each of the three major analyses?
What are the alternative approaches evaluated?
What are the reasons for the choices made?

V. Risk Context

A. What are the qualitative characteristics of the reproductive hazard (e.g., voluntary vs. involuntary, technological vs. natural, etc.)? Comment on findings, if any, from studies of risk perception that relate to this hazard or similar hazards.

B. What are the alternatives to this reproductive hazard? How do the risks compare?

C. How does this reproductive risk compare to other risks?
How does this risk compare to other risks in this regulatory program, or other similar risks that the EPA has made decisions about?
Where appropriate, can this risk be compared with past Agency decisions, decisions by other federal or state agencies, or common risks with which people may be familiar?
Describe the limitations of making these comparisons.

D. Comment on significant community concerns which influence public perception of risk.

VI. Existing Risk Information

Comment on other reproductive risk assessments that have been done on this chemical by EPA, other federal agencies, or other organizations. Are there significantly different conclusions that merit discussion?

VII. Other Information

Is there other information that would be useful to the risk manager or the public in this situation that has not been described above?

6.3.1. Distribution of Individual Exposures

Risk managers are interested generally in answers to questions such as: (1) Who are the people at the highest risk and why? (2) What is the average risk or distribution of risks for individuals in the population of interest? and (3) What are they doing, where do they live, etc., that might be putting them at this higher risk?

Exposure and reproductive risk descriptors for individuals are intended to provide answers to these questions. To describe the range of risks, both high-end and central tendency descriptors are used to convey the distribution in risk levels experienced by different individuals in the population. For the Agency's purposes, high-end risk descriptors are plausible estimates of the individual risk for those persons at the upper end of the risk distribution. Given limitations in current understanding of variability in individuals' sensitivity to agents that cause reproductive toxicity, high-end descriptors will usually address high-end exposure or dose. Conceptually, high-end exposure means exposure above approximately the 90th percentile of the population distribution, but not higher than the individual in the population who has the highest exposure. Central tendency descriptors generally reflect central estimates of exposure or dose. The descriptor addressing central tendency may be based on either the arithmetic mean exposure (average estimate) or the median exposure (median estimate), either of which should be clearly labeled. The selection of which descriptor(s) to present in the risk characterization will depend on the available data and the goals of the assessment.

6.3.2. Population Exposure

Population risk refers to assessment of the extent of harm for the population as a whole. In theory, it can be calculated by summing the individual risks for all individuals within the subject population. That task requires more information than is usually available. Questions addressed by descriptors of population risk for reproductive effects would include: What portion of the population is within a specified range of some reference level, e.g., exceeds the RfD (a dose), the RfC (a concentration), or other health concern level?

For reproductive effects, risk assessment techniques have not been developed generally to the point of knowing how to add risk probabilities, although Hattis and Silver (1994) have proposed approaches for certain case-specific situations. Therefore, the following descriptor is usually appropriate: An estimate of the percentage of the population, or the number of persons, above a specified level of risk or within a specified range of some reference level (e.g., exceeds the RfD, RfC, LOAEL, or other specific level of interest). 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 "may be 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. This descriptor must be obtained through measuring or simulating the population distribution.

6.3.3. 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 or benchmark dose from the most appropriate or sensitive species to the estimated human exposure level from all potential sources (U.S. EPA, 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. 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 database, including the strengths and weaknesses of the toxicity and exposure data, the number of species affected, and the information on dose-response, route, timing, and duration. 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 database used, and the needs of the decisionmaker. 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.

6.3.4. Distribution of Exposure and Risk for Different Subgroups

A risk manager might also ask questions about the distribution of the risk burden among various segments of the subject population such as the following: How do exposure and reproductive risk impact various subgroups? and What is the population risk of a particular subgroup? Questions about the distribution of exposure and reproductive risk among such population segments require additional risk descriptors.

6.3.4.1. *Highly Exposed*

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; Browner, 1995) 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.

Highly exposed subgroups can be identified and, where possible, characterized, and the magnitude of risk quantified. This descriptor is useful when there is (or is expected to be) a subgroup experiencing significantly different exposures or doses from those of the larger population. These subpopulations may be identified by age, sex, lifestyle, economic factors, or other demographic variables. For example, toddlers who play in contaminated soil and consumers of large amounts of fish represent subpopulations that may have greater exposures to certain agents.

If population data are absent, it will often be possible to describe a scenario representing high-end exposures using upper percentile or judgment-based values for exposure variables. In these instances, caution should be taken not to overestimate the high-end values if a "reasonable" exposure estimate is to be achieved.

6.3.4.2. *Highly Susceptible*

Highly susceptible subgroups also can be identified and, if possible, characterized, and the magnitude of risk quantified. This descriptor is useful when the sensitivity or susceptibility to the effect for specific subgroups is (or is expected to be) significantly different from that of the larger population. Therefore, 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 pregnant or lactating women, women with reduced oocyte numbers, men with "borderline" sperm counts, or infants. To calculate risk for these subgroups, it will be necessary sometimes to use a different dose-response relationship; e.g., upon exposure to a chemical, pregnant or lactating women, elderly people, children of varying ages, and people with certain illnesses may each be more sensitive than the population as a whole.

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

It is important to consider, however, that the Agency's current methods for developing reference doses and reference concentrations (RfDs and RfCs) are designed to protect sensitive populations. If data on sensitive human populations are available (and there is confidence in the quality of the data), then the RfD is based on the dose level at which no adverse effects are observed in the sensitive population. If no such data are available (for example, if the RfD is developed using data from humans of average or unknown sensitivity), then an additional 3- to 10-fold factor may be used to account for variability between the average human response and the response of more sensitive individuals (see Section 4).

Generally, selection of the population segments to consider for high susceptibility is a matter of either *a priori* interest in the subgroup (e.g., environmental justice considerations), in which case the risk assessor and risk manager can jointly agree on which subgroups to highlight, or a matter of discovery of a sensitive or highly exposed subgroup during the assessment process. In either case, once identified, the subgroup can be treated as a population in itself and characterized in the same way as the larger population using the descriptors for population and individual risk.

6.3.5. Situation-Specific Information

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 who may be exposed to levels above that value?

"What if ...?" questions, such as those that follow, can be used to examine candidate risk management options:

- What are the reproductive risks if a pesticide applicator applies this pesticide without using protective equipment?
- What are the reproductive risks if this site becomes residential in the future?
- What are the reproductive risks if we set the standard at 100 ppb?

Answering such "what if?" questions involves a calculation of risk based on specific combinations of factors postulated within the assessment. The answers to these "what if?" questions do not, by themselves, give information about how likely the combination of values might be in the actual population or about how many (if any) persons might be subjected to the potential future reproductive risk. However, information on the likelihood of the postulated scenario would be desirable to include in the assessment.

When addressing projected changes for a population (either expected future developments or consideration of different regulatory options), it usually is appropriate to calculate and consider all the reproductive risk descriptors discussed above. When central tendency or high-end estimates are developed for a scenario, these descriptors should reflect reasonable expectations about future activities. For example, in site-specific risk assessments, future scenarios should be evaluated when they are supported by realistic forecasts of future land use, and the reproductive risk descriptors should be developed within that context.

6.3.6. Evaluation of the Uncertainty in the Risk Descriptors

Reproductive risk descriptors are intended to address variability of risk within the population and the overall adverse impact on the population. In particular, differences between high-end and central tendency estimates reflect variability in the population but not the scientific uncertainty inherent in the risk estimates. As discussed above there will be uncertainty in all estimates of reproductive risk. These uncertainties can include measurement uncertainties, modeling uncertainties, and assumptions to fill data gaps. Risk assessors should address the impact of each of these factors on the confidence in the estimated reproductive risk values.

Both qualitative and quantitative evaluations of uncertainty provide useful information to users of the assessment. The techniques of quantitative uncertainty analysis are evolving rapidly and both the SAB (Loehr and Matanoski, 1993) and the NRC (1994) have urged the Agency to incorporate these techniques into its risk analyses. However, it should be noted that a probabilistic assessment that uses only the assessor's best estimates for distributions of population variables addresses variability, but not uncertainty. Uncertainties in the estimated risk distribution need to be evaluated separately. An approach has been proposed for estimating distribution of uncertainty in noncancer risk assessments (Baird et al., 1996).

6.4. SUMMARY AND RESEARCH NEEDS

These Guidelines summarize the procedures that the EPA 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 interpretation of data and interspecies extrapolation is needed. This research includes studies that: (1) more completely characterize and define female and male reproductive endpoints, (2) more completely characterize the types of developmental toxicity possible, (3) evaluate the interrelationships among endpoints, (4) examine quantitative extrapolation between endpoints (e.g., sperm count) and function (e.g., fertility), (5) provide a better understanding of the relationships between reproductive toxicity and other forms of toxicity, (6) explore pharmacokinetic disposition of the target, and (7) examine mechanistic phenomena related to pharmacokinetic disposition. These types of studies, along with further evaluation of a nonlinear dose-response for susceptible populations, should provide methods to more precisely assess risk.

7. 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. New York Academy of Science, New York. pp. 21-32.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994a) Dose-response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with NOAELs. Fundam. Appl. Toxicol. 23:487-495.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994b) Dose-response assessment for developmental toxicity: III. Statistical models. Fundam. Appl. Toxicol. 23:496-509.

Amann, R.P. (1981) A critical review of methods for evaluation of spermatogenesis from seminal characteristics. J. Androl. 2:37-58.

American Academy of Pediatrics Committee on Drugs. (1994) The transfer of drugs and other chemicals into human milk. Pediatrics 93:137-150.

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.

Auger, J., Kunstman, J.M., Czyglik, F., Jouannet, P. (1995) Decline in semen quality among fertile men in Paris during the past 20 years. N. Engl. J. Med. 332:281-285.

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. JAMA 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.

Baird, S.J.S., Cohen, J.T., Graham, J.D., Shlyakhter, A.I., Evans, J.S. (1996) Noncancer risk assessment: a probabilistic alternative to current practice. Human Ecol. Risk Assess. 2:79-102.

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.

Biava, C.G., Smuckler, E.A., Whorton, D. (1978) The testicular morphology of individuals exposed to dibromochloropropane. Exp. Mol. 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. Fundam. 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. and 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.

Boyd, J.A., Clark, G.C., Walmer, D.K., Patterson, D.G., Needham, L.L., Lucier, G.W. (1995) Endometriosis and the environment: biomarkers of toxin exposure. Abstract from Endometriosis 2000 Workshop, May 15-17.

Boyers, S.P., Davis, R.O., Katz, D.F. (1989) Automated semen analysis. Curr. Probl. 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.

Brouwer, A., Ahlborg, U.G., Vandenberg, M., Birnbaum, L.S., Boersma, E.R., Bosveld, B., Denison, M.S., Gray, L.E., Hagmar, L., Holene, E., Huisman, M., Jacobson, S.W., Jacobson, J.L., Koopmanesseboom, C., Koppe, J.G., Kulig, B.M., Morse, D.C., Muckle, G., Peterson, R.E., Sauer, P.J.J., Seegal, R.F., Smitsvanprooije, A.E., Touwen, B.C.L., Weisglaskuperus, N., Winneke, G. (1995) Functional aspects of developmental toxicity of polyhalogenated aromatic hydrocarbons in experimental animals and human infants. Eur. J. Pharmacol. 293:1-40.

Browner, C.M. (1995) EPA risk characterization program. U.S. EPA Memorandum, March 21, 1995. Available from the EPA Air docket.

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

Bujan, L., Mansat, A., Pontonnier, F., Mieusset, R. (1996) Time series analysis of sperm concentration in fertile men in Toulouse, France between 1977 and 1992. Br. Med. J. 312: 471-472.

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.

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. Endocrinology 96:576-586.

Byskov, A.G., Hoyer, P.E. (1994) Embryology of mammalian gonads and ducts. In: Knobil, E., Neill, J.D. The Physiology of Reproduction. Raven Press, New York. pp. 487-540.

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., Filler, R.S., Gulati, D., Heindel, J.J., Katz, D.F., Mebus, C.A., Obasaju, F., Perreault, S.D., Russell, S.R., Schrader, S., Slott, V., Sokol, R.Z., Toth, G. (1992) Methods for assessing rat sperm motility. Reprod. Toxicol. 6:267-273.

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. Fundam. 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. Fundam. 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.

Clegg, E.D. (1995) Reversibility of effects: overview and reproductive systems. Inhal. Toxicol. 7:881-889.

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

Colie, C.F. (1993) Male mediated teratogenesis. Reprod. Toxicol. 7:3-9.

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.

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., Vandenbergh, 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.

Cooper, R.L., Barrett, M.A., Goldman, J.M., Rehnberg, G.R., McElroy, W.K., Stoker, T.E. (1994) Pregnancy alterations following xenobiotic-induced delays in ovulation in the female rat. Fundam. Appl. Toxicol. 22:474-480.

Cooper, R.L., Stoker, T.E., Goldman, J.M., Parrish, M.B., Tyrey, L. (1996) Effect of atrazine on ovarian function in the rat. Reprod. Toxicol. 10(4):257-264.

Crisp, T.M. (1992) Organization of the ovarian follicle and events in its biology: oogenesis, ovulation or atresia. Mutat. Res. 296:89-106.

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

Csapo, A.I., Pulkkinen, M. (1978) Indispensability of the human corpus luteum in the maintenance of early pregnancy: lutectomy 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.

Darney, S.P. (1991) In vitro assessment of gamete integrity. In: Goldberg, A.M. In Vitro Toxicology: Mechanisms and New Technology. Mary Ann Liebert, Inc., New York. pp. 63-75.

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.

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.

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.

Egeland, G.M., Sweeney, M.H., Fingerhut, M.A., Wille, K.K., Schnorr, T.M., Halperin, W.E. (1994) Total serum testosterone and gonadotropins in workers exposed to dioxin. Am. J. Epidemiol. 139:272-281.

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. Am. 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. Endocrinology 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.

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.

Fisch, H., Goluboff, E.T., Olson, J.H., Feldshuh, J., Broder, S.J., Barad, D.H. (1996) Semen analyses in 1,283 men from the United States over a 25-year period: no decline in fertility. Fertil. Steril. 65:1009-1014.

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.

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. Fundam. 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. Am. 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 and 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.

Galletti, F., Klopper, A. (1964) The effect of progesterone on the quantity and distribution of body fat in the female rat. Acta Endocrinol. 46:379-386.

Gardner, M.J., Hall, A.J., Snee, M.P., Downes, S., Powell, C.A., Terrell, J.D. (1990a) Methods and basic data of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. Br. Med. J. 300:429-434.

Gardner, M.J., Snee, M.P., Hall, A.J., Powell, C.A., Downes, S., Terrell, J.D. (1990b) Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. Br. Med. J. 300:423-429.

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

Gellert, R.J. (1978) Kepone, mirex, dieldrin, and aldrin: estrogenic activity and the induction of persistent vaginal estrus and anovulation in rats following neonatal treatment. Environ. Res. 16:131-138.

Generoso, W.M., Piegorsch, W.W. (1993) Dominant lethal tests in male and female mice. In: Chapin, R.E., Heindel, J.J. Methods in Toxicology: Male Reproductive Toxicology. Academic Press, San Diego. pp. 124-139.

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.

George, F.W., Wilson, J.D. (1994) Sex determination and differentiation. In: Knobil, E., Neill, J.D. The Physiology of Reproduction. Raven Press, New York. pp. 3-28.

Gerhard, I., Runnebaum, B. (1992) Grenzen der hormonsubstitution bei Schadstoffbelastung und fertilitatsstorungen. Zentralbl. Gynakol. 114:593-602.

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.

Gill, W.B., Schumacher, F.B., Bibbo, M., Straus, F.H., Schoenberg, H.W. (1979) Association of diethylstilbestrol exposure in utero with cryptorchidism, testicular hypoplasia and semen abnormalities. J. Urol. 122:36-39.

Ginsburg, J., Okolo, S., Prelevic, G., Hardiman, P. (1994) Residence in London area and sperm density. Lancet 343:230.

Giusti, R.M., Iwamoto, K., Hatch, E.E. (1995) Diethylstilbestrol revisited: a review of the long-term health effects. Ann. Intern. Med. 122:778-788.

Giwercman, A., Carlsen, E., Keiding, N., Skakkebaek, N.E. (1993) Evidence for increasing incidence of abnormalities of the human testis: A review. Environ. 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.

Gorski, R.A. (1986) Sexual differentiation of the brain: a model for drug-induced alterations of the reproductive system. Environ. Health Perspect. 70:163-175.

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., Ostby, J.S. (1995) In utero 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) alters reproductive morphology and function in female rat offspring. Toxicol. Appl. Pharmacol. 133:285-294.

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. Fundam. 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. Fundam. Appl. Toxicol. 15:281-297.

Gray, L.E., Ostby, J.S., Kelce, W.R. (1994) Developmental effects of an environmental antiandrogen: the fungicide vinclozolin alters sex differentiation of the male rat. Toxicol. Appl. Pharmacol. 129:46-52.

Gray, L.E., Kelce, W.R., Monosson, E., Ostby, J.S., Birnbaum, L.S. (1995) Exposure to TCDD during development permanently alters reproductive function in male Long Evans rats and hamsters: reduced ejaculated and epididymal sperm numbers and sex accessory gland weights in offspring with normal androgenic status. Toxicol. Appl. Pharmacol. 131:108-118.

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. Epidemiol. 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. Fundam. 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. Available from the EPA Air docket.

Hales, B., Crosman, K., Robaire, B. (1992) Increased post-implantation loss and malformations among the F2 progeny of male rats chronically treated with cyclophosphamide. Teratology 45:671-678.

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. Fundam. Appl. Toxicol. 19:186-196.

Harris, R.Z., Benet, L.Z., Schwartz, J.B. (1995) Gender effects in pharmacokinetics and pharmacodynamics. Drugs 50:222-239.

Harrison, P.T.C., Humfrey, C.D.N., Litchfield, M., Peakall, D., Shuker, L.K. (1995) IEH Assessment on Environmental Oestrogens: Consequences to Human Health and Wildlife. MRC Institute for Environment and Health. Leicester, UK.

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.

Hattis, D., Silver, K. (1994) Human interindividual variability: a major source of uncertainty in assessing risks for noncancer health effects. Risk Analysis 14:421-431.

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.

Hemminki, K., Vineis, P. (1985) Extrapolation of the evidence on teratogenicity of chemicals between humans and experimental animals: chemicals other than drugs. Teratogenesis Carcinog. 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. Fundam. 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.

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

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. Fundam. 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.

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

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

Hugenholtz, 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 Perspect. 78:171-175.

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

Imagawa, W., Yang, J., Guzman, R., Nandi, S. (1994) Control of mammary gland development. In: Knobil, E., O'Neill, J.D. The Physiology of Reproduction. Raven Press, New York. pp. 1033-1063.

International Conference on Harmonization of Technical Requirements of Pharmaceuticals for Human Use. (1994) ICH harmonized tripartite guideline, Detection of Toxicity to Reproduction for Medicinal Products. Federal Register 59(1831):48746-48752.

Irvine, S., Cawood, E., Richardson, D., MacDonald, E., Aitken, J. (1996) Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men in Scotland over 11 years. Br. Med. J. 312:467-471.

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.

Kavlock, R.J., Allen, B.C., Kimmel, C.A., Faustman, E.M. (1995) Dose-response assessment for developmental toxicology: benchmark doses for fetal weight changes. Fundam. Appl. Toxicol. 26:211-222.

Kelce, W.R., Stone, C.R., Laws, S.C., Gray, L.E., Kemppainen, J.A., Wilson, E.M. (1995) Persistent DDT metabolite p,p'-DDE is a potent androgen receptor antagonist. Nature 375:581-585.

Kesner, J.S., Wright, D.M., Schrader, S.M., Chin, N.W., Krieg, E.F. (1992) Methods of monitoring menstrual function in field studies: Efficacy of methods. Reprod. Toxicol. 6:385-400.

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

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

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 Liaison 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. (1995) Reproductive toxicity testing: a risk assessment perspective. In: Witorsch, R.J. Reproductive Toxicology. Raven Press, New York. pp. 75-98.

Kissling, G. (1981) A generalized model for analysis of nonindependent observations. Dissertation. University of North Carolina.

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

Kline, J., Stein, Z., Susser, M. (1989) Conception to Birth: Epidemiology of Prenatal Development. Oxford University Press, New York.

Klinefelter, G.R., Laskey, J.W., Kelce, W.R., Ferrell, J., Roberts, N.L., Suarez, J.D., Slott, V. (1994a) Chloroethylmethanesulfonate-induced effects on the epididymis seem unrelated to altered Leydig cell function. Biol. Reprod. 51:82-91.

Klinefelter, G.R., Laskey, J.W., Perreault, S.D., Ferrell, J., Jeffay, S., Suarez, J., Roberts, N. (1994b) The ethane dimethanesulfonate-induced decrease in the fertilizing ability of cauda epididymal sperm is independent of the testis. J. Androl. 15:318-327.

Knobil, E., Neill, J.D., Greenwald, G.S., Markert, C.L., Pfaff, D.W. (1994) The Physiology of Reproduction. Raven Press, New York.

Ku, W.W., Chapin, R.E., Wine, R.N., Gladen, B.C. (1993) Testicular toxicity of boric acid (BA): relationship of dose to lesion development and recovery in the F344 rat. Reprod. Toxicol. 7:305-319.

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.

Lamb, J.C. (1985) Reproductive toxicity testing: evaluating and developing new testing systems. J. Am. 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 comparision of results from two laboratories. J. Am. 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. (1992) Occupational exposures and effects on male and female reproduction. In: Rom, W.N. Environmental and Occupational Medicine. Little, Brown, Boston, MA. pp. 147-170.

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., Selevan, S.G. (1993) Toxic exposures and reproduction: a view of epidemiology and surveillance. In: Scialli, A.R., Zinaman, M.J. Reproductive Toxicology and Infertility. McGraw-Hill, New York. pp. 307-321.

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.

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

Lindbohm, M.L., Hemminki, K., Bonhomme, M.G., Anttila, A., Rantala, K., Keikkila, P., Rosenberg, M.J. (1991) Effects of paternal occupational exposure on spontaneous abortions. Am. J. Public Health 81:1029-1033.

Linder, R.E., Hess, R.A., Strader, L.F. (1986) Testicular toxicity and infertility in male rats treated with l,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 two ages of young adult male rats. Fundam. Appl. Toxicol. 14:284-298.

Linder, R.E., Strader, L.F., Slott, V.L., Suarez, J.D. (1992) Endpoints of spermatotoxicity in the rat after short duration exposures to fourteen reproductive toxicants. Reprod. Toxicol. 6:491-505.

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.

Liu, D.Y., Baker, H.W.G. (1992) Tests of human sperm function and fertilization in vitro. Fertil. Steril. 58:465-483.

Loehr, R.A., Matanoski, G.M. (1993) Letter to Carol M. Browner, EPA Administrator, re: quantitative uncertainty analysis for radiological assessments. U.S. EPA Science Advisory Board, July 23, 1993 (EPA-SAB-RAC-COM-93-006).

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. (1994) Test methods for assessing female reproductive and developmental toxicology. In: Hayes, A.W. Principles and Methods of Toxicology. Raven Press, New York. pp. 989-1037.

Mason, H.J. (1990) Occupational cadmium exposure and testicular endocrine function. Hum. Exp. Toxicol. 9:91-94.

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.

McDonald, A.D., McDonald, J.C., Armstrong, B., Cherry, N.M., Nolin, A.D., Robert, D. (1989) Father's occupation and pregnancy outcome. Br. J. Ind. Med. 46:329-333.

McGregor, A.J., Mason, H.J. (1991) Occupational mercury vapour exposure and testicular, pituitary and thyroid endocrine function. Hum. Exp. Toxicol. 10:199-203.

McKinney, J.D., Waller, C.L. (1994) Polychlorinated biphenyls as hormonally active structural analogues. Environ. Health Perspect. 102:290-297.

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. Radiat. 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.

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., Schwetz, B.A., Teague, J.L., Morris, R.W. (1988a) Association of sperm, vaginal cytology, and reproductive organ weight data with results of continuous breeding reproduction studies in Swiss (CD-1) mice. Fundam. Appl. Toxicol. 11:359-371.

Morrissey, R.E., Schwetz, B.A., Lamb, J.C., Ross, M.D., Teague, J.L., Morris, R.W. (1988b) Evaluation of rodent sperm, vaginal cytology, and reproductive organ weight data from National Toxicology Program 13-week studies. Fundam. Appl. Toxicol. 11:343-358.

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. Fundam. 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.

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, DC.

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

National Research Council. (1994) Science and Judgment in Risk Assessment. 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.

Nett, T.M. (1989) Hormonal evaluation of testicular function: species variation. J. Am. Coll. Toxicol. 8:539-549.

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

Oberlander, G., Yeung, C.H., Cooper, T.G. (1994) Induction of reversible infertility in male rats by oral ornidazole and its effects on sperm motility and epididymal secretions. J. Reprod. Fertil. 100:551-559.

Organization for Economic Cooperation and Development. (1983) First addendum to OECD guideline 415 for testing of chemicals, "One-Generation Rreproduction Toxicity." OECD, Paris, pp. 1-8.

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. OECD, Paris.

Organization for Economic Cooperation and Development. (1993b) First amendment to OECD guidelines 416, "Two Generation Reproduction Toxicity". OECD, Paris, pp. 1-8.

Oskarsson, A., Hallen, I.P., Sundberg, J. (1995) Exposure to toxic elements via breast milk. Analyst 120:765-770.

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. Endocrinology 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.

Paul, M. (1993) Occupational and Environmental Reproductive Hazards. Williams and Wilkins, Baltimore.

Paulsen, C.A., Berman, N.G., Wang, C. (1996) Data from men in greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. Fertil. Steril. 65:1015-1020.

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.

Perreault, S.D., Jeffay, S.C. (1993) Strategies and methods for the functional evaluation of oocytes and zygotes. In: Heindel, J.J., Chapin, R.E. Methods in Toxicology: Female Reproductive Toxicology. Academic Press, San Diego. pp. 92-109.

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.

Qiu, J., Hales, B.F., Robaire, B. (1995) Damage to rat spermatozoal DNA after chronic cyclophosphamide exposure. Biol. Reprod. 53:1465-1473.

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.

Ratcliffe, J.M., Schrader, S.M., Clapp, D.E., Halperin, W.E., Turner, T.W., Horning, R.W. (1989) Semen quality in workers exposed to 2-ethoxyethanol. Br. J. Ind. Med. 46:399-406.

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.

Rier, S.E., Martin, D.C., Bowman, R.E., Dmowski, W.P., Becker, J.L. (1993) Endometriosis in rhesus monkeys (Macaca mulatta) following chronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Fundam. Appl. Toxicol. 21: 433-441.

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.

Rowland, A.S., Baird, D.D., Weinberg, C.R., Shore, D.L., Shy, C.M., Wilcox, A.J. (1992) Reduced fertility among women employed as dental assistants exposed to high levels of nitrous oxide. N. Engl. J. Med. 327:993-997.

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.

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. Tissue 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.

Safe, S.H. (1995) Modulation of gene expression and endocrine response pathways by 2,3,7,8-tetrachlorodibenzo-p-dioxin and related compounds. Pharmacol. Ther. 67:247-281.

Sakai, C.N., Hodgen, G.D. (1987) Use of primate folliculogenesis models in understanding human reproductive biology and applicablity to toxicology. Reprod. Toxicol. 1:207-222.

Samuels, S.J. (1988) Lessons from a surveillance program of semen quality. Reprod. Toxicol. 2:229-231.

Savitz, D.A., Harlow, S.D. (1991) Selection of reproductive health end points for environmental risk assessment. Environ. Health 90:159-164.

Savitz, D.A., Sonnenfeld, N.L., Olshan, A.F. (1994) Review of epidemiologic studies of paternal occupational exposure and spontaneous abortion. Am. J. Ind. Med. 25:361-383.

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. (1993) Chemically Induced Birth Defects. Marcel Dekker, New York.

Schrader, S.M., Chapin, R.E., Clegg, E.D., Davis, R.O., Fourcroy, J.L., Katz, D.F., Rothmann, S.A., Toth, G., Turner, T.W., Zinaman, M. (1992) Laboratory methods for assessing human semen in epidemiologic studies: a consensus report. Reprod. Toxicol. 6:275-279.

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) Menstrual dysfunction. In: Gold, J.J., Josimovich, J.B. Gynecologic Endocrinology. Harper and Row, Hagerstown, MD.

Seed, J., Chapin, R.E., Clegg, E.D., Darney, S.P., Dostal, L., Foote, R.H., Hurtt, M.E., Klinefelter, G.R., Makris, S.L., Schrader, S., Seyler, D., Sprando, R., Treinen, K.A., Veeranachaneni, R., Wise, L.D. (1996) Methods for assessing sperm motility, morphology, and counts in the rat, rabbit and dog: a consensus report. Reprod. Toxicol. 10:237-244.

Selevan, S.G. (1980) Evaluation of data sources for occupational pregnancy outcome studies. Thesis. University of Cincinnati.

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. (1991) Environmental exposures and reproduction. In: Keily, M. Reproductive and Perinatal Epidemiology. CRC Press, Boca Raton. pp. 115-130.

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., Keye, W.R. Reproduction: The New Frontier in Occupational and Environmental Research. Alan R. Liss, Inc., New York. pp. 15-48.

Sharpe, R.M. (1994) Regulation of spermatogenesis. In: Knobil, E., Neill, J.D. The Physiology of Reproduction. Raven Press, New York. pp. 1363-1434.

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. Pharmacol. Ther. 38:269-304.

Slott, V.L., Perreault, S.D. (1993) Computer-assisted sperm analysis of rodent epididymal sperm motility using the Hamilton-Thorne motility analyzer. In: Chapin, R.E., Heindel, J.J. Methods in Toxicology: Male Reproductive Toxicology. Academic Press, San Diego. pp. 319-333.

Slott, V.L., Suarez, J.D., Simmons, J.E., Perreault, S.D. (1990) Acute inhalation exposure to epichlorohydrin transiently decreases rat sperm velocity. Fundam. 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.

Slott, V.L., Jeffay, S.C., Suarez, J.D., Barbee, R.R., Perreault, S.D. (1995) Synchronous assessment of sperm motility and fertilizing ability in the hamster following treatment with alpha-chlorohydrin. J. Androl. 16:523-535.

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. Neuroendocrinology 14:129-138.

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.

Snell, L.M., Little, B.B., Knoll, K.A., Johnston, W.L., et al. (1992) Reliability of birth certificate reporting of congenital anomalies. Am. J. Perinatol. 9:219-222.

Sonawane, B.R. (1995) Chemical contaminants in human milk: an overview. Environ. Health. Perspect. 103:197-205.

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

Stein, A. and 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.

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.

Stoker, T.E., Goldman, J.M., Cooper, R.L. (1993) The dithiocarbamate fungicide thiram disrupts the hormonal control of ovulation in the female rat. Reprod. Toxicol. 7:211-218.

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. Fundam. 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 comparision of pregnancy history recall and medical records: implications for retrospective studies. Am. J. Epidemiol. 121:269-281.

Toppari, J., Larsen, J.C., Christiansen, P., Giwercman, A., Grandjean, P., Guillette, L.J., Jegou, B., Jensen, T.K., Jouannet, P., Keiding, N., Leffers, H., McLachlan, J.A., Meyer, O., Muller, J., Rajpert-De Meyts, E., Scheike, T., Sharpe, R., Sumpter, J., Skakkebaek, N. (1995) Male Reproductive Health and Environmental Chemicals with Estrogenic Effects. Miljoprojekt nr. 290. Danish Environmental Protection Agency.

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. Fundam. Appl. Toxicol. 13:16-25.

Toth, G.P., Stober, J.A., George, E.L., Read, E.J., Smith, M.K. (1991a) Sources of variation in the computer-assisted motion analysis of rat epididymal sperm. Reprod. Toxicol. 5:487-495.

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

Toth, G.P., Wang, S.R., McCarthy, H., Tocco, D.R., Smith, M.K. (1992) Effects of three male reproductive toxicants on rat cauda epididymal sperm motion. Reprod. Toxicol. 6:507-515.

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.

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.

Tucker, H.A. (1994) Lactation and its hormonal control. In: Knobil, E., O'Neill, J.D. The Physiology of Reproduction. Raven Press, New York. pp. 1065-1098.

Tyl, R.W. (1987) Developmental toxicity in toxicologic research and testing. In: Ballantyne, B. Perspectives in Basic and Applied Toxicology. John Wright, Bristol. pp. 203-208.

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. Integrated 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. Federal Register 56(234):63798-63826.

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

U.S. Environmental Protection Agency. (1995a) Proposed Guidelines for Neurotoxicity Risk Assessment. Federal Register 60(192):52032-52056.

U.S. Environmental Protection Agency. (1995b) The Use of the Benchmark Dose Approach in Health Risk Assessment. EPA/630/R-94/007.

U.S. Environmental Protection Agency. (1996a) Health Effects Test Guidelines OPPTS 870.3800: Reproduction and Fertility Effects (Draft). Federal Register 61(43):8282-8283.

U.S. Environmental Protection Agency. (1996b) Proposed Guidelines for Carcinogen Risk Assessment. Federal Register 61(79):17960-18011.

U.S. Environmental Protection Agency. (1996c) Benchmark Dose Technical Guidance Document. EPA/600/P-96/002A.

Van Waeleghem, K., De Clerq, N., Vermeulen, L., Schoonjans, F., Comhaire, F. (1996) Deterioration of sperm quality in young healthy Belgian men. Hum. Reprod. 11:325-329.

Vierula, M., Niemi, M., Keiski, A., Saarikoski, M., Suominen, J. (1996) High and unchanged sperm counts of Finnish men. Int. J. Androl. 19:11-17.

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

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. Endocrinology. 80:714-718.

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

Weinberg, C.R., Baird, D.D., Wilcox, A.J. (1994) Sources of bias in studies of time to pregnancy. Stat. Med. 13:671-681.

Weir, P.J., Rumberger, D. (1995) Isolation of rat sperm from the vas deferens for sperm motion analysis. Reprod. Toxicol. 9:327-330.

Welch, L.S., Schrader, S.M., Turner, T.W., Cullen, M.R. (1988) Effects of exposure to ethylene glycol ethers on shipyard painters: II. Male reproduction. Am. J. Ind. Med. 14:509-526.

Welch, L.S., Plotkin, E., Schrader, S. (1991) Indirect fertility analysis in painters exposed to ethylene glycol ethers: sensitivity and specificity. Am. J. Ind. Med. 20:229-240.

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.

Wilcox, A.J., Weinberg, C.R., O'Connor, J.F., Baird, D.D., Schlatterer, J.P., Canfield, R.E., Armstrong, E.G., Nisula, B.C. (1988) Incidence of early pregnancy loss. N. Engl. J. Med. 319:189-194.

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. Fundam. Appl. Toxicol. 15:651-665.

Wilson, J.G. (1973) Environment and Birth Defects. Academic Press, New York.

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. Teratology 11:169-178.

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

Witorsch, R.J. (1995) Reproductive Toxicology. Raven Press, New York.

Wolff, M.S. (1993) Lactation. In: Paul, M. Occupational and Environmental Reproductive Hazards. Williams and Wilkins, Baltimore. pp. 60-75.

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.

World Health Organization. (1992) WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction. Third edition. Cambridge University Press, Cambridge.

Wright, D.M., Kesner, J.M., Schrader, S.M., Chin, N.W., Wells, V.E., Krieg, E.F. (1992) Methods of monitoring menstrual function in field studies: attitudes of working women. Reprod. Toxicol. 6:401-409.

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 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.

Yeung, C.H., Oberlander, G., Cooper, T.G. (1992) Characterization of the motility of maturing rat spermatozoa by computer-aided objective measurement. J. Reprod. Fertil. 96:427-441.

Zeger, S.L., Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121-130.

Zenick, H., Blackburn, K., Hope, E., Baldwin, D.J. (1984) Evaluating male reproductive toxicity in rodents: a new animal model. Teratogenesis Carcinog. Mutagen. 4:109-128.

Zenick, H., Clegg, E.D., Perreault, S.D., Klinefelter, G.R., Gray, L.E. (1994) Assessment of male reproductive toxicity: a risk assessment approach. In: Hayes, A.W. Principles and Methods of Toxicology. Raven Press, New York. pp. 937-988.

Zinaman, M.J., Clegg, E.D., Brown, C.C., O'Connor, J., Selevan, S.G. (1996) Estimates of human fertility and pregnancy loss. Fertil. Steril. 65:503-509.

Zuelke, K.A., Perreault, S.D. (1995) Carbendazim (MBC) disrupts oocyte spindle function and induces aneuploidy in hamsters exposed during fertilization (meiosis II). Mol. Reprod. Dev. 42:200-209.

PART B: RESPONSE TO SCIENCE ADVISORY BOARD AND PUBLIC COMMENTS

1. INTRODUCTION

A notice of availability for public comment of these Guidelines was published in the *Federal Register* (FR) in February 1994. Seven responses were received. These Guidelines were presented to the Environmental Health Committee of the Science Advisory Board (SAB) on July 19, 1994. The report of the SAB was provided to the Agency in May 1995, with further communication from the SAB Executive Committee provided in December 1995.

The SAB and public comments were diverse and represented varying perspectives. Many of the comments were favorable and expressed agreement with positions taken in the proposed guidelines. A number of the comments addressed items that were more pertinent to testing guidance than risk assessment guidance or were otherwise beyond the scope of these Guidelines. Some of those were generic issues that are not system specific. Others were topics that have not been developed sufficiently and should be viewed as research issues. There were conflicting views about the need to provide additional detailed guidance about decision-making in the evaluation process as opposed to promoting extensive use of scientific judgment. Also, comments provided specific suggestions for clarification of details.

2. RESPONSE TO SCIENCE ADVISORY BOARD COMMENTS

In general, the SAB found "the overall scientific foundations of the draft guidelines' positions to be generally sound." However, recommendations were made to improve specific areas.

The SAB recommended that EPA retain separate sections for identification and dose-response assessment in the draft guidelines. In subsequent meetings involving the SAB Executive Committee, members of the Clean Air Scientific Advisory Committee, and the Environmental Health Committee, this issue was explored further. After discussion, the SAB agreed with expanding the hazard identification to include certain components of the dose-response assessment. The resulting hazard characterization provides an evaluation of hazard within the context of the dose, route, timing, and duration of exposure. The next step, the dose-response analysis, quantitatively evaluates the relationship between dose or exposure and severity or probability of effect in humans. EPA has revised these Guidelines to reflect that position which is consistent also with the 1994 NRC report, *Science and Judgment in Risk Assessment*. The SAB suggested an alternative scheme for characterizing

health effects data in Table 5. The Agency's intent for Table 5 is not to characterize the available data, but rather to judge whether the database is sufficient to proceed further in the risk assessment process. The text has been modified to clarify the intended use of this table and to ensure that it is consistent with the reorganization of the Guidelines into separate hazard characterization and quantitative dose-response analysis sections.

The SAB supported the concept of using a gender neutral default assumption, but indicated that more discussion to support this assumption was needed. In particular, the Committee indicated that a fuller discussion is needed on "information to the contrary" (to obviate the need for making this default assumption), as well as additional guidance for using this and other default assumptions in risk characterization. The Agency agrees with this recommendation and provides further guidance on the use of the gender neutral default assumption. In keeping with recent Agency guidance on risk characterization, discussion on the use of default assumptions has been expanded in the risk characterization section of these Guidelines.

The SAB in its reviews of the reproductive toxicity and neurotoxicity risk assessment guidelines discussed assumptions about the behavior of the dose-response curve. The SAB's advice has been that the Agency examine available data first, and only use nonlinear behavior as a default if available data do not define the dose-response curve. The SAB also recommended that the benchmark dose method be considered as a possible alternative to the NOAEL/LOAEL approach. The Agency agrees.

The SAB recommended that more discussion be devoted to the issue of disruption of endocrine systems by environmental agents. The section on Endocrine Evaluations has been expanded to include endocrine disruption of the reproductive system during development in addition to effects on adults.

The SAB supported the principle in the Guidelines that more than one negative study is necessary to judge that a chemical is unlikely to pose a reproductive hazard. That principle has been retained and, as recommended by the SAB, an explicit statement included that data from a second species are necessary to determine that sufficient information is available to indicate that an agent is unlikely to pose a hazard.

The SAB recommended that the topic of susceptible populations be expanded and that the Guidelines should indicate that relevant information be incorporated into risk assessments when possible. To address this issue, the Agency has emphasized potential differences in risks in children at different stages of development, females (including pregnant and lactating females), and males, and indicated that relevant information on differential risks for susceptible populations should be included in the risk characterization section when available. When specific information on differential risks is not

available, the Agency will continue to apply a default uncertainty factor to account for potential differences in susceptibility.

The SAB recommended that the Agency provide more specific guidance for exposure assessment issues that arise when characterizing exposure for reproductive toxicants. The Agency agrees and has indicated that an exposure assessment: include a statement of purpose, scope, level of detail, and approach used; present the estimate of exposure and dose by pathway and route for individuals, population segments, and populations in a manner appropriate for the intended risk characterization; and provide an evaluation of the overall level of confidence (including consideration of uncertainty factors) in the estimate of exposure and dose and the conclusions drawn. The SAB recommended that the MOE discussion be modified to address specific circumstances where the administered dose and the "effective dose" are known to be different. The discussion has been modified to emphasize that pharmacokinetic data, when available, be utilized to address such instances.

The SAB recommended that the Agency expand substantially the discussion of overall strategy to evaluate exposure from mixtures, exposures to multiple single agents, and exposures to the same agent via different routes. It is anticipated that this type of information will be addressed in the Agency's upcoming revisions to the chemical mixture guidelines.

3. RESPONSE TO PUBLIC COMMENTS

In addition to numerous supportive statements, several issues were indicated although each issue was raised by a very limited number of submissions. Use of the benchmark dose was supported along with the suggestion that the amount of text could be reduced on that subject. The text has been reduced and reference made to the report, *The Use of the Benchmark Dose Approach in Health Risk Assessment* (U.S. EPA, 1995b). A request was made for increased emphasis on paternally mediated effects on offspring. The text in that section has been expanded to provide additional discussion and references. Concern was expressed about the existence of constraints on the use of professional judgment in the risk assessment process, particularly in determining the relevance and sufficiency of the database, in evaluating biological plausibility of statistically different effects, and in the determination of uncertainty factors. Requests also have been made to provide additional criteria for when and under what conditions the risk assessment process will be used. These Guidelines emphasize the importance of using scientific judgment throughout the risk assessment process. They provide flexibility to permit EPA's offices and regions to develop specific guidance suited to their particular needs. The comment was made that the exposure assessment and risk characterization sections were

not developed as well as the rest of the document. In 1992, EPA published *Guidelines for Exposure Assessment* (U.S. EPA, 1992) that were intended to apply generically to noncancer risk assessments. These Guidelines only address aspects of exposure that are specific to reproduction and have been developed sufficiently. The risk characterization section has been expanded substantially to reflect the recent guidance provided within EPA for application in all risk assessments.