Chapter 2

Policies for Testing, Assessing, and Regulating Carcinogens

CONTENTS

Boxes

Figures

Table

Chapter 2

Policies for Testing, Assessing, and Regulating Carcinogens

INTRODUCTION

Over the last dozen years, health, safety, and environmental regulatory agencies have issued guidelines and policies on how they intend to identify, evaluate, and regulate carcinogens. Some guidelines and requirements address the design of toxicity tests in animals. Other policies describe the kinds of evidence, human or animal, that the agencies will use to identify and evaluate carcinogens. In these policies, agencies have given considerable attention to methods for predicting the nature and extent of possible human health risks based on human and animal data.

Some of the important issues in assessing potentially carcinogenic chemicals turn on the interpretation of test data, others on the use of assumptions (or "inference options"). These assumptions are derived from theories about cancer causation and decisions about appropriate public policy. OTA has identified four important kinds of assumptions:

1. assumptions used when data are not available in a particular case;
2. assumptions potentially testable, but not yet tested;[1]
3. assumptions that probably cannot be tested because of experimental limitations; and
4. assumptions that cannot be tested because of ethical considerations.

The lack of data and use of risk assessment assumptions, especially in conjunction with underlying political disputes about the desirability of government regulation, make this area of research the subject of lively debates.

This chapter will describe and compare the Federal agency policies that attempt to resolve certain issues in identifying carcinogens and assessing human risks. These policies include the guidelines on the design of animal bioassays for carcinogenicity, the guidelines governing the regulatory use of human epidemiologic data, animal toxicology tests and other information on toxicity, and the procedures for combining all this information in risk assessments.

The study of carcinogenesis is advancing rapidly. In this chapter, OTA has not attempted to summarize current scientific understanding, but only to describe and compare Federal agency policies on testing, assessing, and regulating carcinogenic chemicals. In addition to following procedures described in this chapter, agencies must also, prior to regulatory action, meet certain other statutory requirements. Depending on the statute, these may involve determining that the estimated risk is unreasonable or significant, that exposure reduction is technologically achievable, that the costs of control are economically achievable or proportionate to the benefits anticipated, and that the relevant statute authorizes regulatory activity for that hazard. These additional steps are not discussed in this chapter.

Types of Evidence

Four kinds of evidence may be used for qualitatively identifying carcinogens: epidemiologic studies, long-term animal bioassays, short-term tests, and structure-activity relationships. (See ref. 217 for a more detailed discussion of methods for identifying carcinogens.)

Epidemiologic studies collect information about human exposures and diseases. Reports of individual cases or clusters of cases are very often used to generate hypotheses for later study. In fact, many of the chemicals now determined to be human carcinogens were first identified in case reports by astute physicians. Larger epidemiologic studies are divided into descriptive, or correla-

[1] One important area of research is testing such assumptions and developing new experimental methods. Such work is taking place at the National Institute of Environmental Health Sciences and National Center for Toxicological Research.

tional, studies and analytic studies. Descriptive epidemiologic studies correlate risk factors (including exposures) and diseases or causes of death in populations. They are useful in generating hypotheses for further study and in providing clues about potential hazards. Analytic epidemiologic studies use comparison populations. In cohort studies, the comparison is made between a group exposed to the agent of interest and a group that is not exposed. For case-control studies, the comparison is made between people with a given disease and those without the disease.

Long-term animal bioassays are laboratory studies in which animals are exposed to a suspected hazard (for about 2 years in the case of rodents). The animals are examined for the presence of tumors and other signs of disease throughout the study. At the end of the study, the surviving animals are sacrificed. Tissues from these animals and from those that died during the study are given gross and microscopic examinations and tumors are diagnosed. The incidence of tumors in exposed and control groups is then compared.

Short-term tests examine genetic changes in laboratory cultures of cells, or in humans or other animals, or in lower organisms. These tests take relatively little time to perform. Short-term tests can be completed in days, weeks, or a few months, rather than requiring the several years needed to complete a bioassay in rodents.

Structure-activity relationships (SARs) in this context refer to associations between chemical structures and carcinogenicity. In a sense, judgments about them are "paper chemistry, " because predictions are made about the carcinogenicity of substances based on previously observed associations between structure and toxicity, but without additional toxicity testing. The predictive value of using SARs is highest for chemicals within a class of closely related chemicals for which extensive carcinogenicity testing has already been conducted. Predictions based on SARs are less certain for classes of chemicals less extensively tested.

Many of the Federal carcinogen guidelines discuss the different roles to be played by the different kinds of evidence, as is discussed below. All of these policies value positive epidemiologic studies as the most conclusive evidence for human carcinogenicity, they generally presume that substances carcinogenic for animals in long-term bioassays should be treated as carcinogenic for humans, and they treat short-term test results as supporting information.

In practice, regulatory activity maybe initiated based on positive human or long-term animal data. In most cases, if the only evidence consists of short-term test results, agencies will not initiate regulatory action to reduce exposures, although such test results might be the basis for requiring further animal testing. SARs are used mostly when no other data are available, for example, to identify new chemicals for which further testing is warranted prior to large-scale manufacture.

The relative ranking of these types of evidence is often an academic issue because for many types of chemicals, there are often few toxicity data of any sort, whether from human epidemiology, long-term animal bioassays, or short-term tests (138). In these situations, Federal agencies may be hampered in their efforts to protect public health.

Risk Assessment and Risk Management

It is common now to distinguish between risk assessment and risk management (111). This language was adopted in the report of a Committee on the Institutional Means for the Assessment of Risk to Public Health convened by the National Research Council of the National Academy of Sciences (NAS) (137). This committee described risk assessment as the process of characterizing the adverse health effects of human exposures to environmental hazards. Risk assessment relies on information from epidemiologic, clinical, toxicologic, and environmental research. Risk management, on the other hand, is the process of evaluating and choosing among regulatory options, based on information on economic, social, political, and engineering factors, as well as information on risk.

Some of the agency policies described below also outlined distinctions between risk assessment and risk management, predating the NAS report.

The Environmental Protection Agency (EPA) (1976) (293) describes two decisions: whether a substance poses a cancer risk and what regulatory action, if any, should be taken to reduce risk. The National Cancer Advisory Board (NCAB) (1977) (348), the oldest of the policy documents considered here in detail, argues that scientists play a major role in evaluating benefits and risks by providing and interpreting data, but "the final decision . . . must be made by society at large through informed governmental regulatory and legislative groups." Thus, the division, real or perceived, between "scientific data and interpretation" and "political decisions" has been noted for some time.

Risk assessment determines the qualitative nature of the risk posed by particular exposures to chemical or physical agents and quantifies the dimensions of that risk. The term "risk" has been used in many ways. OTA uses "risk" to mean the combined effects of the intrinsic hazard presented by the agent in question and the degree of exposure. Thus, an inherently very toxic agent may pose little risk when exposure levels are very low. Conversely, an agent of low intrinsic toxicity may be an important public health problem because a large number of people are exposed at fairly high levels.

Qualitative and Quantitative Risk Assessments

One distinction, frequently made in discussing policies on carcinogen regulation, is between the qualitative determination of a hazard and the quantitative evaluation of risk. The qualitative determination is a "yes" or "no" answer to the question: Does substance X cause cancer? These decisions may be difficult and may even include some quantitative analysis. For example, statistical techniques are used to determine whether an exposed group of people or animals have a significantly higher than expected incidence of tumors. In addition, qualitative determination depends on some interpretation, such as views on whether animal carcinogens are presumed to be human carcinogens, or whether benign tumors in animals indicate a hazard for humans.

Quantitative risk assessment starts with the qualitative determination that a substance does cause cancer and then goes on to ask: To what extent does exposure to a particular agent cause tumors? The answer involves four separate analytic exercises: developing a mathematical description of the dose-response relationship, extrapolating from animal data to human effects, developing information on human exposure levels, and using all this information to estimate individual risks and the number of expected cases in the human population.

Instead of "qualitative" and "quantitative" risk assessment, the NAS Committee on Risk Assessment used the terms hazard identification, dose-response assessment, exposure assessment, and risk characterization (137). These terms more clearly describe the separate analytic steps in a risk assessment, although the older terms will also be used in this background paper.

Hazard identification determines whether exposure to an agent increases the incidence of an adverse condition, for example, cancer in test animals. Dose-response assessment describes the relationship between the level of exposure or the dose and the incidence of disease. The two most important aspects of this step are extrapolating from information on incidence at high doses to predict incidence at lower doses and, in the case of risk assessments based on animal data, converting animal doses into equivalent human doses. Exposure assessment estimates the frequency, duration, and intensity of human exposures to the agent in question. Finally, risk characterization uses information from both dose-response and exposure assessments to estimate the expected incidence of the adverse health effect.

Inference Guidelines and Policies

All the steps described above involve uncertainties, some owing to the lack of data on particular agents, some to lack of knowledge concerning the causes and mechanisms of toxicity. Where the science is uncertain, inferences must be made. The NAS Committee used the term "components" to refer to the various points in the process where the risk assessor must choose among "scientifically plausible options. " For example, one component

63-986 0 - 87 - 2

would be the number of animal studies needed to be sure that the substance in question is truly a carcinogen. Some people are willing to act based on a single study in a single species, others want confirmation in a second species.

An "inference guideline" consists of assumptions that must be made to estimate human risk. The NAS Committee defined "risk assessment policy" as "the analytic choices that must be made in the course of a risk assessment. Such choices are based on both scientific and policy considerations" (137).

An agency might also adopt a risk management policy for choosing among regulatory options. In the committee's view, risk management policies should not be allowed to control risk assessment policy. While risk assessment and risk management are commonly distinguished, both are based on policy choices.

In addition to risk assessment guidelines and risk management policies, agencies have developed guidelines for conducting and evaluating animal toxicity tests. To some extent, these testing guidelines overlap with risk assessment guidelines. For example, both might specify whether benign tumors are to be considered with malignant tumors when evaluating the results of animal tests.

Agency policies and guidelines have varied in the degree of formality and in the basic approach they take toward evaluating evidence for risk assessments. Some policies, notably the cancer policy issued by the Occupational Safety and Health Administration (OSHA) and the sensitivity of method (SOM) guidelines proposed by the Food and Drug Administration (FDA), are intended to be binding regulations and were subject to notice-and-comment rulemaking. Other guidelines have been developed more informally by agency staff, printed, and made available to the public.

Rushefsky has classified agency carcinogen policies into three types: presumption-rebuttal, weight-of-the-evidence, and leave-it-to-the-scientists (180). The OSHA policy (276) represents the presumption-rebuttal approach. This policy approach uses the regulatory process, establishes "presumptions" and sets stringent conditions on when and how these presumptions may be "rebutted." Other policies, particularly the latest policies of the White House Office of Science and Technology Policy (OSTP) (351) and the carcinogen risk assessment guidelines of EPA (284) take a weight-of-the-evidence approach, in which all relevant data are used. A weight-of-the-evidence approach is more flexible, and in implementation by the agencies, is more open to considering negative, as well as positive, data on carcinogenicity. The OSHA policy, on the other hand, restricted the circumstances in which negative data could be considered. In the third approach, leave-it-to-the-scientists, a separate body for conducting risk assessments is established. This represents the clearest separation of risk assessment from risk management. According to Rushefsky, only one agency policy, a paper prepared by OSTP staff in 1979 (23), adopts this approach, although other proposals for creating centralized science panels for developing or reviewing risk assessments are of this type (for a discussion of these proposals see ref. 217).

Interest groups have differed in their preferences for the different approaches. Industry groups have often supported various proposals to centralize risk assessments, while labor, public interest, and environmental organizations have opposed such proposals. Industry groups have also strongly endorsed the weight-of-the-evidence approach. Because of the importance labor, public interest, and environmental organizations place on the potential for harm to health, these groups want regulatory agencies to act on limited positive evidence and often when industry thinks the weight of the evidence does not support action.

Policies also vary in length, amount of detail, and complexity. Some, like EPA's "interim" guidelines of 1976 are only a few pages long, while, for example, the explanation of OSHA's policy occupies nearly 300 pages in the *Federal Register*.

Utility of Policies

The NAS committee cited above recommended that agencies adopt uniform risk assessment policies. Such guidelines have the advantages of promoting quality control, consistency, predictability, public understanding, administrative efficiency, and improvements in risk assessment

methods. Potential disadvantages include oversimplification, inappropriate mixing of scientific knowledge with risk assessment policy, misallocation of agency resources to guideline development, and the freezing of science (137). An important use of guidelines within the agencies is in training junior staff in agency practices and procedures.

Some have hoped that risk assessment might be conducted as a neutral, nonpartisan, scientific enterprise. However, inference choices are necessary, and these, although often based on scientific understanding, are not empirically tested. Some hypotheses are extremely difficult to test experimentally; for instance, determining the doses that cause an increase in cancer risk of 1 percent would demand the use of 1,600 laboratory animals.[2] Others raise ethical issues, for example, evaluating the predictive value of animal test data by exposing human subjects to suspect carcinogens to follow them prospectively. Political and social values may also be reflected.

Policies may also reflect agency judgments on the acceptability of errors. From a regulatory perspective, two risks must be balanced:

> The first is the risk of taking precautionary action for a safe chemical (a regulatory false positive). The second is the risk of not controlling an unsafe chemical . . . (a regulatory false negative) **(154).**

The appropriate evaluation of an agency policy would not then be seen in whether the agency correctly identified every carcinogen and every noncarcinogen and placed them into the correct categories. Evaluation should be based on the overall success of the policy in improving public health. An important part of this is considering the costs of delaying public health protection (174). Some, however, argue that agency efforts to adopt "conservative" assumptions for developing risk assessments are misguided, leading to substantial overestimates of actual risks and distorting agency priorities (149,178).

In addition, agency guidelines are not, by themselves, sufficient to surmount two regulatory hurdles: the different perspectives of various interested parties and the importance of case-by-case interpretation.

In regulatory proceedings, the opinions of industry, labor, environmental groups, public interest organizations, and government are often substantially different. These groups place different values on the harm caused by unnecessarily regulating a chemical that later turns out to be safe and the harm caused by not regulating a chemical that turns out to be harmful. In a survey, Frances Lynn found evidence that there are links between political values, place of employment, and scientific beliefs. For example, industry scientists in Lynn's sample were less willing to accept animal data on carcinogenicity and more likely to believe in the existence of no-effects thresholds for carcinogens (see the discussion below) than were government scientists (112). In describing the history of Federal "cancer policies," Rushefsky points to the importance of political values in explaining some of the features of risk assessment policies (179,180). Another source of different perspectives is the various disciplines participants are trained in and the various scientific paradigms they work under **(86).**

Even when the agencies have policies, there will always be issues of interpretation in particular cases, especially for "flexible" policies. For example, if a policy establishes five categories, questions will arise on which category applies to a particular chemical. Even accepting the value of animal data in general, much regulatory debate on carcinogens centers on whether a particular animal study is reliable and on whether the data apply in particular cases. Arguments on particular cases are not likely to disappear, especially for commercially important chemicals.

[2] With a 95-percent confidence limit ranging from 0.5 percent to 1.5 percent (64).

HISTORY OF AGENCY POLICIES

The Food and Drug Administration

FDA was the first agency to set guidelines for toxicity assessment. FDA has responsibility for regulating the safety of foods, drugs, cosmetics, and medical devices. The 1958 Food Additives Amendment to the Food, Drug, and Cosmetic Act includes the Delaney clause, which proscribes the

intentional use of food and color additives determined to be carcinogenic in either humans or animals. This clause does not apply to all food ingredients because some were considered to be "generally recognized as safe" or had been federally sanctioned prior to the 1958 amendment. Nevertheless, the general FDA policy (until recently) has been to ban food and color additives whenever they were determined to be carcinogenic. FDA has not explicitly specified any guidelines on interpreting carcinogenicity data.

FDA has specified the protocols for developing the animal data necessary to evaluate the safety of food and color additives. FDA first published toxicity testing guidelines, consisting of a series of papers by staff scientists, in a 1955 journal article (109). A revised version was published as a book in 1959 (267).

In 1970, an FDA advisory committee on protocols for safety evaluation prepared a report on designing experiments and on using animal data. FDA also made recommendations, specifying the use of at least two species, the maximum tolerated dose, and a two-generation bioassay design, in which exposure begins prior to conception and continues throughout the lifetime of the offspring. Reflecting the state of the science then, the committee concluded that "at the present time there is not enough information available to provide a basis for recommending any rapid [i. e., short-term] test for carcinogenicity" (247). In 1982, FDA updated its guidelines on conducting animal toxicity tests (in the FDA "Red Book") (248).

As described later in this chapter, in the 1970s FDA began using quantitative risk assessments for certain environmental contaminants found in food. In the 1980s, FDA began applying these techniques to food and color additives, both when color additives are contaminated with small amounts of carcinogenic impurities and when the additive itself is determined to be carcinogenic. (FDA procedures for using such risk assessments are discussed in ch. 3.)

For animal drug residues in human food, the "DES proviso, " part of the drug amendments of 1962, prohibits carcinogenic drug residues that can be detected by analytic methods approved by FDA. For years FDA has been working on a regulatory definition of what these approved methods would entail. The general label for these regulatory requirements is "sensitivity of method" or "SOM. " SOM procedures were first proposed in 1973, finalized in 1977, challenged in court, withdrawn in 1978, and reproposed in 1979.

The 1973 proposal suggested use of a modified Mantel-Bryan procedure for extrapolating from effects at high doses to those at low doses (see the discussion later in this chapter on extrapolation models) and a risk cutoff of 1 in 100 million. This number means that exposures at the permissible limit would be associated with an upper bound estimate of 1 in 100 million people exposed.[3] In 1977, FDA issued a final rule keeping the Mantel-Bryan procedure but changing the risk cutoff to 1 in 1 million. The reproposal in 1979, kept this cutoff figure, but adopted linear extrapolation.

Subsequently, the responsibility for these regulations was transferred from FDA's Center for Food Safety to its Center for Veterinary Medicine. The guidelines were then reproposed in October 1985 (246). An approved analytic technique is defined as one that could detect residue concentrations as low as the level associated with an upper-bound human risk estimate of 1 cancer for every 1 million persons exposed. Of course, this technique requires a risk assessment to estimate what residue levels correspond to this particular risk level.

In 1968 and 1973, FDA published guidelines on required toxicity information for investigating and marketing new human drugs (67). These guidelines specified an 18-month rat study and a 12-month study in dogs or monkeys, which was intended to cover both chronic toxicity and carcinogenicity. A 12-month rat study and a mouse carcinogenicity study could be substituted for the 18-month rat study (67).

In the 1970s, FDA and the Pharmaceutical Manufacturers' Association (PMA) convened a workshop to discuss toxicity testing for drugs, in-

[3]Agencies have not always distinguished clearly between risk estimates based on all cases of cancer and those based only on cancer deaths. Depending on the tumor site, the two estimates can differ (124). In this case, the FDA proposal referred only to "a minimal probability of risk to an individual (e.g., 1/100,000,000) . . ." (246).

eluding the length of carcinogenicity studies. While the workshop had been convened with the expectation that new guidelines would be issued, FDA decided not to update its own guidelines at that time. PMA, however, published guidelines in 1977 that reflected the workshop's consensus in requiring longer duration studies in two species (67). For carcinogenicity study designs for human drugs, FDA staff also refer to the "Red Book" guidelines for toxicity testing of food and color additives, the documents published by OSTP, and the report of the National Toxicology Program (NTP) Ad Hoc Panel on study design (258). No new formal guidelines for testing drugs have been issued, although FDA staff state that they are being developed (249).[4]

Nevertheless, in reviewing new drug applications, there is general understanding between FDA and industry about the evidence needed to obtain approval. The kinds of tests needed depend on the stage of clinical investigation and approval process, and the expected duration of human use of the drug (e. g., several days, up to 2 weeks, up to 3 months, 6 months to unlimited use). (For a summary, see ref. 218.)

For drugs expected to be continuously administered for 6 months or more, an application to conduct a Phase I or Phase II clinical investigation must include the results of 3-month animal toxicity studies conducted in two species. To initiate a Phase 111 trial, there must be information from two species given the drug for 6 months or more as part of ongoing studies of chronic toxicity and carcinogenicity. A New Drug Application (for a drug intended for chronic or repeated use in the general population) must now include the results of 18- to 24-month chronic studies in two rodent species (usually rats and mice) and a 12-month chronic study in a rodent species and a nonrodent species (e.g., dogs or monkeys).

FDA evaluates the evidence in the New Drug Application for therapeutic efficacy and potential risks of the drug. If FDA judges that the risks outweigh the benefits, the drug is not approved for marketing. If the benefits are thought to outweigh the risks, the drug is approved, but the labeling for the drug will discuss potential hazards, including any animal evidence for carcinogenicity (66). For any particular drug, the final decision depends on how "persuasive" or "alarming" the tumorigenic finding is, expected use of the drug, and the nature of alternative therapies (69).

[4]FDA has provided guidance for statistical analysis of data for studies of human drugs, but this will not be discussed here.

The Environmental Protection Agency

EPA began developing carcinogen assessment guidelines during regulatory proceedings on the suspension and cancellation of several pesticides. In legal briefs written at the end of those proceedings, EPA attorneys summarized the expert testimony that the agency had received on evaluating carcinogenicity. These summaries were referred to as "cancer principles. " (See box *2-A.)*

Partly in response to criticism of these cancer principles, EPA established a permanent organizational unit, the Carcinogen Assessment Group (CAG), within EPA and developed a new set of guidelines (9,122,137). In May 1976, EPA published "interim" guidelines for assessing the health risks and economic impacts of suspected carcinogens (3,293). The text and explanation of these guidelines occupied less than four pages in the Federal *Register.*

In November **1977,** the Environmental Defense Fund petitioned EPA to establish a policy on classifying and regulating carcinogenic air pollutants. In October 1979, EPA published its proposed airborne carcinogen policy. This policy has never been issued in final form, although agency staff indicate that they follow the outlines of this policy (103).

EPA issued water quality criteria documents under the Clean Water Act in response to a court order to assess the hazards and risks posed by a large group of substances. (See ch. 3 for details on the development of this list.) In March 1979, EPA made available a methodology for assessing human risk (methods for assessing other aspects of water quality, e.g., the hazard to aquatic life forms, had been prepared earlier). In November 1980, EPA announced the availability of the water quality criteria documents and published sum-

Box **2-A.—Development** of "Cancer Principles" at the Environmental Protection Agency

The substance of EPA "Cancer Principles" originated in the work of a group of scientists assembled by National Cancer Institute (NCI) scientist Umberto Saffiotti. In **1970,** the group prepared a report to the Surgeon General, "Evaluation of Chemical Carcinogens." This report responded to another report prepared by the Food Protection Committee of the National Research Council of the National Academy of Sciences. This committee had suggested that regulators might allow potential carcinogens to be added to foods at "toxicologically insignificant levels." The committee also suggested that some substances might be considered safe without undergoing testing, if they had "been in commerical production for a substantial period" and that a "no carcinogenesis level" might be shown for an animal species, although there were no generally accepted ways of translating this threshold level to humans (122).

The 1970 report (250) by the Surgeon General's ad hoc committee represents one of the first "guidelines" for evaluating potential carcinogens:

> Any substance which is shown conclusively to cause tumors in animals should be considered carcinogenic and therefore a potential cancer hazard for man . . .
>
> No level of exposure to a chemical carcinogen should be considered toxicologically insignificant for man. For carcinogenic agents a "safe level for man" cannot be established by application of our present knowledge. The concept of "socially acceptable risk" represents a more realistic notion . . .
>
> No chemical substance should be assumed safe for human consumption without proper negative lifetime biological assays of adequate size. The minimum requirements for carcinogenesis bioassays should provide for: adequate numbers of animals of at least two species and both sexes with adequate controls, subjected for their lifetimes to the administration of a suitable dose range, including the highest tolerated dose, of the test materials by routes of administration that include those by which man is exposed . . .
>
> Evidence of negative results, under the conditions of the test used, should be considered superseded by positive findings in other tests . . .
>
> The implication of potential carcinogenicity should be drawn from both tests resulting in the induction of benign tumors and those resulting in tumors which are more obviously malignant. . . .
>
> The principle of zero tolerance for carcinogenic exposures should be retained in all areas of legislation presently covered by it and should be extended to cover other exposures as well. Only in the cases where contamination of an environmental source by a carcinogen has been proven to be unavoidable should exception be made to the principle of zero tolerance. Exceptions should be made only after the most extraordinary justification, including extensive documentation of chemical and biological analyses and a specific statement of the estimated risk for man, are presented. All efforts should be made to reduce the level of contamination to the minimum. Periodic review of the degree of contamination and the estimated risk should be made mandatory.
>
> No substance developed primarily for uses involving exposure to man should be allowed for wide-spread human intake without having been . . . tested for carcinogenicity and found negative. . . . Any substance developed for use not primarily involving exposure in man but nevertheless resulting in such exposure, if found to be carcinogenic, should be either prevented from entering the environment or, if it already exists in the environment, progressively eliminated . . .
>
> A unified approach to the assessment and prevention of carcinogenesis risks should be developed in the federal legislation; it should deal with all sources of human exposure to carcinogenic hazards . . .
>
> An ad hoc committee of experts should be charged with the task of recommending methods for extrapolating dose-response bioassay data to the low response region . . .

At the EPA hearings on canceling registration of DDT, Saffiotti included parts of the ad hoc committee report in his testimony.[1] In the brief, which summarized the evidence in the DDT cancellation decision (315), EPA attorneys listed seven "general principles" for determining carcinogenic hazards that were drawn from the ad hoc committee report:

1. Any substance shown conclusively to produce tumors in animals should be deemed potentially carcinogenic in man, except when the effect is caused by physical induction, or where the route of administration is grossly inappropriate in terms of human exposure.
2. Carcinogenic data *on* man is acceptable only when it presents critically evaluated results of adequately conducted epidemiological studies.

[1]The NCI ad hoc committee report was also used by OSHA in justifying its "14-carcinogen standard. "

3. No level of exposure to a chemical carcinogen should be considered toxicologically insignificant for man.
4. Carcinogenic bioassays should include two species of animals of both sexes, with adequate control animals, subject to lifetime administration of suitable doses, including highest tolerated doses, by routes of administration including those by which man is exposed.
5. Negative results should be considered superseded by positive results, which should be deemed definitive, unless new evidence conclusively proves that the positive results were not causally related to exposure.
6. An implication of potential carcinogenicity should be drawn both from tests which induce benign tumors and those resulting in tumors more obviously malignant.
7. The principle of zero tolerance is valid and should be expanded.

In a subsequent proceeding concerning the pesticides aldrin and dieldrin, the EPA brief listed nine "cancer principles":

1. A carcinogen is any agent which increases tumor induction in man or animals.
2. Well-established criteria exist for distinguishing between benign and malignant tumors; however, even the induction of benign tumors is sufficient to characterize a chemical as a carcinogen.
3. The majority of human cancers are caused by avoidable exposure to carcinogens.
4. While chemicals can be carcinogenic agents, only a small percentage actually are.
5. Carcinogenesis is characterized by its irreversibility and long latency period following the initial exposure to the carcinogenic agent.
6. There is great variation in individual susceptibility to carcinogens.
7. The concept of a "threshold" exposure level for a carcinogenic agent has no practical significance because there is no valid method for establishing such a level.
8. A carcinogenic agent may be identified through analysis of tumor induction results with laboratory animals exposed to the agent, or on a post hoc basis by properly conducted epidemiological studies.
9. Any substance which produces tumors in animals must be considered a carcinogenic hazard to man if the results were achieved according to the established parameters of a valid carcinogenesis test (quoted in 122).

In its notice proposing to suspend registration of the insecticides chlordane and heptachlor, EPA set forth principles very similar to these nine statements. Organizations, particularly from industry, and individuals outside EPA expressed concern about the principles' substantive content, and EPA staff scientists became concerned that these scientific principles had been formulated by EPA attorneys.

Later, Saffiotti prepared a draft summarizing 17 principles of carcinogenesis that had been used in previous proceedings. EPA attorneys attempted to have these principles included as "officially noticed facts" in the proceedings concerning the pesticide Mirex, Apparently a storm of protest followed, after which the 17 principles were reduced to "three basic facts":

1. There is presently no scientific basis concluding that there is a "no effect" level for chemical carcinogens.
2. Experimental data derived from mouse and rat studies can be used to evaluate whether there is a cancer risk to man.
3. All tumorigens must be regarded as potential carcinogens. For purposes of evaluating carcinogenicity hazard, no distinction should be made between the induction of tumors diagnosed as benign and the induction of tumors diagnosed as malignant (quoted in 122).

In April 1976, the Administrator of EPA decided that, while these proposed "facts" represented the best available evidence and were valid for supporting regulatory action, he wasn't prepared to designate them as "officially noticed facts" (122).

In 1975, while the effort to transform the principles into "officially noticed facts" was pending, an EPA scientist asked NCI to review EPA's cancer principles. This question was referred to a Subcommittee on Environmental Carcinogenesis of the National Cancer Advisory Board, which was asked in September 1975 "to develop general criteria for use in the assessment of whether specific environmental agents constitute a carcinogenic hazard in humans. " EPA later withdrew its request for this effort, but the subcommittee, chaired by Phillipe Shubik, met in November 1975. The subcommittee finished a document in June 1976 that covered issues related to the identification of carcinogens (348). The report cautions that evidence of hazards must be evaluated case by case and that "criteria appropriate for one agency may not necessarily apply to another. " In other respects, the conclusions of this report were similar to those in the other lists of "principles. "

maries of the documents in the *Federal Register*. The policy described later in this chapter is found in an appendix to that announcement. The major change from the 1979 methodology to that described in 1980 was EPA's adoption of the linearized multistage model for extrapolating from high to low doses. The appendix describing carcinogen risk assessment was prepared by the staff of CAG. This publication was the most extensive explanation of their procedures available at that time.

In 1984, EPA published a proposed revision of its carcinogen assessment guidelines (309). EPA's purpose was "to promote quality and consistency of carcinogen risk assessments within the EPA and to inform those outside the EPA about its approach to carcinogen risk assessment." The guidelines were to "provide general directions for analyzing and organizing available data" and were not intended to alter risk management policies established under the various statutes administered by EPA. Also in November 1984, EPA published proposed guidelines for exposure assessment (310) and for mutagenicity and developmental toxicants risk assessments (311,313). Shortly thereafter, it published proposed guidelines for risk assessments of chemical mixtures (312). After making revisions and waiting for the Office of Management and Budget (OMB) to complete its review, EPA published the final version of these guidelines in September 1986 (284,285,286,287).

These last guidelines on carcinogen risk assessment consist of 10 pages in the *Federal Register*, describing the "general framework" to be used in assessing carcinogenic risk and "some salient principles to be used in evaluating the quality of data and in formulating judgments concerning the nature and magnitude of the cancer hazard from suspect carcinogens" (284). This policy outlines the various steps of risk assessment: hazard identification, dose-response assessment, exposure assessment, and risk characterization. Finally, the policy presents a "weight-of-the-evidence" classification system, with five basic categories. A chemical will be classified based on the nature (sufficient, limited, inadequate, etc.) of the evidence from human and animal studies. This classification has acquired important regulatory implications because EPA's Office of Drinking Water uses it to set recommended limits in drinking water and EPA's Office of Emergency Response uses it as part of a ranking system for adjusting reportable quantities of hazardous substances covered by the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA, commonly known as Superfund).

The Consumer Product Safety Commission

The Consumer Product Safety Commission (CPSC) published carcinogen assessment guidelines in 1978 and made them effective immediately (229). At the same time, CPSC provisionally classified perchloroethylene as a suspect carcinogen using the policy. Dow Chemical Company sued, claiming that even such a provisional classification harmed Dow. The court held that CPSC could not use the cancer policy in this manner until it was adopted in rulemaking procedures (45).

Subsequently, CPSC formally withdrew its cancer policy from the rulemaking process and decided to use the guidelines adopted by the Interagency Regulatory Liaison Group (IRLG), and more recently the guidelines issued by OSTP. (Even though CPSC's policy was withdrawn, its contents are still interesting in light of other Federal agency policies, and for this reason, the policy is discussed further below.)

The Occupational Safety and Health Administration

OSHA published a proposed regulation governing identification and regulation of carcinogens on January 20, 1977. OSHA held hearings, accumulated an extensive record, and published a final regulation in 1980. One important purpose of the policy was to improve the efficiency of the standards-setting process. OSHA officials argued that the slowness in setting standards was partly related to the many discussions, arguments, and lawsuits involved in every regulatory proceeding on carcinogens (365). The proposed policy generated considerable controversy about OSHA's identification and regulation of carcinogens, and dispute about the fraction of cancer incidence in the United States that can be attributed to occupa-

tional exposures. (See refs. 217 and 159 for discussions of this second issue.)

OSHA published its carcinogen policy as a binding regulation. Its intent was to collect evidence and testimony on "generic" issues in carcinogen identification and regulation, make decisions on these issues, and then rely on these decisions in future proceedings. (The policy uses a "presumption-rebuttal" approach (180).) The framers of this policy hoped that its use would speed the regulation of carcinogens by limiting debate about generic issues in the regulatory proceedings on individual carcinogens.

In contrast to other agencies' adoption of quantitative risk assessments for setting standards, the OSHA cancer policy stated that quantitative risk assessments would be used only to set priorities. Originally, some of the provisions of OSHA's carcinogen policy concerning risk management stated that once OSHA determined a substance to be a carcinogen, it would then set an exposure standard based only on feasibility. In 1981, OSHA amended its carcinogen policy to conform to the Supreme Court decision on OSHA's benzene standard, which provided that OSHA could only regulate exposures posing a "significant risk" to the health of workers and only if the regulation would significantly reduce the risk. The amendment allowed OSHA to consider the significance of estimated risk and feasibility in setting health standards for carcinogens. Regarding the specifics of risk assessment, OSHA did not change its judgments on the science of identifying carcinogens, although it did indicate that certain types of evidence and arguments that it had originally hoped to exclude from specific proceedings might be relevant to determining whether there was a significant risk. OSHA policy was also amended to reflect this conclusion (274). In 1982, OSHA suspended parts of the policy that required publication of lists of candidate carcinogens (after one list had been published in 1980) and requested public comment on more general issues concerning the substance of its policy (275). As yet, no changes have been made in the policy based on those comments. The policy was legally challenged by industry groups shortly after it was published in 1980, although the case was never argued and the suits have been dismissed.

Other Agencies and Interagency Efforts

OSTP and several interagency committees have worked on carcinogen assessment guidelines and regulatory policies. The Carter Administration established the IRLG, which initially had representatives from EPA, CPSC, OSHA, and FDA. Later, the Food Safety and Quality Service of the Department of Agriculture (USDA) joined. An IRLG working group, consisting of scientists from the IRLG agencies, the National Cancer Institute (NCI), and the National Institute of Environmental Health Sciences (NIEHS), published "Scientific Bases for Identification of Potential Carcinogens and Estimation of Risks" in July 1979. The report is noteworthy because it represents the first joint attempt of regulatory agencies to develop a consistent approach to identify carcinogens. However, some differences of opinion remained, especially concerning the desirability of quantitative risk assessment (137).

Another interagency group in the Carter Administration, the Regulatory Council, also prepared a document on carcinogen regulation. The Regulatory Council's conclusions on the science of identifying carcinogens relied heavily on the IRLG document, which was published as an appendix to the Council's document (354).

In 1979, several staff members of OSTP prepared a document to "stimulate development of a uniform decision-making framework to assure consistent Federal action regarding the identification, characterization, and control of potential human carcinogens." Making a distinction between "scientific data collection and analysis" and "regulatory decision-making, " it examined only the former. This document was relatively short, consisting of short discussions of particular areas and giving the authors' recommendations for improvements in Federal decisionmaking. In particular, they suggested the coordination of Federal risk assessment activities under the aegis of NTP (23).

In 1981, the new Reagan Administration articulated a strong opposition to most government regulation. Even before the inauguration, David Stockman, the first Director of OMB under the Reagan Administration, had published a list of

regulations he thought were undesirable. Within 2 months of taking office, Reagan created a task force on regulatory relief, chaired by the Vice President, and issued an Executive order providing for OMB review of agency regulatory proposals and final rules. The order stated that agencies could regulate only when the benefits of regulation exceeded its cost, except when this was prohibited by law. Administration officials asked affected businesses to inform them about regulations the businesses wanted changed. In 1981 the Reagan Administration also dissolved IRLG and the Regulatory Council.

In 1982, several events suggested the beginnings of a decidedly different approach to assessing carcinogenicity. In that year, EPA decided not to designate formaldehyde for priority review under the Toxic Substances Control Act (TSCA). In a memo to EPA Administrator Anne Burford, John Todhunter, EPA's Assistant Administrator for Pesticides and Toxic Substances, concluded that while formaldehyde appeared to be carcinogenic in rats, the carcinogenic potential of formaldehyde also seemed to "vary significantly with species and route." Moreover, although in certain exposure situations, formaldehyde could pose a human risk, the available epidemiologic information "supports the notion that any human problems may be of low incidence or undetectable." Quantitative risk estimates fell into a range that Todhunter considered low. For these reasons, he did not think formaldehyde should be subject to an accelerated review under TSCA (155,196).

A second EPA document appearing in 1982 was a draft of guidelines for assessing carcinogenicity, specifically for developing water quality criteria. The draft described a weight-of-the-evidence stratification scheme, modeled after the scheme of the International Agency for Research on Cancer (IARC),[5] and suggested that regulatory distinctions be made between carcinogens that act by causing gene mutations and those that act by different mechanisms. For the latter, the draft suggested development of water quality standards using the "no observable effect level" (NOEL) (180,279). In 1983 it was also revealed that Rita Lavelle had written a memo urging that trichloroethylene (TCE) be reevaluated and that EPA develop a "threshold model risk assessment for nongenotoxic chemicals such as TCE" (117).

A third draft document represented an administrationwide effort to revise agency practices on carcinogenicity risk assessment. In 1982, as part of the Reagan Administration's efforts to reduce the burden of government regulations and to develop a "scientifically sound basis for identifying and characterizing potential human carcinogens," OSTP convened an interagency committee to update the information contained in the 1979 IRLG document. The committee developed a "rough first draft" statement on "the current state of the science" (105).

The draft, which criticized many of the existing procedures used by regulatory agencies including the use of high-dose testing in animals and linear non-threshold extrapolation models, suggested distinctions based on mechanisms of action (e.g., between epigenetic and genotoxic agents) and the greater use of pharmacokinetic information in risk assessments (180).

The draft was circulated among a number of scientists and generated considerable controversy. Criticism especially focused on a chapter by John Todhunter, which suggested distinctions based on mechanisms. Congressional committees held hearings on this and other aspects of the Administration's regulatory policies in 1982 and 1983. In the wake of several revelations not directly related to the ongoing effort to develop "science principles" for carcinogenesis, most of the top EPA officials left office (362).

Review of the scientific basis for carcinogen risk assessment continued under Ronald Hart, the Director of FDA's National Center for Toxicological Research (NCTR). Drawing on scientists from NIEHS, NCI, NCTR, OSHA, CPSC, FDA, EPA, USDA, and OSTP, another draft was prepared and published for public comment and was generally received favorably (116).

A final version, "Chemical Carcinogens: A Review of the Science and Its Associated Principles, February 1985," was published by OSTP in March 1985 (351). It was republished in the journal Envi-

[5]EPA later adopted such a scheme in its 1986 policy, as discussed below.

romnental *Health Perspectives* in 1986 under the authorship of the U.S. Interagency Staff Group on Carcinogens, reflecting the contributions of staff from all the agencies involved. This document is an extensive summary of the state of various scientific fields underlying risk assessment: mechanisms of carcinogenesis, short-term tests, long-term bioassays, epidemiology, and exposure assessment. The document concluded with a discussion of the assumptions used in the process of risk assessment and included a series of summary principles.

GUIDELINES FOR CONDUCTING CARCINOGENICITY TESTING

Toxicity testing and interpreting test results are important features of several Federal regulatory and research efforts aimed at preventing exposures to carcinogens. In some circumstances, toxicity testing may include long-term bioassays to determine directly whether substances cause cancer in animals. Several laws that provide for carcinogen regulation allow Federal agencies to order regulated industries to conduct toxicity tests. The Federal Government's own carcinogenicity bioassay program was once housed at NCI, but is now coordinated by NTP. (See ch. 4.)

Required Carcinogenicity Testing

FDA requires carcinogenicity testing for some substances that are proposed as new, direct food or color additives. Decisions on whether a substance must be tested are made using a complex scheme based on the chemical structure of the substance (e. g., its chemical relation to known carcinogens) and the expected concentration of the substance in food. These two factors are used to classify the substance into one of three "concern levels." Only for the highest level are lifetime carcinogenicity bioassays required. For the other two levels, FDA requires that the substance be tested in a battery of short-term tests. The results of the short-term tests may alter the concern level of the substance and thus lead FDA to require a long-term bioassay.

For animal drugs that may leave potentially harmful residues in food, FDA uses three kinds of information to decide whether to require carcinogenicity testing:

- the potential toxicity of the drug, which is evaluated based on chemical structure and short-term and subchronic tests;
- the estimated level of use in food-producing animals; and
- Ž the amount of drug residue expected to be consumed by a person during a single exposure.

For human drugs, FDA requires carcinogenicity testing for any new drugs expected to be used for chronic or repeated use, although these requirements are not found in any written guidelines or regulations. These requirements apply to new drugs and not to drugs that were approved prior to 1968 when FDA began to require chronic tests.

Under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), EPA may require animal carcinogenicity studies for registering new pesticides and reregistering existing pesticides. A carcinogenicity bioassay is required for these substances in three specific circumstances:

1. when the active ingredients, metabolizes, degradation products or impurities are structurally related to recognized carcinogens, cause mutations in short-term tests, or produce a worrisome effect in subchronic studies;
2. when use of the pesticide will require that EPA or FDA issue a food tolerance limit or food additive regulation; and
3. when use of the pesticide will result in significant human 'exposure (e.g., in fabric treatments, insect repellents, and indoor pesticides).

Under TSCA, EPA may require testing either for new chemicals entering the market or for existing chemicals. In the latter case, EPA must conclude, first, that the chemical may present an unreasonable risk to health or the environment or that it may or will enter the environment in large quantities *or* present significant *or* substantial hu-

man exposures; and, second, that testing is necessary to provide more information about the chemical.

Analysis of Test Designs

In this section, OTA compares the carcinogenicity bioassay study designs required for food and color additives (245), for chronically used human drugs (161), for pesticides (332), and for tests ordered under TSCA (318). For animal drug carcinogenicity studies, FDA has not adopted separate guidelines but instead refers drug sponsors to the NCI test guidelines (251) and to the 1971 and 1982 versions of the guidelines for food additive testing (245,247). NCI developed a standard design for federally funded tests and published it in 1976 (251). NTP conducts most of its tests through the use of contract laboratories and, through its contractual "statement of work, " sets the design of these studies (256). In addition to study designs used by NCI and NTP, the comparison covers the recommendations of a group of scientists whose findings were published by IARC (59) and of a panel of outside scientists convened by NTP (258). This comparison covers only written requirements and suggestions. OTA has not attempted to determine how well the conducted bioassays comply with these guidelines.

Not all regulatory agencies have specified guidelines for test design. OSHA, for example, rejected the specification of test protocols and data analysis in favor of reliance on informed scientific judgment. In part, as OSHA pointed out, this reflects its own regulatory purposes-OSHA uses whatever test data are available and does not require toxicity testing. FDA and EPA, on the other hand, can require industry to test.

General Provisions for Test Designs

Today there is relatively little controversy about the general design of carcinogenicity studies, and guidelines are relatively consistent in their requirements for study design. The basic study design uses two different animal species. Because of the relatively low cost and long experience using rats and mice, these two species are usually used. The animals must be free of disease and quarantined, then they are randomly assigned to different groups.

Exposure routinely begins by the time the animals are 6 weeks old and the study ends usually after 2 years of exposure. Exposure is preferably through the route that most closely imitates human exposure. For example, food additives should be tested by adding the suspect additive to the animals' feed, while airborne toxic substances should be tested by mixing the substance into the air the animals breathe.

Animals are randomly assigned to two or three treatment or exposure groups and a control group, which is not exposed at all. Care must be taken to ensure that the exposed animals and the control animals live under the same conditions, except for the exposure to the suspect substance.

Animals that die during the study are examined for signs of toxicity and for tumors. At the end of the study, all the surviving animals are killed and necropsy is performed. The various guidelines specify a complete examination for visible lesions and tumors, and list the organs that are to be prepared for microscopic examination, although they differ on the extent of microscopic examination required. After tumors have been diagnosed, statistical analyses are used to determine whether the exposed groups had a higher incidence of tumors than the control group.

Issues in Test Design

Table 2-1 presents several main issues in the design of carcinogenicity bioassays and the way they are handled by the various guidelines. As outlined in OTA's 1981 report, Tehnologies *for Determining Cancer Risks From the Environment,* the principal issues in study design are the following:

- *study plan,* including the selection of animal species, the number of animals for each dose level, and the dose levels themselves;
- *dosing regimen,* including the age at which to begin exposure, when to terminate exposure, and whether there should bean observation period between the end of exposure and the sacrifice of the animals;
- *pathology,* including the nature of the autopsy examination of the animals and the extent of microscopic examination; and
- *personnel qualifications.*

In addition, some of the guidelines discuss survival criteria, such as the number of animals that must survive to have a valid positive or negative study.

Some differences in terminology exist, but the guidelines are generally consistent about the important issues in study design. All the guidelines require that two different species be tested. NCI and NTP guidelines specify the strains of rats and mice to be used by testing programs. EPA's guidelines and PMA's drug testing guidelines specify the rat and the mouse as test animals, while the FDA "Red Book, " IARC scientists, and the NTP Ad Hoc Panel suggest considering hamsters. The Ad Hoc Panel further encourages the search for other species for carcinogenicity testing.

All the guidelines specify that testing shall be done in both males and females, and all but 1 set of guidelines specify the size of each test group as 50 animals per dose. The NTP Statement of Work specifies **60** animals per dose, including 10 animals scheduled for interim sacrifice between the 12th and 18th month of the study. Several other guidelines mention that the number of animals should be increased if the researchers want to conduct an interim sacrifice, although no other guidelines require interim sacrifice. The number of animals needed for the chronic phase of testing depends on the number of doses.

NCI and PMA guidelines require at least two dose groups in addition to the control group.' All the other guidelines suggest the use of three dose groups and the unexposed control group. With 50 male rats, 50 female rats, 50 male mice and 50 female mice for each exposure level, a study using 2 exposure levels and controls uses 600 animals, a study with 3 exposure levels and controls requires 800.

All the guidelines provide that the highest dose level should be based on information gathered in a subchronic toxicity study (usually lasting about 90 days). However, slightly different terminology is used to refer to this dose level. The most common term in the toxicologic literature is "maximum tolerated dose" or MTD. Most of the guidelines refer instead to the "high dose level" or "high dose, " perhaps to avoid the controversy that "maximum tolerated dose" has engendered. (As discussed below in the section on agency risk assessment policies, the reason for high-dose testing is to enable a study to best detect a carcinogenic response.)

In general terms, the high dose should be as high as possible without shortening the animals' lives from noncarcinogenic toxic effects. FIFRA, TSCA, and "Red Book" guidelines specify that the dose should be minimally toxic without substantially altering the normal lifespan of the animal. The NTP Statement of Work and NTP Ad Hoc Panel documents also state that the high dose should not affect the animals normal lifespan from effects other than carcinogenicity. NCI guidelines give more detail: the MTD should neither alter the lifespan (other than from carcinogenicity), clinical signs of toxicity, or pathological lesions (other than neoplasms) that shorten the animals' lives nor should it lead to more than a 10-percent decrement in weight gain in experimental animals relative to controls. PMA guidelines specify that the highest dose should be "slightly below toxic dose," without providing any further guidance. Important to the success of a bioassay is the professional judgment of the researchers conducting the study and analyzing the data from prechronic studies. To some extent, setting the highest dose requires an educated guess.

The low doses are often defined as fractions of the highest dose. For example, NCI guidelines set the second dose as one-half or one-fourth of the MTD. (This formula is also given by IARC scientists for studies aimed at only a qualitative determination of a substance's carcinogenicity.) PMA guidelines specify that the second dose should be greater than or equal to the expected equivalent human dose, but less than or equal to half the high dose. The FDA "Red Book" and TSCA guidelines specify the lowest of three doses to have "no indication of toxicity" and, generally, to be 10 percent of highest dose (FDA) or not less than 10 percent of the highest dose (TSCA). IARC scientists suggest, for studies gathering quantitative information, that the doses be scaled by factors of 3, 5, or 10, The NTP Ad Hoc Panel and TSCA guidelines mention that researchers should

"Today, however, virtually all drug carcinogenicity studies are conducted with three dose groups (249).

Table 2.1 .—Test Design Issues

	NCI (251)	NTP-statement of work (256)	NTP Ad Hoc Comm. (258)	IARC (59)	FDA-food & color additives (248)	FDA/PMA-drugs (161)	TSCA (318)	FIFRA (332)
Animal species	'2, NCI used primarily B6C3F1 mice & Fischer 344 or Osborne-Mendel rats	2, B6C3F1 mice & Fischer 344 rats (unless NTP specifies otherwise)	Recognizes rats, mice, hamsters as most popular, encourages search for other species	2; choose among rats. mice, hamsters	2, rodents: rats mice, hamsters	2, rats & mice (pick strains w/low background incidence)	2. rats & mice	2, rats & mice preferred
Number at each dosage	50 male, 50 female	60 male, 60 female (allows for interim sacrifice of 10 animals)	50 male, 50 female; for special studies, number of groups & distribution may be altered	50 male, 50 female	50 male, 50 female (increase for interim sacrifice)	50 male, 50 female	50 male, 50 female (increase to allow for interim sacrifice)	50 male, 50 female (Increase to allow for interim sacrifice)
Dosages	At least 2 PIUS control, MTD & MTD/2 or MTD/4, MTD defined as "highest dose that can be predicted not to alter the animal longevity from affects other than carcinogenicity' i. e , no more than 10% weight loss, no mortality or clinical signs of toxicity that shorten animal's life, desirable to have positive control group	3 plus control, high dose is "predicted not to alter normal longevity of the animals from effects other than carcinogenicity'	3 plus control. MTD identified in prechronic studies as dose which will not Impair normal longevity from effects other than induction of tumors; use metabolic/pharmacokinetic studies to select lower doses; route should be same as human exposures, if using gavage, conduct pharmacokmetlc studies to back up, explore alternates to vegetable oil gavage	2 plus control for qualitative studies, 3 or more for studies to be used for quantitative assessment, select high dose "as one that produces some toxicity but not appreciable cell death or organ dysfunction, toxicity that impairs lifespan (other than tumors) or more than 10% decrement in weight gain compared to controls; lower doses. qualitative MTD/2 to MTD/4, quantitative: scale by factors of 3,5,10	3 plus control, high dose should elicit minimal toxicity w/o substantially altering lifespan (other than effects related to tumors); lowest dose should induce no signs of toxicity (generally 10% of high dose); intermediate dose should be approx midway, depending on pharmacokinetics	2 or more plus control, high dose–slightly below toxic dose, low dose–greater than or equal to human dose level & less than or equal to one-half of high dose	3 plus control, high-dose level–minimally toxic w/o substantially altering normal lifespan, lowest dose–should not interfere w/normal growth or show any other signs of toxicity, should not be less than 10% of high dose, intermediate—m between, depending on toxicokinetic properties, if known	3 plus control, highest dose level should be sufficiently high to elicit signs of minimal toxicity w/o substantially altering the normal lifespan
Required number surviving/Termmatlon criteria	Terminate when survival reaches 10% within group	May terminate when cumulative mortality jeopardizes ability to draw conclusions on carcinogenicity: contractor must consult with NTP	—	Not satisfactory if mortality exceeds 50% before week 104 for rats, week 96 for mice; end study when mortality in control or low-dose groups equals 75%	Survival must be at least 50% at 24 months (rats) & 18 months (mice), no more than 10% lost due to autolysis, cannibalism, or management problems, may terminate 'under special circumstances' if there are only 10 survivors m any group after 24 months (rats) or 18 months (mice), but minimum survival criteria must be met	Terminate if mortality reduces control group to less than 40% of original number of animals per sex	For valid negative study, survival must be greater than 50% in all dose groups at 24 months (rats) or 18 months (mice) & less than 10% loss due to autolysis, cannibalism, or management problems	Survival must be at least 50% at 18 months (rats) or 15 months (mice) & at least 25% at 24 months (rats) or 18 months (mice)

continued on next page

Start of dosing	Weanings if possl-ble, no older than 6 weeks	6-7 weeks old	Consider starting ex-posure in utero m certain cicumstances	–	Begin exposure to parents prior to mat-ing, exposure m utero & for life of offspring for non-nutritive additives & certain other sub-stances	"Weaning animals have often been used, but suggests prenatal design may be more sensitve, recommends data de velopment on this	As soon as possible after weaning, Ideally before 6 weeks not later than 8 weeks	As soon as possible after weaning, Ideally before 6 weeks, not after 8 weeks
Duration of dosing	"Greater part" of the animals' lifespan–24 months	103 weeks (plus up to 2 more weeks to schedule necropsles)	NTP should do studies to determine optimal endpoint	See termination cri-teria above, in any case, no longer than 130 weeks for rats, 120 weeks for mice	At least 104 weeks, up to 130 weeks	24 months for rats, 18 months for mice	24-30 months for rats, 8-24 months for mice	At least 24 months for rats, at least 18 months for mice
Observation period	May be desirable to hold animals for ad-ditional 3-6 months	1 week	Recommends against ending exposure prior to sacrifice un-less there IS concern about exposure to technicians	Prefers no observa-tion period, if desired, treat to 104 weeks for rats 96 weeks, for mice	–		–	24-30 months for rats, 18-24 months for mice
Dosing frequency	7 days/week for food/water exDosure, otherwise based on human exposures	–	–	–	7 days per week, use oral exposure route	–	Continuous, 7 days per week IS prefer-able; 5 days per week IS acceptable	Ideally 7 days per week; 5 days per week acceptable, oral route preferred provided substance is absorbed m GI tract
Organs & tissues examined	All animals gwen gross exam; histopathology for: 1 gross lesions/sus-pect tumors, 2) list of organs for all treated & control animals	All animals given gross exam; full histopathoiogy for all treated & control animals	All animals gwen thorough examina-tion; consider alter-natives to reduce burden of histo-pathology, such as inverse pyramid & selected inverse pyramid	Gross exam of all; detailed histopathol-ogy on high dose & control groups; if no difference is found, then histology can be restricted to examin-ing gross lesions & sites where signifi-cant lesions are ob-served in high dose group; suggests dis-tinguishing between fatal & incidental tumors, especially for statistical analysis	Gross exam for all animals, microscopic exam for: 1) all visi-ble tumors, 2) all animals that died during study, 3) high dose group & con-trols. If significant difference is seen, then examine the particular or-gans/tissues in all animals	Microscopic exam of 1) all gross lesions, 2) complete set of tissues of all high dose survivors & controls; if ques-tionable, then exam-ine other exposure groups	**Full** histopathology on: 1) all animals in control & high dose groups & all that died during study, 2) all gross lesions, 3) target organs in all animals; if there were excessive early deaths or problems with high dose group, use next lower dose group for full histopathology	Full histopathology on: 1) all animals m control & high dose groups & all that died during study, 2) all gross lesions, 3) target organs, 4) lungs, liver & kid-neys m all animals, if there were prob-lems with high dose group, use next lower group for full histopathology
Study director	–	Doctorate m toxicol-ogy, pathology, vet-erinary medicine, biochemitry, or chemistry	–	–	–	–		Appropriately edu-cated, trained, & ex-perienced toxicologist
Pathologist	Board-certified w/ex-perience in laboratory animal pathology	Formal training & ex-perience required, board cerification desirable	–			[4] "Individual possess-ing expertise m lab-oratory animal pathology'	–	Board-certified or board-eligible or per-son w/equivalent training w/expertise
Histology technicians	Supervised by HT/ASCP technician	ASCP-registered technicians	–			–	–	Certified by HT/ASCP or having equivalent traimng & capability

SOURCE **office** Of Technology Assessment based on recorded documents

use information on metabolism and pharmacokinetic studies, if available, to help set dose levels.

The control group should be completely untreated or sham treated, and should otherwise be handled by the lab workers in the same way as the treated animals. Sometimes control animals are treated with the "vehicle" used to administer the test compound, such as the corn oil used in gavage studies (vehicle controls). Sometimes, researchers will also include a group of animals to be exposed to a known animal carcinogen (positive control), to be sure that the animals being used are in fact sensitive to a known carcinogen. Except for the NCI guidelines, the guidelines that mention this possibility generally include it for routine studies. IARC scientists state information should be collected on control animals to evaluate any changes over time.

The basic laboratory alternatives for dosing the animals include adding the substance to the animals' food or water, exposing the animals by contaminating their air in special inhalation chambers, painting the substance onto the animals' skin, or delivering the substance, usually dissolved in corn oil, directly into the animals' stomachs using a special tube (gavage studies). With regard to the dosing regimen, the guidelines provide that the route of administration be as close as possible to the human exposure route, recognizing that sometimes this is not possible, for example, when the suspect compound is so unpalatable that animals will not eat the treated feed. For dosing in food or water, exposure is generally 7 days per week. For inhalation or gavage studies, laboratories generally expose the animals five times per week to match the schedules of laboratory personnel.

The guidelines also specify the age of the animals at the start of the study. The NCI, NTP Statement of Work, FDA "Red Book," TSCA, and FIFRA guidelines all require dosing to begin shortly after the animals have been weaned and before the animals reach 6 to 8 weeks. For food and color additives, FDA often requires the manufacturer to conduct the carcinogenicity study in at least one rodent species with in utero exposures. Parents are exposed to the test compound prior to mating, and exposure continues through pregnancy and throughout the lives of the animals. It is argued that this design is particularly sensitive in detecting carcinogenic effects and is especially appropriate for substances in the food supply because exposures may be continuous for parents and children. PMA guidelines also mention the in utero design as a possibility, although data on this design are lacking. The guidelines suggest use of the design to develop data, especially for drugs that may be used in childbearing women. The NTP Ad Hoc Panel also suggests that the in utero design be considered under certain circumstances.

The guidelines differ concerning when a study should end. The basic principle is that carcinogenicity studies should expose the animals for the "greater part of the animals' lifespans" (251). For rodents, this is generally considered to be 2 years. Thus, the NCI guidelines provide for exposures of 24 months, the NTP Statement of Work provides for 103 weeks (plus up to 2 more weeks to schedule autopsies), and the FDA "Red Book" requires at least 104 weeks (24 months), though such a study may last up to 130 weeks (30 months). PMA, TSCA, and FIFRA guidelines provide for a shorter exposure for mice: 18 months for PMA, at least 18 months for FIFRA, and 18 to 24 months for TSCA. For rats, these guidelines provide for studies of at least 24 months.

NCI guidelines suggest it "maybe desirable to hold animals for an additional period of 3-6 months" after exposure has stopped. This time was termed an "observation period." More recently, however, the NTP Ad Hoc Panel concluded that, except when there is concern about exposure to lab personnel, "it does not seem wise to terminate exposure prior to sacrifice." The NTP Statement of Work sets a l-week period for observation. FIFRA guidelines say that rat studies may be 24 to 30 months long and mouse studies 18 to 24 months long.

IARC scientists define the length of a study in terms of the animals' survival. According to them, a study should be terminated when 75 percent of either the control or low-dose group have died. But in no case should the study extend beyond 130 weeks (30 months) for rats or 120 weeks (28 months) for mice. IARC prefers no observation period, but suggests that if one is desired, treat-

ment can continue 104 weeks for rats and 96 weeks for mice.

NCI guidelines provide for the termination of a study when survival drops to 10 percent in any group. The NTP Statement of Work allows for termination when "cumulative mortality jeopardizes ability to draw appropriate conclusions on carcinogenicity," but requires consultation with NTP before a contractor can sacrifice the animals. The FDA "Red Book" provides that a study may be terminated "under special circumstances" if there are only 10 surviving animals in any group (because these groups contain 50 animals, this is a survival rate of 20 percent) after 24 months in rats and 18 months in mice.

The FDA "Red Book," TSCA guidelines, FIFRA guidelines, and IARC scientists specify the minimum survival necessary for a valid negative study. The FDA "Red Book" requires at least 25 animals per sex at 24 months for rats and at 18 months for mice. In addition, no more than 10 percent of the animals should have been lost due to autolysis (tissue destruction before necropsy), cannibalism, or management problems. TSCA guidelines require at least 50-percent survival in all groups, while FIFRA guidelines specify that survival must not be less than 50 percent at 18 months for rats and 15 months for mice and not less than 25 percent at 24 months for rats and 18 months for mice. IARC scientists suggest that a study is not satisfactory if mortality is greater than 50 percent before week 104 (24 months) for rats and 96 weeks (22 months) for mice.

All the guidelines describe the nature of the necropsies that should be conducted after animals die during the experiment or are sacrificed at the end. In general, all the animals should be examined carefully, including gross visual examination of a number of specified tissues. Instructions on preserving tissues are given, and tissue portions are prepared for microscopic examination to discover tumors and their types. The guidelines differ with regard to the extent of this microscopic examination. In general, it is required for all observed gross lesions and for sections of major body tissues. NCI guidelines require microscopic examination of these tissues for all exposed and control animals. The other guidelines allow for less comprehensive microscopic examinations, specifically, of all animals that died during the study, all animals in the control group, and all animals in the highest exposure group. Microscopic examination should also be conducted for animals in lower exposure groups on the specific organs (target organs) in which tumors were discovered in the highest exposure group, and in some guidelines, on the lungs, livers, and kidneys of all animals. Microscopic examination of animals in the lower exposure groups may also be necessary if there are excessive early deaths or other problems in the highest dose group.

A large part of the costs of long-term carcinogenicity bioassays owes to examining or reading the large number of microscopic slides. For example, EPA has estimated that a bioassay conducted to meet the requirements of TSCA regulations will generate about 40,000 slides, requiring about three-quarters of a year of a pathologist's time in addition to the costs of technicians and materials. Because of these costs, the NTP Ad Hoc Panel suggested that alternatives be considered to reduce the burden of conducting microscopic pathology. NTP has tried to implement such an approach to reduce the pathology requirements. But when using a reduced pathology system, NTP often found it necessary to go back to the original tissues to obtain additional slides for diagnosis. Consequently, NTP has now returned to examining all tissues at all dose levels. While the reduced pathology system decreased the costs of pathology, it increased the time necessary to complete the study (95,121).

Several of the guidelines also detail the qualifications necessary for principal study personnel. NCI guidelines required that the study pathologist be board-certified and that histology technicians be supervised by a registered technician. The NTP Statement of Work requires that the study director have a doctorate in a relevant discipline, that the pathologist have formal training and experience in animal pathology, and that histologists be registered with the relevant accrediting organization. PMA guidelines specify only that the pathologist should have expertise in laboratory animal pathology.

63-986 0 - 87 - 3

ANALYSIS OF AGENCY CARCINOGEN ASSESSMENT POLICIES

The following comparison will examine policies' that have been issued by the major regulatory agencies:

- EPA Interim Guidelines—EPA (1976) (ref. 293),
- EPA Water Quality Criteria–EPA (1980) (ref. 323),
- EPA Standard Evaluation Procedure for Pesticides—FIFRA (1985) (ref. 328),
- EPA Carcinogen Risk Assessment Guidelines–EPA (1986) (ref. 284),
- CPSC Interim Carcinogen Policy —CPSC (1978) (ref. 228),
- OSHA Carcinogen Policy –OSHA (1980) (ref. 276), and
- FDA Sensitivity of Method Policy (proposed)–FDA SOM (1985) (ref. 246).

There have also been several interagency collaborative efforts, and efforts by nonregulatory bodies:

- National Cancer Advisory Board, Subcommittee on Environmental Carcinogenesis—NCAB (1977) (ref. 348),
- Interagency Regulatory Liaison Group—IRLG (1979) (ref. 347),
- Office on Science and Technology Policy—OSTP (1979) (ref. 23), and
- Office on Science and Technology Policy—C)STP (1985) (ref. 351).[7]

These policies were issued under a variety of circumstances and are organized in several different ways. In some cases, they appear to have been adopted as relatively informal statements of scientific understanding on how carcinogens might be identified. In other cases, they are formally adopted agency regulations, specifying how the agency will identify carcinogens and attempting to limit the kinds of arguments and evidence to be considered in any specific regulatory proceeding. In between these two extremes, some documents outline an agency's standard procedures and discuss problematic areas of interpretation, often including the inference options the agency will generally use.[a]

This chapter focuses on formal written policies, with only limited attention to actual agency practices on carcinogen risk assessment. The policies themselves will be referred to using agency acronyms and the year of the policy, for example, OSHA (1980), OSTP (1985).

Definitions

Not all of these policies propose a formal definition of "carcinogen," although in most cases the text of the Policy outlines the various criteria and considerations that will be used to identify and classify carcinogens. In its simplest form, a carcinogen may be defined as a substance that causes cancer (217). Two more complete definitions of a carcinogen are those of OSHA (1980) and OSTP (1985). OSHA gave the following definition of a potential occupational carcinogen:

> . . . any substance, or combination or mixture of substances, which causes an increased incidence of benign and/or malignant neoplasms, or a substantial decrease in the latency period between exposure and onset of neoplasms in humans or in one or more experimental mammalian species as the result of any oral, respiratory or dermal exposure, or any other exposure which results in the induction of tumors at a site other than the site of administration. This definition also includes any substance which is metabolized into one or more potential occupational carcinogens by mammals.

A more recent definition is offered by OSTP:

> . . . a substance which is capable under appropriate test conditions . . . of increasing the incidence of neoplasms (combining benign and malignant when scientifically defensible) or decreasing the time it takes for them to develop.

OSTP (1985) added the qualification that, before concluding that a chemical is carcinogenic, gen-

[a]See reference numbers in the following two bulleted lists for complete policy citations. Policies are only cited by year elsewhere in this section.

[7]This discussion does not cover one other policy that was prepared by the Committee to Coordinate Environmental and Related Programs (CCERP) of the Department of Health and Human Services (232).

[8]For another comparison of agency policies, see Rushefsky (179,180).

eral principles in evaluating animal test results should be followed (e.g., eliminating experimental artifacts).

Qualitative Risk Assessment: Hazard Identification

Use of Human Epidemiologic Data

Most policies declare that well-conducted positive epidemiologic studies provide conclusive evidence for carcinogenicity, The FIFRA (1984) evaluation procedure did not discuss epidemiology at all, and FDA SOM (1985) treated it only cursorily. In evaluating new pesticides and residues of new animal drugs, it is not likely that there will be relevant epidemiologic studies.

EPA (1986), IRLG (1979), and OSTP (1985) all discuss in some detail several of the important factors to consider in evaluating epidemiologic studies, including the strength of association, level of statistical significance, information on dose-response relationships, biological plausibility, temporal relationships, confounding factors and bias, accuracy of exposure and cause-of-death classifications, adequacy of followup, and whether sufficient time has elapsed to allow for latent effects.

One epidemiologic issue provoking special attention in many of these policies is the role to be played by negative human studies in evaluating chemical hazards. All Federal policies addressing this issue state that negative human studies can only set an upper bound on risk estimates. A negative study cannot prove the absence of a carcinogenic hazard. A negative study indicates, at most, that the true risk is unlikely to exceed the specified upper bound. The magnitude of this upper bound depends on the size of the study and the background incidence of the cancer in question.

OSHA (1980) went even further than this, generally referring to negative studies as "nonpositive" studies. In characteristic fashion, OSHA (1980) also set down explicit and stringent criteria for when a "nonpositive" study would be acceptable evidence for an OSHA rulemaking. Such a study will be considered only if:

1. the study involved at least **20** years of exposure and at least **30** years of observation after initial exposure,
2. documented reasons are provided for predicting human cancer site(s) at which the substance would induce cancer if it were carcinogenic in humans, and
3. the exposed group was large enough to detect a 50 percent excess risk at the predicted sites.

To use a "nonpositive" study to set an upper limit on risk, both of the first two criteria must be met and, in addition, there must be reliable human exposure data.

OSHA (1980) pointed out that there have been negative studies for arsenic, benzene, coke oven emissions, petroleum refinery emissions, and vinyl chloride, substances and mixtures now generally believed to be carcinogenic on the basis of other epidemiologic studies. Even the epidemiologic evidence of the association between asbestos exposure and lung cancer among nonsmokers was "nonpositive" for a long time. Selikoff found no excess of lung cancer among nonsmoking asbestos workers for the first 30 years after exposure, though 5 more years of followup demonstrated a positive effect.

Use of Long-Term Animal Bioassay Data

All policies accept the use of animal data as predictive for human beings. Explicitly or implicitly, all the policies acknowledge that substances shown to be carcinogenic in animals should be presumed to present a carcinogenic hazard to humans.

An often-quoted statement on the value of animal data in assessing human risk is that of IARC. Their principle is based on two points: that a number of chemicals were first identified as animal carcinogens, and then evidence confirmed carcinogenicity in humans. Second, all chemicals accepted as human carcinogens that have been adequately studied in animals are positive in at least one species. (See the discussion in ch. 4.) IARC concluded:

> Although this association cannot establish that all animal carcinogens also cause cancer in humans, nevertheless, in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is *suffi-*

cient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans (99).

However, determining exactly what evidence will be considered sufficient to demonstrate a substance to be an animal carcinogen is a little more complex.

OTA identified the following major issues on use of long-term animal bioassay data for hazard determination in agency policies:

- use of the maximum tolerated dose,
- route of administration,
- criteria for a valid negative study,
- classification of tumors as benign or malignant and deciding which should count as evidence for carcinogenicity,
- evaluation of certain problem tumor types and commonly spontaneous tumors,
- use of historical control data,
- statistical evaluation, and
- performance of overall qualitative evaluation.

Some of the agency policies also gave guidance for the design of bioassays. These points have generally been covered in the earlier section on testing.

Use of **Maximum** Tolerated Dose.-For reasons of economics and practicality, long-term animal bioassays are much too small to provide experimental data on the hazards of low exposures. Therefore, to maximize the sensitivity of animal bioassays for detecting carcinogenic effects, agency guidelines for designing tests specify use of the MTD. This position was also affirmed by the Ad Hoc Panel convened by NTP to consider issues in carcinogenicity testing (258).

In bioassays, the power of a study to detect a tumor increase reliably depends on the number of animals in each exposure group, spontaneous incidence of the particular tumor increased, and the magnitude of the increase. The probability of missing an increase even though the substance is truly carcinogenic (a false negative) is fairly high. For example, with a standard bioassay design using 50 animals per exposure level, the probability of not detecting an increased tumor incidence from 1 percent in controls to 10 percent in the exposed group is 73 percent or nearly three-quarters of the time (258).[9]

With one exception, all agency policies on interpreting test results accept positive test results using the MTD. One policy (FIFRA 1984) raised the concern that exposures at the MTD may represent a toxic insult qualitatively different than those at much lower exposures. FDA (1985) indicated when test levels turn out to have exceeded the MTD (after conducting the 2-year study), negative results do not remove suspicion about possible carcinogenicity. Several other policies address how to interpret positive results at levels that exceed the MTD or show noncarcinogenic toxic effects.

EPA (1986) suggested that studies be carefully reviewed to determine whether the high exposure levels induce effects that would not be seen at lower levels. OSHA (1980) generally accepts positive results from high-dose testing. OSHA will entertain arguments that high doses are not relevant to human exposures only if documentation shows that:

1. at high doses the test animals produce metabolizes that are produced only at high doses,
2. these high-dose metabolizes are the ultimate carcinogens and the ones produced at low doses are not, and
3. the carcinogenic metabolizes are not produced by humans exposed to low doses.

FDA SOM (1985) required "convincing evidence" to rule out carcinogenic effects seen at exposures above the MTD. OSTP (1979) suggested accepting these results only if the noncarcinogenic toxic effects have not altered metabolism or immune system responses in a way that could have caused the carcinogenic effects, while OSTP (1985) declares it is appropriate to consider animal test results that use exposures exceeding human exposures although, as mentioned above, OSTP (1985) also requires that possible organ damage

[9]Using Fisher exact test with a one-sided significance level < 0.025.

and metabolic saturation (experimental artifacts that may occur at high doses) be considered before concluding that a chemical is carcinogenic.

Thus, agency policies accept the use of high-dose testing, but many policies raise concerns about how to interpret test results at levels that exceed the MTD. One difficulty in conducting these tests is making a guess concerning the MTD for a 2-year study based on the results from a 13-week study. Sometimes the researchers estimate poorly what the MTD will be. The result is that the study is conducted with doses that are substantially above or below the MTD.

The decisions about how to use results from these studies are a problem. The policies generally appear to reject use of negative results from such studies, but differ in how to handle positive results. FDA SOM (1985) accepted positive studies, even if the MTD was exceeded, "unless there is convincing evidence to the contrary. " OSHA (1980) set a policy of entertaining arguments that high-dose results are not relevant to humans only if documented evidence is presented that shows that the ultimate carcinogenic metabolizes are produced only at high doses and not in humans exposed at low doses.

Other policies are more restrictive in interpreting such studies. EPA (1986) asked for careful review of studies at levels above the MTD to determine if there was a response that does not occur at lower exposures. FIFRA (1984) went further in arguing that use of the MTD was "interjecting biases of considerable importance" in evaluating animal studies.

As will be discussed later in this chapter, there are other difficulties in applying test results from high doses in animals to predict human risk, even using high doses that do not exceed the MTD.

Route of Administration. -In animal bioassays, the substance under test may be administered in any of several ways: it may be incorporated into the animals' diet, or into their drinking water; the animals may inhale the substance as they breathe; the substance may be dissolved in corn oil (or similar vehicle) and then administered through a feeding tube directly into their stomachs (in gavage); or it may be injected or implanted in the animals or painted onto their skin. Although it is desirable that the exposure route used in the animal study be similar to human exposures, this is not always possible. For example, some substances when mixed with feed or water will alter the taste, leading the animals to refuse to eat or drink enough to receive the desired dose.

The two major issues of interpretation are these: If an exposure route that differs from the human route is used, are the results applicable to humans? If tumors are found only in tests that use "unusual" administration routes (such as injection or implantation) or the tumors are found only at the site of administration, are these results applicable to humans?

One view the policies express is that if the resultant tumors are found in organs or tissues distant from the site of application, then the substance should be considered a carcinogen, irrespective of the route of exposure. EPA (1980), OSHA (1980), and IRLG (1979) clearly express this view. On the other hand, tumors found only at the site of administration or by unusual methods may "raise the possibility" of carcinogenicity (NCAB 1977) or might be used as "concordant evidence" (OSHA 1980). Other policies caution that these results merit additional evaluation in assessing their human relevance (EPA 1984, CPSC 1978) or suggest that more testing is needed to resolve safety concerns (FDA SOM 1985). OSHA (1980) treated as indicative of carcinogenicity any contact tumors (those occurring at the site of exposure), from oral, respiratory, and dermal exposure routes and noncontact site tumors (those found at sites away from the exposure site), regardless of the route of exposure. OSHA (1980) will consider arguments that a tumor response only at the site of administration is not predictive of human hazard if: 1) the exposure route is not oral, respiratory, or dermal (i. e., is through injection or implantation); and 2) the tumor induction is related to physical configuration or formulation of material and not to its chemical properties.

Criteria for Valid Negative Study .—Two of the policies provide criteria that must be satisfied for an animal study to be considered a negative study. (As discussed above, some of the test protocols also provide minimum standards for negative studies.) For the IRLG, negative test results can only be considered evidence of no effect when "minimum requirements have been met. " "Accepted procedures" include: a) the observation of all animals in the study . . . until their spontaneous death, b) the sacrifice of animals that show clinical signs of severe illness or impending death . . ., and c) terminal sacrifice at a scheduled date near the end of the lifespan (e.g., after 24 months on test). None of the other policies provide criteria for a valid negative animal study, although, as discussed below, many of these policies do specify the weight to be given to "negative" studies.

EPA's CAG guidelines (1986) do not specify minimum test design requirements, but provide that a substance will be classified in the "no evidence" category if there is no tumor increase in "at least two well-designed and well-conducted animal studies of adequate power and dose in different species."

Use of Data on Benign and Malignant Tumors. —When diagnosing cancer in a human patient, microscopic examination of tumor tissue yields a classification of the tumor as benign or malignant. Cancer is a disease of malignant tumors, that is, ones that have the potential to invade other tissues and spread throughout the body. The cells of benign tumors, on the other hand, remain together and do not invade other parts of the body. The clinician can then formulate a prognosis and develop a therapy based on whether the tumor is benign; malignant, having the potential to spread, but localized; or already widespread. Benign tumors may still be of concern, however, because if they develop in vital organs (e.g., the brain), they can cause serious disability and death.

This classification is also used when examining animals exposed during a bioassay. While there are difficulties in the precise classification of some tumors, *10 the more general* Controversy concerns how to count the tumors diagnosed as "benign." If there is an increase in benign tumors in the exposed groups compared to the control group, with no increase in the number of malignant tumors, should that increase in benign tumors be sufficient to classify a substance as a carcinogen? If there is no statistically significant increase in the frequency of malignant tumors, but benign and malignant tumors together increase, should that serve to classify a substance as a carcinogen?

Most of the policies have taken the position that it is appropriate to count both benign and malignant tumors when evaluating the carcinogenicity of a substance, although EPA and NTP policies now provide that benign tumors will not be grouped with malignant tumors when they are of a type that is not known to progress to a malignant stage. This principle of grouping tumors based on their potential for progression appears to be gaining general acceptance. The burden, however, is to demonstrate that progression is not likely. In the absence of such evidence, the benign tumors will be grouped with malignant ones.

As an example of an early policy, NCAB (1977) argued that because compounds that induce benign tumors frequently also induce malignant ones, that because benign tumors may represent a stage in transformation to malignancy, and that because benign tumors may themselves endanger health, "if a substance is found to induce benign neoplasms in experimental animals it should be considered a potential human health hazard which requires further evaluation. " If the increase in malignant tumors is of questionable significance, the NCAB (1977) policy provided that a parallel increase in benign tumors in the same tissue adds weight to evidence for carcinogenicity. CPSC (1978), IRLG (1979), OSHA (1980), and FDA SOM (1985) all provide for grouping benign and malignant tumors together. EPA (1976) and EPA (1986) provided for combining benign and malignant tumors, unless the benign tumors are not considered to have the potential to progress to malignancy. FIFRA (1984) cited the 1984 proposal of the EPA (1986) guidelines and also includes a list, prepared for the NTP Ad Hoc Panel, of specific tumor types that should and should not be combined. OSTP (1979) apparently dropped a

1NCAB (1977), FIFRA (1984), and OSTP (1985) pointed to the lack of standard nomenclature for classifying tumor types and the need for professional judgment in examining tissue slides.

discussion of benign and malignant tumors in their final document because a "wide range of opinion" had been expressed, making it "clear that no consensus exists on this issue. "

For cases in which the animal response consists of only "benign" tumors, EPA (1976) and EPA (1986) would classify an increased incidence of benign tumors alone as only "limited" evidence of carcinogenicity. OSHA (1980) did allow for the possibility of a benign-tumor-only response, although it required a substantial amount of proof that this response is truly limited to benign tumors and that the tumors will not progress to malignancy,

For evaluating its bioassays, NTP considers chemically induced benign tumors to be an "important toxicological indicator of a chemical's carcinogenic potential in rodents, " and includes these findings in its evaluations. A substantially increased incidence of benign tumors alone may serve to place an experimental result in the category of clear evidence for carcinogenicity category, as discussed below.

Although the weight to be placed on benign tumors remains controversial, the fact remains that very few chemicals testing positive in NCI/NTP studies induced only benign neoplasms. Of 113 chemicals studied by the NTP and reviewed by the NTP Peer Review Panel at the time of the analysis, **56** were found to have evidence of carcinogenicity in rodents. Of these, only four (7 percent) were based entirely on the finding of benign neoplasia. Moreover, none of these four were placed in the category of "clear evidence. " Of the **56** chemicals, 20 were carcinogenic in all experiments (17 in all **4** experiments, 1 in 3 of 3 experiments, and 2 in 2 of 2 experiments). Of these 20, all caused malignant neoplasms (92).

Evaluation of Certain Problem Tumor Types, Common Spontaneous Tumors, and Historical Control Data.—Another difficulty in evaluating bioassay results arises when an increase in the frequency of relatively common, "spontaneous" tumors is found. "Spontaneous" tumors are ones that arise, generally for unknown reasons, in animals that are not being deliberately exposed to a carcinogenic agent. The difficulty in interpretation occurs because the incidence of some spontaneous tumors is relatively high and variable. This variability can create an especially problematic situation for evaluation when the control group in a particular bioassay happens to have had a spontaneous tumor incidence that is on the low side of the range for the particular species or strain of laboratory animal. (This may occur simply as a result of random variations among groups of animals: some will be above average, some below average in the incidence of tumors.) For a particular bioassay, the exposed groups may then have a tumor incidence that is similar to what could be expected for purely spontaneous tumors, but there would appear to be a significant excess just because the control group had an abnormally low incidence. In addition, background frequency may in some cases be affected by the animals' diet and metabolic state. On the other hand, if the study has been conducted properly, the animals will have been assigned randomly to treatment and control groups. A control group with an abnormally low incidence would be accompanied by exposed groups that would be expected to have had a similar incidence if they had not been exposed.

Historical control data consist of information on the incidence of tumors in groups of control animals of particular species and strains within a given laboratory or from a particular source of animals. Historical control data can be useful in two circumstances:

1. judging the likelihood that the difference in tumor incidence between the exposed and concurrent control group (the animals used during the actual study) can be explained as random variations within an unexposed population of animals; and
2. judging whether a rare tumor maybe of concern, even though the comparison between exposed and control groups has low statistical power because there are very few cases.

Care must be taken when using historical control data because spontaneous tumor incidence can change over time as a result of genetic changes in the animal population over generations or changes in pathology and tumor diagnosis. In addition, there may be differences of opinion among the pathologists who examined the different sets of animals. Because concurrent controls

have been treated to the same conditions, except for chemical exposure, as the exposed group, and have been examined by the same pathologists, they are the best source of information on the spontaneous tumor incidence in the group of animals being studied.

In their policies, OSHA (1980) and CPSC (1978) simply declared that a significant increase in spontaneous tumors would serve to identify carcinogens. NCAB (1977) concluded that an increase in spontaneous tumors would "raise the possibility" that a substance was carcinogenic, and IRLG (1979) urged caution in interpreting bioassay results when the observed increase in a spontaneous tumor type is within the range observed in historical controls from the same colony of animals.

OSTP (1985) acknowledged the problems of evaluating increases in spontaneous tumors, emphasized the need to consider other biological evidence, and suggested that historical controls can aid in evaluation, although "care should be exercised when combining different control groups. "

FIFRA (1984) suggested that while the occurrence of spontaneous tumors "complicates" evaluation, "judicious use" of historical control data can be of assistance, although it should not substitute for concurrent control data. However, in an example given in the text, the authors of FIFRA (1984) discounted an apparent tumor excess because the incidence in the exposed group was within the range found in historical controls. FIFRA (1984) also described how historical control data could be used for interpreting the observation of rare tumors, while it cautioned that underlying spontaneous frequency can change, depending, for example, on changes in pathology technique.

According to IRLG (1979), the occurrence of rare tumors raises suspicion and is worthy of careful review, but this occurrence by itself is "not necessarily evidence" of carcinogenicity without additional supporting evidence.

Because an increase in the occurrence of mouse liver tumors is a subject of continuing controversy, some policies have special provisions for dealing with this result. According to EPA (1986), a bioassay that shows an excess only in mouse liver tumors should be considered "sufficient" evidence of carcinogenicity when other conditions for classification of "sufficient" evidence occur. Classification of evidence could be downgraded to "limited," however, if a number of factors among the following are observed:

- the liver tumors occur only in the highest dose group or only at the end of the study,
- there is no substantial dose-related increase in proportion of tumors that are malignant,
- the tumors that occur are predominantly benign,
- there is no dose-related shortening of the time to tumor appearance,
- short-term tests are negative, or
- excess tumors occur only in one sex.

Statistical Evaluation. —Only a few of the policies actually devote any substantial discussion to the topic of statistical analysis of bioassay results. OSHA (1980), EPA (1986), and OSTP (1985) tated that such analyses shall be performed. OSHA (1980) did provide that these analyses would not be used exclusively to evaluate evidence for carcinogenicity. EPA (1986) provided, on the other hand, that evidence for carcinogenicity should be based on statistically significant response in specific organs or tissues, although the weight given to level of significance and other information is "a matter of overall scientific judgmerit. " The policies have not specified the level of statistical significance to be used in evaluation.

Overall Evaluation of Bioassay Results.—Some of the policies provide a list of some general principles for overall evaluation of bioassay results. EPA (1986) stated that the strength of positive evidence increases with:

- an increase in number of tissue sites affected;
- increase in number of species, strains, and sexes showing response;
- "occurrence of clear-cut dose-response relationships" and high level of statistical significance of tumors in treated compared to control animals;
- dose-related shortening of time to tumor occurrence or time to death with tumor; and
- dose-related increase in proportion of malignant tumors.

OSTP (1985) listed several factors that increase confidence in the conclusion that a substance is a carcinogen. These include:

- an observed dose-response relationship,
- a marked increase in tumor incidence in treatment groups,
- tumors being found at multiple sites,
- significant reduction in latent period, and
- information comparing the neoplastic stages of tumors in treatment and control groups.

In addition, information on preneoplastic lesions, target organ effects in prechronic studies, and chemical activity at physiological, cellular, and molecular levels may help.

But OSTP (1985) added a caution:

> . . . the carcinogenic effects of agents maybe influenced by non-physiological responses (such as extensive organ damage, radical disruption of hormonal function, saturation of metabolic pathways, formation of stones in the urinary tract, saturation of DNA repair with a functional loss of the system) induced in the model systems.

Tests that produce these responses need to be evaluated for human relevance, according to OSTP (1985). While there has always been concern that high-dose testing might not be relevant to human exposures, the OSTP (1985) caution represents a more explicit discussion of these potential problems than had been seen in most earlier policies.

Use of Short-term Tests and Structure-Activity Relationships

While none of the agency policies provided for using positive short-term test results as the sole basis for identifying *or* regulating carcinogens, they all indicated that such test results may be used either as supporting information (EPA 1980, EPA 1986, OSHA 1980, IRLG 1979, OSTP 1985) or as an indication that further testing may be warranted (CPSC 1978, NCAB 1977, OSTP 1979).

Few of the policies directly discussed use of structure-activity relationships (SARS) to identify carcinogens. Of those that do, none will use SARS as the sole basis for such identification, although IRLG (1979) states they might be used as corroborative evidence along with other data, or in the absence of other data, as limited, suggestive evidence. EPA (1980) states that SARs will not be used as the sole basis for quantitative risk assessments, while CPSC (1978) suggests that if related chemicals are carcinogenic, the chemical in question should be tested prior to being used in consumer products.

Evaluating Conflicting Data

There are major issues in evaluating substances for carcinogenicity when the evidence is mixed, such as both positive and negative *results* in animal bioassays, or negative data in human studies and positive animal data. Many of the policies have come to conclusions about what to do in these situations, and provide guidance for the overall qualitative evaluation.

Conflicting Animal Data.—Animal evidence may differ in two ways: it may be positive and negative studies in different species, or positive and negative studies in the same species. Agency policies generally hold that positive results in one species outweigh negative results in another. Thus, in theory, the policies imply that agencies can regulate chemicals based on positive results in a single species. This principle, that positive results supersede negative results, may be extended to cover not just conflicts between results in different species, but also conflicting results in different strains of the same species or between sexes. Of course, arguments will still occur over what constitutes convincing positive evidence. Conflicting positive and negative results in the same species will, in general, provoke a case-by-case evaluation. For example, EPA (1986) provided:

> Positive responses in one species/strain/sex are not generally negated by negative results in other species/strains/sexes. Replicate negative studies that are essentially identical in all other respects to a positive study may indicate that the positive results are spurious.

Conflicting Animal and Human Data.—Similarly, the policies generally provide that positive animal data will outweigh negative human data. OSTP (1979) points out that limitations in the power of human epidemiologic investigations to detect an effect can often explain the apparent dis-

crepancy between positive animal data and negative human data. As noted elsewhere, EPA (1980) and IRLG (1979) provide that in these situations the negative human studies can be used to set an upper bound on the estimate of human risk.

Quantitative Risk Assessment: Dose-Response Determination

The NAS committee divided the more quantitative aspects of risk assessment into dose-response determination, exposure estimation, and risk characterization. Quantitative estimation has been a particularly vexing area because of the need to make a series of often untestable assumptions to perform quantitative extrapolation from animal to human cases, the inadequacy of historical information from epidemiologic studies, and a frequent lack of data on current human exposures. During the 1970s, many argued that quantitative risk assessment was far too imprecise for use by regulatory agencies except for the relative ranking of different hazards for setting priorities. OSHA (1980) in fact adopted this position in its cancer policy, although because of court interpretations of its regulatory authority, OSHA now conducts quantitative risk assessments.

Long-term animal bioassays provide information relating the dose administered to the animals (the exposures) to the proportion of animals that are diagnosed with tumors. In these bioassays, exposure levels are deliberately set at high levels to maximize the probability of detecting a carcinogenic effect.

Information from epidemiologic studies may be used to relate some measure of exposure with the proportion of people incurring cancer, although there are often significant inaccuracies in exposure estimates made many years ago. For epidemiologic studies, the population examined is often a group of workers who were exposed to relatively high levels, often much higher than the levels workers are currently exposed to or that the public may be exposed to.

In many ways, the problems of extrapolating from effects of high doses to those of low doses differ based on whether one is using epidemiologic data or animal data. For example, an epidemiologic study may have examined the incidence of lung cancer among workers exposed to relatively high levels of arsenic inside a plant. Extrapolating from those exposures to the lower exposures found among community residents outside the plant poses difficulties, but the magnitude of the range from high worker exposure to lower ambient environmental exposure is less than that encountered when extrapolating from animal study results. As mentioned earlier, the highest exposed animals are to be exposed at the MTD. The exposures in human study populations are high compared to general environmental exposures, but are not at levels approaching the MTD. Of course, for regulating worker exposures, frequently the exposure levels for the study population are close to those found in workplaces that are to be regulated.

When using either animal or human data, the first issue is to extrapolate from the estimated probability of harm at these higher exposure levels to estimate what the probability of harm is at the lower exposure levels of interest. This high- to low-dose extrapolation is often very uncertain and controversial because there are usually few observed data on health outcomes at the lower levels.

This problem is most severe in the case of animal data because of the use of the MTD in those studies. The dose levels that the animals are exposed to are often several orders of magnitude greater than exposures that most people experience in the general environment .11

Risk assessors using epidemiologic data of exposed worker groups must also extrapolate from relatively high exposures to low exposures. Past exposures, before the agent's toxic effects were recognized or before the beginning of concerted public and private efforts to reduce workplace hazards, were often much higher than current exposures, and nearly always much higher than proposed new exposure limits. For example, workers exposed to benzene during the 1950s and 1960s at levels substantially above the current OSHA

[11] Orders of magnitude refer to differences that can be expressed in powers of 10. Thus the difference between 10 and 100 is one order of magnitude, while the difference between 10 and 1,000 is two orders of magnitude.

standard of 10 ppm incurred a significantly increased risk of leukemia. But how large is the risk of leukemia among workers currently exposed below 10 ppm?[12]

Unfortunately, there are very often few data on historical exposure levels because quantitative industrial hygiene measurements may never have been taken. In these cases, the risk assessors must make guesses about what the exposure levels were.

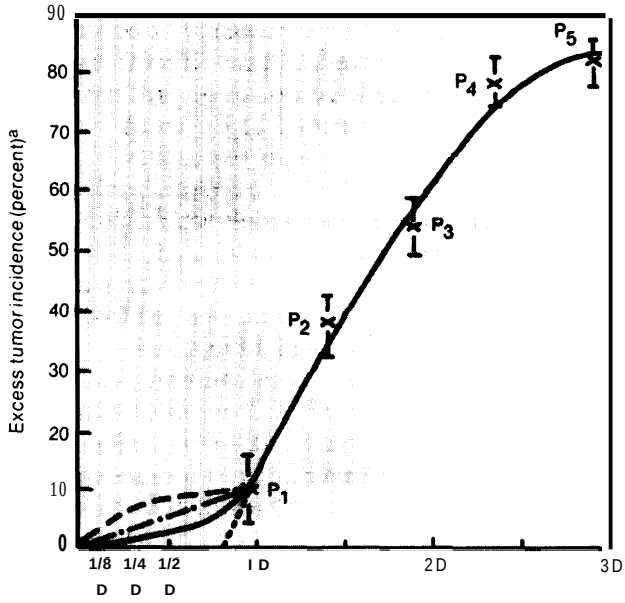
The relationship between dose (or exposure) and biological response (in this case the induction of cancer) is one of the most fundamental in the fields of toxicology and epidemiology. If data on exposures and responses are available, they may be plotted on a graph. The line joining these plotted points is called a dose-response curve. Generally, the dose-response curve has a positive slope, that is, the greater the dose, the larger the response.

Even so, the dose-response curve may have several different shapes, ranging from a straight line to differently shaped curves. Figure 2-1 shows several possible dose-response curves. In an ideal world, dose-response assessment would ascertain the shape of the curve and thus make estimates of what human risk is likely to be. In this ideal world, there would be data on the response in the range of human exposures, and there would be enough data points at different levels to distinguish between different possible dose-response curves. Alas, in the real world, the data from animals usually represent dose levels substantially higher than the range of human exposure, often several orders of magnitude higher.

A number of methods have been proposed for extrapolating from high to low doses. They range from the simple technique of drawing a straight line on a graph to sophisticated computer programs that fit the available data to develop a mathematical equation relating exposure to response. As discussed below, the Federal agencies have adopted linear no-threshold models, which, while allowing for nonlinear dose-response curves at higher exposures, extrapolate to low doses using an assumption of low-dose linearity.

[12] See ref. 172 for a recent study of the dose-response for benzene.

Figure 2-1.—A Stylized Dose-Response Curve and Some Extrapolated Curves



aExcess tumor incidence (percent) is defined as

$$\frac{\text{tumors in exposed population}}{\text{number of exposed population}} - \frac{\text{tumors in control Population}}{\text{number of exposed population}} \times 100$$

a sigmoid dose-response curve; infralinear between O and P,
-.-- linear extrapolation
— — · supralinear extrapolation
------ line projected to a threshold

SOURCE: U.S. Congress, Office of Technology Assessment, *Assessment of Technologies for Determining Cancer Risks from the .Environment* (Springfield, VA: National Technical Information Service, June 1981).

Agency Policies on Dose= Response Assessment

The different policies of the Federal agencies contain various degrees of discussion concerning some of the problematic areas of quantitative risk estimation. These policies include a series of assumptions about how to estimate human risk based on animal data. Some of the assumptions discussed in this section apply only to risk assessments based on animal data (choice of animal species and species conversion factors), while the others apply to use of both human and animal data.

Do Thresholds Exist?

The first issue in dose-response determination concerns whether there might be a “no effects” or “safe” level of exposure to carcinogens. This

particular issue has generated intense regulatory debates.

For noncarcinogenic toxic agents, toxicologists have generally believed that no-effect thresholds could be determined. To do so, several groups of animals would be exposed at different levels to a toxic agent. At the higher exposures, most of the animals might suffer toxic effects, while at the lowest levels none of the animals would show such effects. The researcher would then determine which of the various exposure levels was associated with no toxic effects and declare that to be a NOEL or a "no observed adverse effects level" (NOAEL). To estimate "safe" human exposures, the highest NOEL or NOAEL would be divided by a safety factor, often by 10 or 100. The underlying premises were that human response was similar to that in the tested animals, that humans had some ability to detoxify the harmful agent or recover from its effects, and that the safety factor would provide sufficient protection from incorrect guesses about the degree of toxicity.

However, for carcinogenic effects the general belief in the scientific community is that it is not possible to determine a no-effect threshold for carcinogens. This belief is based on observations, experimental limitations, and theoretical considerations.

Dose-response data from many epidemiologic and toxicologic studies of carcinogens fit mathematical models that are linear without an apparent threshold. Also, because of experimental limitations inherent in the size of bioassays typically used, it is not possible to demonstrate conclusively the existence of a no-effect threshold. This is especially true for risk levels of interest to regulators, which are much smaller than those detectable in these studies.

Certain theoretical considerations about cancer causation also imply the absence of no-effect thresholds. Cancer is a disease of self-replicating cells. Tumors can begin from a single cell, the DNA of which has been damaged by a small amount of a carcinogenic chemical. Unless that damage is repaired, the genetic material of the cell's "daughters" will have been altered. These daughter cells are then irreversibly "initiated" and may eventually develop into a tumor. Various cellular repair processes do exist, but these repair processes will lead to a no-effect threshold only if their efficiency is 100 percent.

Moreover, even if thresholds for individuals could be determined, genetic differences among people would make it difficult to demonstrate a no-effect threshold for a population. Finally, if exposure to a carcinogenic agent is contributing to the background incidence of cancer, the additional effect of the new carcinogen is approximately linear at low doses, without a no-effect threshold.

On the other hand, there are data from epidemiologic and toxicologic studies that are curvilinear. Some scientists interpret these data as revealing possible thresholds. Some scientists also believe that a threshold may exist if the carcinogen acts through an indirect mechanism, although it is currently difficult to distinguish among carcinogens based on mechanism. In addition, if the carcinogen contributes to the background of cancer, the additional effects will be linear and would not show a threshold. Finally, even if carcinogens lack a theoretical no-effects threshold, many believe that at some finite exposure level the additional risk is so low that it may be regarded, for practical purposes, as safe. In this case, the carcinogen might be considered to have a practical threshold.

For policy reasons, in addition to these scientific considerations, the agencies have generally assumed that carcinogenic chemicals lack no-effect thresholds. This reflects the agency's desire to be conservative in assessing risk, that is, to err on the side of safety in protecting public health. Most of the agency policies endorse the view that no-effects threshold levels do not exist for carcinogens. EPA (1980) presents this view:

> Because methods do not now exist to establish the presence of a threshold for carcinogenic effects, EPA's policy is that there is no scientific basis for estimating "safe" levels for carcinogens. The criteria for carcinogens, therefore, state that the recommended concentration for maximum protection of human health is zero.

EPA (1986), CPSC (1978), OSHA (1980), FDA (1985), IRLG (1979), and OSTP (1985) also set forth the position that there is no safe exposure

level for carcinogens. EPA (1976) stated that the linear model derived from study of radiation effects is also applicable to chemical carcinogens, but added that the costs of prohibiting exposures to some chemicals might be socially unacceptable.

Two of the policies, however, did not strongly endorse the no safe threshold principle: OSTP (1979) and FIFRA (1984), OSTP took no stand on the issue, citing disagreement within the scientific community. FIFRA argued that the no-effect threshold concept is contrary to the toxicologic principles for other kinds of toxicity, suggests that this concept has inhibited scientific discussion in this field, and provides that if an EPA evaluator believes a threshold exists, this fact should be stated along with its rationale.

Mathematical Models for Fitting Data and Extrapolating From High to Low Doses

Beyond the issue of whether safe thresholds might be discoverable, the agencies have issued some guidance concerning the techniques that will be used to describe, in mathematical terms, the dose-response relationship. This process is important because carcinogens vary a great deal—by a factor of more than 10 million—in their potency (70). Quantitative dose-response estimation is designed to produce estimates of carcinogenic potency, which can then be used to estimate the degree of human risk associated with a given exposure level or the exposure level that corresponds to a preselected level of human risk, for example, what exposure would lead to an increased risk of 1 in 1 million? Information on the degree of risk for given exposure levels can be further combined with information on exposure levels to estimate the number of cases occurring or expected to occur in a population. These estimates could then be used to decide on a particular regulatory action.

The basic issue at this step in risk assessment is to develop estimates of response rates at low exposure levels using dose-response information from the generally much higher doses of animal bioassays or epidemiologic studies. Put another way, the problem involves extrapolating from the high-dose region of the dose-response curve where animal tumor rates are in the range of 5 to 50 percent to the low-dose region corresponding to an estimated human incidence range of between 1 case for every 100,000 people (10^{-5}) to 1 case for every 100 million people (10^{-8}). '3

A variety of mathematical models have been proposed for this analytic step. These models fall into three basic types: mechanistic models, tolerance distribution models, and time-to-tumor models (351). The major ones used by the agencies include the multistage, one-hit, multihit, logit, probit, and Weibull models.

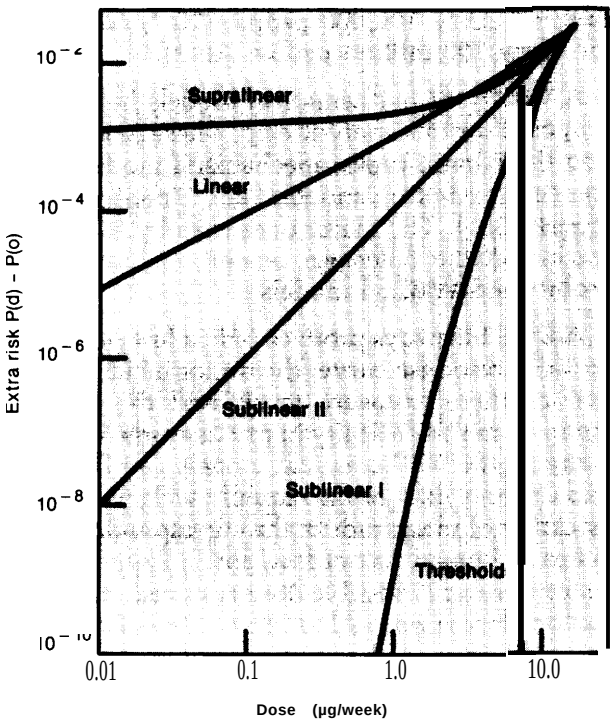
Animal bioassays produce only a few points on the dose-response curve: the tumor incidence for two or three exposure levels and the control group. A major difficulty in risk assessment is that many of the mathematical models can fit these two or three data points equally well, yet differ by orders of magnitude in the corresponding estimates of human risk at low doses. Figure 2-2 illustrates several possible dose-response curves in the low dose region. All of these curves fit the actual experimental data reasonably well, but the models underlying them utilize different assumptions about the nature of the biological processes that may be involved.

Choosing among these competing models involves questions of risk assessment policy. The major issues include crucial assumptions concerning whether there might be a no-effect threshold, whether the risk of exposure adds to the background cancer rate, whether the dose-response curve at low doses is linear, and the precise mathematical techniques for the calculation algorithm.

As a rule, the agency policies endorse models that assume the absence of no-effect thresholds and that the low-dose portion of the dose-response curve is linear. The assumption of linearity is based on several considerations and is generally thought to be conservative in the sense that it is unlikely to understate the true risk. Recent research, however, indicates that this assumption is not always true. In these cases, the dose-

lsF$_{or}$ e$_{nv}$i$_{ronmen}$t$_{i}$exposures, there is no consensus on what level of risk might clearly be considered "desirable," "acceptable," or "unacceptable, " but agencies often act to regulate exposures that pose a cancer risk greater than 1 death for every million people exposed (or a risk of 1O').

Figure 2=2.—Possible Types of Dose-Response Curves in the Low-Dose Region



SOURCE: National Academy of Sciences/National Research Council, Risk Assessment *in the Federal Government: Managing the Process* (Washington, DC: National Academy Press, 1963).

response curve may be supralinear and the use of linear models may actually underestimate the true risk (13). On the use of models, OSTP (1985) states:

> No single mathematical procedure is recognized as the most appropriate for low-dose extrapolation in carcinogenesis. When relevant biological evidence on mechanism of action (e.g., pharmacokinetics, target organ dose) exists, the models or procedures employed should be consistent with the evidence. However, when data and information are limited, and when much uncertain y exists regarding the mechanisms of carcinogenic action, models or procedures which incorporate low-dose linearit, are preferred when compatible with the limited information.

Over time the agencies have become more convinced that quantitative extrapolation is possible and useful. NCAB (1977) emphasized the uncertainties of extrapolating based on animal data. OSHA (1980) rejected an approach of setting regulatory standards based on quantitative risk assessment. On the other hand, EPA, with the formation of CAG in the mid-1970s, began applying quantitative extrapolation and the agency has now built up an extensive background in risk assessment.

One important change in the late 1970s was EPA's shift away from the "one-hit" model to the linearized multistage model developed by Kenneth Crump (discussed in EPA 1980). While this model is able to fit dose-response curves in the high-dose region that are very curvilinear, the upper confidence limit of this model is effectively linear in the low-dose region. This model was chosen because it was compatible with the multistage theory of carcinogenesis and because it uses all the data from most animal experiments (123).

Some of the policies also indicated that a variety of models should be used in any particular case (EPA 1976, OSTP 1985). The use of several different plausible models allows the risk assessor to characterize the potential uncertainty that is related to the choice of model. Some policies, in particular OSTP (1985) and EPA (1986), emphasized that selection of the extrapolation model must be chosen case by case. The chosen model should be the one that has the most correspondence with other evidence which relates to the expected mechanism of action and to the biological activity of the chemical in question. The selection should also be based on statistical considerations. In both of these policies, however, there was a preference for models which incorporate an assumption of low-dose linearity. FDA (1985), in contrast, stated that using a variety of models is not likely to provide useful information.

Should Dose-Response Estimates Be Upper Confidence Limits or the Maximum Likelihood Estimate?

There has been some debate on which of two different risk estimates-maximum likelihood estimates or upper confidence limits—should be emphasized in risk assessments. For a given model, a maximum likelihood estimate is the estimated risk at low doses that corresponds to the maximum likelihood curve, which is defined as the mathematical curve that best fits the given high-dose data. The upper confidence limit (often designated as the 95-percent confidence limit), for a

given model, is calculated under certain assumptions about the dose-response curve and is linear in the low-dose region.

For example, the results of a bioassay for exposures at 100 and 200 ppm can be fed into the computer. The program generates a risk estimate for exposures at 0.1 ppm. The maximum likelihood estimate might be 1 chance in 10,000 or 1 x 10^{-4}, while the upper confidence limit might be 1 chance in 100 or 1 X 10^{-2}.

The estimated risk at the upper confidence limit will be higher than that for the maximum likelihood estimate. For certain dose-response data, the difference will not be important. For other data, the differences may be large. Of course, if the underlying model is wrong, then both the maximum likelihood estimate and upper confidence limit will also be wrong.

Because the upper confidence limit is forced to be linear in the low-dose region, it allows for the dominant view that carcinogens lack no-effect thresholds and have dose-response curves that are linear at low doses. It is usually stated that the true risk is unlikely to exceed the upper confidence limit, and it is possible that the true risk is actually less. Although this is the case for most bioassay data, for some data sets this is not always true.

Some argue that the maximum likelihood estimate is the "best estimate" for a given model and ought to be used in preference to the upper confidence limit, which is possibly too high and thought to represent unnecessary conservatism.[14] In fact, the maximum likelihood estimate maybe misnamed. It is the extrapolated risk estimate for a particular dose on the maximum likelihood curve, which, under certain assumptions about the mathematical form of the dose-response curve, is the curve that most closely fits the actual experimental data.

But it is possible, with certain high-dose animal data, to develop a maximum likelihood estimate for the low doses of regulatory interest that does not have a linear term and would thus not exhibit low-dose linearity. Because, for reasons discussed above, it is generally presumed that the dose-response curve for carcinogens is linear at low doses, the maximum likelihood estimate in these cases would systematically understate the "most likely" risk based on our understanding of cancer causation. Moreover, small fluctuations in the underlying data at high doses in the bioassay can dramatically change the maximum likelihood estimate at the low doses of interest. The upper confidence limit is a more stable number. Finally, because in most cases the true risk is not likely to exceed the upper confidence limit, a regulation based on this estimate will be sufficiently protective. Thus, the estimates based on upper confidence limits are used not only to be conservative in assessing risk, but because a dose-response curve that is linear in the low-dose region is plausible on biological grounds.

Only two of the policies explicitly chose between the maximum likelihood estimate and the upper confidence limit. Both of these are in discussions prepared by CAG. For EPA's water quality criteria documents (323), the discussion supports the adoption of the linearized multistage procedure developed by Kenneth Crump. Although EPA's CAG still uses the upper confidence limit, EPA (1986) seems to respond to critics of the upper confidence limit with a different view:

> Such an estimate [based on the upper confidence limit] . . . does not necessarily give a realistic prediction of the risk. The true value of the risk is unknown, and maybe as low as zero. The range of risks, defined by the upper limit given by the chosen model and the lower limit which may be as low as zero, should be explicitly stated.

EPA (1986) argued that current data and procedures do not allow calculations of "best estimates, " but promises to use them if they become available, most likely "when human data are available and when exposures are in the dose range of the data. "

Choice of **Data:** Using the Most Sensitive Species

If data on carcinogenic response are available from more than one animal study or from both animal and human studies, which of these data

[14] Because different models give different estimates, the best estimate depends on the model selected.

sets should be used? The general principle in the agency policies is to use the data from the most sensitive species, that is, the study showing the highest response for a given exposure level, provided the study is of acceptable quality (EPA 1980, OSHA 1980, IRLG 1979).[15] The general rationale is that little is known about the relative sensitivities of different species and in the absence of evidence on the effects of a chemical in human beings, it is not possible to know human sensitivity. Because it is possible that humans are in fact the most sensitive species, the approach taken is to use data for the species, strain, and sex that is most sensitive, and not to reduce the risk estimates by combining data from a less sensitive species, strain, and sex.

EPA (1986) suggests use of animal data from a species that responds most like humans, if information on this correlation exists. In the more likely event that it does not exist, then the policy suggests use of all biologically and statistically acceptable data from all animal studies to identify a range of risk estimates. However, emphasis is to be placed on results from the "animal studies showing the greatest sensitivity . . . with due regard to biological and statistical considerations." When human and animal exposure routes differ, the policy states that the risk assessment should consider uncertainties about doses delivered to target organs and outline the assumptions used.

Species Conversion Factors

The second important mathematical step in dose-response assessment is estimating human doses or exposures equivalent to those used in the animal studies (converting "from mouse to man"). There are several different ways this can be done: using the ratio of body weights, the ratio of body surface areas, daily or lifetime doses; or by assuming equivalence in terms of exposure concentrations in food, air, or water. Depending on the method and whether rat or mouse data are being used, the resulting risk estimates can vary by up to a factor of 40 (217).

The two most debated methods for cross-species scaling use the ratio of body weights and the ratio of body surface areas.[16] The assumption that equivalent doses may be calculated using body surface areas was based on studies of the effects of certain drugs in different species. Studies of some other drugs and chemicals show effects that are proportional to the ratio of body weights. The choice of the body weight conversion (using mg/kg/day), will lead to risk estimates that are one-fifth (using rat data) to one-twelfth (using mouse data) of those developed using the surface area conversion (mg/m^3). Thus the risk estimates scaled using body surface area will indicate higher estimated risks for a given exposure level than those based on body weights. Both methods have their proponents.

EPA's policies (1980 and 1986) assumed that response is proportional to daily dose per unit of body surface area, although the latter policy allows for the use of other methods if information is available.[17] IRLG (1979) argued that there is no one single factor to capture the differences in animal and human susceptibility and suggests that "several species-conversion factors should be considered in estimating risk levels for humans . . . ." The other policies did not specify what species scaling factor to use.

Agency Practice of Quantitative Risk Assessment

In practice, many of the choices in performing a risk assessment for a particular chemical are specific to that chemical. These choices focus on which particular study to use; for instance, if there are several animal bioassays, which are of acceptable scientific quality, and of those that are acceptable, which particular study should be used? Also at issue is which data set to use; for example, if several tumor sites are aff ected, which one should be used? While several policies suggest using the most sensitive animal and tumor sites, often the

[15] If however, human data are also available, those data may be used to set an upper limit on the risk estimates, as discussed above.

[16] In practice, surface area is approximated by raking the body weight to the 2/3rd power.

[17] "Specifically, comparative toxicologic, physiological, metabolic and pharmacokinetic information might be useful for directly developing a cross-species extrapolation (58). However, most of the time, these data are limited or not available.

policies also suggest incorporating information on the human relevance of these tumor sites, if that information is available.

Choices like these are inescapable in risk assessment. General guidelines will not obviate these choices. It may not be desirable to do so either. If the process is made routine and a "cookbook" approach has been formulated, talented individuals with the most to contribute toward developing new approaches may be dissuaded from entering the process (86).

In discussions with OTA, agency staff often indicated that they use a flexible approach to risk assessment, incorporating new knowledge when it becomes available, and allowing choices to be made case by case to develop a risk assessment that is appropriate for a given chemical. Still, the agencies also tend to use certain "default" assumptions in the absence of other data or considerations." Some of these default assumptions are discussed in the written policies, although not all of them are. To develop an understanding of the agencies' approaches to these issues, OTA asked agency staff about the default assumptions used for extrapolating from high to low doses and for converting animal data to estimate human risk.

To extrapolate from high to low doses, EPA, OSHA, and CPSC all use the multistage model, and more specifically, the same computer program for the actual mathematical manipulations. This program uses a specific mathematical algorithm to develop an equation for the dose-response curve and then uses that equation to estimate the risk at low doses. In addition, these three agencies often run other models (e.g., one-hit, multihit, probit, logit, or Weibull models) to obtain a range of possible estimates (33,58,118).

Again, these three agencies differ in whether they use the upper confidence limit or maximum likelihood estimate. As discussed above, EPA uses the upper confidence limit of the multistage model, also known as the "linearized multistage model, " in the absence of information that would indicate the use of another model. In published risk assessments using the multistage model, OSHA has used the maximum likelihood estimate (118). CPSC also uses the maximum likelihood estimate, although only if the data appear to be linear at low doses (33).

FDA uses a different procedure, which usually gives results similar to those of the linearized multistage model. This method represents a modification of the Gaylor-Kodell linear interpolation method. In the FDA modification, the estimated response rate at the lowest dose, where curvilinearity is no longer discernible from the data, is extrapolated linearly to the background response rate (the tumor incidence when the exposure level is zero) to estimate risk at low doses. With data that are considered insufficient or of questionable quality, the upper confidence limit on the response rate is used as the starting point for this linear extrapolation. FDA has used upper limits on other models (e.g. multihit, logit, and Weibull models) and found the results comparable to the Gaylor-Kodell procedure (188).

In setting a species conversion factor for converting animal data to human risk estimates, the four agencies split evenly on the default factor to be used—EPA and CPSC convert on the basis of body surface area, while FDA and OSHA convert on the basis of body weights (33,58,118,182).

For this step in risk assessment, agency staffs are actively exploring the use of pharmacokinetic models; with appropriate data for both humans and animals, it is theoretically possible to perform the cross-species conversion directly rather than by relying on an assumption. CAG has developed a draft risk assessment for tetrachloroethylene that uses pharmacokinetic information (58). The other agencies are exploring ways to use this kind of information, both qualitatively and quantitatively, but have not yet published risk assessments that used pharmacokinetic modeling.

Thus, the four main regulatory agencies have chosen four different approaches to these three issues: the method of extrapolating from high to low doses, use of the Upper Confidence Limit or the Maximum Likelihood Estimate, and converting data from one species to predict another specie's response. How much these choices affect the

"Computer users will recognize this use of the term "default. " In computer terminology, "default" usually refers to the variables, values, or parameters that will be used unless the user specifies otherwise.

resulting risk assessments depends on the precise nature of the data used, such as the shape of the animal dose-response curve. In some cases, these different approaches could lead to important differences in estimated risk; in other cases, there would be little difference. However, compared to the use of models that assume a no-effects threshold, the models chosen by the four agencies will tend to be relatively close to each other.

Agency Policies on Human Exposure Estimation

After the dose-response characterization, the next step is to combine the resulting mathematical representation of hazard with data on actual human exposures. According to OSTP (1985), a risk assessment is only as good as the human exposure estimates, although far less attention has been paid to this aspect of the risk assessment process. In fact, exposure is often the crucial step in determining whether human risk is substantial or trivial. But this is often the area where the data are weakest. The agency policy documents generally devote only limited attention to the issues of exposure assessment, often merely presenting questions that should be answered.

OSTP (1979) recommended that research be done on exposures and CPSC (1978) stated that its staff analyses consider the nature and extent of human exposure to products containing the regulated substance and their potential for human uptake.

OSTP (1985) gave some attention to exposure assessment. Exposure assessments rely largely on monitoring data (e. g., actual measurements of chemical concentrations in water) and modeling (e.g., computer programs designed to predict exposure levels under a variety of assumptions). In the appendixes to a chapter on exposure (OSTP 1985), participating agencies presented descriptions of a number of the data bases, such as on food consumption, food additive use, modeling techniques, and other pertinent topics.

In its summary principles, OSTP (1985) argued that "a single generally applicable procedure for a complete exposure assessment does not exist. " Exposure assessments should be tailored to provide information relevant for the risk assessment and should describe the "strengths, limitations, and uncertainties of the available data and models and should indicate the assumptions made to derive the exposure estimates. " A range or array of exposure values is generally preferred to a single numerical estimate.

EPA (1976) presented a list of exposure variables to identify and factors to consider in risk assessment: known and possible exposures, data on factors relevant to effective dose, physical and chemical parameters, possible interaction of agents, likely exposure levels, both time pattern and weighted averages for total population and subgroups with different exposures, size of groups (and whether exposures involve children and pregnant women), adequacy of exposure estimation methods, and uncertainty.

EPA's Carcinogen Risk Assessment Guidelines (EPA 1986) called for a case-by-case selection of methods to match data and level of required sophistication and, unless there is evidence to the contrary, for basing risk estimates on cumulative doses received over lifetimes, expressed as average daily exposure prorated over a lifetime. Furthermore, analysts should assess the level of uncertainty in exposure assessment.

EPA has also issued separate guidelines on exposure assessment (285). The intent was that these guidelines, "by laying out a set of questions to be considered in carrying out an exposure assessment, should help avoid inadvertent mistakes of omission" (285). Thus, consistency among exposure assessments would be promoted and the information developed would be in a form compatible with dose-response assessments. The text of the guidelines, only 11 pages in the Federal *Register,* is largely an outline of points that analysts should cover in exposure assessment. These include information on the properties of the chemical in question, the sources of production and distribution, exposure pathways and environmental fate, information on measured or estimated concentrations, a description of the exposed population, and an "integrated exposure analysis. " The last item consists of the actual calculation of exposures, information on human dosimetry, development of exposure scenarios (occupational,

consumer, transportation, disposal, food, drinking water, ambient), and discussion of uncertainty. An issue of emerging importance for exposure estimation involves potential human exposures to toxic chemicals from different routes. For example, a carcinogenic chemical in drinking water might also present a dermal and inhalation hazard when people are taking showers.

The EPA guidelines are general and do not specify particular methods, procedures, or assumptions to use in the absence of data. The guidelines express a preference for measured data, but recognize the need to use mathematical modeling in many cases. In actual practice, exposure estimation has involved extensive use of computer models in the absence of exposure measuremerits. '9 The EPA guidelines encourage "the development of realistic assessments based on the best data available," rather than the use of "worst-case assessments." But EPA "will err on the side of public health when evaluating uncertainties, when data are limited or nonexistent" (285). EPA's Office of Toxic Substances has developed nine volumes presenting methods for assessing exposure in the ambient environment, chemical disposal, drinking water, occupational exposure, consumer exposure, food contamination, transportation-related spills, and on methods for enumerating and characterizing exposed populations (342).

Two of the guidelines present specific assumptions used in risk assessment calculations. EPA (1980), for the water quality criteria documents, provided an assumed average drinking water consumption of 2 liters of water per day and average fish consumption of 6.5 grams of fish per day. FDA (1985), for the SOM guidelines, states that allowable drug residue level will be set after correcting for food intake in total human diet. For these calculations, FDA specifies various "food factors" which imply that up to one-third of diet might be cattle, pig, sheep, or poultry muscle or poultry eggs, and 100 percent of diet might be cow's milk.

'"EPA, for example, has been criticized by its Science Advisory Board and others for an overreliance on computer modeling for exposure assessment (156),

Risk Characterization

The first issue in risk characterization is whether to make the characterization quantitative. The Federal agencies are all using quantitative risk assessment today, although some observers urge caution in the use of these quantitative approaches (8,156).

Different approaches to whether quantitative estimates are possible or desirable are found in the policies. CPSC (1978) and OSHA (1980) argued that quantitative risk assessment would be used at those agencies only for setting priorities. Court decisions and the general regulatory environment have superseded that stand, and both agencies today prepare quantitative risk assessments.

OSTP (1979) admitted that "extrapolation from the animal model to humans represents something of a leap of faith, " but nevertheless recommends that quantitative potency estimates should be used in determining the human risk posed by a carcinogen and that assessment of relative potencies "will aid agencies in the establishment of regulatory priorities and in the selection of appropriate regulatory action. " EPA (1976, as well as subsequent EPA policies) accept quantitative estimation as important for setting regulatory standards, but indicate it "should be regarded only as rough indications of effect. "

EPA (1986) presents several options for quantitative characterization of risk:

- unit risk—''excess lifetime risk due to continuous constant lifetime exposure of one unit of carcinogen concentration, "
- . dose corresponding to a given level of risk,
- individual risks—excess individual lifetime risk, and
- population risk—excess number of cancers in an exposed population.

Individual and population risks are those most used in policy debates. Both of these estimates incorporate information on potency from the dose-response characterization with estimates of exposures. Individual risk is the increase in the probability of disease or death for an individual in a lifetime. It is often expressed as the number of deaths per thousand or million similarly exposed

persons. Thus, exposure to a particular carcinogen might present a 1 in 1,000 lifetime risk. Population risk involves the number of excess cases of disease found in the exposed population. Thus among 70,000 people exposed to a 1 in 1,000 lifetime risk, there would be 70 excess cancer deaths associated with this exposure. Some exposures might present high individual risks among a relatively small subgroup in the population, yet also present a low population risk (because of the small number of people exposed). What to do in these situations is an important regulatory issue. Most of the policies are silent on this issue, although OSTP (1985) suggests that agencies consider identifying high-risk populations.

EPA (1986) also cautioned against using more than one significant figure in the quantitative estimates, although their published potency estimates in the past frequently had three significant figures. (See table 3-24 in ch. 3.)

EPA (1980) used a boilerplate for its risk summary of the water quality criteria documents: [name of chemical] is a carcinogen, exposures to carcinogens should be zero, but may not be attainable. "Therefore, the levels which may result in incremental increase of cancer risk over the lifetime are estimated at 10^{-5}, 10^{-6}, 10-7. " The corresponding estimates for the particular chemical were then presented.

Treatment of Uncertainty

It is also important to describe uncertainties in the characterization of the risk. There is little opposition to discussing uncertainties and assumptions in risk assessments. The agency policies differ, however, on the utility of developing a range of estimates using different extrapolation models. [20] For example, EPA (1976) urges that "where appropriate, a range of estimates should be given on the basis of several modes of extrapolation. " OSTP (1985) states that it is important to discuss the various sources of uncertainty, including statistical uncertainty, variability introduced by the chosen extrapolation model, and variability associated with interspecies scaling. The uncertainty in the choice of model can be characterized by indicating the range of estimated risks that can be developed using different plausible models.

EPA (1980) and FDA (1985) on the other hand argued that little is gained by adding estimates from several different models. EPA (1980) went so far as to state that this would "add no additional scientific information while at the same time would create confusion and thereby undermine the utility of risk estimates." OSTP (1985) advises agencies to distinguish clearly among facts, consensus, assumptions, and science policy decisions. Although it is not clear that this effort will actually reduce regulatory controversies, it may clarify the issues in dispute and outline the areas of greatest uncertainty.

New Areas for Risk Assessment Policy

Two topics have been given increased attention in recent years: possible distinctions among carcinogens based on their mechanisms of action and consideration of the pharmacokinetics of toxic chemicals within the body. The former received considerable attention and argument early in the Reagan Administration when suggestions were made that regulatory distinctions could be made based on carcinogenic mechanisms.

Distinctions Based on Mechanism

The development of cancer consists of stages. These are typically called initiation, promotion, and progression. Initiation involves an alteration in a cell's genetic material, an alteration that can remain latent (without apparent disease) for years. Promotion involves the expression of genetic information and the transformation of latent initiated cells into tumors. Progression consists of the growth of tumors and the development of metastasis in distant tissues.

Some chemicals are primarily initiators; others act only as promoters. Many chemicals are both initiators and promoters and are termed complete carcinogens.

Because they can directly damage genetic material, leading to creation of initiated cells, and

[20] Although there are other sources of uncertainty in risk assessments, much of the discussion in policies on this point concerns the choice of extrapolation model.

because such damage might be from an interaction with a very small amount of the chemical and may not be reversible, it is generally felt that initiators would not exhibit a no-effects threshold.

It is possible that the mechanism of promotion involves alteration in body chemistry, cellular growth and repair, and other processes. Because these alterations may be reversible and may not be harmful at relatively low doses, it has been suggested that there may be safe thresholds for these agents (358,359).

Weisburger and Williams have suggested the terms genotoxic and epigenetic to distinguish carcinogens based on their mechanisms (358,359). Other distinctions made, though with different meanings, are genotoxic and nongenotoxic, and direct and indirect carcinogens. But while such distinctions have been hypothesized, most scientists do not believe that chemicals in these groups can be reliably distinguished (98,157,258,351,356). This is particularly the case because many carcinogenic chemicals affect both initiation and promotion. While research on mechanisms is moving rapidly, there is currently no accepted group of tests for determining the mechanism of action for carcinogenic chemicals.

Moreover, promoters may themselves be very potent. "Dioxin" (2,3,7,8-TCDD), the most potent animal carcinogen known, may be acting as a promoter. Finally, some argue that the dose-response curve for indirect carcinogens will be linear at low doses, just as it is for direct carcinogens (86). An important argument on this point is that if it is assumed that exposure to a carcinogen adds to the background risk of cancer, the dose-response curve will be linear in the low-dose region (37).

While the issue has received enhanced attention in recent years, it is one that has been addressed in some of the earlier policies as well. In general, agency policies have refused to give a blanket endorsement of regulatory distinctions based on mechanism, although several appear willing to entertain an argument, with supporting documentation, that a substance acts only through an indirect mechanism.

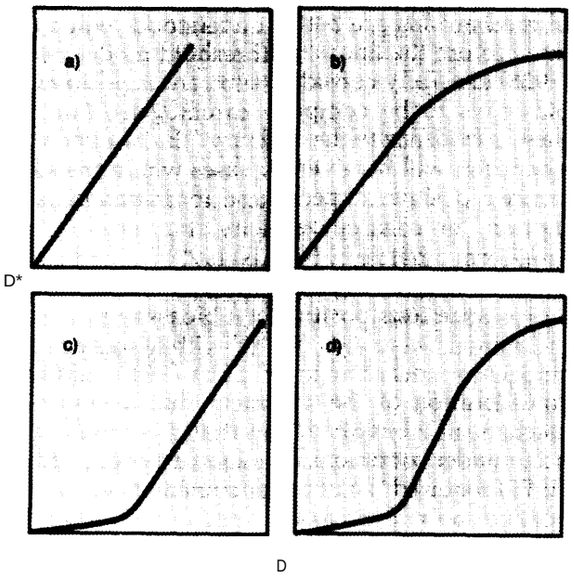
EPA (1980) saw no currently satisfactory way of estimating risk for "epigenetic" agents and until mechanisms are better understood, will continue to use the linear no-threshold model. EPA (1986) generally considers substances positive in bioassays to be complete carcinogens, "unless there is evidence to the contrary. " Individual consideration will be given to cases where the substance is positive in special tests for initiation, promotion, or cocarcinogenicity, but negative in long-term bioassays. NCAB (1977) wanted additional tests before extrapolating to humans when a substance tests positive in a bioassay in which the animals were also treated with a known carcinogen or cocarcinogen. OSHA (1980) allowed consideration of the argument that an indirect mechanism is involved and that this would not occur under conditions of human exposure. IRLG (1979) required "rigorous documentation" that a positive animal bioassay does not represent a complete carcinogenic process, but is due solely to an enhancing factor. IRLG (1979) also noted that in considering indirect mechanisms, promoters, and metabolic pathways, it would consider evidence, but expressed concern that false-negative judgments be avoided.

Incorporation of Pharmacokinetics

Exposure to a drug or chemical can lead to a variety of chemical and biological reactions in the body. The substance can be absorbed into the body, metabolized into other substances, distributed to other organs and tissues, or removed from the body. The general term for all of these processes is pharmacokinetics (171).

The various biochemical pathways in the body that activate, metabolize, detoxify, transport, and excrete chemicals will determine the relationship between the administered doses and the effective dose. As shown in figure 2-3, this relationship might be linear, but may not be. Either activation or detoxification mechanisms could become saturated or overwhelmed. In these cases, the administered dose (the dose externally administered to the animal) will not be proportionate to the effective dose (the amount of the chemical or its metabolizes) that actually reach the target tissue, and qualitatively different effects may occur. The shape of the curve relating administered dose

Figure 2-3.—Possible Types of Relationships Between Administered and Effective Dose



Possible relations between administered dose, D, **and effective dose at the target (e.g.** DNA), D', **for several kinetic** models: a) simple first-order kinetics; b) saturation of the activation system; c) saturation of detoxification or repair systems; and d) combination of b) and c).

SOURCE: D.B. Heal, N.L. Kaplan, and M.W. Anderson, "implication of Nonlinear Kinetics on Risk Assessment in Carcinogenesis," Science 219:1032-37, 1983.

with effective dose depends on the particular pathways used and the levels at which each becomes saturated .2*

While some observers hope that analysis of pharmacokinetics will refine risk assessments and improve regulatory decisions, there is much that is not understood, and there are practical difficulties in undertaking the experiments and analyzing the data. Until recently, the agencies have not attempted to include quantitative modeling of a chemical's pharmacokinetics in their quantitative risk assessments, However, even some of the earlier policies place a value on information concerning the metabolism of carcinogenic compounds.

EPA (1976) stated that a risk analysis should describe a substance's metabolic characteristics and similarities to other known classes of carcinogens at high and low doses and in different species. EPA (1980) stated that pharmacokinetic information can be useful for interspecies comparisons and estimating human risk. Further, EPA stated that relevant metabolic differences among species should be considered. EPA (1986) mandated a summary of relevant metabolic and pharmacokinetic data and suggested that this information might affect choice of a high to low dose extrapolation model. With this information or other evidence on the cancer mechanism, an extrapolation using a model other than the linearized multistage model "might be considered more appropriate on biological grounds. "

FIFRA (1984) also desired a summary of available pharmacokinetic data. In an example, the policy suggested that an evaluator should then consider the available metabolic and pharmacodynamic data for an explanation of the shape of the dose-response curve. Although not stated in the text, the data in this example appear to be from a study of rats exposed to formaldehyde. FIFRA (1984) suggested that a "threshold dose has been exceeded" in describing a very curvilinear dose-response curve.

OSHA (1980) set forth detailed requirements for arguments that metabolic differences between animals and humans justify the conclusion that a positive animal carcinogen does not pose a human health risk. This policy stated that OSHA would also consider a substance to be a carcinogen if it is metabolized into one or more potential occupational carcinogens. IRLG (1979) stated that while knowing the dose at the target organ is the ideal and that study of a substance's pharmacokinetics can in theory provide that information, there are still many uncertainties about metabolic pathways in humans and considerable variation within the human population. OSTP (1985) urged that the extrapolation model chosen needs to be consistent with available information on pharmacokinetics and target tissue dose.

Use of pharmacokinetic modeling is frequently hampered by lack of data and by our incomplete knowledge about which metabolizes are the ultimate carcinogens. Obtaining these data in humans

[21]Note that each of the four curves in figure 2-3 assumes the absence of a no-effects threshold.

for known animal carcinogens would require deliberately exposing people to suspect carcinogens —an enterprise that is ethically objectionable.

Classification of Carcinogens

Some policies provide for classifying substances by the nature or strength of the evidence for carcinogenicity. Regulatory agencies that provide such classifications on major statements include OSHA (1980), CPSC (1978), and EPA (1986, and the proposed EPA airborne carcinogen policy in 1979). NTP also has a classification of "levels of evidence" for the results of the long-term carcinogenicity studies. Finally, there is the carcinogen classification scheme of IARC.

Boxes 2-B through 2-G summarize the various classification systems. Except for the NTP "levels of evidence, " which is designed solely for evaluating animal test results, all of the systems accept human evidence and accord the highest overall classification to substances shown by human epidemiologic studies to be carcinogenic. For example, for IARC this is Group 1, for EPA (1986), Group A.

All of the policies accept the use of animal data alone for suggesting a carcinogenic hazard to humans, although there are some differences in the nature of the required evidence. A significant increase in malignant tumors in two or more species or two or more independently conducted studies in the same species is considered "sufficient" evidence by IARC and EPA (1986), "strong" evidence by CPSC (1978), and enough to bring the substance into "Category I" of OSHA (1980). Positive results in only one animal species are considered to be "limited" evidence by IARC and EPA (1986), but would be adequate to place the substance in the "high probability" category of the EPA airborne carcinogen policy (1979). Both CPSC (1978) and OSHA (1980) indicated that positive results in one species with supporting or concordant short-term test results would also lead to classification as a carcinogen (CPSC: strong evidence; OSHA: Category I). In addition, both agencies indicate that in certain circumstances they could decide to classify a substance as a carcinogen even without such supporting data. Thus, for both CPSC and OSHA, as well as EPA (1979), a substance could be classed as a carcinogen of regulatory interest based on a positive bioassay in single species.

All the schemes will classify substances as carcinogens based on both benign and malignant tumors. Both IARC and EPA (1986) include the possibility of downgrading a substance to the "limited evidence" category if the response is an increase in tumors that have a high spontaneous background rate. The classic example of this is an increased incidence of mouse liver tumors.

The IARC classification scheme presented in box 2-B represents the results of a recent modification. The major changes involved creation of a new Group 4 for agents probably not carcinogenic to humans and the criteria for Group 2. Group 2A (probably carcinogenic) will generally be used for agents with limited evidence in humans and sufficient evidence in animals. Group 2B (possibly carcinogenic) will be used for agents that have only limited evidence in humans without sufficient evidence in animals, inadequate or no data in humans but sufficient evidence in animals, and inadequate or no data in humans and limited evidence in animals when there is other supporting data. Generally agents will be classified in Group 4 based on combined evidence from animals and humans which indicate a lack of carcinogenicity (99).

The EPA weight-of-the-evidence classification system was developed as an adaptation of an earlier version of the IARC classification system. EPA (1986) quotes extensively the IARC definitions of "sufficient" and "limited" evidence. The EPA classification has regulatory meaning as well. The regulation of chemicals in drinking water depends on the weight-of-the-evidence classification. In addition, the weight-of-the-evidence classifications were used in developing adjustments of the reportable quantities of chemicals covered under CERCLA. **(See** ch. 3.)

Box 2-B.—1987 Classification of Carcinogens by the International Agency for Research on Cancer (99)

Degree of Evidence for Carcinogenicity to Humans and to Experimental Animals and Supporting Evidence

It should be noted that these categories refer only to the strength of the evidence that these agents are carcinogenic and not to the extent of their carcinogenic activity (potency) nor to the mechanism involved. The classification of some agents may change as new information becomes available.

Human Carcinogenicity Data.—The evidence relevant to carcinogenicity from studies in humans is classified into one of the following categories:

Sufficient Evidence of Carcinogenicity. —The Working Group considers that a causal relationship has been established between exposure to the agent and human cancer. That is, a positive relationship has been observed between exposure to the agent and cancer in studies in which chance, bias, and confounding could be ruled out with reasonable confidence.

Limited Evidence of Carcinogenicity. —A positive association has been observed between exposure to the agent and cancer for which a causal interpretation is considered by the Working Group to be credible, but chance, bias, or confounding could not be ruled out with reasonable confidence.

Inadequate Evidence of Carcinogenicity. —The available studies are of insufficient quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of a causal association.

Evidence Suggesting Lack of Carcinogenicity.—There are several adequate studies covering the full range of doses to which human beings are known to be exposed, which are mutually consistent in not showing a positive association between exposure to the agent and any studied cancer at any observed level of exposure. A conclusion of evidence suggesting lack of carcinogenicity is inevitably limited to the cancer sites, circumstances and doses of exposure, and length of observation covered by the available studies. In addition, the possibility of a very small risk at the levels of exposure studies can never be excluded.

In some instances, the above categories may be used to classify the degree of evidence for the carcinogenicity of the agent for specific organs or tissues,

Experimental Carcinogenicity Data.—The evidence relevant to carcinogenicity in experimental animals is classified into one of the following categories:

Sufficient Evidence of Carcinogenicity. —The Working Group considers that a causal relationship has been established between the agent and an increased incidence of malignant neoplasms or of an appropriate combination of benign and malignant neoplasms . . . in (a) two or more species of animals or (b) in two or more independent studies in one species carried out at different times or in different laboratories or under different protocols.

Exceptionally, a single study in one species might be considered to provide sufficient evidence of carcinogenicity when malignant neoplasms occur to an unusual degree with regard to incidence, site, type of tumor, or age at onset.

In the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is *sufficient evidence* of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans.

Limited Evidence of Carcinogenicity. —The data suggest a carcinogenic effect but are limited for making a definitive evaluation because, e.g., (a) the evidence of carcinogenicity is restricted to a single experiment; or (b) there are unresolved questions regarding the adequacy of the design, conduct, or interpretation of the study; or (c) the agent increases the incidence only of benign neoplasms or lesions of uncertain

neoplastic potential, or of certain neoplasms which may occur spontaneously in high incidence in certain strains.

Inadequate Evidence of Carcinogenicity. —The studies cannot be interpreted as showing either the presence or absence of a carcinogenic effect because of major qualitative or quantitative limitations.

Evidence Suggesting Lack of Carcinogenicity. — Adequate studies involving at least two species are available which show that, within the limits of the tests used, the agent is not carcinogenic. A conclusion of evidence suggesting lack of carcinogenicity is inevitably limited to the species, tumor sites, and doses of exposure studied.

Supporting Evidence of Carcinogenicity .—The other relevant data judged to be of sufficient importance as to affect the making of the overall evaluation are indicated.

Overall Evaluation

Finally, the total body of evidence is taken into account; the agent is described according to the wording of one of the following categories, and the designated group is given. The categorization of an agent is a matter of scientific judgment, reflecting the strength of the evidence derived from studies in humans and in experimental animals and from other relevant data.

Group 1: The Agent Is Carcinogenic to Humans.—This category is used only when there is *sufficient evidence* of carcinogenicity in humans.

Group 2.—This category includes agents for which, at one extreme, the degree of evidence of carcinogenicity in humans is almost sufficient, as well as agents for which, at the other extreme, there are no human data but for which there is experimental evidence of carcinogenicity. Agents are assigned to either 2A (probably carcinogenic) or 2B (possibly carcinogenic) on the basis of epidemiological, experimental, and other relevant data.

Group *2A; The Agent Is Probably Carcinogenic to* Humans.-This category is used when there is *limited evidence* of carcinogenicity in humans and *sufficient evidence* of carcinogenicity in experimental animals. Exceptionally, an agent may be classified into this category solely on the basis of *limited evidence* of carcinogenicity in humans or of *sufficient evidence* of carcinogenicity in experimental animals strengthened by supporting evidence from other relevant data.

Group 2B: The Agent Is Possibly Carcinogenic to Humans.—This category is generally used for agents for which there is *limited evidence* in humans in the absence of *sufficient evidence* in experimental animals. It may also be used when there is *inadequate evidence* of carcinogenicity in humans or when human data are nonexistent but there is *sufficient evidence* of carcinogenicity in experimental animals. In some instances, an agent for which there is inadequate evidence or no data in humans but *limited evidence* of carcinogenicity in experimental animals together with supporting evidence from other relevant data may be placed in this group.

Group *3:* The Agent Is Not Classifiable as to Its Carcinogenicity to Humans.—Agents are placed in this category when they do not fall into any other group.

Group 4: The Agent Is Probably Not Carcinogenic to Humans.—This category is used for agents for which there is *evidence suggesting lack of carcinogenicity* in humans together with *evidence suggesting lack of carcinogenicity* in experimental animals. In some circumstances, agents for which there is *inadequate evidence* of or no data on carcinogenicity in humans but *evidence suggesting lack of carcinogenicity* in experimental animals, consistently and strongly supported by a broad range of other relevant data, may be classified in this group.

Box **2-C.–1986** Classification of Carcinogens by National Toxicology Program (255)

- Clear Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing a dose-related: (i) increase of malignant neoplasms, (ii) increase of a combination of malignant and benign neoplasms, or (iii) marked increase of benign neoplasms if there is an indication from this or other studies of the ability of such tumors to progress to malignancy.
- Some Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing a chemically related increased incidence of neoplasms (malignant, benign, or combined) in which the strength of the response is less than that required for clear evidence.
- Equivocal Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing a marginal increase of neoplasms that may be chemically related.
- No Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing no chemically related increases in malignant or benign neoplasms,
- Inadequate Study of Carcinogenic Activity is demonstrated by studies that because of major qualitative or quantitative limitations cannot be interpreted as valid for showing either the presence or absence of carcinogenic activity.

When a conclusion statement for a particular experiment is selected, consideration must be given to key factors that would extend the actual boundary of an individual category of evidence. This should allow for incorporation of scientific experience and current understanding of long-term carcinogenesis studies in laboratory animals, especially for those evaluations that may be on the borderline between two adjacent levels. These considerations should include:

- the adequacy of the experimental design and conduct;
- occurrence of common versus uncommon neoplasia;
- progression (or lack thereof) from benign to malignant neoplasia as well as from preneoplastic lesions;
- some benign neoplasms have the capacity to regress but others (of the same morphologic type) progress. At present, it is impossible to identify the difference. Therefore, where progression is known to be a possibility, the most prudent course is to assume that benign neoplasms of those types have the potential to become malignant;
- combining benign and malignant tumor incidence known or thought to represent stages of progression in the same organ or tissue;
- latency in tumor induction;
- multiplicity in site-specific neoplasia;
- metastasis;
- supporting information from proliferative lesions (hyperplasia) in the same site of neoplasia or in other experiments (same lesion in another sex or species);
- the presence or absence of dose relationships;
- the statistical significance of the observed tumor increase;
- the concurrent control tumor incidence as well as the historical control rate and variability for a specific neoplasm;
- survival-adjusted analyses and false positive or false negative concerns;
- structure-activity correlations; and
- in some cases, genetic toxicology.

These considerations together with the definitions as written should be used as composite guidelines for selecting one of the five categories. Additionally, the following concepts (as patterned from the International Agency for Research on Cancer Monographs) have been adopted by the NTP to give further clarification of these issues:

The term *chemical carcinogenesis* generally means the induction by chemicals of neoplasms not usually observed, the induction by chemicals of more neoplasms than are generally found, or the earlier induction by chemicals of neoplasms that are commonly observed. Different mechanisms may be involved in these situations. Etymologically, the term *carcinogenesis* means induction of cancer, that is, of malignant neoplasms; however, the commonly accepted meaning is the induction of various types of neoplasms or of a combination of malignant and benign neoplasms. In the Technical Reports, the words *tumor* and *neoplasm* are used interchangeably.

Box 2-D.–1986 Classification of Carcinogens by the Environmental Protection Agency (284)

Group A—Human Carcinogen:

This group is used only when there is sufficient evidence from epidemiologic studies to support a causal association between exposure to the agents and cancer.

Group B—Probable Human Carcinogen:

This group includes agents for which the weight of evidence of human carcinogenicity based on epidemiologic studies is "limited" and also includes agents for which the weight of evidence of carcinogenicity based on animal studies is "sufficient." The group is divided into two subgroups. Usually, Group Bl is reserved for agents for which there is limited evidence of carcinogenicity from epidemiologic studies. It is reasonable, for practical purposes, to regard an agent for which there is "sufficient" evidence of carcinogenicity in animals as if it presented a carcinogenic risk to humans. Therefore, agents for which there is "sufficient" evidence from animal studies and for which there is "inadequate evidence" or "no data" from epidemiologic studies would usually be categorized under Group B2.

Group C—Possible Human Carcinogen:

This group is used for agents with limited evidence of carcinogenicity in animals in the absence of human data. It includes a wide variety of evidence, e.g., (a) a malignant tumor response in a single well-conducted experiment that does not meet conditions for sufficient evidence, (b) tumor responses of marginal statistical significance in studies having inadequate design or reporting, (c) benign but not malignant tumors with an agent showing no response in a variety of short-term tests for mutagenicity, and (d) responses of marginal statistical significance in a tissue known to have a high or variable background rate.

Group D—Not Classifiable as to Human Carcinogenicity:

This group is generally used for agents with inadequate human and animal evidence of carcinogenicity or for which no data are available.

Group E—Evidence of Non-Carcinogenicity for Humans:

This group is used for agents that show no evidence for carcinogenicity in at least two adequate animal tests in different species or in both adequate epidemiologic and animal studies.

The designation of an agent as being in Group E is based on the available evidence and should not be interpreted as a definitive conclusion that the agent will not be a carcinogen under any circumstances.

Box 2-E.— 1979 Classification of Carcinogens in the **EPA Airborne Carcinogen Policy** (324)

Identify carcinogens based on EPA guidelines, supplemented by IRLG guidelines, judgments based on quality and weight of evidence, classify into high, moderate, or low, based on probability of human carcinogenicity

- high probability —''best" or "substantial" evidence exists from epidemiologic and/or at least one mammalian study;
- moderate probability —''suggestive" evidence exists from epidemiologic, animal, or "short-term" studies; and
- low probability—only "ancillary" evidence exists, such as from structural correlations, or for which epidemiologic or animal results are judged to indicate low probability.

Box 2-F.—1982 Classification of Carcinogens by the Occupational Safety and Health Administration (276)

Category I Potential Carcinogens. -If substance meets definition of potential occupational carcinogen in: 1) humans, 2) a single mammalian species in a long-term bioassay where the results are in concordance with some other scientifically evaluated evidence of a potential carcinogenic hazard, 3) in a single mammalian species in an adequately conducted long-term bioassay, in appropriate circumstances where [OSHA] determines the requirement for concordance is not necessary. Evidence of concordance is any of the following: positive results from independent testing in the same or other species, positive results in short-term tests, or induction of tumors at injection or implantation sites.

Category II Potential Carcinogens.–1) Meets criteria for category I, but the evidence is only "suggestive" or 2) meets criteria for category I in a single mammalian species without evidence of concordance.

The requirement for concordance may be waived in "cases where the evidence has been carefully scrutinized and found to be unusually compelling," including the induction of many unusual tumors or early deaths for most of the exposed animals.

Box **2-G.—1978** Classification of Carcinogens by the CPSC (228)

Category A—Strong Evidence:

1. NCI has issued a finding that the substance is an animal or human carcinogen,
2. substance significantly increases incidence or reduces time to onset of benign or malignant neoplasms in humans in exposed compared to nonexposed, or
3. substance significantly increases incidence or reduces time to onset of one or more types of benign or malignant neoplasms in treated compared to control groups [of experimental animals].

Ordinarily, positive animal results must derive from systemic distribution of substance and must be obtained in:

- two animal species; or
- one species when replicated in a second experiment using independent control groups; or
- one species of test animal when supported by a battery of well designed and soundly conducted relevant short-term tests; or

Ž CPSC finds that there is other evidence sufficiently compelling to classify substance in Category A. Thus classification may be based on single, unreplicated long-term animal study.

Category B—Evidence is Suggestive:

1. human or animal data are suggestive but not conclusive because they are statistically inconclusive or methodologically deficient but nonetheless tend to support carcinogenicity;
2. positive results in one or more short-term tests, but not confirmed in human or animal studies;
3. positive results in only one unreplicated long-term animal study which is not compelling enough to classify in Category A; absence of positive short-term results.

Category C:

Substances which are members of chemical classes which include known carcinogens and other substances about which questions have been raised, but with very limited evidence.

Category D:

Reclassified substances.

AGENCY POLICIES ON CARCINOGEN RISK MANAGEMENT

Regulatory Procedures

Compared to the problems of identifying and assessing carcinogenic risk, relatively little attention is given in the various agency policies to the topic of how to reduce, eliminate, or control the risks posed by carcinogens. Only CPSC (1978), EPA (1979), OSHA (1980), and FDA (1985) give any details beyond a summary of the agency's statutory mandate on the kind of regulatory action that can be anticipated after identifying a carcinogen. In addition, OSTP (1979) suggests focusing regulatory action on particular exposures to improve the ratio of benefits to costs.

Most of these statements are in documents that have the official status of proposals (EPA 1979 and FDA 1985), or that have been suspended (OSHA) or withdrawn (CPSC 1978). However, in communication with OTA, EPA staff suggest that the airborne carcinogen policy (1979) reflects the broad outlines of their approach, FDA staff stated in their proposed SOM (1985) that they would follow the procedures outlined in it until it was published in final form. OSHA has not taken action to revoke their cancer policy.

The basic policy outlined by three of the documents is a very protective approach for eliminating or substantially reducing carcinogen exposures. CPSC (1978) states a general policy of not permitting "known carcinogens to be intentionally added to consumer products if they can be absorbed, inhaled or ingested . . . ." CPSC (1978) required the use of substitutes for identified carcinogens or reduction to the lowest attainable level" until substitutes can be found. However, actual practice at CPSC since 1978 demonstrates that CPSC has followed a less conservative course in addressing chemical hazards. As discussed in chapter 3, CPSC has often deferred to voluntary industry action and CPSC-mandated labeling of hazardous consumer products to reduce exposures (81).

For Category I carcinogens *(for which* the evidence is clear), OSHA (1980) originally required that exposures be reduced to the lowest feasible level, using engineering controls and work practices (and not through use of respirators). If suitable, safer substitutes are found, OSHA will set permissible exposures at zero to encourage substitution. The EPA airborne carcinogen policy (288) mandates that for identified carcinogens, the standards issued under section 112 of the Clean Air Act will, at a minimum, require use of best available technology. If the risk remaining after application of the best available technology is still unreasonable, EPA will consider mandating further control.

In all three cases, the agencies state that they do not believe that safe thresholds exist for carcinogens. This view, combined with their interpretations of statutory mandates, leads them to require that exposures be reduced as much as possible through the use of technology and substitution.

In its SOM paper, FDA argues that it follows a protective approach in setting a maximum lifetime risk cutoff of 1 in 1 million and in using upper bound risk estimates for determining the added risk of one permitted animal drug residue. In the U.S. population of 240 million, assuming that everyone is exposed, this would imply a maximum of 240 deaths. FDA suggests that because of the assumptions behind the development of their risk estimates, the actual risks are lower than that. In fact, they argue that it is likely that no one will actually die as a result of these exposures. The FDA policy is in fact the only one to adopt a risk level that it deems to be "safe. " Other agencies may do this informally and implicitly; FDA alone has done this explicitly.

OSHA's 1980 policy is the most detailed in describing regulatory procedures that will be followed if OSHA identifies a substance as an occupational carcinogen. In addition to regulating Category I carcinogens to the lowest feasible level (see above), OSHA will regulate what it calls Category II carcinogens (for which the evidence is only suggestive) in a manner consistent with statutory requirements. OSHA does not seem willing to force substitution, set a permissible exposure limit, or require compliance plans, hygiene facilities, and regulated areas in the case of substances with only suggestive evidence of carcinogenicity. For Category II substances, OSHA is

also to ask for additional research from the appropriate agency.

The OSHA policy further provided for the publication of lists–annual candidate lists of substances under scientific review and semiannual priority lists of the substances with the agency's highest priority ranking. Only one of these lists was ever published (277). In 1981, OSHA suspended the publication of these lists.

Finally, in an attempt to increase the timeliness of agency action, the OSHA policy set a number of specific deadlines for agency action. For example, a final standard was to be published within 120 days from the end of a hearing or 90 days from the end of a posthearing comment period, whichever is earlier. This time may be extended for one more 120-day period, unless important new evidence is found:

Regulatory and Research Priorities

Some of the policies also give guidance on setting regulatory and research priorities. In particular, the EPA airborne carcinogen policy and the OSHA cancer policy provide some general statements about setting priorities. These are very general and probably provide only limited insight into decisionmaking at either agency. The magnitude of the exposed population and the availability of controls or the low cost of applicable controls appears in both policies. The EPA airborne carcinogen policy explicitly refers to estimated carcinogenic potency and the upper bound incidence associated with exposures, both presumably derived from EPA's quantitative risk assessments. The policy refers to risk estimates for the most highly exposed individuals and the population risk.

OSHA (1980) rejected quantitative risk assessment for setting the exposure level in an occupational health standard; instead the policy was to regulate down to the lowest level feasible. However, it did suggest that quantitative risk assessment could be used in setting priorities. As discussed in other sections of this background paper, this policy was changed in response to court decisions. OSHA (1980) also mentioned that they would consider a substance's molecular structure, the potential for controls to prevent other adverse occupational and environmental effects, pending actions by other agencies, and OSHA's other responsibilities before taking action on a chemical.

OFFICE OF MANAGEMENT AND BUDGET POLICY ON CARCINOGEN RISK ASSESSMENTS

Since 1981, the Office of Management and Budget (OMB) has had an important, if not central, role in many decisions on regulatory policy through its review of agency proposals and final rules. Each of those actions takes place case by case as OMB interprets the requirement for cost-benefit analysis contained in Executive order 12291 and judges the desirability of particular regulations. Summarizing such case-by-case interpretation is difficult, in part because much of it takes place in private meetings between OMB officials, agency officials, and others. In this section, OTA will not attempt such a summary.

OMB has publicly indicated its concern over several areas of carcinogen risk assessments, and its expressed opinions on these matters are contrary to the general consensus that has evolved in the agency policies. OMB's general position is that the use of many "conservative" assumptions (in these cases, assumptions designed to err on the side of caution and minimize the chances of understating the true risk), will compound each other, guaranteeing that the estimated risk is overstated. If the estimated risks are overstated, regulatory decisions will not be as efficient or as cost-effective as they could be.[22]

For hazard identification and dose-response assessment, OMB concern involves assumptions about the treatment of benign tumors, selection of the most sensitive species and sex for risk assess-

[22]For an academic discussion of this issue, see ref. 149.

ments, and the use of conservative high- to low-dose extrapolation techniques. On benign and malignant tumors, OMB suggests that because not all benign tumors become malignant, use of the benign tumor data in the risk assessment "can overstate the real risk present. " Regarding choice of data for the risk assessment, OMB argues that use of the most sensitive species and sex will bias the risk assessment and that a "more accurate" estimate would be derived from a "weighted average of all the scientifically valid, available information. " On the choice of extrapolation technique, OMB is worried about use of the upper confidence limit, suggesting that "such an extrapolation has a 95 percent chance of overstating the true risk. " In fact, OMB is misinterpreting the meaning of the upper bound estimates prepared by the regulatory agencies.

In regard to exposure assessment, OMB expresses opinions on the use of worst-case environmental scenarios, the assumption of lifetime exposure, and the focus of regulation being placed on the most highly exposed individuals. The worst-case scenarios are used to simplify the tasks of estimating risks and setting standards. OMB is worried that these worst-cage conditions are not representative of actual conditions throughout the Nation. It is usually assumed that people might be exposed to a lifetime of drinking contaminated water or might spend their entire lives working with the same hazards. OMB thinks that this assumption can bias upward the estimates of risk because people move and change jobs. Finally, OMB expresses concern about basing regulations on the maximally exposed individual who has the worst combination of exposures (350).

One example of the implementation of OMB's approach may be found in OMB's comments to OSHA about OSHA's proposed regulation of occupational formaldehyde exposures. In those comments, OMB surveyed the epidemiologic literature on formaldehyde exposures and cancer and concluded that formaldehyde is not a human carcinogen because there is little consistency in the tumor sites among 19 different studies and little evidence that the observed excesses are actually related to the level of exposures. They cited, in addition, a recent epidemiologic study conducted by NCI, which was also negative. The approach embodied in most agency policies is to use negative epidemiologic studies only to estimate an upper bound on estimated risk and not to use these studies to dispute positive animal evidence. In fact, the position of the agency policies is that positive animal evidence outweighs negative human evidence. In contrast to this, OMB used the human evidence to cast doubt on the validity of the animal evidence. OMB also argued that the pharmacokinetics of formaldehyde exposure predict that carcinogenic effects will occur only at high doses that overwhelm the body's protective mechanisms at the exposure sites.

For developing its own quantitative dose-response assessment, OMB follows an approach to selecting animal data and extrapolating from high to low doses that is different from that used by the regulatory agencies. Instead of selecting the most sensitive sex, species, and strain or choosing one with the biologically most plausible response, OMB combined the animal responses from six different studies in three different species. In the case of formaldehyde this has a large effect on estimated risk because the studies are clearly positive only in rats; for mice and hamsters, the studies are largely negative. Thus, the OMB approach uses both positive and negative data and calls it a weight-of-evidence approach.

OMB then recalculated the extrapolation from high to low doses using several mathematical models—the multistage model and several models that do not incorporate low-dose linearity assumptions—and also incorporating different exposure scenarios and estimates of effective or delivered dose. On the basis of these calculations OMB concluded that the "carcinogenic risk of formaldehyde to workers *is* likely to be de minimis, " that is, less than 1 in a billion (349).

OSHA has not issued its final formaldehyde rule. In this particular instance, it is not clear how persuasive OMB's arguments will be to OSHA. Nor is it clear how strongly OMB will push its position about risk assessment assumptions and de minimis risks on the agencies in particular cases or if it will expect the agencies to rewrite their general guidelines.

CONCLUSIONS

Federal agency guidelines are generally consistent on major features of animal bioassay design, specifying testing in two animal species and generally requiring use of three dose groups and a control group. The guidelines agree that a study must set the highest dose as high as possible without shortening the animals' lives because of noncarcinogenic toxic effects.

Agency policies value epidemiologic studies as the most conclusive evidence for human carcinogenicity, generally presume that substances carcinogenic to animals in long-term bioassays should be treated as presenting a hazard to humans, and treat short-term test results as supportive information. Analyses of structure-activity relationships are used mostly when no other data are available.

The policies state that the agencies will use animal data derived from use of the maximum tolerated dose and will treat the appearance of malignant or benign tumors as evidence for carcinogenicity, except when the benign tumors are of a type that does not progress to malignancy. Policies usually state that positive results in animals outweigh negative epidemiologic results, and that positive results in one species outweigh negative results in another.

During the 1970s and 1980s, the agencies began using quantitative risk assessments for carcinogens. Today, while there are still considerable uncertainties in quantitative risk assessment, all the agencies use it. They assume that human risk estimates can be derived from animal data, that carcinogenic chemicals lack no-effect thresholds, and that risk estimates should be based on the most sensitive animal species. All the agencies use extrapolation models that assume low-dose linearity, although they differ on the mathematical technique to use, whether the focus should be on the "upper confidence limit" or "maximum likelihood estimate, " and the method to convert animal doses into human doses. The agency policies do not distinguish chemicals based on their mechanisms of action and are only beginning to explore the use of pharmacokinetic modeling techniques.

Agency policies give much less detailed guidance on estimating human exposures to specific chemicals. Instead, they rely on case-by-case evaluations. Nevertheless, the lack of detailed guidelines does not diminish the great importance of this factor in estimating human risk.

Despite some differences, the approaches of the regulatory agencies to carcinogen risk assessment have many similarities. OMB, on the other hand, has indicated that it does not agree with parts of the common approach.