Reviewer Guidance

Evaluation of Human Pregnancy Outcome Data

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication of the *Federal Register* notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers La, Rm. 1061, Rockville, MD 20857. All comments should be identified with the docket number listed in the notice of availability published in the *Federal Register*.

For questions about this draft document, contact (CDER) Rose Cunningham, 301-594-5468; (CBER) Toni Stifano, 301-827-3028.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

Clin #
June 1999

Draft — Not for Implementation

Reviewer Guidance

Evaluation of Human Pregnancy Outcome Data

Additional copies are available from:

Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

or

Office of Communication,
Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm
(Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

Clin #
June 1999

Draft — Not for Implementation

Table of Contents

Draft — Not for Implementation

Reviewer Guidance[1]

Review of Human Pregnancy Outcome Data

I. INTRODUCTION

This guidance for FDA reviewers is intended to provide reviewers with a standardized approach to the general assessment of the potential risks related to exposures of pregnant women to drugs and biological products. In particular, it places standard considerations of risk assessment from a variety of data sources in the scientific context of pregnancy and human pregnancy outcomes.[2]

When new products come to the market human data on risks to the pregnant woman and her developing fetus normally are not available. Usually, animal reproductive and developmental toxicology form the basis of ascertaining risk. After products have been marketed for some time (months to years), information on human pregnancy outcomes begins to become available and may be brought to the reviewer's attention from a variety of sources that do not necessarily converge in time (e.g., individual spontaneous reports of congenital anomalies associated with exposure; literature reports of epidemiology

[1] This guidance has been prepared by the Pregnancy Registry Working Group of the FDA Pregnancy Labeling Taskforce and the Women's Health Subcommittee of the Medical Policy Coordinating Committee (MPCC) in the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration. This guidance represents the Agency's current thinking on FDA reviewers' evaluation of human pregnancy outcome data. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

[2] The Agency is developing a series of good review practices documents to provide a standardized framework for Agency review of data submitted in a new drug application. This guidance can be considered one part of this analytical framework and will be incorporated into the good review practices documents when they are completed.

studies; pregnancy registry reports). The review of such data requires consideration of clinical teratology, factors relevant to the setting of obstetric medical care, and integration of these areas with principles of epidemiology. [3]

This guidance addresses the FDA reviewer assessment of human data as opposed to animal reproductive toxicology data.[4] Each of the three major types of postmarketing human pregnancy outcome data is introduced, followed by general descriptions of factors to consider when assessing them. Detailed guidance is provided in later sections regarding potential types of human pregnancy outcome data that the reviewer might encounter within the context of pregnancy-related outcomes. A section is devoted to pregnancy registries, as they constitute an increasingly popular means of collecting data that the reviewer is least likely to be familiar with.[5]

The traditional focus of most data related to pregnancy, and the focus of this guidance, is fetal outcomes. These data tend to be the most frequently available and are also of most concern for reviewers and clinicians. Reviewers should also consider product- and pregnancy-related information more broadly, including information that relates to toxicities that may be unique or altered in pregnant patients, and especially pharmacokinetic data in pregnant patients. Additional areas of interest include reproductive endpoints, such as fertility and premature ovarian senescence, as these may be relevant to potential viability and growth of the fetus.

II. BACKGROUND

FDA clinical reviewers, like most practicing clinicians, have little experience in systematically evaluating data related to pregnant women. There may be occasional inadvertent pregnancy exposures during clinical trials of new products, but available data are usually insufficient to permit robust analysis and the collection systems inadequate to allow for long-term follow-up. When clinical studies are conducted in this special population, the trials tend to be small and are not informative for assessing treatment group differences in pregnancy outcomes and fetal effects. Data typically emerge during the postmarketing phase of product development, but are limited in scope and detail. Nonetheless, these data cannot be ignored and must be analyzed on their own merit in the context of what is already known about a product from animal and other human studies. This analysis is less daunting given a basic understanding of the types of human data that one is likely to encounter and some general principles for interpretating the data.

---

[3] Throughout this document, the term *drug* includes therapeutic biological agents and vaccines.

[4] A guidance is under development on preclinical reproductive toxicology assessments.

[5] A guidance for industry is under development on establishing pregnancy registries and should be available in June 1999.

Integrating the many considerations outlined in this document and reaching a confident estimation of risk to the pregnant patient and her developing fetus of a particular exposure are often challenging tasks. Important resources within FDA that the clinical reviewer can access to assist in this effort include:

- The Center for Drug Evaluation and Research (CDER) Pharmacology Toxicology Coordinating Committee.

- CDER's Office of Post Marketing Drug Risk Assessment (OPDRA)

- CBER's Division of Biostatistics and Epidemiology.

A large FDA cooperative agreement program is managed by OPDRA, allowing the Agency to participate in the design of pharmacoepidemiology studies to answer questions about drug and biologic product safety and exposure outcomes, which may include pregnancy outcomes.

At the time of any new drug approval, these offices should receive direct communication from the new drug review division regarding concerns already identified from animal or human data related to pregnancy outcomes, especially when extensive use of a product by women of reproductive age is expected.

III. TYPES OF POSTMARKETING DATA – SOURCES AND UTILITY

FDA receives postmarketing data from a variety of sources and in many different forms. By far the most common source is the *case report*, which, in the case of pregnancies, is usually a report of a congenital anomaly. Such reports come to FDA directly from medical care providers or consumers through MedWatch or from product sponsors through the postmarketing Adverse Experience Reporting System (AERS) or Vaccine Adverse Event Reporting System (VAERS). Unfortunately, most estimates are that fewer than 10 percent of all adverse drug events ever reach FDA's attention or the public domain. From 1994 to 1996 the term *congenital anomaly* was coded on fewer than 1 percent of the MedWatch adverse event reports that FDA received.

Such case reports suffer from many limitations, including lack of information on the number of exposed patients (the denominator with which the cases need to be compared) and lack of a control group for comparison. Reports of abnormal pregnancy outcomes are further limited by the often prolonged interval between the mother's exposure to a product and the birth. Despite these shortcomings, well-described case reports can be useful and have been important in leading to the identification of agents that are now well known to be harmful to the developing fetus, such as angiotensin-converting enzyme inhibitors.

Formal epidemiology studies (cohort and case control studies) are another source of data. These studies which are sometimes used to evaluate the effects of drugs on pregnancy outcomes. They may be conducted independently by a sponsor or planned cooperatively among a sponsor, the Agency, and reputable pharmacoepidemiology groups. When well conceived and well conducted, epidemiology studies provide data sources of adequate size to examine relatively rare events. However, these studies have a potential for bias in patient selection and ascertaining outcomes and exposures.

Pregnancy registries are the third most common source of data. Registries have been established for some products to systematically collect postmarketing data on exposures of pregnant patients to a particular product and the outcomes of their pregnancies.[6] Sponsors may develop pregnancy registries as a response to safety concerns identified at the time of a new product's marketing approval, either on their own initiative or when requested by FDA as a phase-4 commitment. Groups other than pharmaceutical sponsors may develop registries, such as teratology information services[7] or academic institutions, either independently or under contract from a pharmaceutical firm. Although essentially a type of cohort study, the elements and application of such registries during postmarketing surveillance warrant careful consideration.

Under the best of circumstances, registries can provide detailed numerator- and denominator-based information to allow for the estimation of risks of individual or groups of outcomes. To be most useful and subject to the least bias, registries should enroll pregnant patients exposed to a product before the outcome is known. Outcome data should then be obtained in a complete and unbiased manner. Operationally, meeting such ideals may not be possible. However, to help interpret the data that registries provide, it is useful for the reviewer to be cognizant of various methods of designing and maintaining them. Some special considerations for design and interpretation of registry data are discussed in Section V.

IV. CRITICAL FACTORS IN EVALUATING PREGNANCY OUTCOME DATA

[6] Examples of products for which registries have been established include acyclovir/valcyclovir, antiretrovirals, antiepileptic drugs, fluoxetine, varicella vaccine, isotretinoin, and lyme disease vaccine.

[7] Teratology information services are independent counseling services, often linked to state health departments or universities, poison control, or genetic counseling centers, that provide information to pregnant women and/or healthcare providers on the risk of exposure to various drugs and biological products and environmental agents. Some of the 35 services located in North America collect data into registries on particular exposures of concern and follow women to determine pregnancy outcomes.

A. Background Incidence of Adverse Pregnancy Outcomes

When evaluating whether an exposure increases the risk for any adverse pregnancy outcome, it is critical to understand the background rate of the outcome in the population of concern. Incidences of various pregnancy outcomes within the general population of the United States are shown in Table 1. These percentages vary somewhat by year, by data source, and by subgroup of the population, including geographical, socioeconomic, and maternal age differences.

However, the ranges provided in Table 1 have remained constant for many years across many sources. Beyond major malformations (those that are of surgical, medical, or cosmetic importance), which are estimated to occur spontaneously in approximately 4 percent of live births in the general population, less specific and variably identifiable abnormalities may occur (Oakley[8]). General estimates are that at least an additional 5 percent of children have minor malformations, such as short fingernails or wide-spaced eyes. These minor malformations occur spontaneously and are usually not considered medically or cosmetically important.

Table 1. Background Rates of Adverse Pregnancy Outcomes*

Spontaneous Abortions (recognized pregnancies)	15%
Premature Delivery	6-10%
Children with Major Malformations[a, b]	4%
Additional Births with Minor Malformations[c]	5%

* These are general estimates accumulated from a variety of sources.
[a] Oakley 1996
[b] March of Dimes 1996
[c] Holmes 1998

To assess whether the therapeutic use of a product results in fetal developmental toxicity, it is common to compare the rate of the outcome estimated from extrapolations of exposure frequency and reporting rates to the background rate of the event. Occasionally, background rates such as those in Table 1 are appropriate initial comparisons. However, determining whether there is an

[8] See also, March of Dimes Birth Defects Foundation, *Birth Defects and Infant Mortality. A National and Regional Profile*, 1996.

\\

increased incidence is difficult if the adverse effect is uncommon or the apparent increase in risk is small. Exposures that produce abnormal outcomes at a high rate (e.g., 40 percent spontaneous abortion compared to a baseline estimated rate of 20 percent) are easier to identify than those that produce abnormal outcomes infrequently (e.g., an increase in the incidence of neural tube defects from 0.1 percent to 0.2 percent), even though both may represent a doubling of the background rate. Also, within the general population of the United States, the rates of abnormal outcomes may vary across demographic groups and be affected by maternal history and other factors.

B. Combined vs Individual Rates of Malformations

A population that experiences a ten-fold increase in a rare congenital anomaly (e.g., transposition of the great vessels) as a result of an exposure may still have a total congenital anomaly rate (e.g., all malformations or cardiac malformations) that is not measurably different from that in a reference population. Therefore, it is important to look at rates of individual malformations to identify possible increases in exposure-associated abnormalities. Table 2 lists the ranges of reported international rates for some individual abnormalities as reported to the International Clearinghouse for Birth Defects Monitoring System in 1991.[9]

Table 2. Individual Malformation Rates As reported to the International Clearinghouse for Birth Defects, 1991

(Per 10,000 births)

anencephaly[a]	0.8-18.4
spina bifida[a]	2.5-17.5
microtia	0.1-6.4
transposition great vessels[b]	0.1-5.4
hypoplastic left heart[b]	0.1-3.4
cleft palate	26-10.1
cleft lip	5.2-16.3
esophageal atresia	0.4-3.6
limb reduction	3.1-8.0
omphalocele	0.8-3.9

[a] Rate will be influenced by antenatal diagnosis and abortion rates at various data collection sites.

[b] Rate will be influenced strongly by length of follow-up and availability of diagnosis.

[9] March of Dimes Birth Defects Foundation, 1996.

C. Major vs. Minor Malformations

Most case reports and studies of congenital malformations focus on major malformations. These defects are either obvious at birth or serious enough to require medical or surgical interventions, or they are considered of substantial cosmetic importance. It should be kept in mind, however, that cardiovascular defects with no immediate physiologic consequences to the infant may not be detected at birth (since newborns are not routinely screened for these defects). However, they remain within what are considered major malformations.

The term *minor anomaly* generally refers to minor physical defects that are well documented in many studies to occur in at least 4 to 6 percent of infants, but which do not have obvious medical, surgical, or cosmetic consequences. The importance of minor malformations remains controversial, but they should not be summarily dismissed, especially when they occur in a particular pattern or in combinations of defects. For example, the presence of certain minor anomalies alerts physicians to look for more serious syndromes in an infant, such as the presence of a single transverse palmar crease, Downs syndrome, and the classic findings of fetal alcohol syndrome. Some studies have shown that infants with three or more minor congenital anomalies are at increased risk for also having a major malformation (Leppig et al., 1987).

D. Timing of Exposure

When interpreting pregnancy-related data, the reviewer should consider the timing and duration of maternal exposure and its relation to windows of developmental sensitivity. For example, when an exposure has occurred after the critical development time for the malformed organ has passed, such as in a fetus with transposition of the great vessels who was exposed to a drug only in the third trimester, the exposure is an unlikely cause of the malformation. When considering the timing of exposure, one may want to incorporate information from animal studies of the product of concern, as well as detailed information about the human exposure data being evaluated. Agents that produce adverse effects on the fetus typically do so during discrete sensitive periods of gestation, and animal data can provide important clues to such adverse effects. Knowledge of the sensitive period for human target organ development facilitates optimal data interpretation (Table 3). Note that data may be presented differently from one study to the next in describing timing of exposure or sensitivity windows (e.g., days after fertilization; days or weeks of gestation).

Table 3. Human Embryo Susceptibility to Teratogens*

Organ	Week Since Conception	Weeks from LMP** (Gestational Age)
Central Nervous System	3-7	5-9
Heart	3-7	5-9
Special senses (ears, eyes)	4-12	6-14
Limb	4-8	6-10
Genitalia	5-12	7-14
Teeth	6-10	8-12

* Moore and Persaud 1998.

** LMP = First day of last menstrual period.

For example, consider a medication that causes an increase in neural tube closure defects from 0.1 percent to 1 percent. Neural tube closure occurs between 21 and 28 days (3 – 4 weeks) after fertilization in human pregnancy. If one identifies 10,000 women known to have been exposed to the medication during the sensitive period and 100 of the resulting children have a neural tube defect, the effect rate is 1 percent. Alternatively, if one identifies 10,000 women exposed during the first trimester of pregnancy, but only 1,000 of them were exposed during the sensitive period, they might give rise to only 10 affected children. The other 9,000 pregnancies will manifest the background 0.1 percent rate (9 affected children). The total 19 affected children from 10,000 exposed pregnancies will produce a 0.19 percent rate, which might not be appreciated as different from the background and which would underestimate the true effect rate.

As a practical matter, the sensitive period for exposure to a medication or biologic is often unknown, making it common to globally assess risk from first trimester exposures. This is usually because no clear toxicity has been identified and because organogenesis occurs in the first trimester. There are two potential sources of error in using this global approach. First, as seen in the above example, sensitive time periods for different organs may make up a small portion of the first trimester (see Table 4). Second, toxicities may not be limited to the first trimester. Agents that can lead to adverse fetal effects in later trimesters include the angiotensin-converting enzyme inhibitors. In addition, some agents may produce abnormalities during more than one exposure

window. Nonetheless, if a sensitive period for a toxic effect is unknown, global trimester assessments may be used as a crude screen.

Table 4. Examples of Critical Timing of Exposure for Some Products.

- Thalidomide can produce limb and other defects following exposure between days 20-36 after fertilization.
- Diethylstilbestrol exposure in utero before week 9 of gestation leads to a risk of vaginal adenosis in female offspring of >70 percent. For exposure after week 17, the risk is less than 10 percent.
- ACE inhibitor exposure in the second and third trimester of pregnancy is associated with fetal and neonatal hypertension, anuria and oligohydramnios.

E. Intensity of Exposure

The amount of an agent reaching the fetus is a critical element of toxicity governed by actual dose taken and its pharmacokinetics in the maternal, placental, and fetal circulations. Almost all exposures can be toxic to the fetus if the dose is high enough, even if only by virtue of maternal toxicity. Most agents will cross the placenta to some extent, the exceptions being highly charged or certain very large molecules like heparin and insulin. Maternal changes in absorption, volume of distribution, metabolism, plasma protein binding, and excretion will affect the extent to which the fetus is exposed. These parameters are dynamic over the course of pregnancy. For example, some products may be more readily transported across the placenta during late gestation than during early gestation because of an increased unbound fraction in maternal circulation, increased utero-placental blood flow, increased placental surface area, or a more acidic fetal circulation to trap basic compounds. Agents that undergo relatively little metabolism in the fetus, but are excreted into the amniotic fluid by the well-developed fetal kidney in the third trimester, may have greater exposure as the fetus continually swallows amniotic fluid.

F. Variability of Response

Just as adverse effects related to a given drug do not occur in all exposed individuals, exposures that are known to increase the incidence of adverse pregnancy outcomes may do so only in a fraction of those exposed. For many therapeutic products that are associated with congenital anomalies, it is common to see a rate of affected offspring on the order of 1 to 10 percent of those

exposed, even when the conditions of exposure appear to be identical. Furthermore, products that interfere with fetal development may produce manifestations along a continuum of response, ranging from no effect at one extreme through growth impairment, identifiable anomalies, and death at the other extreme. These responses may be closely connected with a dose-response relationship. For example, alcohol at low doses throughout pregnancy is associated with slightly decreased birth weight. At higher doses, it has effects on fetal neurologic development, and at progressively higher doses, it is associated with microcephaly and other visible anatomic effects. Although generally predictable from a population perspective, the nature and extent of effects are not necessarily possible to predict as a function of dose in individual patients. Even at the same dose in the same gestational window, there can be a range of possible outcomes. Table 5 shows some other examples of this unpredictable relationship between exposure and response.

Table 5. Examples of Variability of Response

- Isotretinoin – exposure during the first 28 days of gestation may produce abnormalities of the external ear and jaw; cardiac malformations; neurobehavioral alterations; spontaneous abortions; or no effect on development under similar conditions and timing of exposure.

- Phenytoin – Identical exposure throughout pregnancy can lead to single or combinations of malformations such as cleft lip and/or cleft palate; heart defects; umbilical and inguinal hernias; hypospadias; other craniofacial defects; nail hypoplasia; growth deficiency or intellectual impairment.

It is important to remember that the concept of variability extends not only to toxic responses, but also to baseline attributes of populations. Birth weight, for example, varies by race and gender. Normative growth curves developed in Caucasian populations that are used to evaluate African-American or Asian-American babies may result in over-diagnosis of growth impairment. Genetic differences in metabolism of a drug by the mother, the placenta, and the fetus may contribute to variation in how much of a drug's metabolites reach fetal tissue and thereby lead to variable rates and types of toxicity manifested.

G. Biologic Plausibility

The most critical factor in the global assessment of a relationship between a drug and an adverse pregnancy outcome is biologic plausibility. In addition to the principles described above, biologic plausibility involves the potential mechanisms of the toxicity. For example, in addition to the variability of fetal and developmental responses to particular agents and windows of sensitivity to target effects, there is likely to be a dose or systemic level of an agent below which no effect can be detected (i.e., threshold level of effect). Thus, a finding of several cases of exposed pregnancies with normal outcomes does not ensure that a particular drug does not cause an effect. A product that is known to alter cardiac tissue development, but not known to affect vasculative, skeletal muscle, or bone development, is unlikely to have been responsible for limb reduction defects in the case of a mother exposed only in the third trimester. The type of defect seen and the timing of exposure make this implausible. Assessment of biologic plausibility requires integration and consideration of the dose of a potentially toxic agent, its pharmacology, mechanism and targets of toxicity, variability of response and windows of likely effect, much of which can be learned from careful review of preclinical toxicology data.

V. TRADITIONAL SOURCES OF PREGNANCY OUTCOME DATA

A. General Considerations

Data on pregnancy outcomes related to a drug or other exposure are sought because of concerns about whether an exposure leads to or causes adverse outcomes. Studies to ascertain relationships between exposure and a pregnancy outcome are usually not possible in humans for ethical reasons. The next best method of estimating risk is through the use of observational epidemiology studies.

Analyses of pregnancy study outcomes in association with a particular exposure are similar to other observational or epidemiology studies. Considerations include, but may not be limited to, the following:

- Why was the study undertaken?
- Was the study well designed?
- Was the study protocol followed?
- Does the study sample represent the population of interest?
- Was the study retrospective or prospective?
- Was the outcome assessment thorough and of high quality?
- What were the extent and effect of losses to follow-up?
- Are there other sources of bias? If so, what were their impact on data and their interpretation (e.g., selection or enrollment bias; recall or reporting bias; misclassification)?

- Are there no other exposures likely to account for the observations?

1. Association vs. Causality

Observational data may demonstrate an association between exposure and outcome, but sometimes the exposure may not have caused the outcome. Determination of causality requires rigorous, prospective evaluation. Often the reviewer must balance the desire to confirm causality with further study and the desire to prevent harm, such as through a strongly worded label that warns against use of an agent in pregnancy. To find a reasonable balance, one must evaluate the strength of the association being considered. If the following criteria have been met, a strong case of cause-and-effect relationship can be made between a given exposure and a pregnancy outcome. For pregnancy outcomes in association with particular exposure, conditions that are considered supportive of a cause-and-effect relationship include the following:

- Animal toxicology data have a scientific link to the human outcomes observed.
- The association under review is a strong one.
- Timing of exposure is relevant to the fetal development stage of concern.
- A dose response relationship is evident.
- The association is reversible (e.g. one woman has more than one pregnancy with differential exposure in each).
- There is consistency of findings across studies.
- Biologic plausibility is present.
- The outcome of note is unique or otherwise rare.

2. Bias and Confounding

One of the greatest challenges presented by observational data of all types is assessing the degree to which bias has entered into data collection. Pregnancy outcome data are as susceptible to bias and confounding as any type of data. Systematic bias may lead to the conclusion that a strong relationship exists between an exposure and an effect, or it may work against showing that a relationship exists. This contrasts with the random error that occurs in any study by chance alone. Recall and nonresponder biases and confounding by indication may warrant special attention in epidemiology studies addressing pregnancy-associated risks.

Recall bias is most likely in studies requiring practitioners to report prior prescribing or in requiring women to remember what medicines they took during pregnancy. Where there has been an untoward outcome of a pregnancy, patients and healthcare providers are more likely to recall exposures.

Nonresponder bias, in which the likelihood of response to questioning is related to the exposure of concern, may be subtler. For example, a case control study on intestinal atresias was conducted in which the authors sought to explore whether there was any association with maternal work outside the home during pregnancy. The first data analysis suggested an increased odds ratio for working outside of the home during pregnancy. When the status of nonresponders was evaluated using census data, it became clear that housewives in the case group were more likely to be nonresponders than housewives in the control group. This raised the issue that perhaps the mothers who stayed home during pregnancy and had an abnormal baby felt the investigators would be less interested in their response because, in their view, they had not been exposed to anything (Erickson et al., 1998).

Confounding by indication is common and has substantial implications for pregnancy outcome considerations. Congenital anomalies are known to occur with increased frequency in infants born to women with certain medical conditions, such as diabetes mellitus, epilepsy, and congenital heart malformations, independent of the pharmaceutical agent used to treat the disease. These risks may or may not be increased by medications used to treat them, as is the case for some antiepileptic medications. It is essential to assess the role of the underlying maternal condition in any data set that focuses on pregnancy outcomes. Some common examples of risk for fetal abnormality related to maternal conditions are listed in Table 6.

Table 6. Examples of Risk for Fetal Abnormalities Related to Maternal Condition

Maternal Condition	Increased Fetal Risk
Diabetes, types I or II	Increased body weight; decreased body weight; stillbirth; cardiac, neural tube and skeletal defects
Congenital heart defects	Congenital heart defects
Epilepsy	Increased overall rate of anatomic abnormalities

B. Case Reports of Adverse Pregnancy Outcomes

Postmarketing case reports provide the most common type of human data that associate maternal exposures with pregnancy outcomes, but these data can be the most difficult to interpret. Reports that reach FDA's attention range from voluntary submissions to FDA, to manufacturer reports of individual cases, to reports from medical literature. The reports themselves vary substantially in quality. Their anecdotal nature rarely permits determination of a causal link between a product and an outcome. On the other hand, the appearance of several case reports of a distinct abnormality or group of abnormalities associated with an exposure (a *cluster*) may be very useful. A classic example is the association between thalidomide exposure during pregnancy and phocomelia, which was initially based on the observation of a very large cluster of phocomelia and the implausibility that such an unusual abnormality would be seen so often by chance alone. Another less dramatic example of the value of several cases coming to the attention of an investigator is that of trimethadione. This anticonvulsant was established as a human teratogen predominantly on the basis of case studies with similar abnormalities in nine exposed families (Feldman et al., 1977).

The most important bias in case reports of any type is the tendency for adverse outcomes to be disproportionately reported. Healthcare providers and patients will try to explain adverse outcomes by looking to medication and other exposures. When outcomes are normal, there is no incentive to recall, much less to report, exposures. Despite this bias, overall, the phenomenon of under-reporting of effects is well described. Most sources estimate that, at best, fewer than 10 percent of all effects related to exposures are ever reported in any venue. Reasons for this are multiple, with one of the most commonly cited being physician fear of blame and legal reprisals (and their emotional burdens) resulting from prescription of a drug that is perceived to have led to an adverse outcome (Inman 1985). This is especially pertinent where adverse pregnancy outcomes are of concern.

Failure to include important information in a case report is also common and can make its interpretation challenging if not erroneous. For example, a reporter may not have considered the role of the mother's underlying illness as a predisposition to the outcome and, therefore, did not include any information about it. Similarly, reporters may fail to ascertain, consider, or report concomitant exposures. Consequently, careful follow-up of individual reports often becomes an essential part of an FDA review.

When considering anecdotal data from a broader perspective, cases of effects or outcomes in pregnant patients exposed to a particular agent may be countered with anecdotal reports of normal outcomes. These, too, should be viewed cautiously. Given the variability in response and windows of organ susceptibility with any exposure in pregnancy, it should not be expected that a measurable risk of adverse pregnancy outcome could be excluded based on small samples. For example, the identification of the association in humans between valproic acid used during pregnancy and spina bifida was probably delayed because of an anecdotal report of 12 normal pregnancies exposed to this drug (Scialli 1992). Twelve pregnancies were not enough to detect

the 1 percent rate of affected children after exposure to this medication. On the other hand, it is possible to calculate a statistical bound for the magnitude of a rate of toxicity that is consistent with a finding of zero affected pregnancies in an exposed sample of a given size.

As for any source of information about a given product, the basic data in case reports that describe the circumstances of the exposure and the outcome are the most important data. This includes information about the exposure itself including dose, duration, and gestational time at which it occurred and the presence and timing of additional treatments and interventions. Descriptions of situations in which discontinuation of a drug leading to reversal of effect (dechallenge) and reinstitution led to recurrence of the adverse effect (rechallenge) are classically considered helpful. When a mother has had several pregnancies and those during which a specific exposure occurred all had the same adverse outcomes while those in which there was no exposure were unaffected, this might serve as a model for dechallenge and rechallenge.

Case reports of effects of all types can provide information that is of sufficient concern to warrant regulatory action (e.g., a labeling change), particularly when biologic plausibility is high. Specifically, features that can make a case or case series highly suggestive of a true relationship between an exposure and an outcome include an unusual outcome; outcomes with features reproduced in more than one case; and when an outcome in a human case reflects what is known about windows of sensitivity and target organ susceptibility from animal studies.

C. Epidemiology Studies

Relationships between maternal exposures and pregnancy outcomes are often addressed in cohort and case-control studies. Their large size and ability to control for potential confounders make both types of study potentially valuable in identifying and quantifying associations between exposures and outcomes. This section discusses some of the advantages and difficulties in assessing relationships between exposures and pregnancy outcomes in epidemiology studies. Many of the points here intentionally illustrate and expand on general considerations introduced earlier in this document.

Cohort studies are considered the most appealing type of epidemiology study, but tend to require very large sample sizes. For example, a study of the effect of antihypertensives on pregnancy outcome might compare a group of hypertensive women on one drug with a group of women receiving a different drug, or no therapy. Typically, in such a study the choice of drug treatment is not determined by the study protocol. In other words, the women are not randomly assigned to their exposures, but are instead identified for inclusion in the study by their previously designated exposure status. To ascertain whether a drug is a human teratagen in such a study, one would typically need in excess of 10,000 treated and untreated women, depending on the baseline rate of the event of concern in the population. For example, to assess a possible relationship between a new drug and cleft lip with or without cleft palate, it must be considered that this abnormality

occurs in approximately 1 in 100 live births. Unless the use of the drug in question markedly increases this incidence, thousands of exposed women and a similar number of controls would be required to give the study sufficient power to address the question in a meaningful way.

It is often more practical to address questions about rare adverse outcomes by using a *case-control study*. For example, spina bifida occurs at a rate of approximately 0.1 percent in the United States and so may not occur at all in a cohort of several hundred pregnancies. To ascertain whether a specific drug increases the risk of spina bifida, it might be preferable to compare children with spina bifida to those without spina bifida and explore differences in their in utero exposure to medications, as was done in establishing the association between valproic acid and spina bifida. Case control studies require fewer patients, overall, but their retrospective nature makes misclassification of exposure the most challenging hurdle to overcome.

1. Exposure information

Exposure status in a prospective cohort study is not difficult to ascertain because the exposure can be verified by asking women directly what they are taking or checking prescription records. However, it can be challenging to identify and control for potential confounders of exposure status. For example, in a study of pregnancy outcome in women with epilepsy on anticonvulsant medication compared to women with epilepsy on no medication, it would be difficult to know whether adverse outcomes in the medication exposed group were associated with medication or with severity of maternal epilepsy.

Ascertainment of exposure can be problematic for case control studies as well, because subjects must recall a previous exposure. If women coming to the hospital to deliver, for example, are asked about analgesic exposure during their pregnancies, exposed and unexposed cohorts can be identified. The accuracy of group assignment depends solely on the accuracy of the subjects' recall. Patients rarely recall details of medication courses and their timing, as has been shown in numerous studies, including some of postpartum women (Rubin et al., 1993), making recall bias potentially problematic. Prescription fill data may offer additional information when available, but do not necessarily reflect actual exposure or its timing (e.g., the patient may not have taken the medicine, taken only some of it, or delayed taking it). For over-the-counter products, even these types of confirmatory data are unavailable.

2. Outcome ascertainment

Reliable sources of outcome ascertainment are a critical element of any epidemiology study. There are hundreds of possible birth defects and complications of pregnancy, so even ensuring basic consistency of terminology within a dataset can be a challenge. Loss to follow-up (i.e., no reliable outcome information) also can undermine the best study design. The use of birth certificate information was at one time a common method of ascertaining adverse pregnancy

outcome. However, individuals who have not examined the babies often complete these forms. Diagnoses on the face sheet of the newborn's hospital record may come closer to being reliable, because they are often completed by someone who has been involved in the care of the child. However, diagnoses may arise or be subject to subsequent modification as a child is more thoroughly evaluated or has additional testing after discharge.

The presence or absence of the outcome of interest defines membership in a case-control study; therefore, ascertaining outcomes is less problematic than in cohort studies. Still, considerable planning is necessary to avoid diluting the subjects with the outcome of interest by inclusion of those with the same outcome due to a factor unrelated to the exposure of interest. For example, a study of children with congenital heart disease might evaluate an association with a particular maternal exposure. If children with Down Syndrome are included in the case group, any association between heart defects and the exposure will be difficult to determine, because of the known association of heart defects and Down Syndrome.

3. Evaluation and interpretation of epidemiology study findings

Most epidemiology studies that deal with pregnancy outcomes are designed to answer a question about risk for adverse outcomes in relation to a particular exposure. Risk itself is a reflection of incidence (i.e., number of events/cases arising in a defined population during a given period of time). Results are commonly expressed as a *relative risk* (i.e., how many more times likely exposed individuals are to have a particular adverse outcome than nonexposed individuals). Case-control studies allow for an estimate of risk due to exposure by starting from an assessment of outcomes of interest that have already been identified. They address the probability of having one condition (the outcome of interest) if another condition (i.e., exposure) is present, expressed as an *odds ratio*. Thus, case-control studies do not measure incidence, but prevalence of exposure.

Two potentially controversial issues of data interpretation relevant to epidemiology studies are important to mention. The first is whether to focus concern on a point estimation of risk or the confidence intervals around that estimate. For example, in a case-control study investigating the association between spontaneous abortion and painting as a maternal occupation, Heidam obtained an odds ratio of 2.9, estimating that women with spontaneous abortion were 2.9 times more likely to have been painters than women without abortion (Heidam 1994). The 95 percent confidence interval was 1.0 to 8.8 and, because it included unity, it was concluded that no statistically significant association was demonstrated. Whether the confidence interval includes unity may be less important than its size and the magnitude of the point estimate. In other words, the result from the Heidam study may be considered by assessing the clinical relevance of the association between spontaneous abortion and painting, given the magnitude of this odds ratio of 2.9 and the 95 percent confidence interval of 1.0 to 8.8. Ultimately, all three factors must be considered together in light of other information available about the exposure and outcome of interest.

The second analysis controversy in studies of pregnancy outcomes relates to multiple comparisons. For pregnancy outcomes in particular, epidemiology studies may be designed as exploratory exercises in which a large number of outcomes or exposures are evaluated. When large groups of exposed and nonexposed pregnancies are followed for a variety of adverse outcomes, some differences in outcome are likely to occur by chance alone. If a large population of mothers of children with a particular kind of malformation (e.g., heart defects) are surveyed about hundreds of different exposures, some portion of the exposures will probably give rise to significant odds ratios. These exploratory studies, which pepper the medical literature and tend to incite alarm, are best regarded as generating hypotheses suitable for testing in further studies.

4. Linkage methods

For completeness, it is important to mention the concept of *linkage studies*. Computer technology has enabled studies in which exposure to a product is identified by examining computerized medical files in a medical records database. Pregnancy outcomes can be determined by assessing data from newborn medical records. Associations between exposure and outcome may then be tested by electronically linking maternal and infant files within the larger database. Such linkage studies may be considered a type of cohort study or case control study, depending on the query employed.

As in other epidemiology studies, linkage studies are dependent on the accuracy of identification of exposure and outcome status. For example, birth certificate information in the United States is notoriously inaccurate, but is often used as a basis for entering information into billing or other automated systems. Without an ability to access medical records to confirm data when an apparent association appears from such systems, linkage studies offer little advantage to more traditional epidemiology studies and run the risk of providing seriously unreliable information. (Table 7 provides examples of linkage studies in pregnancy outcome research.) On the other hand, where supporting data are shown to be reliable and verifiable, they offer a rapid means of testing associations.

Table 7. Example of Linkage Studies in Pregnancy Outcome Research

- Scandinavian countries maintain national databases wherein a social security-like number identifies citizens. Birth certificate registries include birth defect information and the citizen numbers of the parents. Comparisons of parental occupations for children with and without specific malformation diagnoses may be conducted to identify potential associations.

- The Michigan Medicaid billing database was used to study women exposed to medications in pregnancy and their infants' outcomes. Outcome status was based on the appearance of a malformation diagnosis on two occasions within the billing codes for the child. A number of associations were identified, as expected by chance alone. To further test associations, the exercise was repeated for a different time period, based on the presumption that if any association arose twice in two independent time periods it would be less likely due to chance alone (Piper et al., 1987).

VI. DATA FROM PREGNANCY REGISTRIES

Pregnancy registries are systematic observational epidemiology studies that assess the health effects of exposure to a particular drug, vaccine, or other exogenous agent during pregnancy. Pregnancy registries are becoming an increasingly likely data source for reviewers. For the most part, general considerations already outlined for epidemiology studies are applicable to the review of data from registries. Attachment A provides a reviewer worksheet that may assist the clinical reviewer when examining data from a pregnancy registry.

FDA may ask the sponsor of an approvable product to provide data on the potential risks of the product in human pregnancy under a phase-4 commitment. Alternatively, the sponsor of a marketed product may wish to obtain additional data via registry studies on potential risks and negative findings associated with the use of the product in human pregnancies to update the product label. Pregnancy registry studies are recognized as one method of identifying major risks associated with a drug or biologic exposure during pregnancy. For products known to adversely affect pregnancy outcomes or the developing fetus, registries may be established to estimate the magnitude of risk. A registry may also be used to identify factors that modify risk and to identify and quantify long-term effects such as delayed development, other neurological impairments, or effects that might be detected in older children exposed in vitro.

Pregnancy registries should be designed to obtain information on the risk of the product to the mother and child with an expected time frame for completion. Examples of products for which it may be appropriate to establish pregnancy registries include:

- New molecular entities for which there is an expectation of use in women of reproductive potential.

- Live, attenuated vaccines or other products with the potential to cause subclinical infection in the mother.

- A product expected to be used commonly by women of reproductive potential, including perimenopausal women.

- Products whose use is continued during pregnancy because they are necessary for conditions associated with high morbidity or mortality.

- Products suspected of adverse effects in human pregnancy based on structure, pharmacologic activity, pharmaceutical class, findings from laboratory animal studies, or spontaneous human case reports.

- Products known to be harmful if used during human pregnancy, but where the magnitude or other risk characterization is unknown.

Pregnancy registries are unlikely to be appropriate in the following situations: (1) there is no systemic exposure, (2) the product is not intended for use in women, or (3) the product is not intended for use in women with reproductive potential.

When establishing a pregnancy registry, participants usually are recruited for voluntary reporting of exposure and outcomes either before or after the pregnancy outcome is known. The person contacting a registry is called *the reporter*. This may be the pregnant woman or the obstetrician caring for her, but may also be a family member, a pharmacist, a nurse, or a pediatrician. The sophistication, accuracy, and detail of the information collected depend on the reporter and that individual's reason for contacting the registry as well as the person's access to the case and medical records. Other operational aspects of the registry can substantially affect its ability to ultimately be informative.

- **Recruitment strategies:** Active recruitment for enrollment in registries may be attempted through product circulars, mailings to physicians, advertisements in medical journals or lay magazines or through pharmaceutical representatives, depending on who the sponsor of the

registry is. The more active the recruitment effort, the more likely that the registry will have vigorous enrollment, hence, a larger denominator upon which to base risk estimates.

- **Enrollment:** The enrollment strategy of a registry is an important determinant of the generalizability of its findings to the population at large. For example, to not miss potentially important individual cases and to increase the size of the registry, many registries collect information on exposures identified prospectively and retrospectively (i.e., after the pregnancy outcome is known). Exposures identified prospectively are inherently more useful and less subject to bias. Those identified after pregnancy outcome is known should be analyzed separately.

- **Identification systems:** Registries must employ a system to allow the woman who is exposed to a product of interest to be followed and the outcome of her pregnancy assessed. This may be through use of a study identification number or another mechanism. Registries should also have a specific system for identifying and recontacting the reporter, whether that reporter is the patient herself or a healthcare provider. Registries that establish contact and solicit prospective reports directly from the woman herself (either solely or in addition to contact with her healthcare provider) have had excellent recruitment with minimal losses to follow up (Piper et al., 1987).

- **Written protocol:** Although it may seem obvious, pregnancy registries should have and follow a standard protocol for enrollment, data collection, and follow-up. This helps to ensure the use of standard methods of obtaining information ensure that all key data are collected.

A. Representativeness and Power

Under ideal circumstances, one advantage of a pregnancy registry is its ability to capture numerator and denominator data (i.e., the number of exposed pregnancies from which affected and normal/unaffected children arise). Most registry conditions are not ideal, however, and some degree of selection or referral bias is likely. Recruitment strategies interface with this in some respects. The registrant (usually considered the patient) or reporter (whether that is the patient herself or her physician or other healthcare provider) must have knowledge of the registry to enroll. Such awareness may be more likely in subsets in certain subsets of the population. For example, studies of registries conducted by teratology information services have shown that enrolled individuals tend to be Caucasian, well educated, and of moderate to upper income brackets, raising concerns about generalizability of the data. This issue of representativeness is a consideration in any type of study, not only registries, as no single study can represent the universe of clinical experience. In registries, the issue of representativeness can be ameliorated, though not eliminated, by choosing an appropriate comparison group. On the other hand, in the

vast majority of situations, representativeness is unlikely to substantially undermine the registry's findings.

In the context of a registry study, power refers to the likelihood of detecting a difference between the exposed group and a reference group, if an actual difference exists. As in any study, power depends on how frequently the adverse outcome occurs in the reference group, how much of an increase needs to be detected, and how many subjects are in the study.

B. Exposure Information

A major determinate of a registry's usefulness is its ability to confirm and specify gestational timing of actual exposures. Exposures that occur outside the window of fetal susceptibility to the toxicity of interest may not be informative. This is one reason that exploratory registries (i.e., those not set up to test a specific association) can be difficult to interpret unless a cluster of events with similar exposure intensity and timing can be documented. An obvious, but easily overlooked aspect of exposure information in the context of pregnancy registries is whether the timing, dose, formulation and duration of exposure can be determined at all. The more access the registry has to the patient directly, and the closer in time that contact is to the actual exposure, the more reliable the recall. In addition, the better the registry's access to and employment of additional sources such as medical records, the more reliable the information.

> *Example:* Teratomycin is a hypothetical new agent marketed for the treatment of infections that are common in young women. The manufacturer sets up a pregnancy registry and promotes it widely. During the first three years, 10,000 first-trimester exposures are registered prospectively. Outcome data show a major malformation rate of 3 percent, identical to that in the general population. The sponsor concludes that there is no increase in risk of major malformations.
>
> In reviewing the data, the FDA reviewer notes that in animal studies, teratomycin seemed to target cardiac valves, and a few cases of human tricuspid abnormalities in infants whose mothers were exposed early in pregnancy were reported several years ago in Europe. The target window for tricuspid valve susceptibility in human development is very short, including only a few days during the sixth week after fertilization. The reviewer returns to the registry data to learn that the number of cases that have exposures that would be informative about the incidence of cardiac malformations is far fewer than 10,000. Also, the relevant background rate of malformed cardiac valves is 0.01 percent, not 3 percent.

C. Follow-Up Systems

Follow-up systems directly affect any registry's usefulness and the necessary number of patients who must be enrolled. The greater the expected loss to follow-up, the more patients that will

need to be enrolled to reach the registry's critical event rate. A rigorous follow-up system is critical to any registry's success. The most successful registries are likely to be those with staff dedicated specifically to manage them and who employ consistent and aggressive procedures to minimize loss to follow-up.

Follow-up systems may be brief in duration (e.g., when the outcome of concern is related to delivery itself) or prolonged and complex (e.g., when the outcome of concern requires neonatal follow-up or assessment of infant development). Planning follow-up requires consideration of how and where patients receive medical care. For example, obstetricians care for pregnant women, but do not care for or follow up on the medical care of their infants. Once pregnancy is completed, the mother's healthcare may be turned over to providers other than an obstetrician. For this reason, when follow-up beyond delivery is desirable, a data collection system that includes the enrolled pregnant women themselves is especially attractive and likely to minimize loss to follow-up.

How long to follow up patients and what amount and type of outcome data to collect are important decisions for any registry in development. Anticipating questions that will arise in the course of data analysis based on what is known from pharmacology and animal data or anecdotal human reports is critical to making these decisions. Sponsors of registries must balance between what would be interesting to know and what is essential, and then decide what data are reasonable to collect. The FDA reviewer faced with the study's results must then assess whether there was appropriate and vigorous pursuit to capture, document, and confirm the relevant data.

D. Outcome Information

Specific outcomes to be measured in a pregnancy registry will depend on findings of concern from laboratory animals and human sources, as well as characteristics of the patient population. For most registries, data on maternal adverse events, labor and delivery, and major categories of pregnancy outcome (e.g., spontaneous abortion, elective termination, fetal death/stillbirth, and live births) are sought. Congenital anomalies identified at birth are also routinely recorded. It may be appropriate to also seek information beyond the birth itself, such as regarding neonatal conditions like hyperbilirubinemia or growth patterns. These decisions will need to be balanced in light of feasibility. Detailed case definition of all outcomes to be measured and how they will be assessed should be specified in the protocol.

Once again, completeness and accuracy of pregnancy outcome reports depend on the reporter and on the registry's access to additional information. Although desirable for tracking, reliance on the patient/mother for outcome information presumes that mothers will know the diagnoses their children have received, or at least whether or not they are presumed to be normal. Verbal reports are less reliable than data confirmed by medical records. Who gives the report is also important for written or medical records. For example, the newborn description in the obstetrician's

delivery note is less reliable than that in the newborn physical performed by the attending pediatrician in the nursery.

Some of the most rigorous registries rely on evaluations by a single individual or group of medical professionals following a specific protocol for examination of outcomes. This may permit more detailed evaluations, such as identification of minor as well as major malformations and allows for the examiner to be *blinded* to the maternal exposure of interest. Although costly in terms of initial resources required, this technique may allow for more informative registry data from a smaller number of patients than a large registry with more limited data collection. For example, Chambers et al., (1996) used this rigorous approach in their registry of pregnancy exposures to fluoxetine, in which all babies were examined for minor malformations by one highly trained individual, a pediatric dysmorphologist (Holmes 1998). Another much larger registry study on the same medication could make no comment about minor malformations because its reporter-based outcome information would not have been sensitive to such diagnoses (Chambers et al. 1996).

Table 8. Example of Registries with Different Outcome Ascertainment Sources

Drug	Outcome Ascertainment Source
Acyclovir	Registry reporter (physicians)[a]
Ca Channel Blockers	Registry reporter (mothers) [b]
Fluoxetine	Selected pediatric examiners[c]

[a] Goldstein et al. 1997
[b] Andrews et al. 1992
[c] Holmes 1998

E. Additional Factors and Influences on Outcome

In addition to the product of interest for a given registry, multiple potential determinants of pregnancy outcome should be considered in enrollment and comparison group selections, including maternal age, obstetric history, family history, medical history, and concomitant exposures. For example, maternal age is a well-described risk factor for trisomy 21, or Down Syndrome, a condition that itself is associated with multiple physical anomalies. A family history of certain anomalies is also a risk factor for the same in any given pregnancy. Already mentioned

is the association of certain maternal medical conditions with abnormal birth outcomes and complications. Information such as this about exposed pregnant women and their medical histories may greatly facilitate interpretation of registry data.

F. Analysis and Interpretation

Each registry will have its own approach and challenges in analysis, depending on what questions it has been designed to answer. There are approaches that should be applied in analysis of most pregnancy registries, however. Most important, registry reports should be stratified into prospective and retrospective status. Prospective reports might be further stratified according to timing of exposure and/or whether the outcome was a spontaneous abortion, elective termination, still or live birth. Criteria used to classify reports that are pending or lost to follow-up should be stated. Also, at least initially, maternal and fetal effects should be analyzed separately. At a minimum, proportions of each with 95 percent confidence intervals should be calculated and comparison to an appropriate reference group made.

Although retrospective reports cannot be used in the calculation of outcome rates, they can occasionally be used in other ways to evaluate potential associations between exposure and pregnancy outcome. For example, retrospective reports can confirm trends observed in prospective reports. They can also be compiled into case series allowing appropriate experts to help identify unique anomalies or syndromes.

All of the factors operative in establishing a registry must be considered in interpreting its findings. For registries that are exploratory and not set up to test a specific association, interpretation can be very difficult. Some associations are likely to be found due to chance, bias or confounders, which may or may not be identifiable.

Ultimately, the reason for a registry's existence may be a clue to potential sources of bias in its enrollment of cases, data collection and outcome ascertainment. For example, unless a registry protocol is designed to specifically capture minor malformations and the use of trained examiners to identify them, they will go undetected. Similarly, a registry that is not established to capture exposures within windows of greatest developmental risk to exposure will be unlikely to detect an effect. For these reasons it is important that an exploratory registry operate meticulously to evaluate clusters of individual types of malformations and outcomes in addition to their overall frequency and patterns of exposure.

Registry data can be most helpful in suggesting areas for further study, whether in a further more targeted registry study, animal studies, or clinical trials. What is most important is that the information they generate be used rationally and responsibly. Above all else, the data generated should be subjected to assessments of the same critical factors used to evaluate all human pregnancy outcome data, especially biologic plausibility. This can be facilitated by engaging the

collective consultation of experts in human teratology, clinical genetics and epidemiology in their interpretation.

VII. OVERALL ASSESSMENT OF POSTMARKETING HUMAN DATA

No formula exists by which an association between a drug, biological, or other product and an adverse reproductive outcome can be considered to prove a cause-effect relationship. Evidence from a number of sources, including human and nonhuman data, must be considered collectively to determine the strength of their relationship. It is particularly useful to evaluate the biological plausibility of a putative association, based on the known pharmacologic and toxicologic properties of the medication and perhaps on structural similarities to other entities with better characterized reproductive toxicity. In addition, the pharmacologic effect of a medication may have important toxicologic ramifications for susceptible individuals such as newborns. Even when many elements of relationship appear strong, they still may not prove causality. Assessing those elements and coming to a judgment of their value and need for regulatory action (e.g., labeling change, obtaining a sponsor's commitment to conduct a formal study) is where the judgment of the medical reviewer is critical.

In coming to conclusions regarding reproductive and developmental toxicity, the value of experimental animal models should not be underestimated. Direct extrapolation to human risk assessment may not be possible or appropriate, but animal studies often assist in the understanding of human data. An overall interpretation of reproductive risk associated with a chemical exposure relies on an integrated evaluation of data from all sources, experimental and clinical, human and nonhuman. If, for example, an exposure produces no developmental toxicity in appropriately performed studies in two experimental species and if there is no apparent increase in adverse pregnancy outcome among 600 patients in a registry, the conclusion of lack of adverse effect of the medication is better supported than if the registry data were the only available information.

For drug and biological products, conclusions regarding risks associated with exposure should be communicated in the package insert or labeling. There are no medications for which the labeling could say that all exposed pregnancies would be adversely affected, and only a few for which the label could say that most pregnancies would be affected. If a relationship between medication use and adverse outcome exists, it is more typical for the relationship to represent a relatively small increase in background risk. Therefore, it is probably not useful to approach the review of data with the question: Is this medication teratogenic? What is useful is to consider the question: What is the apparent magnitude of risk of adverse outcome associated with this medication and is that risk demonstrably greater than the incidence among unexposed pregnancies?

As a final note, it is also important that the reviewer consider maternal toxicity as well as other potential reproductive endpoints, such as fertility and premature ovarian senescence, which could

reflect germ cell toxicity or effects on the hypothalamic-pituitary-gonad axis and could be relevant to potential pregnancy effects such as viability and growth of the conceptus. In addition, experimental animal studies in this field contain segments in which reproductive parameters in addition to pregnancy are evaluated. For a more complete integration of the experimental animal and human data, comparable endpoints need to be considered. Pharmacologic and general toxicologic features of the product of interest remain paramount in evaluating the potential for human reproductive effects. Furthermore, if a pharmacologic effect occurs in an adult, it should be considered that a comparable effect in the fetus might occur and might produce toxicity. Consequently, there is a direct link between consideration of exposure in pregnancy, with effects of a product on the fetus and what is known about its toxicity profile in newborns and infants, all of which should be communicated with a clear connection in the package insert whenever possible.

BIBLIOGRAPHY

Andrews, E. B., B. C. Yankaskas, J. F. Cordero, K. Schoeffler, S. Hampp, and the Acyclovir in Pregnancy Registry Advisory Committee, 1992, "Acyclovir in Pregnancy Registry: Six Years' Experience," *Obstet Gynecol.*, 79:7-13.

Bjerkedal, T., A. Czeizel, J. Goujard, B. Källén, P. Mastroiacova, N. Nevin, G. P. Oakley, and E. Robert, 1982, "Valproic Acid and Spina Bifida," *The Lancet*, 2:1096.

Chambers, C. D., K. A. Johnson, L. M. Dick, R. J. Felix, and K. L. Jones, 1996, "Birth Outcomes in Pregnant Women Taking Fluoxetine," *N Engl J Med.*, 336:1010-5.

Ebert S.N., X. Liu, and R.L. Woosley, 1998, *Journal of Women's Health*, Vol 5, No. 5.

Erickson, A., B. Kallen, O. Merik, and P. Westerholm, 1992, "Gastrointestinal Atresia and Maternal Occupation During Pregnancy, *J Occup Med.*, 24, 515.

Feldman, G.L., D.D. Weaver, and E.W. Lourien, 1977, "The Fetal Trimethadione Syndrome. Report of an Additional Family and Further Delineation of This Syndrome," *Am J Dis Child.*, 131, 1389.

Food and Drug Administration, 1997, *E2C Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs*, (FDA 5/19) May 19, 1997.

Goldstein, D. J., L. A. Corbin, K. L. Sundell, 1997, "Effects of First Trimester Fluoxetine Exposure on the Newborn," *Obstet Gynecol.*, 89:713-8.

Heidam, L. Z., 1994, "Spontaneous Abortion Among Dental Assistants, Factory Workers, Painters, and Gardening Workers: A Follow Up Study," *J Epidemiol Comm Health*, 38:149-55.

Heinonen, O.P., D. Slone, and S. Shapiro, 1997, "Birth Defects and Drugs in Pregnancy," *Publishing Sciences Group*, Littleton, MA.

Hiilesmaa, V. K., A. H. Bardy, M. L. Granstrom, and K. A. Teramo, 1980, "Valproic Acid During Pregnancy," *The Lancet*, 1:883.

Holmes L., 1998, North American Registry for Epilepsy and Pregnancy: A Unique Public/Private Partnership of Health Surveillance, abstract, Drug Information Association Annual Meeting.

Inman W., 1985, "Under-Reporting of Adverse Drug Reactions," *Br Med J.*, 290 (6478):1355.

International Clearinghouse for Birth Defect Monitoring Systems, 1991, *Congenital Malformations Worldwide,* New York: Elsevier.

Leppig, K.A., M.M. Werler, C.I. Cann, C.A. Cook, and L.B. Holmes, 1987, "Predictive Value of Minor Anomalies, I, Association With Major Malformation," *J. Pediatr.,* 110: 530-537.

Magee, L. A., B. Schick, A. E. Donnenfeld, S. R. Sage, B. Conover, L. Cook, P. R. McElhatton, M. A. Schmidt, and G. Koren, 1996, "The Safety of Calcium Channel Blockers in Human Pregnancy: A Prospective Multicohort Study," *Am J Obstet Gynecol.*, 174:823-8.

March of Dimes Birth Defects Foundation, N.Y., 1996, *Birth Defects and Infant Mortality: A National and Regional Profile*.

Moore, K.L. and T.V.N. Persaud, 1998, *The Developing Human: Clinically Oriented Embryology*, 6th edition, W.B. Saunders: Philadelphia.

Oakley, G.P., 1986, "Frequency of Human Congenital Malformation," *Clin Perinatol,* 13:545.54.

Piper, J. M., C. Baum, and D. L. Kennedy, 1987, "Prescription Drug Use Before and During Pregnancy in a Medicaid Population," *Am J Obstet Gynecol.*, 157:148-156.

Robert, E., 1996, "Treating Depression in Pregnancy," *N Engl J Med.*, 335:1056-8.

Rubin, J. D., C. Ferencz, C. Loffredo, and the Baltimore-Washington Infant Study Group, 1993, "Use of prescription and Non-Prescription Drugs in Pregnancy, *J Clin Epidemiol.,* 46:581-589.

Scialli, A.R., 1992, *A Clinical Guide to Reproductive and Developmental Toxicology*, CRC Press: Boca Raton, FL.

DRAFT - Not for Implementation

ATTACHMENT

Reviewer's Worksheet for Pregnancy Registries

The checklist below is provided to serve as a worksheet for medical officers when pregnancy registry data are reviewed. Its format does not directly correspond to examples in the Appendex. Also, some of the points in the checklist may require collaboration with other FDA professionals, such as a statistician. Nonetheless, this checklist may assist the medical reviewer in organizing what might otherwise be complex data.

Criteria	**Ranking**
A. Participant enrollment • Was all enrollment prospective? • Were women actively recruited or did they initiate contact? • When was the exposure to the target drug ?	
B. Data collected at enrollment • What was the timing of enrollment from exposure? From identification of pregnancy? • What was the timing of enrollment? • Was there accurate assessment of dose, timing and extent of exposure? • What additional information was collected, (e.g., indication, medical history, past pregnancies and their outcomes, socio-demographics, smoking, alcohol consumption)?	
C. Data collected during follow-up • What was the method and extent of follow-up? How frequent were the contacts? • Were drug exposures and other information properly ascertained at each contact? • Was loss to follow-up minimized? • Could differential follow-up have biased the results?	
D. Determination of pregnancy outcome • How and when were major anomalies ascertained for births and abortions? • How and when were minor anomalies ascertained? • How were additional and secondary outcomes defined and ascertained? • Were physical or pathological examinations conducted? How and by whom?	
E. Analysis and Interpretation • Was the risk of each outcome properly calculated with confidence intervals?	

- Was the minimal detectable risk (properly) estimated?
- Was variation in risk assessed?
- Are groupings and analyses appropriate?
- How do the results change if all lost had adverse outcome or none had adverse outcome?
- Are the study results valid in light of the proportion lost to follow-up?
- Does the medical condition elevate the risk of the adverse outcome? Could the drug further elevate that risk?
- Was the comparison group appropriate?
- Did the investigators compare anomaly rates among women with the same medical condition, pregnancy history, socio-demographics?