A SURVEY OF METHODS FOR CHEMICAL HEALTH RISK ASSESSMENT AMONG FEDERAL REGULATORY AGENCIES*

Report Prepared for the
National Commission on Risk Assessment and Risk Management

by

Lorenz R. Rhomberg, Ph.D.

Assistant Professor of Risk Analysis and Environmental Health
Harvard Center for Risk Analysis
Harvard School of Public Health
718 Huntington Avenue
Boston, MA 02115

*Reprinted with permission of the Risk Commission on Risk Assessment and Risk Management.

TABLE OF CONTENTS

INTRODUCTION....1
The Maximally Exposed Individual....3
PURPOSE AND SCOPE....5
METHODS OF STUDY....7
RISK ASSESSMENT AND REGULATION....9
Risk Assessment....9
Risk Assessment and Risk Management....9
Legislation, Regulation, and the Administrative Procedures Act....10
Program-By-Program Review Of Risk Mandates And Assessment Methods....13
FOOD AND DRUG ADMINISTRATION....15
The FFDCA and its Mandates....15
Risk Mandate....18
Implementation....19
Hazard Identification....20
Dose-Response Analysis....21
Exposure....24
Risk Characterization and Regulation....28
OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION....29
The OSHAct and its Mandates....29
Risk Mandate....31
Implementation....32
Hazard Identification....33
Dose-Response Analysis....34
Exposure....36
Risk Characterization and Regulation....37
CONSUMER PRODUCT SAFETY COMMISSION....41
The CPSA, The FHSA and their Mandates....41
Risk Mandate....44
Implementation....44
Hazard Identification....45
Dose-Response Analysis....47
Exposure....49
Risk Characterization and Regulation....*50*
ENVIRONMENTAL PROTECTION AGENCY....53
Overview of Risk Assessment at EPA....54
Organizational Structure and Risk Assessment at EPA....54
Coordination, Consistency, and Oversight....56
Risk Assessment Methodology Common to EPA Programs....60
Hazard Identification for Carcinogens....62
Hazard Identification for Effects Other Than Cancer....67
Quantitative Risk Assessment for Carcinogens....68
Quantitative Analysis for Effects Other than Cancer....77
Exposure....80
Risk Characterization....81
Risk Management Considerations....83
Risk Assessment in EPA Regulatory Programs....83

PESTICIDES - EPA OFFICE OF PESTICIDE PROGRAMS 85
FIFRA and its Mandates 85
The FFDCA and its Mandates Regarding Pesticides 87
Risk Mandate 87
Implementation 88
Hazard Identification 89
Dose-Response Analysis 90
Exposure 91
Risk Characterization and Regulation 95
TOXICS - EPA OFFICE OF POLLUTION PREVENTION AND TOXICS 97
TSCA and its Mandates 97
Risk Mandate 99
Implementation 100
Hazard Identification 101
Dose-Response Assessment 101
Exposure 103
Risk Characterization and Regulation 104
EPA OFFICE OF AIR AND RADIATION 105
The CAA and its Mandates 109
Implementation for Criteria Pollutants 112
Implementation for Air Toxics 113
Hazard Identification 114
Dose-Response Analysis 115
Exposure 116
Risk Characterization and Regulation 119
EPA OFFICE OF WATER 121
The CWA and its Mandates 122
The SDWA and its Mandates 123
Risk Mandate 124
Implementation 125
Hazard Identification 125
Dose-Response Analysis 126
Exposure 126
Risk Characterization and Regulation 128
HAZARDOUS WASTE - EPA OFFICE OF SOLID WASTE 131
RCRA and its Mandates 131
Risk Mandate 132
Implementation 133
Hazard Identification and Dose-Response Analysis 135
Exposure 136
Risk Characterization and Regulation 137
SUPERFUND - EPA OFFICE OF EMERGENCY AND REMEDIAL RESPONSE 139
CERCLA and its Mandates 139
Risk Mandate 141
Implementation 141
Hazard Identification and Dose Response Analysis 144
Exposure Assessment 144
Risk Characterization 146
CONCLUSIONS 149
REFERENCES 153
TABLES 157
APPENDIX A 165
ACKNOWLEDGMENTS 173

INTRODUCTION

According to its charter, the Commission on Risk Assessment and Management is charged with investigating "the policy implications and appropriate uses of risk assessment and risk management in regulatory programs under various Federal laws." Current practices in these areas vary among Federal agencies and even among regulatory programs within the EPA. Some of this variation is attributable to different requirements among the Federal laws authorizing regulatory activity, either in the form of explicit methodological requirements that assessments must follow or as differently mandated regulatory responsibilities that the assessments must support. Other differences reflect variations in policy among organizations, adopted as a matter of differing scientific and policy judgment or simply because of the independent establishment of varying precedents and preferences.

This array of methodology reflects the fact that there is no single, agreed upon scientific procedure for the assessment of health risks from chemical exposures. The primary reason is that the needs of the risk assessment process, to make projections of possible human health risks for the variety of types and levels of exposures that may arise, far outstrip the ability of scientific investigation to give firm answers. The practical need remains, however, to make characterizations of the risk consequences (including the uncertainty about those consequences) of various potential actions and activities by industries, by government, by individuals, and by society as a whole.

Faced with this practical problem, regulatory agencies have arrived at practical methodology. This methodology includes reliance on procedures that, while attempting to embody information from the available data, of necessity rely on uncertainty-bridging principles derived from a combination of general knowledge about chemicals, their behaviors in the environment and their toxic effects, a desire to maintain internal case-by-case consistency in how uncertainties are resolved, and a desire to ensure that regulatory decisions are likely to fulfill the legislative mandates about public health protection.

The basic issues of chemical health risk assessment and the role of risk assessment methods, default assumptions, and conservatism have been discussed in the National Academy of Sciences Report, *Science and Judgment in Risk Assessment* (NRC, 1994). This document builds on earlier works taking a comprehensive view of risk assessment and the principles underlying its conduct, especially *Risk Assessment in the Federal Government: Managing the Process* (NRC, 1983), widely known and herein referred to as the "NAS red-book," and *Chemical Carcinogens: A Review of the Science and Its Associated Principles* [50 FR 10371-10442], widely known as the "OSTP Principles."

These documents epitomize an ongoing discussion that has largely succeeded in defining a common framework and structure for risk assessment. Within this framework, however, there continues to be vigorous debate about the most appropriate risk assessment approaches, the bearing of various kinds of data on risk projections, and the degree and appropriateness of conservatism in risk assessment methods. This larger debate is beyond the scope of the present report.

What is important, however, is that, faced with this continuing disagreement about methods, various Federal regulatory agencies have adopted somewhat different procedures. In part, this diversity can be attributed to the different questions being asked of the risk assessment process in different regulatory contexts by different environmental statutes. In part, it reflects different institutional judgments about the most appropriate methods and different scientific judgments about matters with high scientific uncertainty. And in part. it reflects simple policy choice made for the sake of consistency within each organization (which, owing to independent histories, becomes inconsistent among organizations).

The effect of this diversity of methods among Federal regulatory agencies is to make it difficult to compare risks, or the actions taken to mitigate those risks, from one regulatory program to another. One program's concern for a one-in-a-million cancer risk, say, may be based on an upper bound low-dose extrapolation to an average person in the exposed population extrapolated from mice based on a presumption of equal toxicity when daily doses are scaled by surface area, while another program's one-in-a-million is for a hypothetical person exposed to an agent at the regulatory limit for 45 years based on a maximum likelihood low-dose extrapolation and the presumption that equitoxic doses are proportional to body weight.

Although defaults and standard methods are necessary in the face of uncertainty and lack of case-specific knowledge, variation from group to group in these defaults enhances the sense of arbitrariness in risk assessment analyses. In cases where regulatory responsibilities overlap or when different groups have cause to assess the same exposures, differences in assessment outcome can lead to conflict and confusion among the public and the regulated community.

Despite the importance of the diversity of risk assessment methodology, a comprehensive survey and comparison among regulatory agencies has not been compiled. It is the purpose of the present report to provide such a survey, examining the risk assessment methods for potential chemically induced health effects among Federal regulatory agencies in the context of each regulatory program's enabling legislation.

Several previous surveys have addressed particular aspects of this question for particular sub-sets of Federal agencies (Rosenthal, Gray, and Graham, 1992; Schierow, 1994; Sadowitz and Graham, 1995; Hattis and Minkowitz, 1995). The present report has benefited from these papers in a way that is difficult to document with point-by-point citation, and the overall debt that this survey owes these authors is hereby gratefully acknowledged.

THE MAXIMALLY EXPOSED INDIVIDUAL

Many of the methods of quantitative risk assessment, in the face of usually incomplete case-specific data, make conservative assumptions, on the grounds that "worst-case" analyses will at least not underestimate the true human risks. This practice is criticized for leading to unrealistic and exaggerated statements of potential human health impact from low chemical exposures, which then lead to unnecessarily restrictive regulation.

An application of the worst-case principle that has received considerable attention is the emphasis on risks calculated for the "maximally exposed individual" or MEI. The notion is that, in order for a regulatory action to protect the entirety of an exposed population, it must protect the person with the most exposure; hence, the most exposed person's potential risk serves as a benchmark for the adequacy of a proposed strategy to control, restrict, or ameliorate environmental concentrations of a chemical agent.

The MEI concept has been criticized on two grounds. First, it is difficult accurately to estimate such a maximum, since the estimate will hinge on assumptions about rarely seen extremes in the habits, behaviors, and actions among those exposed, as well as of the distribution of the factors determining variation in exposure concentrations. Factors may themselves be estimated by worst-case analysis of their components, perhaps in an exaggerated way, owing to lack of empirical data, and the resulting combination of factors will compound the over-estimate, leading to a characterization of a hypothetical MEI exposure that may be far higher than the actual exposure of even the most exposed real individual in the population.

The second grounds for criticism of the MEI concept comes from a risk management perspective. It is that, even if the maximum exposure were to be accurately characterized, undue emphasis on the risks potentially posed by chemical exposure at this top end of the distribution provides no picture of the actual impact on the population as a whole. If almost all of the people in an exposed population are exposed to ten- or a hundred-fold less of the agent than is the MEI, then focusing on the MEI's risks alone may obscure the fact that only those individuals with particular uncommon habits, lifestyles, or experiences may be at any meaningful risk. Conversely, dismissing an exposure to a wide population because no single individual is at great risk may ignore a total health impact on the population that is considerable. According to critics, these problems tends to skew consideration of both monetary and social costs and benefits of regulatory strategies and may inappropriately focus attention on remedies that lower *average* exposures rather than those that reduce extreme exposures. In the view of these critics, analyses that focus on the distribution of exposures (and of potential risks), rather than on the maximum alone, will lead to more effectual, appropriate, and efficient regulatory strategies.

Science and Judgment in Risk Assessment (NRC, 1994), discussed many of the technical and assessment issues but pointed out that, having estimated the distribution of exposures, the further issue of which point (or points) of the distribution to use in analysis of regulatory options—the maximum exposure or the median, the average, a certain percentile, or the whole distribution—is a risk management decision. This latter question has been put to

the Commission, which has been asked to consider the use of risk to the MEI and other measures of potential health impact on exposed populations and their appropriate use in the risk management process.

In considering these issues, the question arises How often in current EPA practice and policies does the risk to the MEI actually form the basis of a regulatory decision? One may further ask whether any such use follows from specific mandates in the regulatory statutes, from interpretations of the regulatory intent of such statutes, from a general policy of "conservatism," or from other grounds. It is also of interest how and whether the technical difficulty of estimating the maximum exposure in a population (and the tendency, perhaps, to overestimate it) is factored in to the risk management decision process. Accordingly, this report will focus particular attention on the question of how various programs characterize exposure, on how individual risk versus population risk play in setting regulatory levels, and in particular on the role of estimates of the high end of individual exposure in this process.

PURPOSE AND SCOPE

This report comprises a survey of chemical health risk assessment methodology among the Federal agencies primarily charged with regulating the production, use, emissions, and disposal of potentially toxic chemicals. The primary focus is on differences in standard methodology among these agencies and regulatory groups, examined in the context of each group's legislative mandates. The groups included are the Food and Drug Administration (Center for Food Safety and Applied Nutrition), the Occupational Safety and Health Administration, the Consumer Product Safety Commission, and the Environmental Protection Agency, with special attention given to the various regulatory programs within the last agency.

The focus is on assessment of potential chemically induced health effects, with a particular emphasis on chronic health effects. Radiation-induced risk is not discussed. Moreover, assessment of effects other than those to human health are not examined, although the Environmental Protection Agency in particular has focused a lot of attention on environmental and ecological risks as well as non-health aspects of human welfare (e.g., the effects of acid deposition on infrastructure deterioration).

The term "risk assessment" means different things to different people. This report follows the widely followed definitions employed in the "NAS red book" (NRC, 1983). That is, methodology is discussed for hazard identification, dose-response analysis, exposure assessment, and risk characterization. Assessment of carcinogens and agents causing chronic health effects other than cancer are both included. Risk management methods are considered insofar as they interact with and represent a basis for differences in risk assessment methodology. That is, risk management issues are noted in the analysis of the public health mandates and regulatory powers of the various environmental statutes, but a comparisons of risk management analytical methodology among agencies is beyond the scope of this report.

The scope is limited to Federal agencies. Risk assessment methods by international organizations, other nations, and the various state and local governments are considered only in passing. Moreover, the large use of risk assessment as a decisionmaking tool by non-government organizations is not considered. Even within the Federal government, consideration is limited to risk assessment in support of regulatory action. Assessments conducted elsewhere in the government, for instance in planning Federal waste site cleanup, are not considered.

Much of the discussion is directed at examination of each regulatory program's enabling legislation, the statutes that mandate regulatory activity, in order to determine how their legislative purposes, mandates, and the nature of the regulatory powers they grant affect the conduct of risk assessment by particular groups. Special attention is focused on the laws' requirements about who in the exposed population is to be protected, how the distribution of

exposures among people comes into play, and how sufficiently protective standards are defined.

METHODS OF STUDY

The Federal regulatory programs covered in this survey are listed in Table 1, together with the acronyms by which they are commonly known. The primary Federal statutes under which these organizations conduct risk assessment in support of chemical regulation were examined for their statements pertaining to public health goals, key statutory language on risk, and specifications about risk assessment and its conduct. The statutes examined, their legal citations and acronyms, and the Federal offices responsible for implementing them, are listed in Table 2. Together, these laws comprise the main part of the authority to regulate potentially toxic chemicals by the Federal government. (There are a number of minor statutes, also administered by the organizations named in Table 1, that were not specifically considered.)

In addition, each organization's principal documentation on risk assessment policy and methodological guidance was examined. These documents include the EPA risk assessment guidelines, the Risk Assessment Guidance for Superfund, the CPSC chronic toxicity guidelines, and similar documents, which are cited in the body of the report.

As discussed below, however, much of the methodology for risk assessment followed by different organizations is not so clearly codified or documented. There are consistent, office-specific practices, but they are not written down as policies *per se*; rather, they are to be found in the patterns of analyses used in particular cases as documented in specific rulemaking actions.

To develop information on these practices, and to gain a perspective on the operation of each regulatory office and its activities, a series of interviews was conducted with key officials, risk assessors, and scientists in each of the offices covered by this survey. Most interviews were face-to-face, but some were conducted by telephone.

Table 3 lists the Federal officials interviewed for this study, together with the office that they represent. The interviews were designed to develop information each organization's specific practices and on how each organization sees its policies and practices in light of statutory mandates and current science. The interviews also elicited institutional views about the nature and reasons for differences in methodology among Federal organizations. The interviews were partly structured (to ensure coverage of issues) and partly unstructured. The structured discussion was based on a set of written questions made available to the interview subjects ahead of time. This set of questions is presented in Appendix A.

In addition to information gathered from the abovementioned methods, the author applied personal knowledge of the activities and methodology of different organizations gained during nine years as a risk assessor at the EPA in both the Office of Toxic

Substances and the Office of Research and Development, a period during which he participated in a good deal of interagency activity and harmonization efforts.

The policies and procedures for conducting risk assessment are not static, they have evolved over time in the face of new demands and upon the availability of new kinds of information. Indeed, the laws themselves have been amended over time, sometimes significantly. The present time is one of great reexamination of environmental regulation and of risk assessment as a tool in that regulation. Guidelines are being revised, methods are being reexamined, new policies are being debated and implemented. Bills are being debated in Congress that would significantly change the risk mandates and analysis to fulfill those mandates in most major environmental laws. Of necessity, the present report will focus on how things stand at the present moment. Some historical perspectives are given in particular cases, but no attempt has been made in the present survey to trace the history of risk assessment nor to project its future.

An attempt has been made to provide citations to references, statutes, and guidance documents wherever possible. The nature of the material is such, however, that it is frequently difficult to name a written source for specific information. As a check on accuracy and in the desire to avoid misrepresentation of any regulatory program and its practices, drafts of the sections of this report on each organization were sent to a key representative of that organization (chosen from among the interviewees) for comment and review, and the comments received have been addressed.

The statutes cited in this report have their U.S.C.A. citation incorporated into Table 2, and this citation is repeated at the beginning of the text section discussing each statute, but otherwise statutes are simply referred to by the name or acronym listed in Table 2. The numbering of sections of the statutes follows their internal numbering system (rather than the U.S.C.A. numbering) because these are the numbers most familiar to practitioners. Citations to the *Federal Register* and to court decisions are incorporated directly into the text. Other sources, cited as (author, year) in the text, are listed in the References section at the end of the report.

RISK ASSESSMENT AND REGULATION

RISK ASSESSMENT

"Risk assessment," according to the NAS (NRC, 1983), is "the use of the factual base to define the health effects of exposure to individuals or populations to hazardous materials and situations." It has four components; quoting from the red book (with slight punctuation modification) they are:

- "hazard identification—the determination of whether a particular chemical is or is not causally linked to particular health effects;"
- "dose response assessment—the determination of the relation between the magnitude of exposure and the probability of occurrence of the health effects in question;"
- "exposure assessment—the determination of the extent of human exposure before or after application of regulatory controls;" and
- "risk characterization—the description of the nature and often the magnitude of human risk, including attendant uncertainty."

The risk assessment process is seen as distinct from that of "risk management," which is defined as "the process of weighing policy alternatives and selecting the most appropriate regulatory action, integrating the results of risk assessment with engineering data and with social, economic, and political concerns to reach a decision."

Risk assessment is carried out by international agencies (for example, the International Agency for Research on Cancer [IARC] and the World Health Organization [WHO]), by many national governments, including the U.S. Federal regulatory agencies, by other Federal agencies for purposes other than regulation, by state and local governments, and by private industry in support of chemical stewardship, planning, and health and safety programs. The present report focuses on risk assessment by the U.S. Federal government in support of its regulation of potentially toxic chemicals, but it should be borne in mind that this represents but part of the whole risk assessment picture.

RISK ASSESSMENT AND RISK MANAGEMENT

Risk assessment *per se* stresses technical methods for using epidemiological, toxicological, and exposure information to develop a characterization of potential health risks and the uncertainty about those risks. It consists of an armamentarium of inference

tools and assumptions designed to bring the available scientific and technical information to bear on the questions asked by the risk management process, but by itself it does not address what practically must be done to fulfill regulatory responsibilities. Nonetheless, those regulatory responsibilities, and the questions asked by the risk management process about the likelihood of risk consequences associated with various actions or policies, must be framed in terms of the kind of information that risk assessment is capable of providing. That is, there is an important area in which risk assessment and risk management questions are intertwined. What tools are used and how their results are interpreted has policy implications. The key question, then, is how technical risk assessment methods interact with these policy questions, how the goals of a regulatory agency and its regulatory responsibilities are implemented in specific methods of analysis. Who is to be protected? At what level of protection? How certain must analyses be?

LEGISLATION, REGULATION, AND THE ADMINISTRATIVE PROCEDURES ACT

A brief overview of the legal basis for regulation may be useful. The following is based on a fuller account in Hattis and Minkowitz (1995).

Enabling legislation generally expresses the goals and ends of regulation in rather general terms. For example, environmental laws may call upon a Federal agency to "protect the public health with an ample margin of safety" or prevent "unreasonable risk of injury." The statutes then grant some particular and limited powers to the administering agency to accomplish these goals, such as the power to issue permits, limit uses, set performance standards, prohibit certain actions, mandate clean-up levels to be achieved, and so on. That is, the regulatory authority must be exercised through specific control points on activity of regulated parties. The question for the implementing agency is how practically to link its regulatory actions to desired goals. The legal means for doing this is through issuance of *regulations* that implement the laws in terms of specific activities, standards, and methods.

Regulation, although carried out by the Executive Branch of the Federal government, is quasi-legislative in that rules are made that limit, prescribe, and control the activity of the public. Regulation is also quasi-judicial, in that the rules are enforceable, with penalties exacted for lack of compliance. In allowing these powers to the Executive Branch, provision has been made to erect a form of "due process" to ensure that regulatory powers are not applied arbitrarily, that regulatory actions have not exceeded the limits to the powers granted by the statutes, and that there is an opportunity for the public to challenge regulatory decisions.

The **Administrative Procedures Act** (APA, 5 U.S.C.A. §§551 to 559) is the main vehicle for providing for such due process. It concerns questions about the granting and justification of administrative power and specifies processes that must be followed in the exercise of that power. These processes include the procedure for promulgating *rules* (the legislative-like process) and the procedures for adjudication of conflicts or

accusations of violation of those rules (the judicial-like process, including procedures akin to trials). Results of adjudication are expressed as *orders*, which are like judicial decisions and constitute final disposition of a specific case.

The procedures for issuing rules provide for a public and open process in which the affected public has the opportunity to make its arguments and in which the authority to take the regulatory action is stated. There are provisions for announcing intended regulations in the *Federal Register* and opportunity for public comment before final action is taken.

Each statute under which rules are promulgated specifies the limits to judicial review of actions. Currently, only final actions (i.e., finalized rules) are generally subject to judicial review. Risk assessments (and the analyses that go into them) are not final rules. They do not have to go through all the rulemaking procedure and are not themselves subject to judicial review. However, the regulatory actions that have been based on the risk assessments must be carried out by rule, and these rules are subject to judicial review. Once judicial review of a rule is opened, all of the basis for that rule becomes reviewable, including the authority claimed to issue the rule and the analyses that went into the formulation of the rule's specific provisions.

The APA calls for the courts to "compel agency action unlawfully withheld or unreasonably delayed." That is, a suit can be brought (by someone who may claim to have been harmed by inaction) seeking to compel an agency to fulfill a responsibility given to it by its enabling statutes. The APA also calls for the courts to "hold unlawful and set aside" any agency actions that are found to be "arbitrary, capricious, [or] an abuse of discretion" or "in excess of statutory jurisdiction, authority, or limitations." With some restrictions, certain actions can be overturned if "unsupported by substantial evidence" or "unwarranted by the facts."

The standard of evidence is whether the agency has acted in a manner that is "arbitrary and capricious." That is, substantial discretion is given to the agency's position; this includes matters of fact on which the agency has expertise as well as policy decisions made in situations where no particular stance can be deemed right or wrong. In effect, an agency need only show that it has acted in good faith on the information available; it need not show that its decisions are "correct" or even the "best."

By and large, risk assessment methodology used by the Federal agencies has held up under the arbitrary and capricious standard. A full history of litigation over environmental regulations (which is beyond the scope of this report) would reveal that many elements of risk assessment methods have been challenged, including the use of animal data to project human risks, cross-route extrapolation of effects, conservatism in default methods, and others. The agencies' positions have usually been upheld, but there are examples of significant impacts , such as the overturn of CPSC's ban on urea-formaldehyde foam insulation [Gulf South Insulation v. CPSC, 701 F.2d. 1137], which turned in part (among many other parts) on the 5th Circuit Court's opinion that the Commission had not shown a legitimate basis for human risk based on the inhalation

bioassay of rats, and the overturning of 428 OSHA Permissible Exposure Limits [AFL-CIO v. OSHA, 965 F.2d. 962], which questioned the standard use of safety factors in non-cancer risk assessment.

Although the agencies are allowed substantial discretion on risk assessment methodological matters, the fact that the "arbitrary and capricious" standard exists tends to increase the reliance on precedent, defaults, and standard methodology at the expense of case-by-case analysis. In the face of the great uncertainty inherent in risk assessment, it is hard to call any good-faith solution "capricious." But once such a solution has been used a number of times, it becomes harder to justify deviating from it in a particular case unless there are sufficient grounds to why that deviation is needed or appropriate. That is, taking a particular position in the face of uncertainty about the "correct" position is not arbitrary or capricious, but changing that position from one case to the next, without being able to show good reason for such changes, may be seen as such.

An important class of challenges for present purposes is that of challenge to the interpretation of the authorizing statute's mandate. Even if a rule is promulgated according to correct procedure, and the specific action is within the scope of actions permitted the agency, the reasons for action must legitimately flow from the mandates of the law. That is, an agency's interpretation of what kinds of actions are legitimate to fulfill obligations "to protect the public health with an ample margin of safety" or to provide "a reasonable probability of no harm" are subject to challenge. Much of the important judicial action on risk assessment has been in this realm, not challenging the risk assessment methods *per se*, but rather interpretations of the meaning of "reasonably necessary" or "adequate margin" or "safety" as they appear in statutes vis-à-vis the specific regulatory actions that these vague phrases call for and permit. There have been some significant changes in interpretation of such phrases as a result of court cases, notably the "benzene decision," [Industrial Union Department v. American Petroleum Institute, 448 U.S. 607 (1980)], and the "vinyl chloride decision" [NRDC v. EPA, 824 F.2d. 1146 (D.C. Cir.1987)], among others, as discussed in the sections on the agencies in question.

In the sections that follow, the various Federal regulatory agencies and programs addressing potentially toxic chemicals are reviewed. For each, the mandates of the enabling legislation are reviewed, and most of these are phrased in very general terms. The specific interpretation of those mandates, and the uses and methods of risk assessment to accomplish them, may seem very unconnected to the statutory language. But it should be borne in mind that these interpretations have developed over time in the context of processes that demand consistency and justification, and that questions about these matters have often been tested, tempered, and altered through the process of judicial challenge and interpretations by the courts.

PROGRAM-BY-PROGRAM REVIEW OF RISK MANDATES AND ASSESSMENT METHODS

The sections that follow present discussions of each agency and EPA program in the context of each program's regulatory responsibility, the applicable statutes under which that responsibility is given and executed, and the risk assessment methods used. There is a section on each agency (FDA, OSHA, CPSC, EPA). The EPA section is more general, covering risk assessment methods, institutions, and procedures applicable across that agency. This discussion on EPA contains a more thorough explanation of the particulars of some key methodology, against which methods elsewhere can be gauged. There follow sections on each major EPA regulatory program, noting differences and particulars that distinguish each program.

In order to aid the reader in keeping the various risk mandates, risk assessment methods, and program-specific considerations straight, a summary table—Table 4—has been included. This table of necessity simplifies the issues, and the text discussing each program should be consulted for a fuller account and a perspective derived from considering the methodological issues in their contexts.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY.

FOOD AND DRUG ADMINISTRATION

The Food and Drug Administration (FDA), which resides within the Department of Health and Human Services, has a number of divisions. The primary one of interest to this report is the Center for Food Safety and Applied Nutrition (CFSAN); most of the FDA's assessment of potential human health risks from exposure to chemical substances is conducted by CFSAN in conjunction with its regulatory responsibility over additives and contaminants of foods and cosmetics. (Other risk-related issues under FDA authority, such as side-effects from pharmaceuticals and safety of medical devices, are not considered in this report. In the area of pesticide residues on food, the FDA's former authority was transferred to the Environmental Protection Agency upon that agency's creation; some legislative aspects of this area are discussed below, but further considerations are provided in the section on EPA's Office of Pesticides Programs, p.85.)

THE FFDCA AND ITS MANDATES

The principal legislation on which FDA's authority is based is the **Federal Food, Drug, and Cosmetic Act** (FFDCA). Although it has been much amended over the years, the original act dates to 1906, making it by far the oldest among federal laws concerned with the regulation of public health risks from toxic substances. As such, much of the methodology for safety evaluation and risk assessment had its origin and early evolution in the implementation of parts of the FFDCA. The act had its origin in response to widespread scandals and "muckraking" exposés of poisonings from dangerous patent medicines, unwholesome meat packing, adulterated foods, and misrepresentations in labeling. Accordingly, the provisions of the act stress avoidance of "filthy, putrid, or decomposed" ingredients, sanitary conditions for processing and packing, proper identification and labeling, and strict limits to prevent "adulteration" of foodstuffs. It is in these adulteration provisions that toxicological risk assessment issues arise—foods are considered adulterated under the act when they contain "added substances" that are poisonous or injurious to health. The application of the act becomes somewhat arcane because the law distinguishes several categories of added substances: food additives, color additives, pesticides, and animal drugs. The question of pesticides is further complicated by the fact that regulatory authority over pesticides is shared by FDA under the FFDCA and the Environmental Protection Agency under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), as further discussed below and in the section on EPA pesticide regulation [p.85].

"Food additives" (regulated under §409) exclude adequately tested substances listed by the agency to be recognized as safe "among experts qualified by scientific training and experience to evaluate its safety" (§201); otherwise, the safety of additives is established by the agency's granting of a petition by the would-be user (although agency initiative is also allowed and pursued in practice). The petition must contain

experimental and toxicological data bearing on the evaluation together with a statement of the conditions of proposed use. In its response, the agency specifies conditions of permissible use (which may differ from those proposed) and maximal concentrations that may remain in the food when marketed. Section 409 specifies that, in considering what uses are safe, "the Secretary shall consider among other relevant factors...the probable consumption of the additive,...the cumulative effect of such additive in the diet..., taking into account pharmacologically related substances,...[and] safety factors which in the opinion of experts qualified by scientific training and experience...are generally recognized as appropriate for the use of animal experimentation data." (Although this is phrased quite generally, this still ranks as one of the more specific statements about risk assessment methods to be found among environmental laws.) Section 409 also stipulates that tolerances should be set no higher than is "reasonably required to accomplish the physical and other technical effect for which such additive is intended" notwithstanding the fact that higher levels might be deemed safe. "Color additives" are regulated under a separate section of the act (§721); other than some procedural differences, however, the risk assessment provisions are similar to those applying to additives.

This methodologic prescription applies only to non-cancer toxic effects, however, because at §409(c)(3)(A) the FFDCA contains a very specific statement about how the safety of potentially carcinogenic food additives is to be treated. This is the well known "Delaney clause," named after the sponsor of the 1958 amendment under which the provision was included in the act. It states that "no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal, or if it is found, after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man or animal." As discussed further below, the rationale cited at the time of the Delaney Clause's adoption was that carcinogens may be without a threshold concentration of toxic action; thus no exposure level could be declared "safe." This stipulation prohibits consideration of the quantitative level of risk that an additive might pose, effectively avoiding the quandary faced under other environmental laws of defining "acceptable" levels of cancer risk.

Although the Delaney Clause is powerful and far reaching in regulation under the FFDCA, it is not ubiquitous. It is as important to understand what the clause does *not* cover as what it does. There are actually three Delaney Clauses, added at different times to different sections of the FFDCA. The first, quoted above, appears in §409 and applies to food additives, but an essentially identical clause also appear in §512(d)(1)(I) regarding new animal drugs and in §721(b)(5)(B) regarding color additives.

None of these Delaney Clauses applies to regulation of cosmetics, however, which are regulated under §601 according to the standard that a cosmetic may not contain a substance "which may render it injurious to users under...such conditions of use as are customary and usual." (Color additives in cosmetics still fall under §721, however.) Nor do they apply to naturally occurring contaminants of foods that have not been specifically added, such as aflatoxin B1 on peanuts, which are controlled under §402 to a standard that "the quantity of such substance in food does not ordinarily render it injurious to health." The so-called "constituents policy" of FDA, discussed below in

the section on implementation [p.19], provides an important interpretation of the Delaney Clause limiting its scope.

Each of the Delaney Clauses specifically exempts "the use of a substance as an ingredient of feed for animals which are raised for food production" if it is found that "no residue of the additive will be found (by methods of examination prescribed or approved by the Secretary...) in any edible portion of such animal after slaughter...or in any food...derived from the living animal" [§409(c)(3)(A)]. This so-called "DES proviso" was added (in 1962) to allow the use of potentially carcinogenic animal drugs (such as diethylstilbestrol, or DES) as long as they did not harm the treated animals and left "no" residues in the derived food products. The weakness of this formulation became evident as methods for detection of chemical residues became more and more able to detect tiny, even infinitesimal amounts. This led to a quandary: the Secretary could fail to specify the most sensitive existing methods (thereby technically avoiding "detection" of chemicals known scientifically to be present) or he could specify that technical advances in detection should be used (thereby indirectly reversing decisions about "safety" of additives even though knowledge about their safety was not what was changing). Debate about the Sensitivity of Method standards produced the realization that the true issue was not about changing detectability, but about the potential for minute quantities of the agent to cause meaningful risk. This debate led to the development of the first methods for quantitative risk assessment of carcinogens at the FDA, as discussed further in the section on dose-response, below [p.21].

Regulatory authority under the FFDCA provisions about pesticides resides not in FDA but in the EPA's Office of Pesticides Programs. Pesticide residues on raw agricultural commodities are regulated under §408. When on unprocessed agricultural products, pesticides do not count as "additives" under §409 (but processing of the food can cause them to be redefined as such, as noted below). The regulation of pesticides is formally separated into questions of *registration* of pesticides (i.e., licensing for use in agriculture), carried out under the Federal Insecticide, Fungicide, and Rodenticide Act by the EPA, and the establishment of *tolerances* for pesticide residues on food as encountered by the consumer, regulated by EPA under the FFDCA §408. The authority to monitor for violations, and to seize food with excessive residues, however, resides with the FDA.

The processes of registration and of establishment of tolerances both proceed through petition by the manufacturer. Although regulated under separate laws and following different procedures, the two processes have a practical linkage in that the conditions and limitations for use of the pesticide established during registration must clearly lead to residues experienced by the consumer that will be below tolerances that can be approved on health grounds. (Consumer exposures are calculated on residues summed from the entire diet, yet pesticides must be used in certain concentrations on crops in order to achieve their pesticidal effect. Thus, registrants are careful to seek registration for only limited uses under FIFRA so that FFDCA tolerances will not reject the marketed food products. This issue is discussed further in the section on EPA's pesticide regulation [p.85].)

The above statements on pesticide tolerances apply to residues in or on raw agricultural commodities, but if those commodities are processed (by "canning, cooking, freezing, dehydrating, or milling") and the processing results in concentration of the residues such that they come to exceed the raw-product tolerance (on a per-weight basis), then §402 of the act stipulates that the pesticide residues be considered "food additives," and the provisions of §409, including the Delaney clause, apply. (§402 is often misread to stipulate that *any* increase in concentration triggers the redefinition, not just an increase that leads to violation of the tolerance.)

This is the source of the so-called "Delaney paradox" (NRC, 1987). If the pesticide is not sufficiently concentrated, its risks are assessed under §408, which mandates limits "necessary to protect the public health," i.e., allowing cancer risks to be treated in a quantitative framework similar to that used in other environmental regulation. But if the pesticide concentrates a bit more, it is unallowable under Delaney, even though the risk picture has not necessarily changed meaningfully. Although the Delaney provision is much decried, the question of how often it makes a practical difference for pesticides is a matter of some controversy. The National Research Council's report, *Regulating Pesticides in Foods: The Delaney Paradox* (NRC, 1987) sees it as severe, but in her interview for this project, Dr. Penelope Fenner-Crisp, Deputy Director of EPA's Office of Pesticides Programs, said that the Delaney Clause ends up affecting only about 10% of pesticides, since many have no positive carcinogenicity results and many that do are not concentrated sufficiently to trigger §409.

RISK MANDATE

As with most environmental laws, the mandates in the FFDCA about risk are phrased generally and depend on interpretation. Section 409, applying to additives, requires that only uses that may be demonstrated to be "safe" be permitted. Soon after this section's addition to the FFDCA in 1958, the agency officially defined "safe" as meaning "that there is a reasonable certainty in the minds of competent scientists that the substance is not harmful under the intended conditions of use" but recognized that absolute safety could not be definitively guaranteed (21 CFR 170.3). (This has commonly been codified into the phrase "a reasonable certainty of no harm," which is widely regarded as a quotation from §409, although it does not in fact appear in the act.) Under §409, consideration of benefits and costs is not allowed.

Section 408, applying to non-concentrating pesticide residues, requires setting tolerances "to the extent necessary to protect the public health," but also states that "appropriate consideration" be given "to the necessity for the production of an adequate, wholesome, and economical food supply." That is, costs and benefits are to be weighed, albeit in an unspecified way.

The act makes many general references to establishing "safety," but the only specific mandates for how this is to be done are the Delaney Clause (in its several

manifestations in the act) and the considerations specified in §409, mentioned above [p.16]. The latter are notable for specifying the use of appropriate safety factors, requiring assessment of cumulative exposures to an agent (and similar agents) in the entire diet, and mandating methods that are generally recognized by experts as appropriate.

IMPLEMENTATION

As with other environmental laws with generally phrased mandates about risk, the specifics of how risk assessment is conducted in practice at the FDA depends on the particular procedures put in place to implement the mandate. As discussed in the introduction, remarkably little of this implementation is firmly documented in citable policy documents, guidelines, or standard operating procedures. This is particularly true of the FDA. Some ascribe this to a desire to maintain as much flexibility as possible in the face of the rigidity and draconian nature of decisions mandated under the Delaney Clause, but it is perhaps more reasonable to note that the history of risk assessment at FDA is long and represents a period of considerable evolution of the role of risk considerations in regulation, from qualitative, *ad hoc*, and peripheral to quantitative, codified, and central. Much of the methodology was invented in attempts to respond to new and emerging needs from the regulatory process. In any case, the methods are codified largely in the history of evolving practice at the agency and in the documentation of regulatory actions (e.g., in the preambles to rules, laying out methods of analysis, in *Federal Register* notices). The accounts that follow are based largely on personal knowledge and interviews with FDA staff.

To a large degree, the FDA relies on seminal publications outlining risk assessment principles as the grounding for its methods. These include the 1983 National Academy of Sciences report, *Risk Assessment in the Federal Government: Managing the Process* (the so-called "red-book"), and the 1985 report from the Office of Science and Technology, *Chemical Carcinogens: A Review of the Science and Its Associated Principles* (the "OSTP Principles" 50 FR 10371-442, March 14, 1985). To a large degree, these expert consensus documents reflect compilation of insights and approaches first developed at FDA along with their elaboration and further development by the agency and other risk-assessing institutions. Unlike the EPA, however, the FDA has no officially published "guidelines" that establish standard methods for conducting risk assessment.

The FDA has made several attempts over the years to establish a *de minimis* interpretation of the Delaney clause. Under such a doctrine, if exposures are sufficiently minor that any possible cancer risk is too small to be of any legitimate regulatory concern, the exposure could be treated as virtually safe and the Delaney prohibition obviated. These attempts have been rebuffed by the courts, however, on the grounds that the Delaney clause does not speak to the level of risk. Under the DES Proviso, however, a *de minimis* interpretation (allowing up to a 10^{-6} lifetime risk regardless of sensitivity of

detection method) has been in place for residues of animal drugs since the 1979 Sensitivity of Method document (Olin et al., 1995).

An important exception to the judicial failure of *de minimis* interpretations of Delaney, however, is in the so-called "constituents policy." The issue addressed is that many additives that do not cause cancer when tested in animal bioassays nonetheless contain inevitable traces of compounds used in their manufacture which, when tested by themselves, have been found to cause cancer in rodents. Many food colors, for instance, contain traces of carcinogenic benzidine compounds used as precursors. The constituents policy, proposed in the form of an advanced notice of proposed rulemaking (ANPR) in 1982 (47 FR 14464), asserts that the reference in the Delaney Clause to ability of an additive to cause cancer applies to testing of the additive as a whole. So long as the additive as a whole tests negative, the detectable existence of contaminants that by themselves test positive does not, under this policy, trigger the Delaney prohibition. (Instead, the contaminants can be subjected to risk assessment and allowable residues set to ensure that their risk is *de minimis*.) Although no final rule has ever been issued, the policy has been applied to a number of additives (beginning in 1982 with FD&C Green No.6, 47 FR 14146). The policy was challenged in court, but the FDA position was upheld by the 6th Circuit Court in 1984 (Scott v. FDA, 728 F.2d. 322).

HAZARD IDENTIFICATION

The issue of identification of agents as carcinogens looms particularly large under the FFDCA owing to the presence of the Delaney Clause. The clause applies not only to agents that have been declared possible carcinogens through some weight-of-evidence scheme such as EPA's, but to any compound with a positive cancer assay or study, even if uncorroborated. This clearly puts a great burden on the criteria for declaring a cancer study positive. The FDA has interpreted the reference to "cancer" in the Delaney Clause quite literally—benign tumors, lacking the property of malignancy, are not strictly speaking cancers, and so studies causing only benign tumors are not seen as triggering Delaney. In the EPA's consideration of pesticides, however, benign tumors are more likely to be seen as evidence of carcinogenicity, perhaps owing to the explicitly noted role of such data in EPA's guidelines for carcinogen risk assessment, where they may add to the overall weight of the evidence.

Because the role of a single positive animal cancer study is dispositive for much FDA regulation, it has not developed the question of "weight-of-the-evidence" for the potential human carcinogenicity to the extent that other agencies have, especially the EPA. While the findings of other institutions (e.g., IARC) on an agent's carcinogenicity are of interest to FDA, they rely solely on their own determination as to whether the Delaney Clause provision has been triggered. (Hazard identification of carcinogenic pesticides is done by EPA following its criteria even though regulation may come under the FFDCA; this is discussed in the section on the EPA Office of Pesticides Programs [p.85].)

The Delaney Clause specifically prohibits agents found to cause cancer "when ingested," reflecting the food-safety concern that it embodies. However, the clause also names substances that induce cancer in "tests which are appropriate for the evaluation of the safety of food additives." In practice, this has meant that bioassays by various routes of exposure, including inhalation, may trigger the clause so long as systemic tumors (i.e., tumors remote from the site of administration) are caused. Tumors caused at the site of subcutaneous injection, respiratory tract tumors caused by inhalation of an irritant, and other such responses do not trigger the clause. (The ability of local physical injury to induce tumors was recognized in 1958, and the Delaney Clause was framed to exclude such responses.)

Regarding identification of toxicities other than cancer, there are no unique provisions in FDA methodology. Generally, it is the responsibility of the petitioner to ensure that a compound is sufficiently tested for the agency to have sufficient basis to declare that safety has been established. The agency has the power to grant a food additive petition provisionally while requiring the undertaking of more toxicological testing before unconditional approval.

DOSE-RESPONSE ANALYSIS

To a large degree, the early interest among regulatory agencies in dose-response analysis for carcinogens was prompted by FDA's struggle with the Sensitivity of Method question in the context of the DES Proviso, as mentioned above. The history of this process has been outlined by former FDA scientist Robert Scheuplein (Olin, et al., 1995, Chapter 2) and Peter Hutt (1985). The Delaney Clause had been enacted (in 1958) in response to expert opinion that carcinogens could not be assumed to have an exposure threshold below which no risk would be incurred, and thus no "safe" level could be named. Under the DES Proviso, non-detectable residues of potentially carcinogenic animal drugs were allowable in meat, but increasing sensitivity of methods of detection (and differences in detectability from one drug to another) led to shifting and inconsistent application based on the toxicologically irrelevant datum of the compound's limit of detection—the residues (and potential risks) were there irrespective of whether they were detectable by current methods. It was recognized that "no (measurable) residue" was actually a call for "no (significant) risk," and that only a risk-based criterion could produce a consistent application within the presumed intent of the DES Proviso. To implement such a policy, however, meant being able to estimate the very low human risks associated with trace amounts of a chemical, with only high-dose animal experiments to draw on as a basis.

The arguments leveled at the time against such low-dose extrapolation were many, and most would be familiar from today's debate. The matter reached crisis level in 1973 as the carcinogenic effects of DES as a *human* drug (rather than as an animal drug) were becoming known, prompting calls for action on DES residues (Hutt, 1985). By this time, however, the first influential paper on low-dose extrapolation had appeared

(Mantel and Bryan, 1961), arguing that the extrapolation may indeed be difficult, but at least an upper bound, "conservative" extrapolation could be made that would be very unlikely to be exceeded in reality. (The original Mantel-Bryan procedure was based on a conservatively chosen slope for a probit model, a somewhat different method than the low-dose linear approach subsequently adopted.) Under this procedure, it was argued, a risk that appeared trivial in a Mantel-Bryan extrapolation could under the *worst* estimation errors only be yet *more* trivial. It seemed absurd to reject such a method (and fall back on a Delaney ban) on the grounds that risks might truly be even less than a level already conceded to be negligible, and so the method of conservative low-dose extrapolation gradually won a place in regulatory analysis.

This advent raised the question of how small a risk could be considered to be trivial. Initially, a level of one-in-a-hundred million (10^{-8}) lifetime risk was used, on the grounds that, at such a level, no extra cancer cases would be expected even if the whole U.S. population were exposed. Later, when the extrapolation method was changed to the still more conservative Gaylor-Kodell procedure (as described below), the cut-off was changed to one-in-a-million (10^{-6}) (Graham, 1993) which (at least according to oral history) is the origin of the 10^{-6} lifetime risk level as a common choice for what regulation of carcinogens should achieve if the exposures are to be considered "virtually safe."

The initial role of low-dose extrapolation at FDA was to screen for cancer risks that could be considered *de minimis* in those particular settings where such findings were allowable under the Delaney Clause. To a large degree, the same is true today. Cancer potency estimation is conducted by the agency in those areas where Delaney does not apply. This class has more members than is widely recognized: the Delaney Clause does not apply to compounds (other than color additives) in cosmetics, to natural contaminants, and (under the "constituents policy," discussed above) to traces of compounds contaminating additives as an unavoidable consequence of manufacturing processes. (Again, cancer risks from non-concentrating pesticides, although regulated under FFDCA §408, are done by EPA using its own risk assessment methods, as discussed elsewhere.)

The method for dose-response characterization and low-dose extrapolation currently used by FDA is known as the modified Gaylor-Kodell method, a version of a proposal published by Gaylor and Kodell (1980). This method is more extensively discussed in the section on EPA's dose-response analysis [p.71]. Briefly, one of the various existing mathematical dose-response models is statistically fitted to the experimental data on tumor incidences at different dose levels, but the curve is used only within the range of experimental observation, not for extrapolation to low doses. (Consequently, the particular model chosen is not a great issue since in this range models closely agree.) Based on the fitted curve, an upper confidence limit is calculated for risk at a dose within this experimental range. Then, a straight line is drawn from this upper-bound risk down to the origin, and this line is used as the basis for low-dose extrapolation. This line serves as an upper bound in the sense that the "true" curve is expected to be convex to some degree and hence always below the linear low-dose

extension. Although the means for arriving at a low-dose linear extrapolation differs from that used by EPA (the so-called linearized multistage model), in practice the results generated by the two procedures are very similar, almost always within a factor of two.

The FDA calls their approach a "modified" Gaylor-Kodell method because they may make slight changes in the procedure—moving the dose at which an upper bound is calculated or using the maximum likelihood estimate of the curve rather than the upper bound to define the point of departure for extrapolation—on a case-by-case basis depending on what is judged toxicologically reasonable for the data set at hand. As with many FDA procedures, the criteria for such judgments are not recorded in guidelines; each case of modification of methods is defended on its case-specific merits.

The above procedure is used to describe the dose-response curve in the experimental animals. To apply this curve to the estimation of upper bounds on human risk, the resulting carcinogenic potency estimate must be extrapolated to humans. The FDA (like OSHA but unlike EPA and CPSC) makes the default assumption that cancer risks are equal in rodents and humans when doses are similar on a lifetime-averaged mg/kg/day basis. That is, equal cancer risks are presumed when daily amounts of agent are scaled in proportion to body weight. (As discussed in the EPA dose-response section [p.75], the EPA and CPSC use the assumption that daily amounts scaled by surface area lead to equivalent risks; this difference would lead to human cancer potency estimates that are about 13-fold higher (when extrapolating from mice) or 7-fold higher (when from rats) than the FDA body weight-scaling method, all else in the risk assessment being equal [which of course it is not].)

In the realm of non-cancer risk assessment, FDA employs a rather standard version of the "NOAEL/Safety Factor" approach (which is discussed more thoroughly in the general section on EPA's methods [p.67]). Indeed, this fundamental methodology was invented for the FDA in order to enable definition of acceptable daily intakes (ADIs) of food additives and contaminants (Lehman and Fitzhugh, 1954). An acceptable daily intake is intended to be an amount that can be ingested daily for a lifetime without harm. It is typically set at a level 100- to 1000-fold less than a daily dose that led to no elevation of toxic effects in a chronic animal study, with the "safety factors" intended to allow for the possibility that humans might be more sensitive to the agent (on a mg/kg/day dosing basis), that humans are expected to have greater variation among individuals (and hence more especially sensitive individuals) than the experimental animals (which are often inbred strains), and similar considerations. This method is less a means to estimate risks as such, but rather one to determine a dose level unlikely to produce any non-cancer toxic effect, even if the exposure is experienced daily. Exposures higher than this level may or may not lead to toxic reactions.

There are a few features particular to FDA's use of this methodology. They require extrapolation from a dose level without experimentally observed induced toxicity (NOAEL, or no observed adverse effect level). When no available study demonstrates such a dose without evident adverse effect, the FDA usually declines to calculate an ADI, preferring to ask for another study with lower dose levels. Other agencies (such as EPA)

may sometimes extrapolate from the lowest dose showing increased toxicity (a LOAEL, or lowest observed adverse effect level) by including an extra 10-fold uncertainty factor. For developmental toxicity, FDA may use a combined safety factor of 100 for less severe, reversible endpoints and 1000 for more severe, irreversible endpoints, whereas other agencies typically do not make this distinction, at least in the calculation of safety factors. Finally, FDA's choice of safety factors tends to be rather rigidly codified according to a few parameters of the toxicological study being used (chronic vs. sub-chronic, animal vs. human, reversible vs. irreversible, etc.), in contrast to the EPA, where the methodology explicitly allows for some case-by-case modification of the overall safety margin (though additional "modifying factors") to allow for special concerns, particularly great or particularly small uncertainty, availability of pharmacokinetic data, and so on (as discussed in the EPA section [p.77]).

EXPOSURE

The exposure concerns under the FFDCA are principally focused on the amount of regulated compounds ingested as a result of consumption of foods. The exposure assessment has two chief components: determination of the concentrations of the compound in various foodstuffs (residues) and determination of the amounts of various foods that are ingested (consumption). The second is the more uncertain component, and also the one most subject to interindividual and temporal variability.

For additives and colorings, it is usually the case that the compound in or on the various foodstuffs is intentionally added in measured amounts. Even in the case of secondary contamination from packaging, the processes leading to residues are more-or-less uniform and describable. Thus, in comparison to the case of "environmental" regulation, the concentrations in the contaminated medium (i.e., food) can in principle be known quite well; issues of uncertainty in "emissions" and in "fate and transport" processes, which loom large in other contexts, are not of concern.

Several provisions of the FFDCA have particular bearing on how FDA assesses exposures. First, as has been mentioned, §409 explicitly calls for an assessment of "the cumulative effect of such additive in the diet...,taking into account pharmacologically related substances." That is, safety of an additive is not to be assessed use by use, but rather according to the sum of the compound in the diet as a whole. For a new compound, all potential uses that will become approved must be considered, and for an existing additive with a new use, exposure from that new use must be considered in the context of its addition to the existing exposure burden.

Second, the petition process occurs before the substance actually enters use. Thus, at the time of such a pre-market evaluation, all exposures are hypothetical. Information on the anticipated residues is usually provided by the petitioner, although the FDA may do its own analysis instead of or in addition to that of the petitioner.

Third, the regulatory power granted the agency under the act is to prescribe "with respect to one or more proposed uses..., the conditions under which such additive may be safely used." That is, the agency is charged with defining the limits to use that will be allowable; all potential uses up to and including such limits become permissible and must be found to be "safe" within the meaning of the act even though the use of the additive may be less in practice. In other words, the regulatory power is over potential use, not actual use.

Together, these provisions lead the FDA to assume in its exposure assessments that all of the foods in which an additive is permitted actually bear it at the level of the tolerance, i.e., at the maximum permissible residue level. For example, a food coloring may be subject to a rule specifying maximum permissible use in a number of food types, but the actual uses on particular foods may be less than the permitted level, depending on the coloring effects desired by the manufactures for each product. Nonetheless, they are free under the regulation to use more of the coloring, up to the prescribed limits if they so choose. FDA is mandated to set the limits such that, if this permission for use were in fact exercised, the cumulative exposure in the diet would be safe.

This practice is clearly conservative with respect to the actual residue levels, although for two reasons the degree of conservatism is probably not great. First, petitioners generally seek approval of uses (and use levels) they intend to put into actual practice, and second, the prohibition by the act of residues greater than reasonably necessary to achieve the intended effect implies that residues significantly lower than tolerance limits would be avoided as ineffective.

Exposure assessments are needed for a "post-market" analysis when a new use is petitioned for, when it is believed that appreciable changes in intake have occurred, or when new toxicity information prompts reconsideration of established tolerances. Post-market exposure analyses are similar to pre-market analyses except that real (rather than hypothetical) information may be available on residue levels. When a new use is in question, existing uses are part of the cumulative background of exposure that must be considered, and when possible FDA may characterize this background in terms of actual residues currently in the food supply.

Contaminants (as opposed to additives) are never intentionally added and have no necessary or desirable level in food. The regulatory concern is not definition of permissible uses, but rather effectiveness of efforts to avoid the contaminant and its potential risks. Accordingly, exposure assessment for contaminants typically stresses actual, rather than permissible, residue levels.

The preceding paragraphs treated the determination of residues on different food types. The second component of FDA's exposure assessment concerns the determination of the rate of consumption of different foods. Clearly, this is an area of greater uncertainty than the determination of residues. There are also variability questions in terms of interindividual differences in food preferences, seasonal changes in food availability, and changes in diet with age.

The main sources of information on daily intakes of various foods are food-intake surveys, food-frequency surveys, and data from manufacturers and distributors on annual poundage produced. In food-intake surveys, participants record the types and amounts of food they eat over some defined survey period, usually a few days. Food-frequency surveys are similar except that no data are taken on the amounts of each food consumed; instead, the analyst uses information from other sources on portion sizes to convert the number of occasions on which various foods were eaten into information on amounts consumed.

The third type of information, production data, produces only indirect estimates of consumption. The annual production of the foodstuff is estimated from data supplied by manufacturers and distributors. The total may be corrected to adjust for lack of exhaustive reporting, but it is often more difficult to adjust for amounts of food that are exported, lost due to spoilage or damage, or not consumed for other reasons. (For this reason, such data are often referred to as "disappearance data" since one cannot be sure about how much of the food that disappears has been consumed.) Dividing the estimated total annual consumption by the U.S. population size (and by 365 days/year) gives an estimate of per capita daily consumption. Clearly, this figure is only an average over all people and all days. Since many people will not consume the food type at all, the average consumption among those that do will be larger than the calculated per capita amount. This method cannot produce information on the distribution of consumption over individuals, demographic groups, or daily variations in diet. It is used when appropriate survey data are unavailable, and it can provide a useful check on survey-based methods.

Survey-based methods are preferred because they can be used to address questions of variability and the distribution of food consumption levels among individuals and demographic groups. Indeed, exposure estimates are made for a number of groups broken down by age, sex, ethnicity, and (on a broad scale) geographic region. There are difficulties, however, not the least of which is the availability of good and current food consumption survey data. The FDA uses a variety of publicly available data sources (primarily from the USDA), but many of the data come from the 1970's and early 1980's, raising questions about whether national trends in changing food consumption have been captured. There are proprietary data bases from which information may be purchased, and sometimes petitioners supply consumption data. One difficulty is that surveys often report certain food consumptions by categories (such as baked goods or dairy products) that represent a greater degree of aggregation (or a different aggregation) than appropriate for matching with the residue data (which may be for only certain baked goods or dairy products, for instance). The FDA is currently grappling with improving its data sources.

Surveys usually cover a study period of a few days; that is, they represent a "cross-section" sample of consumption behavior. It is problematic to extrapolate the food-use frequencies from such a cross-section to those expected in the long run because, without longitudinal data, one cannot separate variation among individuals in diet from

day-to-day variation in consumption patterns for each individual. For example, if a one-day survey found that 3% of participants ate artichokes, does this mean that 3% of the U.S. population eats artichokes every day (and 97% never eat artichokes) or that every person eats artichokes on 3% of the days of their lives (and no artichokes on 97% of the days)? The former assumes that all the variation is among individual diets but that individual people have no day-to-day variation, and the latter assumes that all the variation is among days, with every person having the identical "average" diet in the long run. Neither extreme is plausible, but with the usual cross-sectional sampling design there is no means to characterize the mix of these effects. (The surveys most used have sampling periods of three or 14 days, so a beginning, albeit an inadequate one, can be made on estimating day-to-day variation.) Typically, during pre-market analyses the assumption is made that, at least within demographic subcategories, all the variation represents variation among individuals; that is, the average daily consumption of a food during the survey period is assumed to apply to that person for his or her whole life, and the results for different survey participants are assumed to reflect differences from one person to the next in each person's chronic consumption. The effect of this assumption is an acknowledged overestimation of high-end chronic exposures and an underestimation of the proportion of the population ever consuming particular foods. (It is noteworthy that the alternative assumption—that the surveyed variation represents day-to-day variation, resulting in an estimate of average chronic consumption but failing to estimate high-end exposures at all—is used by the EPA's Office of Pesticide Programs, as discussed on p.93.) The FDA justifies this procedure on the grounds that the aim of the analysis is to define residue levels that will have a reasonable probability of no harm, which should include protection of frequent users of the foods in question.

This issue is one of several related problems having to do with correlation and interdependence among variables that affect interpretation of food survey data. They are recognized but difficult to correct. Among them are the problem that some additives may generate market appeal (or avoidance) of the food product bearing them, skewing food use frequencies in an additive-specific way. Certain foods may tend to be eaten together (beer and pretzels) or to exclude one another from individual diets (regular and diet soft drinks). To varying degrees, individuals have daily "quotas" for caloric and nutrient intake and for use of certain food groups (e.g., meat, beverages, vegetables), making daily food selections to fulfill those quotas. Consumption of certain foods, therefore, makes the consumption of other foods, especially foods in similar categories, less likely in the short run. All of these issues make it difficult to define unbiased estimates of the distribution of consumption of various foods.

To complete the exposure assessment, residue estimates on each food type are combined with estimates of daily consumption of each food type to give a total estimated daily intake, or EDI. Exposures for various demographic groups may be calculated, and in each case there is an attempt to characterize a mean exposure and a "90th percentile" exposure (although the difficulties of making distributional estimates of consumption has been noted above). When possible, average exposures are calculated for the consuming population only. (When several food types are involved, this can involve assumptions about how consumption is correlated over food types.)

RISK CHARACTERIZATION AND REGULATION

Because of the Delaney Clause, risk characterization as such is obviated for carcinogenic food additives, colorings, and concentrating pesticides. (Non-concentrating pesticides, regulated under §408, are discussed under the section on EPA's Office of Pesticides Programs [p.87].) For agents that are not carcinogenic, risk characterization under the FFDCA focuses on whether the mandate of "reasonable probability of no harm" will be achieved under the proposed set of limits on use and permissible residues. Thus, the main issue is whether the higher end (nominally, the 90th percentile) of the distribution of estimated daily intakes is below the acceptable daily intake calculated from toxicity data. That is, the primary concern is for individual risk rather than population risk, with the mandate being interpreted as requiring that, if use of a food additive is to be declared safe within the meaning of the act, heavy consumers of particular foods should be reasonably assured of protection even if residues were at the maximal level allowed. Despite this focus on the upper end of individual exposures, the concept of the maximally exposed individual is not used as such. FFDCA §409 explicitly requires a health-based standard of regulation; consideration of costs and benefits is not allowed. Secondary to the heath-based requirements are technical requirements that permissible residues be no higher than needed to achieve their intended effect and that they be reduced to the extent feasible, even if higher levels could be declared safe.

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION

The Occupational Safety and Health Administration (OSHA) resides within the Department of Labor. This agency, founded in 1970, is responsible for regulation affecting workplace safety. OSHA has divisions concerned with injury prevention and with exposure to potentially harmful substances; it is the latter that is the focus of this report, comprising the Directorate of Health Standards Programs.

THE OSHACT AND ITS MANDATES

OSHA was created by, and has its regulatory authority under, the **Occupational Safety and Health Act** of 1970. (Since the agency and the act share the same acronym, the act is typically abbreviated as "OSHAct" and the agency itself as "OSHA.") The act's stated purpose is "to assure so far as possible every working man and woman in the Nation safe and healthful working conditions" by several means, including "providing medical criteria which assure insofar as practicable that no employee will suffer diminished health, functional capacity, or life expectancy as a result of his work experience" (OSHAct §2). It was passed during the heyday of public concern about environmental health that also saw the founding of the Environmental Protection Agency. Regulatory decision-making under the OSHAct is formally invested in the Secretary of Labor.

The act mandates in §5(a) that "Each employer...shall furnish to each of his employees employment and a place of employment which are free from recognized hazards that are causing or are likely to cause death or serious physical harm." The regulatory authority of OSHA is provided by §6 of the act, which sets out methods and criteria for issuance of occupational safety and health standards. In particular, §6(b)(5) states that "The Secretary, in promulgating standards dealing with toxic materials or harmful physical agents...,shall set the standard which most adequately assures, to the extent feasible, on the basis of the best available evidence, that no employee will suffer material impairment of health or functional capacity even if such employee has regular exposure to the hazard...for the period of his working life." This paragraph further states that "In addition to the attainment of the highest degree of health and safety protection for the employee," the Secretary must consider "the feasibility of the standard" and that "Whenever practicable, the standard promulgated shall be expressed in terms of objective criteria and of the performance desired."

In other words, the achievement of safe and healthful workplaces is to be brought about by the setting of enforceable workplace standards, in practice framed primarily in terms of allowable limits to employee exposure. For a workplace to be considered healthful, the limits to exposure are to be set so that an employee could be exposed at the limit for an entire working life without suffering harm. The authority is over the

exposure limits, not over how they are achieved. In §6(b)(7), however, it is stated that "Where appropriate, such standards shall also prescribe suitable protective equipment and control or technological procedures to be used in connection with such hazards." (This paragraph goes on to prescribe labels, warnings, and provisions for ongoing monitoring of employee exposure.)

Although the standards are to be primarily health-based, §6(b)(5) calls for this achievement "to the extent feasible." Earlier in the act, in the section on definitions [§3(8)], the term "occupational safety and health standard" is defined as "a standard which requires conditions, or the adoption or use of one or more practices, means, methods, operations, or processes, reasonably necessary or appropriate to provide safe and healthful employment and places of employment." As discussed below, the "reasonably necessary or appropriate" language became very important to the Supreme Court's decision on OSHA's regulation of benzene, which largely determined the role of risk assessment in the agency.

There are actually three sorts of standards under the OSHAct: in addition to OSHA-promulgated new standards [§6(b)], the principal OSHA standards and the main subject of this report, there are national consensus standards [§6(a)], and emergency temporary standards [§6(c)]. The last allows temporary action if it is determined that employees are exposed to "grave danger." The second sort, consensus standards, reflects the fact that, prior to the passage of the OSHAct in 1970, there was a hotchpotch of workplace regulations under various authorities interwoven into a system that was primarily voluntary, based on standards recommended the American Conference of Governmental Industrial Hygienists and the American National Standards Institute. OSHA was created to organize and consolidate this system and to place it under enforceable regulatory authority. Thus, §6(a) of the act called on the new agency to adopt as its own "any national consensus standard, and any established Federal standard, unless [the Secretary] determines that the promulgation of such a standard would not result in improved safety or health." (If prior standards conflicted, OSHA was to adopt the most stringent.) The consensus standards were a "one-time" provision of consolidation and start-up for the new agency. Nonetheless, the provision has had an influence on OSHA practice; until a recent court case overturned them, many choices for OSHA PELs (permissible exposure limits) were influenced by consensus standards.

Standard setting may be initiated by OSHA or upon petition by interested individuals or organizations. Section 7 of the OSHAct provides for the creation of external advisory committees that the Secretary may appoint to advise on standard setting (although no such committee is appointed for many standards). A balance of labor and industry representatives, as well as of independent experts, government representatives, and others is mandated. Under §6, a proposed rule (which may be suggested by such a committee but is proposed by OSHA) is published for comment; if any party calls for it, a public hearing must be held on the proposed rule and its basis. Unlike a simple public meeting, this hearing is conducted by an administrative law judge; the hearing is an informal, fact-finding process, but witnesses can be cross-examined by any participating

party. After the hearing, the Secretary must issue the rule (either as it was or modified) or state why the rule is not to be issued.

RISK MANDATE

The OSHAct does not mention risk assessment as such, nor does it say much about the establishment of safe exposures. It is more explicit than some other laws about what constitutes an adverse health effect, however. In §2 it refers to "diminished health, functional capacity, or life expectancy" while §6 mentions "material impairment of health or functional capacity" as outcomes to be avoided. The mandated focus is on individual risk to a hypothetical employee experiencing an agent at the permissible exposure limit for a working lifetime, with regulation set "to the extent feasible" so that such an employee will suffer no impairment.

The interpretation of these provisions has undergone considerable evolution as the result of some key judicial challenges. A full account is beyond the scope of this report, but the history and issues are reviewed by Graham et al. (1988). The challenges were prompted largely by OSHA's regulatory treatment of permissible exposure limits (PELs) for carcinogens.

Initially, the mandate was interpreted as essentially a health-based standard with an added proviso that health-based regulations could not be set so low as to be infeasible, interpreted as meaning having significant financial impact on the industry. For carcinogens, the lack of demonstrable exposure thresholds for toxic effect was interpreted to mean that no workplace exposure standard, however low, could assure that "no employee will suffer material impairment of health." Accordingly, the "feasibility" provision becomes the limiting factor, and workplace standards for carcinogens were set as low as was deemed to be technically feasible at reasonable cost. (This is similar logic to that used by EPA's Office of Water under the Safe Drinking Water Act, discussed on p.123, where carcinogens have unenforceable "goals" of zero concentration and enforceable limits set on technical and financial feasibility.) Under this interpretation, in a proposed "carcinogen policy" (42 FR 54148, 1977), risk assessment for carcinogens played a rather minor role in OSHA's setting of workplace standards, and OSHA staff generally argued that the uncertainties of quantitative cancer risk assessment precluded its use as a basis for regulation.

A proposed 1 ppm standard for workplace benzene exposure set under this interpretation was challenged in court, eventually leading to a 5-4 Supreme Court decision [Industrial Union Department v. American Petroleum Institute, 448 U.S. 607 (1980)], commonly known as the "benzene decision," which imposed fundamental changes in the interpretation of the OSHAct mandate. The court ruled that, before issuing a standard, OSHA must first demonstrate that the chemical posed a "significant risk." Unless there is this argument that the risk is significant, the material does not become a "toxic material" or "harmful physical agent" controllable under the act, and its presence cannot be said to meaningfully lead to an unhealthy workplace. A key part of

this finding was that the §3(8) definition of a standard as a "reasonably necessary or appropriate" action was taken as grounds that action under §6(b)(5) must be shown to be necessary in some quantitative sense. While stating that "OSHA is not required to support its finding that a significant risk exists with anything approaching scientific certainty," the court ruled that the case for significant risk could in principle be made using quantitative risk analysis. On the question of how large a cancer risk is "significant," Justice Stevens, in his opinion, stated that this was OSHA's responsibility, conceded to be a matter of policy, but that "If, for example, the odds are one in a billion..., the risk clearly could not be considered significant. On the other hand, if the odds are one in a thousand..., a reasonable person might well consider the risk significant and take appropriate steps to decrease or eliminate it."

In effect, the benzene decision prompts OSHA to conduct quantitative risk assessment in order to set standards for carcinogens. The court declined to address the related question about whether the "feasibility" and "reasonably required" standard-setting issues should be interpreted to require cost-benefit analysis of proposed standards. In a later supreme court decision, the "cotton dust decision" [American Textile Manufacturers Institute v. Donovan, 452 U.S. (1981)], the court ruled that OSHA may set a level as protective of health as feasible, even if a less stringent one has a more favorable cost-benefit ratio.

One further court case of note is the recent ruling [AFL-CIO v. OSHA, 965 F.2d. 962] that OSHA must make it's risk case for each chemical according to its own analysis. The practice of adopting outside standards, and of setting standards based on general risk arguments rather than case-by-case demonstration of significant risks, was struck down, invalidation 428 OSHA permissible exposure limits.

IMPLEMENTATION

As with other agencies, as a source of authority and guidance on risk assessment methods and principles, OSHA draws on the major consensus documents, particularly the NAS "red book" (NRC, 1983) and the "OSTP Principles" (50 FR 10371-442). (OSHA officials participated in the preparation of the latter through the U.S. Interagency Staff Group on Carcinogens.) Although the agency has published a "cancer policy" document [42 FR 54148 (1977) and 45 FR 5160 (1980)], much of its content has been affected by the benzene decision, and a recent statement of risk assessment principles and methods has not been compiled. That is, as has been noted for several other risk assessing groups, the particulars of OSHA risk assessment practice are to be found in the documentation of analyses performed in support of specific regulatory actions. It is interesting to note that this documentation sometimes refers to guidelines and principles of other risk assessment institutions. For instance, OSHA's final rule for Cadmium refers to EPA's carcinogen assessment guidelines for guidance on weight-of-evidence for carcinogenicity (57 FR 42174). It is the usual practice for OSHA to present the results and methodological basis of other existing risk assessments for a compound in addition to its

own assessment, and to feature several possible bases for risk calculation in its characterization of risks.

The OSHAct does not specifically rule out prospective regulation, but in practice the agency acts mostly on existing workplace exposures, with priorities on those about which issues about the safety of current practices have arisen, and especially those on which petitions have been filed. (This is especially true since the judicial overturning of 428 PELs; OSHA has of necessity focused on its backlog of existing workplace problems.) Section 6(g) says that priorities should give "due regard to the urgency of the need for mandatory safety and health standards."

Although §6(b)(7) requires labels and warnings to workers of workplace hazards, OSHA only provides a general definition of a "hazardous compound" but does not itself define which compounds constitute such hazards or what (specifically) the labels and warnings must contain in the way of information. These are left to employers, who must show that they are exercising their responsibility to provide adequate information on compounds fitting the definition of "hazardous." (Many chemical-specific standards include provisions on labeling and training, however.)

The call in OSHAct §6(b)(5) for standards framed in terms of "performance" has in practice usually been answered by specifying concentration and duration limits for workplace air concentrations. Although the control program mandated in each standard is complex, the primary thrust is usually based on requirement for engineering and process controls to limit emissions to workplace air rather than on protective equipment to insulate workers from an ambient hazard. This is in keeping with the act's primary mandate for a "healthful workplace," as opposed to a mandate to manage workers' risks. OSHA has a policy of "hierarchy of controls," preferring engineering and work-practice limits before less reliable protective equipment, respirators, and the like. Especially when dermal exposure is an important potential source of risk, however, standards may specify work practices and protective equipment.

HAZARD IDENTIFICATION

OSHA does not regularly use a formal weight-of-evidence ranking scheme or hazard classification method, as do EPA and IARC. The agency's cancer policy document (42 FR 54148, 45 FR 5160) does specify an evidence-ranking scheme, but it is little used in practice. OSHA hazard identification follows the general precepts of the field, however, and employs a weight-of-evidence approach of the usual sort (as discussed on p.Hazard Identification for Carcinogens). Classifications by EPA or IARC (and the arguments made to support them) are considered as part of the body of evidence, but OSHA makes its own determination of what compounds should be considered to be "occupational" carcinogens.

There are no criteria particular to the occupational context in making this finding. That is, there is no requirement that human data be in an occupational context or that

animal tests be done by the same route of administration as workers are expected to experience. If there are grounds for limiting the finding of carcinogenicity of an agent to a specific route of administration, OSHA will do so, but the presumption is that data from all routes are potentially relevant. For example, OSHA's assessment of ethylene dibromide carcinogenicity was based largely on experiments in which rodents were treated by oral gavage, although worker exposure by inhalation was at issue.

More than any other agency regulating exposures to toxic substances, OSHA has frequent availability of human data on which to base its assessments. Occupational exposures are often high and well defined, the periods of exposure are often long and well recorded, and the situation of study is directly relevant to the regulatory situation. Even when risks are assessed based on animal data, the human exposures of interest are often not far removed from the levels tested in the rodent bioassays. Thus, extrapolation of effects is less an issue for OSHA assessments than for many environmental exposure analyses.

DOSE-RESPONSE ANALYSIS

OSHA's methods for dose-response analysis of carcinogens are broadly similar to those of other Federal regulatory agencies, with some important exceptions as noted below. Compared to other agencies, OSHA is much more willing to present a variety of potency estimates, based on different choices of data sets or assessment methods. Perhaps this is attributable to the agency's comparatively late entry into the fray. It might also be ascribed to the fact that, even after the benzene decision, in practice the limiting factor on how low an exposure standard is set is usually technical and financial feasibility. That is, establishing that risks exist within a quantitative range that can be considered "significant" is the main risk question; precisely how big they may be is a secondary one. Showing that significant risk estimates can be made under a variety of assessment methods adds to the robustness of the finding that some risk exists. Since the regulation does not have to be set to achieve (or defended as having achieved) a specific level of risk reduction, a single, falsely precise "risk number" is not necessary or even beneficial.

When faced with several data sets, OSHA tends to choose the one showing the highest sensitivity (i.e., most sensitive sex and species), but will frequently present several alternatives together or do several analyses and present the median result. When animals in one experiment develop tumors of several distinct types, the tumors may be pooled or analyzed individually, depending on case-by-case judgment. There is no standard method for correcting for intercurrent mortality, but time-to-tumor analyses are employed if the data are available for such an approach.

OSHA uses the multistage model for quantitative description of animal cancer dose-response patterns and for extrapolation of these patterns to low exposure levels. The number of stages is set at one less than the number of dose groups (a procedure

unlike that used by CPSC and parts of EPA in their use of the multistage model, although the variations make very little practical difference).

The principal difference with EPA's use of this model is that OSHA features the maximum likelihood estimate (MLE) of the fitted curve (although an upper bound is also presented), while EPA usually presents only the upper bound on low-dose potency, as described in the EPA section, [p.73]. In using the MLE, OSHA employs the full equation (i.e., using all the estimated terms, not just the linear one).

The exposure levels of concern to OSHA are much higher than is the case for many "environmental" exposures assessed by other agencies They will thus be relatively closer to the range of experimentally tested exposures, and the extrapolation problem is less severe than in some other applications. The higher terms in the multistage equation may come into play, and so even when the linear term is estimated as zero, there may be some non-negligible risk at exposure levels of concern.

The default dose metric used for cross-species extrapolation is mg/kg/day, or so-called body-weight scaling, as is also employed by FDA. In contrast, CPSC and EPA use surface-area scaling, which tends to produce extrapolated human risk estimates that are higher by about 13-fold when extrapolating from mice and about 7-fold when extrapolating from rats. Cross-species dose scaling is further discussed on p.75.

This default may be modified in specific cases. For instance, if there are data on the fraction of the dose absorbed, a correction may be made. For formaldehyde, a reactive compound evaluated as a cause of respiratory tract tumors, the assumption was made that equivalent exposures in terms of air concentration (ppm equivalence) are equally potent. In this, OSHA is similar to other Federal agencies. Pharmacokinetic modeling data have been entertained, but to date they have not been used as a basis for cross-species extrapolation by OSHA.

While EPA produces estimates of carcinogen potency that will be used in a variety of regulatory contexts with a variety of exposure levels, OSHA is concerned with the primary mandated scenario of a worker exposed for a 45 year working lifetime to the agent at the exposure level set by the standard. Thus, EPA expresses its potencies as a term to be multiplied by an exposure (risk per mg/kg/day or risk per $\mu g/m^3$ in air, averaged over a lifetime), OSHA tends to present calculations specific to its intended exposure scenario, in terms of risk to a lifelong worker at a given constant daily exposure. (Thus, for example, risks to a "1 ppm" exposure may not be comparable, since EPA intends this to mean a person exposed 24 hours a day for 70 years, while OSHA intends an exposure of 8 hours a day for 45 years.)

In the area of non-cancer risk assessment, OSHA has in the past used methods that are comparable to those used elsewhere. In the PEL decision, however, the court called into question the use of standard safety factors; it noted that the use of safety factors was close to the practice prohibited in the benzene decision, i.e., foregoing specific estimation of the degree of risk and instead prescribing low exposures without

demonstrating that lowering is necessary to avoid risk. The safety factors were applied in the court's view only to allow for the possibility of harm at lower concentrations, not because of any chemical-specific indication that a dose adjustment was needed to avoid harm. The court also stated that "application of such factors without explaining the method by which they were determined...is clearly not permitted" (AFL-CIO v. OSHA, 965 F.2d. 962). As a result, OSHA's methods for non-cancer assessment are in flux, and various alternatives are being considered. According to the OSHA officials interviewed for this report, cancer concerns have tended to predominate over non-cancer issues in the setting of standards, since they are generally such that the standards must be set at the lowest feasible level in any case.

EXPOSURE

Compared to environmental exposures, exposures in the workplace tend to be much better defined. The workplace is a confined setting within which practices and behaviors tend to be standardized. Exposure levels are often high enough to be easily measured, and many workplaces have ongoing monitoring of environmental levels of compounds.

The differences in exposure assessment as practiced by OSHA and by other agencies stem largely from two particular aspects of regulation under the OSHAct. First, regulations are usually performance-based, framed in terms of limits on ambient air concentrations that must be achieved in the workplace rather than of specific means to achieve them. Thus, questions of emissions, engineering controls, fate and transport—the determinants of ambient air concentrations—tend not to enter the risk assessment per se. (OSHA must consider such matters as part of its determination of the feasibility of achieving a proposed standard, however.)

Second, the statute clearly specifies that standards are to be set so as to be protective of the hypothetical worker who spends an entire working life at the exposure level permitted by the standard. In practice, then, the exposure scenario of regulatory importance is 45 years of exposure at a fixed ambient concentration specified in the standard for 240 days a year (basically, 5 days a week) for 8 hours a day. According to the OSHAct, a "healthful workplace," the goal of the act, is not achieved unless it is possible to experience this scenario without harm. (Clearly, many workers will experience less exposure and for a shorter time; this information does enter into OSHA's analysis of costs and benefits associated with various regulatory options, but it is not the basis of a standard as such.)

Fundamentally, then, the regulatory focus is on a hypothetical individual with the maximum hypothetical exposure specified in the act. Such a "maximally exposed individual" is not really comparable to the use of this concept in other contexts (e.g., regulation under the Clean Air Act) because the exposure is for a standardized scenario rather than an estimate of actual exposures. In fact, compared to the real distribution of exposures, this hypothetical exposure scenario is not necessarily very conservative. In

practice, permissible limits are usually based on feasibility of achievement, and so it not uncommon that the distribution of workplace exposures has a rather tight distribution not far below the permitted level. Many industrial workers have long job tenure, and even when they switch employers, their new exposures are often similar, being set by the same standard. Also, many workers now work longer shifts than the standard 8 hours, including overtime and 12hr on/12 hr off patterns, potentially leading to longer total exposure durations than the "maximal" scenario entails.

Data on effectiveness of particular controls, actual ambient levels in real workplaces, actual patterns of worker exposure, and so on enter into analyses of the effectiveness, feasibility of achievement, costs, and benefits of various ambient concentration standards that might be considered, but they are secondary to the mandate on how the standards are to be set.

Standards sometimes specify ongoing monitoring of ambient concentrations in the workplace or exposure monitoring of workers, especially if the lowest feasible standard is thought to be associated with appreciable risk. (Often, the standard will specify an "action level," frequently one-half the PEL, that triggers the requirement for monitoring. There is no monitoring, however, under §6(a) consensus standards.) That is, in situations where it is infeasible to assure the "healthful workplace" to the hypothetical maximally exposed worker, the actual patterns of worker exposure may be monitored to ensure that potentially unsafe exposures can be avoided or limited by workplace practices.

In assessing the exposures and risks from a new proposed standard, the assumption is made that one is considering newly exposed workers who will work under the new standard for their entire working lives. That is, no allowance is made for the fact that current workers may have already had exposures higher than the new standard. If such workers continued to work under the new standard for the balance of their working lives, their total lifetime exposure would be higher than permitted under the new standard.

RISK CHARACTERIZATION AND REGULATION

Since the 1980 Supreme Court benzene decision, risk assessment at OSHA has been dominated by the question of showing "significant" risk from exposure to workplace carcinogens. The question that Justice Stevens threw back to OSHA in his benzene opinion—what constitutes a "significant" risk?—has never been fully answered. Justice Stevens' statement that a lifetime risk of one in a thousand is clearly significant has served as something of a benchmark. In practice risks below 10^{-5} are rarely given much significance, but the lower bound on risks considered significant is hard to define since there is no real case to date where OSHA declined to pursue a standard because cancer risks were calculated to be low. In this case, the "significance" question is one of individual risk (rather than of public health impact on the whole exposed population), since the question is still posed in terms of the hypothetical worker exposed at the

permitted limit. (OSHA has a policy of forbidding rotation of employees through jobs with high carcinogen exposure as a work practice to ensure no employee experiences a PEL for a 45 year working life. The grounds are that this strategy would only increase the number of workers exposed. In essence, this is a population risk argument.)

In practice, the technical and financial feasibility of achieving a standard is usually the limiting factor in choosing a permissible exposure level (Infante interview). That is, limits are usually proposed under which the hypothetical maximally exposed worker would be calculated to experience risk in the upper end of Justice Stevens' range. (This is not to say that real workers with their actual exposures are necessarily suffering significant risk, as discussed in the following paragraph.) Under these conditions, the particular numerical estimate of risk level is not the driving issue in regulation, only the more general argument that "significant" risks could be generated. OSHA is able to entertain a variety of risk analyses based on somewhat different data sets and assumptions without muddying the regulatory decision with questions about which single analysis is the "right" one to choose to set a standard.

In the analyses that in practice drive the permissible levels specified in standards—that is, the determination of what levels are feasible to achieve—the costs and performances of various technical control options are considered. In these analyses, actual worker exposure levels and durations of exposure can be considered, including the resulting changes in residual risk to be expected after various regulatory options. Thus, there is opportunity, albeit indirect, for information on distributions of actual exposure to come into play in determining OSHA regulations. Nonetheless, the key consideration in feasibility is not risk, but rather the costs and technical ability needed to reach various ambient concentration levels.

Although the benzene decision has profoundly affected OSHA's approach to the analysis of risk, the practical result is that decisions are not very different from what would have been done under the pre-1980 carcinogen policy. The benzene decision stated that OSHA could not simply limit exposures according to feasibility of control without first showing that lack of control leads to significant risk. In practice, this is usually shown, at least for the standards that OSHA has pursued since 1980, so controls are set primarily on feasibility all the same. The role of risk assessment in this process is largely to establish (1) that significant risks exist under current exposures, and (2) that reducing the exposure as proposed in the standard will reduce the risk. The major practical impact is that the case for significant risk must be made for each compound, focusing the agency's activities and resources to pursue regulation on those compounds where risk can be clearly shown.

The principal notable features of risk assessment at OSHA are that the size of the risks in question are a good deal larger than those encountered in other regulatory programs. Frequently, risks may be assessed on human data directly relevant to the regulatory interest; in recent years about one-half of OSHA PELs have been based primarily on human data. Even when animal data are used, human exposures of interest are often not far below the tested levels. Real, directly relevant exposure data are often

available, and they are often quite defined and less variable compared to environmental exposures for the general population. As a consequence, OSHA risk assessments have to grapple much less with extrapolation questions, and OSHA's methods have less built-in conservatism. Since PELs are in practice set by feasibility, with risk assessment determining the need for controls, OSHA is able to entertain a variety of risk analyses without settling on a single "number" as the canonical one for its regulatory activities. The regulatory focus is on the risk to a worker exposed to the permitted level for a full working life; although this is a hypothetically defined "maximally exposed individual," in practice and for a variety of reasons, this hypothetical exposure may not be much higher than that actually experienced by many workers, and indeed some workers (those doing overtime or previously exposed under a higher standard, for example) may exceed this theoretical "maximum."

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

CONSUMER PRODUCT SAFETY COMMISSION

The Consumer Product Safety Commission is an independent agency charged with regulatory responsibility over the safety of consumer products (which are defined by law to exclude foods, drugs and pesticides, regulated under FFDCA, as well as tobacco and certain other products regulated elsewhere). The Commission was established by the Consumer Product Safety Act (CPSA) of 1972. The regulatory authority over hazardous substances in consumer products derives from the CPSA and the Federal Hazardous Substances Act (FHSA), which has existed since 1960. The FHSA was formerly administered by the Food and Drug Administration, but authority was transferred to the Commission by §30(a) of the CPSA.

The coverage of these two acts is largely overlapping; the CPSA establishes the Commission, sets certain of its procedural requirements, and in addition provides a somewhat broader version of the regulatory authority contained in the FHSA (focusing not just on risks from hazardous substances, but on potential injuries and risks from consumer products generally) and with less formal procedure. Indeed, in its coverage of hazardous substances it essentially repeats much of the specific language of the FHSA. The CPSA [§30(d)] states that any risk that could be regulated under the FHSA should be so regulated "unless it is in the public interest to regulate such risk of injury under [the CPSA]." (The CPSC's ban on urea-formaldehyde foam insulation under the CPSA was overturned by the 5th Circuit Court partly on the grounds that the regulation should have been under the FHSA [Gulf South Insulation v. CPSC, 701 F.2d. 1137].)

THE CPSA, THE FHSA AND THEIR MANDATES

The **Consumer Product Safety Act** (CPSA) establishes the Consumer Product Safety Commission with the mandate "to protect the public against unreasonable risks of injury associated with consumer products" and "to develop uniform safety standards" [§2(b)]. The agency is run by a 5-member Commission appointed by the President (with the consent of the Senate) for seven-year terms. (In recent years, only three Commissioners have been appointed, and in this circumstance, two constitute a quorum.) Decision-making by the Commission is by majority vote among Commissioners who may have been appointed by different administrations. This makes the development of analyses to support decisions somewhat different at CPSC than at agencies answering to a single administration appointee. Staff develop positions and options for the Commission's consideration, laying information out for a final, publicly held, sometimes contentious debate.

The impetus is on the Commission to promulgate consumer product safety standards when it is deemed necessary to protect the public against unreasonable risks of injury. That is, its task is to identify and act against hazards as opposed to endorsing

products as "safe." Although much of the focus of the CPSA is on acute hazards, there are specially mentioned provisions for chronic toxicity, as discussed below. The Commission has a wide variety of regulatory options that can be applied as deemed necessary, including labeling, mandating other provision of information, endorsement of voluntary standards, manufacturing standards, product performance standards, bans, and recalls [CPSA §§7,8,15].

The **Federal Hazardous Substances Act** (FHSA) defines a hazardous substance (or mixture) as one that is corrosive, an irritant, a strong sensitizer, or flammable, or one that "may cause substantial personal injury or substantial illness during or as a proximate result of any customary or reasonably foreseeable handling or use, including reasonably foreseeable ingestion by children" [FHSA §2(f)(1)(A)]. Section 3 of this act gives authority to "declare by regulation any substance or mixture of substances which...meets the requirements" of this definition to be a hazardous substance. (Section 3 specifies a series of procedures which includes the right to petition for hearings; it is these more extensive procedural requirements, in addition to the focus on chemical hazards, that chiefly distinguishes regulation under the FHSA from that under the CPSA.) Labeling of substances declared to be hazardous is mandated. However, if "notwithstanding such cautionary labeling...the degree or nature of the hazard...is such that the objective of the protection of the public health and safety can be adequately served only by keeping such substance...out of the channels of interstate commerce," the substance can be declared a "banned hazardous substance" [§2(q)(1)].

The two acts set out some similar requirements for standard setting; Section 3 of the FHSA largely parallels §9 of the CPSA, and for convenience, the latter will be discussed more specifically. These acts provide for a three-stage rulemaking procedure, which is commenced by publication in the *Federal Register* of an advanced notice of proposed rulemaking (ANPR). The ANPR explains the problem and possible actions to address it and calls for information on any existing standard, including voluntary standards. In addition, it must invite interested parties to submit plans to develop voluntary standards to address the risk in question. The Commission must defer to such a standard that is in existence, ceasing its own rulemaking, if it determines that the existing standard "is likely to result in the elimination or adequate reduction of the risk" and that "it is likely that there will be substantial compliance" [CPSA §9(b)]. The next step in the rulemaking process is to issue a proposed rule per §9(c) and to solicit and respond to comments from the public.

In promulgating a rule, the Commission must make findings regarding "the degree and nature of risk...; the need of the public for the consumer products subject to such rule, and the probable effect...upon the utility, cost, or availability of such products...; and...any means of achieving the objective of the order while minimizing adverse effects on competition or disruption or dislocation of manufacturing and other commercial practices consistent with the public health and safety" [CPSA §9(f)(1)]. The final regulatory analysis of the rule must contain "A description of the potential benefits and potential costs of the rule, including... [those] that cannot be quantified in monetary terms, and the identification of those likely to receive the benefits and bear the costs"

[§9(f)(2)]. Such analysis must also be included for "alternatives to the final rule which were considered, together with...a brief explanation of the reason why these alternatives were not chosen." The Commission is prohibited from promulgating a rule unless it finds "that the rule...is reasonably necessary to eliminate or reduce an unreasonable risk of injury; that promulgation of the rule is in the public interest;...that the benefits expected from the rule bear a reasonable relationship to its costs; and...that the rule imposes the least burdensome requirement which prevents or adequately reduces the risk of injury" [§9(3)]. It must also find that no currently implemented voluntary standard will suffice and that, if the rule is a ban, no other reasonable rule would protect the public. (As with most risk analyses, these findings are protected from judicial review unless the final rule itself is challenged.)

In sum, perhaps more than any other agency, the CPSC is explicitly required to justify its regulation in terms of costs and benefits. Whereas other cost-benefit balancing laws (e.g., FIFRA) merely make brief mention taking costs and feasibility into account, the consumer product laws lay out a series of specific findings that must be made.

Congress placed further restrictions on consumer product regulation, however, in reserving for itself the right to veto standards promulgated by the Commission. Both the CPSA (§36) and the FHSA (§21) provide for a 90-day window after promulgation of a rule under these acts within which a Congressional resolution can nullify the rule. The constitutionality of these provisions is suspect [I.N.S. v. Chadha, 103 S.Ct. 2764 (1983)]. In any event, no CPSC regulation has been nullified under these provisions.

Many of the provisions of the CPSA and the FHSA apply to both acute and chronic hazards. There is a particular provision in the CPSA regarding chronic hazards, however. Before any rule "relating to a risk of cancer, birth defects, or gene mutations" can be proposed, the Commission must appointment a Chronic Hazard Advisory Panel of independent scientific experts [§28] from nominations by the President of the National Academy of Sciences; "the Commission shall request the Panel to review the scientific data and other relevant information...to determine if any substance in the product is a carcinogen, mutagen, or a teratogen." If so, "the Panel shall include in its report an estimate, if such an estimate is feasible, of the probable harm to human health that will result from exposure to the substance" [CPSA §31(b)].

An important amendment to the FSHA was passed in 1988, known as the Labeling of Hazardous Art Materials Act (LHAMA), which established chronic hazard labeling requirements applicable only to art materials and mandated the Commission to "issue guidelines which specify criteria for determining when any customary or reasonably foreseeable use of an art material can result in a chronic hazard" [FHSA §23(d)(1)]. The guidelines were to include criteria for children and adults and for determining acceptable daily intake levels. The Commission issued these guidelines in 1992 [57 FR 46626], and has applied them to its assessment of all chronic health hazards [16 CFR §1500.14(b)(8)].

RISK MANDATE

The Consumer Product Safety Commission has the mandate to set standards that are "reasonably necessary" [CPSA §9(f)(3)(A)] "to protect the public against unreasonable risks of injury associated with consumer products" [CPSA §2(b)] when those products receive customary or reasonably foreseeable use. That is, foreseeable misuse (such as ingestion by children or loose regard for following instructions) should be protected, but unforeseeable misuse need not be. The question of what standards are "reasonably necessary" has evolved through judicial interpretation to recognize that for more severe injuries, a low rate of occurrence justifies action, but for less severe injuries, a higher frequency is needed to trigger action.

The CPSC's regulatory authority is set up to declare "hazards" and act against them; the Commission is not called upon to declare any product "safe," to issue permits, or to grant clearances. With the exception of some labeling provisions (e.g., art materials containing substances that are hazardous according to the CPSC guidelines and FHSA labeling) there is little in the way of automatic action that takes place without specific Commission initiative.

The requirements of the CPSA for rulemaking to include a statement on "the degree and nature of risk" [CPSA §9(f)(1)] and for each Chronic Hazard Advisory Panel to "include in its report an estimate, if such estimate is feasible, of the probable harm to human health" [§31(b)] constitute a fairly clear statutory call for the conduct of risk assessment.

IMPLEMENTATION

In 1978 a set of guidelines were proposed for evaluating carcinogens [43 FR 25658]. These were overturned in court, however, on a technical question of opportunity for public comment [Dow Chemical, USA v. CPSC, 459 F. Supp. (W.S. La. 1978)] and were never re-issued. It was not until the mandate for chronic health effect assessment guidelines under LHAMA that a guideline effort was renewed, and the Commission published its procedures in 1992 [57 FR 46626]. These guidelines establish principles that are broadly in line with those used in other agencies, with some exceptions as noted in the sections that follow. The document usefully compares and contrasts its carcinogen assessment methods with those of EPA and IARC. The guidelines were issued less for CPSC's use in its own assessments than for the use by manufacturers of art materials in order to determine whether their products must be labeled as containing a chronically hazardous substance. Nonetheless, they serve as the basis and explanation of the Commission's methods in its own assessment activities.

Although CPSC's actions have undergone a number of judicial challenges, and although several actions have been overturned, for the most part the issues have been procedural rather than about risk assessment methodology or its use. In its 1983 overturn of the ban on urea-formaldehyde foam insulation [Gulf South Insulation v. CPSC, 701

F.2d. 1137] the U.S. 5th Circuit Court did question whether the use of the principal rat inhalation bioassay, in which nasal tumors were seen, constituted substantial evidence for a human health hazard, but procedural issues were at issue as well. This case and other factors has had a rather stultifying effect on regulation of chronic health hazards by the Commission.

HAZARD IDENTIFICATION

The FHSA shows its origin in concerns for acute toxicity through a remarkably specific definition of "highly toxic." In what must stand as the clearest example of the specification of risk assessment methodology in a statute, the FHSA [§2(h)] defines as highly toxic a substance that "Produces death within fourteen days in half or more than half of a group of ten or more laboratory white rats each weighing between two hundred and three hundred grams, at a single dose of fifty milligrams or less per kilogram of body weight, when orally administered." (Comparable definitions for inhalation and dermal exposure are also given. The act provides, however, that highly-toxic-by-human-experience data take precedence over these definitions.)

More broadly, however, the FHSA considers as toxic "any substance (other than a radioactive substance) which has the capacity to produce personal injury or illness to man through ingestion, inhalation, or absorption through any body surface" [§2(g)]. Technically, a toxic substance under the FHSA should be considered a hazardous substance only if there is reasonably foreseeable exposure sufficient to cause illness or injury. The aim of the CPSC chronic toxicity guidelines is not only to document risk assessment methodology, but to provide an expanded definition of chronic hazard for use in labeling mandates. The CPSC guidelines contain sections on assessment of carcinogenicity, neurotoxicity, and reproductive and developmental toxicity. There is also guidance on exposure and the assessment of bioavailability.

The scheme in the CPSC guidelines for hazard identification of carcinogens is intentionally very similar to that used by EPA and IARC. For both human and animal data, there are categories for 'sufficient evidence,' 'limited evidence,' and 'inadequate evidence,' that, with the exceptions noted below, are essentially identical to EPA's categories of the same names (as described in the section on that agency [p.62]). There are no categories comparable to EPA's 'No Data Available' and 'No Evidence of Carcinogenicity' for either human or animal evidence characterization in the CPSC scheme, since compounds with evidence falling into such categories would not be considered hazards. CPSC argues that its need is only to define hazards, not to characterize evidence that falls short of what is necessary to designate an agent a cancer hazard. Also, the CPSC weight of evidence includes life-threatening benign tumors, as does EPA's, a difference from the IARC scheme. Finally, CPSC considers increased tumor incidences at independent multiple sites of origin in the same species and study to be separate responses, rather than as a single response as do the EPA and IARC. Thus, a study with several tumor types independently elevated by dosing in a single sex and

species would constitute evidence of a repeated response and hence 'sufficient' evidence of the compound's carcinogenicity in animals.

According to IARC's criteria, increases in certain tumors known to have high spontaneous rates in the strain in question would constitute only 'limited' evidence unless there is a basis to upgrade this finding. Like EPA, CPSC considers significant elevation of tumors with high spontaneous rates to be 'sufficient' evidence (if the other criteria for sufficiency are met) unless there are grounds to downgrade the finding to 'limited.'

A substance is considered to be chronically toxic to humans by virtue of its carcinogenicity when the human evidence is either 'sufficient' or 'limited,' or when the animal evidence is 'sufficient.' This corresponds fairly well to EPA's weight-of-evidence categories A, B1, and B2 and to IARC's 1, 2A, and 2B.

The CPSC guidelines provide evidence categorization schemes for both human and animal evidence for the other chronic endpoints in addition to cancer. In this, they differ from and go beyond the EPA guidelines for reproductive and developmental toxicity, which discuss the evaluation of evidence in general and qualitative terms (as discussed in the section on EPA). Indeed, the CPSC guidelines probably represent the most specific and structured treatment for defining the methods that an agency will use in judging the evidence of non-cancer chronic toxicity.

For each type of chronic toxicity discussed, the CPSC document reviews the nature of human studies and animal testing that can be done and provides criteria for judging study quality, interpretation of results, and assessing the coherence of the evidence. These reviews generally follow the established principals of epidemiology and of the appropriate field of toxicology. Then, again for each type of toxicity, criteria are given for categorizing the body of human data and (separately) the body of animal data into findings of 'Sufficient Evidence,' 'Limited Evidence,' or 'Inadequate Evidence' that the agent poses a chronic toxic risk of the type in question.

For example, 'Sufficient Evidence' of developmental or reproductive toxicity based on human data is obtained when based on "a good quality epidemiology study which meets all the requirements [for inference of causality]...; the results are statistically significant and without identifiable bias or confounding factors" [57 FR 46642]. 'Sufficient Evidence' of developmental or reproductive toxicity based on animal data is obtained under these guidelines when "a good quality animal study" (a judgment based on named criteria) has a "statistically significant ($p<0.05$) treatment-related increase in multiple endpoints...in a single species/strain, or in the incidence of a single endpoint at multiple dose levels or with multiple routes of administration in a single species/strain, or increase in the incidence of a single endpoint in multiple species/sexes/strains/experiments. Evidence from animal studies which has been shown to be not relevant to humans is not used for this purpose" [57 FR 46644]. The other categories are defined in terms of degrees of falling short of these standards. Similar definitions are provided for neurotoxicity.

DOSE-RESPONSE ANALYSIS

The methods used by CPSC for quantitative risk assessment of carcinogens are similar to those used by EPA (see p.68), with some exceptions as noted below.

Like EPA, but unlike IARC, the Commission will do analyses that count benign tumors in with malignant tumors of the same histogenic type and site unless there is evidence that the benign tumors are not expected to progress to malignancy. (Parallel analyses without the benign tumors are usually calculated as well to gauge the effect of their being included.) Barring evidence suggesting the particular appropriateness of a certain animal data set for extrapolation to humans, the data from the sex and species showing the most sensitivity to the agent is usually emphasized, as at EPA. CPSC, however, will average the potency results based on males and females from the same species and experiment if the responses are at the same site, rather than taking the more sensitive choice, as does EPA.

When an animal data set shows that the same sex and species demonstrated elevated incidences for more than one type or site of independently originating tumor, the CPSC will analyze each tumor type separately and then, to obtain a measure of total tumor risk, combine the risks. This differs from EPA's procedure of combining the incidences (i.e., counting animals bearing any one among the types of tumors elevated by dosing) before the dose-response curve is fitted.

If animal data from the routes of administration of interest in humans are not available, data from another route may be used, employing the usual means of dose calculation for route extrapolation. The CPSC guidelines [57 FR 46655] more explicitly prefer estimates based on the same route of administration as seen in humans than do the EPA guidelines, although practice appears quite similar. Intercurrent mortality is usually adjusted for by dropping from the analysis animals dying before the appearance of the first tumor of the type of interest in the data set. When data permit, a time-to-tumor analysis may be done.

The multistage model is used to characterize animal dose-response relationships, the same model as used by EPA and OSHA. Unlike EPA, however, CPSC sets the degree of the polynomial in the multistage model equation to a high number (in the neighborhood of 5 or 6 rather than at 1 minus the number of dose groups) in order to allow the model fitting maximum flexibility to choose the degree of curvature. (The fitting algorithm allows such a choice. Although the total number of parameters estimated to be nonzero is limited by the number of data points, which particular parameters are nonzero, corresponding to which powers of dose affect the curve shape, is flexible.) Risk over background is measured as "extra risk"; that is,

$$R(d) = \frac{P(d) - P(0)}{1 - P(0)},$$

where $R(d)$ is the risk above background at dose d, $P(d)$ is the total modeled risk at dose d and $P(0)$ is the modeled risk at zero dose, i.e., the estimated background rate. This is the same method used by most of EPA with some exceptions as noted in the discussion of the Office of Pollution Prevention and Toxic Substances.

Unlike the EPA, the CPSC typically uses the maximum likelihood estimate (MLE) of the fitted curve (i.e., the "best fitting" curve) rather than an upper bound, provided that the MLE of the q_1 parameter is positive. That is, if the MLE curve is linear at low doses, this estimate is used, but if the MLE is nonlinear, then an upper bound is used in the same manner as in the usual EPA analysis. (In the experience of CPSC, using a high degree for the dose polynomial, as described in the previous paragraph, tends to allow the MLE to be linear, decreasing the frequency with which an upper bound is needed to assure low-dose linearity.) That is, CPSC subscribes to the same policy underlying low-dose extrapolation of carcinogenic effects used at all agencies: in the absence of knowledge about the true low-dose shape of the dose-response curve, a linear extrapolation should be used. This extrapolation is interpreted as it is at EPA—as an upper bound in the sense that a linear extrapolation is unlikely to underestimate low dose risks, and will overestimate them when the true curve is convex, as it may be.

Cross-species extrapolation is accomplished by assuming equal lifetime cancer risks when daily doses are proportional to the 2/3-power of body weight, the same method as EPA's surface-area scaling, set out it its 1986 guidelines. This contrasts to the assumption of body weight scaling employed at FDA and OSHA. Body weight scaling gives projections of human risk that are roughly 4- to 6-fold lower when extrapolating from rats and 12- to 14-fold lower when extrapolating from mice than does surface area scaling. This difference constitutes the largest quantitative factor among the differences in quantitative cancer risk assessment among agencies.

In cases where a carcinogen acts at the site of contact (e.g., in the nasal passages or lung for an inhaled substance), CPSC will usually use concentration in the medium as a basis for equivalence, adjusting for the proportion of a lifetime exposed. (EPA may sometimes do this, as it also considers the matter case by case.)

CPSC participated in the interagency effort to define a uniform default cross-species scaling method for carcinogens. That effort led to a joint proposal put forth by the Interagency Pharmacokinetics Group to harmonize methods on a default scaling approach that presumes equal lifetime risks when daily administered doses are scaled in proportion to the 3/4-power of a species' body weight [57 FR 24152]. CPSC has written into its guidelines that, if the 3/4-power scaling proposal is adopted as a harmonized position, then the Commission will use it [57 FR 46654].

The CPSC guidelines discuss the use of pharmacokinetic data more extensively than guidance documents available from other agencies. When sufficiently documented and validated pharmacokinetic analysis is available, CPSC will use measures of internal dose to modify high-to-low-dose extrapolation. That is, the extent to which high and low exposures differ in the proportionality between administered dose and internal dose is

factored in. The Commission in fact used such an adjustment in its analysis of cancer risks from inhaled methylene chloride.

The Commission specifically declines to use pharmacokinetic information for cross-species extrapolation, however. The guidelines note "At this time, pharmacokinetics should not be used to adjust for differences between species in sensitivity to a carcinogen; briefly, this is because information on sensitivity of various species to a 'target' dose is not currently available" [57 FR 46655]. This differs from the stance of EPA. In the only real example at hand—the assessment of methylene chloride cancer risk by inhalation—the CPSC declined to use the available pharmacokinetic information for cross-species adjustments, but did use it to adjust low-dose extrapolation, as noted above. The EPA used the pharmacokinetic analysis both for low-dose and cross-species extrapolation, although in EPA's interpretation of the latter analysis, the final result did not differ markedly from the CPSC's reliance on the usual administered dose procedure in this case. (The issues involved in this choice are discussed in a document on methylene chloride produced jointly by the EPA, CPSC, and FDA [EPA, 1987].)

For evaluating exposures of different time patterns, the CPSC guidelines specify time-averaged exposures [57 FR 46655], as does EPA, but the CPSC guidelines go on to specify that pharmacokinetic information may be used to account for disproportionality of administered and internal doses arising from dose-rate effects.

For the quantitative analysis of chronic effects other than cancer, the CPSC takes the usual approach of establishing no effect levels from animal studies and applying safety factors. This method is essentially similar to that used at EPA in derivation of reference doses and at other agencies, although the CPSC differs in explicitly setting out the methodology in its guidelines [57 FR 46655].

EXPOSURE

By the nature of its statutes, the regulatory efforts of the CPSC are typically directed at existing hazards. Thus, CPSC exposure assessments have the advantage that they need not deal with novel future exposures, as must regulatory programs that issue permits for new activity. The exposures that the Commission must deal with are those that result from uses of consumer products by consumers (i.e., not in occupational uses or distribution). To protect from "unreasonable risk of injury," exposures to be considered are those that are "reasonably foreseeable." These includes exposures typical of normal use, those resulting from reasonable misuse of products such as errors in handling, lax following of instructions, accidental ingestion by children, and so on, but it can exclude blatant or deliberate misuse.

Consumer product safety standards can take a wide variety of forms, from simple labeling of potentially hazardous products to performance standards to outright bans. The CPSA and FHSA require extensive documentation of the analysis of costs and

benefits under all of the regulatory options considered (except, in certain instances, labeling), and the Commission is generally bound to choose the rule imposing the "least burdensome requirement." Thus, exposure assessments must consider existing or predictable exposures to establish the hazard and also a series of hypothetical exposure situations corresponding to the expected impact of the various regulatory options. These entail not only technical options such as formulation and packaging changes, but also assessment of the effectiveness of labeling, voluntary standards, and so on.

Extensive information useful in conducting such exposure estimates is frequently unavailable, but the best use is made of what information may be obtained. Inevitably, defaults and hypothetical scenarios are needed. The aim in making exposure estimates is to try for a best estimate of exposure levels but to establish some upper and lower bounds to exposure as well. Because the results are used in the assessment of costs and benefits, information on population exposures (and hence, population risks) is needed, not just high-end individual exposures. It is frequently difficult to characterize a distribution of exposure across the whole consumer population, and so information on typical exposure levels, upper and lower bounds on these levels, and estimates of the numbers exposed are generated. "Worst case" estimates may be made for screening purposes.

The CPSC guidelines document contains sections on exposure assessment and the evaluation of bioavailability [57 FR 46644 and 46648]. These comprise mostly a technical discussion of the issues arising in estimating exposures and absorption, but do not give specific guidance on characterizing exposure, determining bounds and distributions, etc. As a default, the degree of absorption is assumed to be 100%, or at least similar in animal experiments and in humans, but measured degrees of absorption should take precedence.

RISK CHARACTERIZATION AND REGULATION

The extensive need under the existing consumer protection statutes to cast regulatory risk analyses in terms of costs, benefits, impact on consumers, and the least burdensome regulatory approach among many options focuses attention of CPSC analyses on typical uses at typical levels under various regulatory options. The mandate for protection against "unreasonable risk" has an element of protecting individuals, but the mandated consideration of the costs and benefits of options means that the main concern is for how the number of users and the typical exposure during use will be affected by the various control options. That is, once the product has been determined to be toxic, the main focus is on population rather than on individual risk.

The statutes make no mention of protection of sensitive subpopulations from injury, although the CPSA [§9(e)] does mandate that the special needs of the handicapped and elderly be taken into account regarding the disruption to consumer convenience resulting from a potential rule.

The FHSA amendments known as the Labeling of Hazardous Art Materials Act of 1988, which mandated the creation of the CPSC's guidelines, also require that the Commission establish criteria for acceptable daily intake (ADI) levels in children and adults for the hazardous substances in art materials. The guidelines contain a discussion of the difficulty of establishing acceptable risk levels for carcinogens, but settle on a level of exposure such that "the exposed individual has an estimated additional one chance in a million during his or her lifetime of developing the deleterious effect, such as cancer. The exposure scenario being evaluated can be one use, one year's use, 'normal product utility,' or anticipated use over a lifetime, depending on the nature of the situation being addressed." The guidelines also note that "the choice of the exposure situation evaluated is important to the concept of what risk is 'acceptable" [57 FR 46655]. For non-cancer effects, the ADI is determined by the typical no-effect level and safety factor approach. In the guidelines the Commission explains why it declined to define a special ADI for children, stating that there were no clear grounds for doing so, and that the notion of a "chronic" toxic effect to children posed difficulty when methods are based on lifetime exposure assumptions [57 FR 46631].

Strictly speaking, these definitions of ADI apply only to the evaluation of whether art materials should be considered to contain hazardous substances. But the CPSC has declared that they reflect general principles of its risk assessment methodology and should be taken as applying to all assessments of chronic toxicity. The CPSC guidelines contain the only clear written statement on what will be considered acceptable risk encountered in the research for this report. They are, however, criteria for individual risk, and, as noted above, the focus of CPSC risk analysis for regulatory options is on population risk, for which criteria are not given.

THIS PAGE LEFT BLANK INTENTIONALLY.

ENVIRONMENTAL PROTECTION AGENCY

The Environmental Protection Agency (EPA) was created by executive order by President Nixon in 1970. The EPA was set up as an independent Federal agency to be the administrative home for a number Federal environmental programs that had previously been scattered over the Executive Branch. The agency was founded at the height of a wave of public concern about the environment and amid widespread political pressure for the Federal government to take action to protect the environment from further pollution, to limit the ability of polluters to discharge wastes and poisons, and to clean up the Nation's air, lakes, and rivers. The consolidation of programs was to forge a coordinated Federal effort at protecting the environment.

The existing programs out of which the EPA was cobbled had their own legislative authorities and histories. Since the consolidation was by executive order (and not through a new environmental act specifying a melding and recasting of these programs), the various components of the new EPA retained their different legislative mandates, regulatory powers, and scopes. Many of the laws were amended during the early years of the EPA, tailoring their treatment of issues of particular concern. In addition, new laws were added to bring additional environmental problems into the ambit of the Federal environmental effort.

The result is that, even twenty-five years later, the EPA represents a collection of environmental programs that has only partly been consolidated and centralized. Risk analysis is used in support of regulation and rulemaking under a half-dozen major environmental laws and a number of minor ones. Even though the role of risk assessment, particularly quantitative risk assessment, has grown largely since EPA's founding, the separation of regulatory programs has had an effect on risk assessment practices in various parts of the agency. The history of risk assessment at EPA has been marked by ongoing issues of consistency versus case-specificity of risk assessment methods and analyses, and consolidation versus dispersion of the conduct of risk assessment.

Accordingly, the present discussion of risk assessment methods at the EPA must likewise be partly consolidated and partly dispersed. In the section that follows, an overview discusses the organizational structure for conducting risk assessment at the agency, the means that have been instituted to coordinate risk assessment activity, maintain consistency, and provide oversight, and the sources of EPA-wide policy and guidance. A subsequent section covers aspects of risk assessment methodology common to all EPA programs (while noting differences from methods employed by other agencies), including a treatment of the key technical prescriptions of various EPA risk assessment guidelines. Finally, separate sections discuss each regulatory program in the context of its legislative mandates and its use of risk assessment methods.

OVERVIEW OF RISK ASSESSMENT AT EPA

In the so-called "red book" (NRC, 1983), the National Academy of Sciences distinguished four main components of risk assessment: hazard identification, dose-response analysis, exposure assessment, and risk characterization. Of these, the first two are concerned primarily with properties of particular chemical agents and the characterization of expected toxic effects under a variety of circumstances. These components will therefore be held in common among all assessments of potential risk from a particular agent, whether the exposure comes through contaminated drinking water, air, or from a leaking hazardous waste site. In contrast, the second two components of the NAS paradigm, exposure assessment and risk characterization, will be particular to the specific exposure context in which the compound is experienced.

EPA regulatory programs are organized around the statutes they implement. The statutes, in turn, are focused on particular environmental media they charge EPA to protect (air, drinking water, ground water, etc.) or around particular uses of an agent (as a pesticide, as a hazardous waste, etc.). In other words, regulatory authority is largely compartmentalized according to the various kinds and sources of exposure. (This is true not just of EPA but of Federal environmental regulation generally; FDA, OSHA, and CPSC have their different missions designed around the particular sources of exposure each is charged with overseeing.)

Thus, even when several programs have an interest in the same chemical, the exposure assessment and risk characterization components of risk assessment are not readily centralized. The exposure questions for each program tend to be unique to the particular focus of regulatory interest. Moreover, the different laws have different mandates about risk and its control, including different regulatory tools (permits, emission standards, performance standards, etc.), and different needs to balance health and economic concerns.

For such reasons, exposure assessment and risk characterization tend to be the province of the various regulatory programs at EPA, but the identification of particular chemical agents as toxic hazards and the characterization of dose-response relationships and potencies of such agents is much more centralized, although not completely so.

ORGANIZATIONAL STRUCTURE AND RISK ASSESSMENT AT EPA

Three offices within EPA are able to focus their efforts chemical by chemical rather than on exposures, and they are the ones carrying out most of EPA's hazard identification and dose-response analysis. They are the Office of Research and Development (ORD), the Office of Pesticides Programs (OPP), and the Office of Pollution Prevention and Toxics (OPPT).

The Office of Research and Development carries out chemical-specific assessments at the request of several of the regulatory program offices, especially the

Office of Air and Radiation, the Office of Water, and the Office of Solid Waste and Emergency Response. These assessments are generally limited to the hazard identification and dose-response components, leaving exposure questions and consideration of resulting risks to further analysis by the regulatory office. (The entity within ORD that is responsible for generating such assessments is now known as the National Center for Environmental Assessment [NCEA], but this center is the lineal descendant of the Office of Health and Environmental Assessment [OHEA], which in turn contains the Environmental Criteria and Assessment Offices [ECAOs], the Exposure Assessment Group [EAG], the former Reproductive Effects Assessment Group [REAG], and the former Carcinogen Assessment Group [CAG].) The results of ORD assessments are transmitted to the requesting offices (and to other parties) in the form of chemical-specific risk assessment documents laying out the data employed, the methods of analysis used, and the findings in terms of toxic effects the compound may have and estimates of potency. There are several series of such documents, with the format and elements of the content tailored to the needs of the requesting program.

In addition to these roles in hazard identification and dose-response assessment, ORD has a role in the development of methods, tools, and mathematical models for exposure assessment. The Exposure Assessment Group (now also folded into NCEA) has been a source of expertise and advice on these matters to the various regulatory programs as they conduct their case-specific exposure assessments.

The Office of Research and Development itself has no regulatory role. The assessments it produces become part of the broader risk analysis carried out by the media-specific programs in support of their regulatory responsibilities. Being chemical-specific analyses, however, the ORD documents become resources for all agency programs that may have regulatory interest in the same agent. Indeed, because ORD documents are divorced from the particular regulatory context, and because they must thoroughly document their data, methods, and calculations in order to pass their analyses on to the regulatory programs, the assessment documents are frequently used as resources by state and local governments, other Federal departments, international organizations, and other entities outside the EPA.

The centralization of much of the chemical-specific elements of risk assessment at EPA has resulted to some degree in the institutional separation of risk assessment and risk management responsibilities, although this was not the primary reason for the organizational structure. Placing the toxicological aspects of risk assessment in ORD brings this analysis into closer contact with research scientists in the EPA laboratories.

In addition to ORD, two of EPA's regulatory programs maintain their own full risk assessment capabilities and carry out most of their own hazard identification and dose-response analysis. Compared to other regulatory programs, these offices have a chemical-specific focus that raises the need for in-house analysis. The Office of Pesticide Programs regulates chemicals that for the most part are used primarily as pesticides, and so other offices are interested only in those cases when pesticides become regulable contaminants elsewhere (e.g., drinking water). Moreover, the Federal

Insecticide, Fungicide, and Rotenticide Act calls on the office to register all pesticide chemicals after considering questions of efficacy and safety. Manufacturers seeking registration must submit toxicological data and information on anticipated production, formulation, and use. The pesticides office is empowered to ask for further data, much of which, as confidential business information, cannot be made public. In sum, for several reasons, it is practical for the Office of Pesticide Programs to carry out its own program of hazard identification and dose-response analysis (in addition to the further risk assessment and management components of the NAS paradigm.) These matters are discussed more extensively below in the section on the pesticides office [p.85].

The Office of Pollution Prevention and Toxics also has a chemical-specific focus. Its primary statute, the Toxic Substances Control Act, requires manufacturers or distributors of new chemical substances to submit toxicological information to the EPA. If warranted, the toxics office is empowered to require further testing or data collection that is needed to make a judgment about the agent's possible risks. For chemicals already in commerce, the act requires record keeping, submission of new toxicological information to the EPA, and other mandates. As with the pesticides office, then, the toxics office must conduct a large body of assessments on specific chemicals, carried out under legally mandated time schedules, using powers to require data that may need to be kept confidential. These needs have led the office to maintain its own full risk assessment capability (as discussed beginning on p.97).

These three offices, ORD, OPP, and OPPT, carry out the bulk of hazard identification and dose-response analysis at EPA. Other offices also may carry out such analyses under some circumstances, however. In addition, all EPA offices have an important role in the coordination and oversight of EPA risk assessment through their participation in the Risk Assessment Forum and the workgroups approving assessments for entry onto EPA's Integrated Risk Information System (IRIS) data base. (These groups are discussed below.) It should also be noted that the Office of Air Quality Planning and Standards (within the Office of Air and Radiation) carries out the substantial part of non-cancer assessments for the so-called criteria air pollutants, using input from ORD. This special case is discussed in the section on the air office [p.112].

COORDINATION, CONSISTENCY, AND OVERSIGHT

The dispersion of risk assessment activity over parts of the EPA makes the issue of coordination and maintenance of consistency particularly important to this agency. There are several means in place toward this end. They include the publication of a series of risk assessment guidelines, development of methodology documents, the chartering of several cross-agency groups to coordinate and harmonize practices and to resolve methodological and policy questions that may arise, the reliance for advice and scientific guidance on external experts through the EPA Science Advisory Board, and the maintenance of a computerized, publicly available data base of agency-wide consensus on risk assessments. These are discussed briefly below.

Risk Assessment Guidelines:

The NAS "red book" (NRC, 1983) recommended that "uniform inference guidelines be developed for the use of federal regulatory agencies in the risk assessment process." Although the proposal was for government-wide guidelines, the only vigorous pursuit of this recommendation was by the EPA, which beginning in 1986 published a series of risk assessment guidelines to "set forth principles and procedures to guide EPA scientists in the conduct of Agency risk assessments, and to inform Agency decision makers and the public about these procedures" (51 FR 33992). (Actually, EPA had set out a publicly available "interim" statement of its principles and methods for carcinogen risk assessment as early as 1976 [41 FR 21402].)

As with other Federal agencies, the EPA endorses as overarching guiding principles the NAS "red book" (NRC, 1983) and the 1985 "OSTP Principles" (50 FR 10371). The EPA guidelines supplement these. The guidelines are intended to provide structure and a common framework to risk assessment while allowing case-by-case flexibility to treat each case in an appropriate way. (The guidelines allow much more flexibility than is generally used in practice.) The aim of setting out these structures and principles is not only to inform but also to guard against the danger that the inevitable ambiguities in science and gaps in knowledge will be settled in an inconsistent, *ad hoc* manner, differently from case to case depending on who is making the decision. (It was in reaction to the perception of this problem that the NAS recommended that guidelines be developed.) For this reason, the guidelines establish "default" methods, practices based on background biological knowledge and general principles to be used in bridging gaps or settling ambiguities in the face of lack of case-specific knowledge. Many of these default methods are conservative, in the sense that they are chosen to be unlikely to underestimate true risk. The role of guidelines, default methods, and conservatism is extensively discussed in the National Academy of Sciences study on risk assessment methods, *Science and Judgment in Risk Assessment* (NRC, 1994).

Five sets of guidelines were finalized in 1986. They are the *Guidelines for Carcinogen Risk Assessment* (51 FR 33992), the *Guidelines for Mutagenicity Assessment* (51 FR 34006), the *Guidelines for the Health Risk Assessment of Chemical Mixtures* (51 FR 34014), the *Guidelines for the Health Assessment of Suspect Developmental Toxicants* (51 FR 34028), and the *Guidelines for Estimating Exposures* (51 FR 34042). The developmental toxicity and exposure guidelines have since been revised as the *Guidelines for Developmental Toxicity Risk Assessment* (56 FR 63798, 1991) and the *Guidelines for Exposure Assessment* (57 FR 22888, 1992). Two new guidelines have been published as proposals: the *Proposed Guidelines for Assessing Male Reproductive Risk and Request for Comments* (53 FR 24850, 1988) and the *Proposed Guidelines for Assessing Female Reproductive Risk, Notice* (53 FR 24834, 1988). (They will likely be merged into a single document before finalization, according to the EPA Risk Assessment Forum.)

The most discussed of these EPA guidelines are the *Guidelines for Carcinogen Risk Assessment*. (Indeed, just as may people imprecisely use the term "risk assessment" when they wish to refer specifically to quantitative dose-response assessment for

carcinogens, they often use the term "EPA guidelines" to refer to these carcinogen assessment guidelines.) The EPA carcinogen guidelines have been undergoing a revision process since 1988. EPA has published documents from meetings held to consider revision issues (*Workshop Report on EPA Guidelines for Carcinogen Risk Assessment* [EPA, 1989a], *Workshop Report on EPA Guidelines for Carcinogen Risk Assessment: Use of Human Evidence* [EPA, 1989b], and the outlines of a revision proposal with the marvelously tentative title, *Working Paper for Considering Draft Revisions to the U.S. EPA Guidelines for Cancer Risk: Review Draft* [EPA, 1992a]. As the present words are being written, however, a proposal for revised carcinogen assessment guidelines is in final preparation and is projected to be released sometime in late 1995. Substantial changes to the current guidelines are expected to be proposed therein. The specific provisions of these various EPA guidelines will be discussed under the appropriate topics below.

In addition to these formal guidance documents, EPA produces documents from time to time on the development or advancement of specific methods. Notable among these are: the Exposure Factors Handbook (EPA, 1989c), containing a compendium of useful data (including distributional data) on human behavior patterns, consumption of air, water, soil, human growth and development, and other data that are useful as parameters in exposure model calculations (a revision and update is in preparation); Interim Methods for Development of Inhalation Reference Concentrations (EPA, 1990), developing new methods for non-cancer assessment of inhaled substances; and Use of the Benchmark Dose Approach in Health Risk Assessment (EPA, 1995). That is, development of the specifics of implementation of the risk assessment guidelines are pursued on an agency-wide basis.

Promoting Coordination and Consistency:

In addition to the publication of guidance and statements of methods, a number of management practices within the EPA have been instituted to promote consistency and to coordinate development of approaches to potentially divisive risk assessment issues that may arise. Three particular institutions should be mentioned: the Science Policy Council, the Risk Assessment Forum, and the IRIS database.

The *Science Policy Council* comprises a group of senior EPA managers representing offices involved in risk assessment. They meet regularly to coordinate approaches to agency-wide risk assessment policy questions at a senior management, policy-setting level. They may issue guidance to the agency on emerging issues in the form of internal memoranda. A recent important example is the "Policy for Risk Characterization at the U.S. Environmental Protection Agency," signed by the EPA Administrator on March 21, 1995. This statement sets out the agency's policy for standards of risk characterization in EPA risk assessments, as discussed further below.

The *Risk Assessment Forum* comprises a group of senior staff-level scientists and mid-level managers concerned with risk assessment programs. Members are drawn from across all agency offices and including representatives of EPA regional offices. The Risk

Assessment Forum is responsible for scientific and science policy analysis of precedent-setting or controversial risk assessment issues, especially those that are cross-cutting and fundamental. Its objective is to promote EPA-wide consensus and the incorporation of this common understanding into guidance. It is not ordinarily a forum for review, nor does it set policy, instead being the vehicle for making policy recommendations to the Science Policy Council. Forum membership is for a fixed term, and members are nominated by the office they represent. The Risk Assessment Forum oversees the development and revision of EPA guidelines. It sponsors workshops and colloquia, drawing on internal and external expertise, on overarching issues. It may publish documents discussing methodological issues (e.g., Thyroid Follicular Cell Carcinogenesis: Mechanistic and Science Policy Considerations [EPA, 1988]), or proposing new policies on specific issues (e.g., Alpha-2u-globulin: Association with Chemically Induced Renal Toxicity and Neoplasia in the Male Rat [EPA, 1991], which developed the recommendation on criteria for ruling certain male rat kidney tumor responses irrelevant to human risk assessment).

Throughout its various parts the EPA carries out many risk assessments. These vary in purpose and in data availability, so there is variation in rigor to be found; some assessments are single-purpose screening assessments receiving limited review, while others are fully developed, documented, and reviewed assessments serving as the basis for major regulatory initiatives. To promote internal consistency and coordination, the agency began listing major assessments, together with a review of the basis for the assessments' findings, in a computerized database, the Integrated Risk Information System, or IRIS. This resource was initially intended for internal agency use only, but it soon grew to become a major resource on EPA risk assessment information used by state and local governments as well as internationally.

Until very recently, there were two internal, ongoing EPA workgroups charged with reviewing information before it was placed on the IRIS database: the *Carcinogen Risk Assessment Verification Endeavor* (CRAVE) examined carcinogenicity assessments, and the *RfD/RfC Workgroup* examined non-cancer assessments. (At this writing, the two workgroups have been disbanded and IRIS changes have been suspended until a new system for approving entries, yet to be fully defined, is implemented.) Like the Risk Assessment Forum, each group was composed of representatives nominated by different program offices or regional offices. Originally, these workgroups were to serve primarily as screeners and editors, examining assessments that had been conducted by one or another EPA office, culling those that did not meet a fairly rigorous minimum standard of completeness, rigor, and review, and adjudicating minor differences among regulatory programs. Those assessments that passed this standard were put in a common and comparable format and made available on the database, with each entry listing the pre-existing, already reviewed documents on which the listed assessment was based. Over time, the role of the workgroups grew to include resolving differences among offices about particular points contained in the assessments. Particularly in the case of the RfD/RfC Workgroup, there were few formally constituted risk assessment documents to draw upon, and the group became a forum for developing proposed IRIS entries de novo, which are in some sense new assessments.

Over time, the presence of an entry on IRIS has taken on more importance as an EPA-wide consensus position on a chemical rather than just an item of information about the existence of an assessment. The extensive reliance by regulatory processes within the agency and without on IRIS as a source of authoritative information on EPA's risk assessment stances has led to concerns that an IRIS listing, far from being simply informational, is in some ways equivalent to a regulation and hence in need of the same level of internal and external review, public comment, and formal approval. A series of reforms of the IRIS process are being undertaken, including peer review, public information submission, and comment, and these are responsible for the current suspension.

In addition to these relatively formal institutions, another important force for cross-EPA coordination and consistency is the interaction among representatives of the 10 EPA regional offices. Much of the issuance of permits, site-specific risk assessment (e.g., for Superfund sites), and impact of risk analysis on local and regional decisions is conducted and coordinated at these regional offices. Representatives of each region have monthly teleconferences and there is an annual Regional Risk Assessors meeting at which risk assessment methods as conducted in the various EPA regions are discussed and compared.

Oversight and Review:

Major risk assessment documents receive internal and external peer review. A major new EPA policy on peer review, including but not limited to review of risk assessments, is currently being implemented.

The prime institution for oversight and review of EPA risk assessments and positions on scientific matters, however, is the EPA Science Advisory Board. The Science Advisory Board (SAB) is a public advisory group providing extramural scientific information and advice to the EPA Administrator and other officials. It consists of outside experts appointed for fixed terms, chosen to provide a balanced expert assessment on the scientific matters at hand. *Ad hoc* members may be added for their special expertise in reviews of particular agency documents or products. There are several committees of the SAB focused on different EPA programs. SAB committees review most major risk assessment documents, proposed guidelines, and science policy statements as well as providing advice on agency science efforts and research programs. The findings of the SAB are transmitted in the form of letters to the EPA Administrator. While the recommendations are taken in high regard, they are not legally binding on the agency.

RISK ASSESSMENT METHODOLOGY COMMON TO EPA PROGRAMS

The risk assessment methods employed by the Environmental Protection Agency have much in common with those used elsewhere, reflecting the general practices,

standards, and precepts of the field. Risk assessment is a practical field, and the principles that have evolved reflect the concerns and ends of practitioners, including regulatory agencies and public health institutions, both national and international. The EPA has been an influential player in this development because of its major role in environmental regulation, the growing role of risk assessment in that regulation, and because the agency has made special efforts to define and develop the underpinnings of its methodology through the promulgation of risk assessment guidelines and promotion of scientific discussions about risk assessment methodology.

In this section, elements of risk assessment methodology that are more or less common to all EPA programs and offices are discussed, focusing on those aspects that may be different from practices at other Federal regulatory agencies. The major principles and methods come from the various EPA risk assessment guidelines. A good deal of the detail and specific practice, however, is not codified in guidelines but rather is established by precedent and ongoing practice at EPA. As with other agencies, the documentation of these practices, and the arguments in favor of their employment, are made in the context of their repeated use and defense in specific risk assessment documents and the rules that they support.

The matters discussed below are not static; the present time is one of great reexamination of risk assessment principles and practices, and several major aspects of risk assessment at EPA are in transition as new policies are developed and implemented. Some of this change is in direct response to recent critical studies of EPA risk assessment methodology, in particular three reports from the National Academy of Sciences, *Science and Judgment in Risk Assessment* (NRC, 1994); *Pesticides in the Diets of Infants and Children* (NRC, 1993a); and *Issues in Risk Assessment* (NRC, 1993b).

In particular, a new proposal for revision of EPA's carcinogen assessment guidelines is announced for appearance sometime in the Autumn of 1995. If the contents are similar to the previously released "working paper" (EPA, 1992a), major changes in hazard identification, dose-response analysis, and risk characterization are to be proposed. At the present juncture it is difficult to discuss these, since the proposal has not yet been made, and the implementation of the anticipated changes has yet to be tried.

There are also new EPA-wide policies that have been announced, but for which the specific implementation and experience of use are just underway. These include a policy on risk characterization and a directive that all EPA programs are to consider exposures by multiple routes (and not just through the medium to be regulated by a particular office) in assessing risks. The 1992 revision of the exposure guidelines (EPA, 1992b) set out standards for consistent descriptions of exposures and the distribution of exposures within populations, recommending that, insofar as possible, all assessments include characterizations of average and "high end" exposures. In every interview with EPA risk assessors, these matters came up and ongoing activities in each office to change practices to address these new policies were alluded to. Of necessity, this report will stress current and past practice.

HAZARD IDENTIFICATION FOR CARCINOGENS

The available evidence on an agent's status as a carcinogen is rarely definitive; for some compounds the case can be clear and compelling, while for others there may be few or conflicting data, leading to unverified concern regarding the agent's possible carcinogenic properties. A pure potency analysis (i.e., a purely quantitative approach to carcinogen risk assessment) is calculated contingent on the assumption that the agent in question is indeed carcinogenic, and such a calculation is frequently possible even for agents that have but suggestive evidence that they are carcinogens at all. Thus, during the development of the 1986 guidelines it was seen as necessary to erect a ranking scheme for weight of evidence to distinguish agents with stronger or weaker evidence that they may pose a carcinogenic threat to humans at some conditions or levels of exposure (with the magnitude of the risk, if it exists, being characterized by the subsequent dose-response analysis).

The process of hazard identification for carcinogens—that is, the qualitative determination that an agent may pose a carcinogenic hazard at some dose—follows a more defined method at EPA than at most other Federal regulatory agencies. The process, criteria, and a scheme for explicitly characterizing the overall weight of the evidence in a hierarchical classification scheme are set out in EPA's 1986 *Guidelines for Carcinogen Risk Assessment* (51 FR 33992). These criteria, and the evidence ranking scheme, are very similar (but not identical) to the ones employed by the International Agency for Research on Cancer (IARC, 1987). The Consumer Product Safety Commission has adopted an essentially similar scheme for the identification of carcinogens (57 FR 46626, 1992); some differences are noted in the section on that agency [p.45].

In common with other approaches, the EPA method recognizes three broad categories of data: (1) human data (primarily epidemiological); (2) results of long-term experimental animal bioassays; and (3) a variety of data on short-term tests for genotoxicity and other relevant properties, pharmacokinetic and metabolic studies, physico-chemical properties, and structure-activity relationships. A rebutable presumption is made that carcinogenic responses following one route of administration indicate potential hazard of the agent when exposure is suffered through other pathways of uptake, although certain site-of-administration responses and non-physiological exposures (e.g., injection-site sarcomas) are usually discounted. Questions on the magnitude of exposure associated with risk are left to the subsequent dose-response analysis.

Human data are the preferred basis for inference, but it is recognized that human studies often have low statistical power to detect effects, uncertainties about levels and classifications of exposure, and difficult to avoid problems with confounding and bias as influences on observed results. The guidelines refer to the standards established in the field of epidemiology for determining when a causal inference is credible. This judgment implicitly subsumes judgments about the soundness and adequacy of design, power, and reporting of individual studies as well as about the body of human evidence

as a whole (which often includes multiple studies, the concordance and consistency of which plays in the interpretation). The body of human evidence is classified as displaying "sufficient" evidence of the agent's carcinogenicity when a causal role of the agent can be considered established; "limited" evidence when "a causal interpretation is credible, but...alternative explanations, such as chance, bias, or confounding, could not adequately be excluded;" or "inadequate evidence" when there are few pertinent data or when the available studies "while showing evidence of association, did not exclude chance, bias, or confounding, and therefore a causal interpretation is not credible." There are also categories for "no data" and "no evidence," which means that no association was found of the agent with increased risk of cancer in well conducted analytical epidemiologic studies.

The EPA scheme for evaluating human evidence differs from the similar one of IARC in that (1) life-threatening benign tumors are included in the evaluation, although excluded by IARC as not representing true cancer; and (2) the "no data" and "no evidence" categories are added to allow for characterization of the *lack* of evidence for carcinogenicity when such is the case.

The second category of evidence, chronic animal bioassays, also has a five-level classification scheme with the same category titles as are used for human evidence. The emphasis is on repeatability of findings and the avoidance of false positive conclusions about the existence of elevated tumor incidence in some studies. That is, the presumption is made that a real (i.e., not spurious) finding of carcinogenic activity in *some* animal test systems will serve as strong evidence that the agent may be carcinogenic in humans as well. The scheme implicitly stresses the reliability of the positive findings in animals; negative results in animal studies that do not cast direct doubt on the positive findings in other tests garner relatively little weight.

The guidelines cite a number of considerations to be brought to bear on individual animal bioassays in judging the adequacy of their design and the reliability of their findings. This includes a statement of outcomes that should be considered to be "positive" in the sense of demonstrating carcinogenic activity by the agent. These criteria are intended to reflect the general precepts of the field of cancer toxicology and are generally similar to those employed elsewhere. (The guidelines heavily cite compendia of such criteria.) In the guidelines they are presented as guideposts for interpretation rather than as strict standards to be met for a study to be ruled in or out of consideration. Since criteria for making judgments are not completely codified, it is difficult rigorously to compare EPA's standards in this regard to those employed in other groups conducting hazard identification. In practice, there is rather wide agreement on the question of which studies (and which endpoints in those studies) are to be considered showing a carcinogenic effect, but it should not be overlooked that disagreements about the sufficiency of a particular study design (e.g., whether a maximum tolerated dose had been achieved) or about how to characterize a study's result could arise as a result of use of different statistical methods, different lumping or splitting of tumor-type categories, and other seemingly minor methodological variants that are beyond the scope of the present study to explore.

"Sufficient" animal evidence is obtained when "there is an increased incidence of malignant tumors or combined malignant and benign tumors: (a) in multiple species or strains; or (b) in multiple experiments...;or (c) to an unusual degree in a single experiment with regard to high incidence, unusual site or type of tumor, or early agent at onset" (51 FR 33999). In practice, this has tended to mean that agents with more than one positive animal bioassay are regarded as having "sufficient" evidence, notwithstanding any negative animal results unless they cast direct doubt as to whether the positive results might be experimentally spurious. "Limited" animal evidence is obtained when positive results are not repeated, when study quality or design is not up to the usual standards, or when benign tumors only are caused. "Inadequate evidence" indicates that "because of major qualitative or quantitative limitations, the studies cannot be interpreted as showing either the presence or absence of a carcinogenic effect." As with the evaluation scheme for human data, the animal evidence classification also includes categories for 'no data' and for 'no evidence of carcinogenicity,' i.e., evidence inconsistent with or arguing against the agent's potential carcinogenicity.

The scheme for characterizing the animal evidence is also similar to those employed elsewhere, specifically to that of IARC. The EPA's method differs primarily in two areas: the interpretation of benign tumors and the treatment of tumor responses with high background rates (i.e., appreciable rates even in control animals). IARC tends to discount benign tumors, but EPA's guidelines state that "An increased incidence of combined benign and malignant tumors will be considered to provide sufficient evidence of carcinogenicity if the other criteria defining the 'sufficient' classification of evidence are met....Benign and malignant tumors will be combined when scientifically defensible." However, "an increase in benign tumors alone generally constitutes 'limited' evidence." (51 FR 33996). That is, under the 1986 guidelines the EPA will combine the counts of benign and malignant tumors of the same histologic type before statistical analysis of bioassay results unless there is evidence that the particular benign tumors are not expected to be able to proceed to malignancy. IARC would tend to analyze the incidences of the malignant tumors only, even if they were accompanied by benign tumors.

IARC views increases in tumors that have high spontaneous background rates as "limited" evidence, while EPA's current guidelines state that such results "generally constitute 'sufficient' evidence of carcinogenicity, but may be changed to 'limited' when warranted by the specific information available on the agent" (51 FR 33996). Criteria are given for such downgrading, including "an increased incidence of tumors only in the highest dose group and/or only at the end of the study; no substantial dose-related increase in the proportion of tumors that are malignant; the occurrence of tumors that are predominantly benign; no dose-related shortening of the time to the appearance of tumors; negative or inconclusive results from a spectrum of short-term tests for mutagenicity; the occurrence of excess tumors in a single sex" (51 FR 33995).

The third category of evidence to be weighed in the overall weight of evidence consideration—data on short-term tests for genotoxicity and other relevant properties,

pharmacokinetic and metabolic studies, physico-chemical properties, and structure-activity relationships—has no explicit method prescribed for its summarization or entry into the overall hazard identification. Instead, these data are to enter into the judgments about classification of human and animal evidence (which are not supposed to follow hard and fast rules) and into the overall judgment when the animal and human data classifications are combined.

The question of the relevance of animal carcinogenic responses to prediction of human risk must come into the picture in this same manner. The current EPA guidelines indirectly recognize this issue but do not explicitly provide for removing animal responses not thought to be relevant to humans before the animal evidence is classified. The guidelines make clear, however, that the separate classification of human and animal evidence is only a means to the end of arriving at a total weight-of-evidence judgment, and that that overall judgment should be made based on case-specific insight and scientific interpretation of the bearing of the total body of data on the question of whether the agent should be considered to be a potential human carcinogen. The guidelines state "the scientific data base will have a complexity that cannot be captured by any classification scheme. Therefore, the hazard identification section should include a narrative summary of the strengths and weaknesses of the evidence as well as its categorization in the EPA scheme" (51 FR 33996).

In the final stage of EPA hazard identification of carcinogens, the animal, human, and "other" evidence is combined to place the body of evidence regarding the agent's potential as a human carcinogen into one of several hierarchic categories. If the human data are by themselves sufficient to demonstrate the causal association of the agent with human cancer, the agent is classified in "Group A - Carcinogenic to Humans," a category intended to correspond to IARC's Category 1. Agents with evidence placed in "Group B - Probably Carcinogenic to Humans" generally have "limited" human evidence (with any classification of animal evidence) *or* "sufficient" animal evidence (with less than "sufficient" human evidence). The former agents are subcategorized into "Group B1" and the latter into "Group B2," corresponding to IARC's 2A and 2B, respectively. (IARC characterizes agents in its group 2A as "probable" human carcinogens and agents in 2B as only "possible," whereas EPA characterizes both of these groups as "probable." That is, despite the differing choice of descriptor, IARC's 2B and EPA's B2 represent the same level of evidence; indeed, until recent years, IARC's descriptor for its category 2B was also "probably carcinogenic to humans.")

EPA's "Group C - Possibly Carcinogenic to Humans" generally comprises agents with "limited" animal evidence and little human data. The guidelines state "This group... 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" (51 FR 34000). A source of risk assessment differences among EPA programs is whether, as a

practical matter, they treat agents placed in Group C as they do agents with more certain carcinogenicity or according to some other procedure, as detailed in the sections on each regulatory program.

EPA has a further "Group D" for agents without adequate data either to suggest or refute the suggestion of carcinogenicity of an agent, and a "Group E" that "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" (51 FR 43000). Thus, somewhat unlike hazard identification procedures elsewhere, EPA's scheme is geared not only to identify carcinogenic hazards, but also to characterize the degree to which available evidence tends to exonerate an agent as a potential human carcinogen.

These last two EPA weight-of-evidence categories, "C" and "D," together more-or-less correspond to Category 3 of IARC. That is, compared to IARC's, the EPA scheme makes provision for recognition of some substances about which concern for carcinogenicity is limited by virtue of sparse or equivocal evidence (the category "C") as distinct from agents with no basis at all for such concern (the category "D"), while IARC lumps these agents into a single category covering agents with insufficient evidence regarding carcinogenicity. This is perhaps the largest difference between the EPA and IARC evidence classification schemes.

Indications are that EPA is planning to propose a significant alteration of this carcinogen hazard identification scheme. Likely features of the new proposal include a great reliance on narrative statements describing the main lines of evidence and their bearing and interpretation in place of pre-defined hierarchical categories with alphabetic designations. It is expected that rather than a three-step process of separate evaluation of human evidence, animal evidence, and melding these judgments into an overall weight of evidence (while bearing in mind the short-term test data), the new guidelines proposal will suggest a single comprehensive evaluation process stressing the explicit consideration of coherence of the various data elements into one scientific interpretation that evaluates (to the extent possible) how well the commonality of mode of carcinogenic action between human beings and the various test systems has been established. Emphasis will also be placed on defining the qualitative conditions under which carcinogenic hazards might be expected; if warranted, limitations to the finding of carcinogenic hazard can be drawn based on route of exposure, necessity of some other toxic reaction to which tumorigenesis is secondary, and doses below which such toxic reactions (and hence elevation of cancer risk) are not expected to occur.

These proposed changes are intended by EPA to address criticisms that have been leveled at its hazard identification of carcinogens. They would implement several recommendations of the NRC (1994) report, *Science and Judgment in Risk Assessment*, but in doing so they would bring EPA's hazard identification process farther from those currently in use by other risk assessing institutions, including IARC.

HAZARD IDENTIFICATION FOR EFFECTS OTHER THAN CANCER

Hazard identification procedures are less formally set out for non-cancer effects than they are for the identification of carcinogens. Nonetheless, the EPA has gone further than most other agencies in setting out guidance, in the form of the guidelines for mutagenicity assessment, developmental toxicity assessment, and (proposed) guidelines for male and female reproductive toxicity assessment, as mentioned above [p.57].

(An ongoing interest at the EPA in publishing guidelines for the assessment of systemic toxicity has never come to fruition. The agency's long history of application of a consistent set of principles in the procedure for identifying reference doses for hundreds of non-carcinogens constitutes a substitute for such guidelines, however. These procedures are documented in the methodology statements associated with the IRIS data base, mentioned earlier on p.59.)

Clearly, the above-mentioned guidelines do not exhaust the spectrum of toxicological endpoints of potential concern for risk assessment. Nonetheless, the principles they set out help to define the methodological approach for a variety of non-cancer toxicities.

Rather than specifying risk assessment methodology, these non-cancer guidelines tend to focus on the proper conduct of testing and the appropriate toxicological interpretation of results of the commonly done assays. Caveats, potential pitfalls, and points to consider in interpretation are listed, but the guidance for hazard identification decisions—including on such questions as what outcomes should be considered adverse effects—is quite general,. For instance, the *Guidelines for the Assessment of Suspect Developmental Toxicants* (51 FR 34028) state that "all data pertinent to developmental toxicity should be examined in the evaluation of a chemical's potential to cause developmental toxicity in humans, and sound scientific judgment should be exercised in interpreting the data in terms of the risk for adverse human developmental health effects."

The *Guidelines for Mutagenicity Risk Assessment* (51 FR 34006) present a scheme for identifying agents that are hazards for human germ-cell mutagenesis. (The guidelines note that somatic-cell mutagenesis is also of concern, especially as a potential mechanism of carcinogenesis and teratogenesis, but that the guidelines' focus is on heritable effects. Somatic-cell effects may constitute evidence of potential germ-cell effects, however.) The guidelines note that point mutations as well as numerical and structural chromosomal aberrations are of concern, and state that "the Agency will place greater weight on tests conducted in germ cells than in somatic cells, on tests performed *in vivo* rather than *in vitro*, in eukaryotes rather than prokaryotes, and in mammalian species rather than in submammalian species." A hierarchical scheme of eight hazard categories is presented, with the highest category to include agents with "positive data derived from human germ-cell mutagenicity studies," and ranging downward as data are restricted to non-humans, to *in vitro* studies, and to demonstration of somatic rather than germ-cell mutagenic effects. Although this scheme is the most worked-out systematic

process for hazard identification outside the realm of cancer assessment, it is seldom used in practice. Instead, the focus of mutagenicity assessment at EPA (as elsewhere) is on the role of somatic mutations as an indicator of carcinogenesis and its mode of action.

QUANTITATIVE RISK ASSESSMENT FOR CARCINOGENS

The issues on which the EPA *Guidelines for Carcinogen Risk Assessment* (51 FR 33992) provide guidance are the selection of data sets on which to perform quantitative analysis, the choice of mathematical extrapolation models, and equivalent exposure units among species. All three of these bear some differences from practices at other agencies.

Choice of Data:

Epidemiologic data are preferred providing the study chosen is well conducted and has adequate exposure information and reporting of tumors for a useful quantitative analysis. ("Negative" human studies can sometimes provide useful upper bounds on risk.) If data from rodent bioassays are to be used, the various available data sets are presented, separated by experiment, sex, and by the site and type of tumors. The guidelines state that "Benign tumors should generally be combined with malignant tumors for risk estimates unless the benign tumors are not considered to have the potential to progress to the associated malignancies of the same histogenic origin." (Other agencies differ in treatment of benign tumors, either tending to drop them from analysis, or combining only when positive evidence of progression is available.)

In the face of choice among analyses of several data sets, the guidelines state that "The range of the risk estimates is presented with due regard to biological relevance (particularly in the case of animal studies) and appropriateness of route of exposure." They further state that "Because it is possible that human sensitivity is as high as the most sensitive responding animal species, in the absence of evidence to the contrary, the biologically acceptable data set from long-term animal studies showing the greatest sensitivity should generally be given the greatest emphasis."

Although a range of potency estimates from the various data sets is presented within the body of a risk assessment document, the regulatory culture as it currently operates requires that a single choice for the estimate of an agent's carcinogenic potency emerge at the end. (This is less the case at some other agencies where the choice among regulatory options hinges less on specific risk estimates. For example, OSHA routinely entertains a number of potency estimates simultaneously, as discussed in the section on that agency [p.37].) In practice, the highest potency estimate—i.e., the one emerging from the "most sensitive sex and species"—often becomes the number chosen. It is important to note that the guidelines themselves do not require this (mandating only "emphasis" on this result among others), nor do they strictly require that a single potency number be chosen.

The emphasis by EPA on the most sensitive sex and species is less absolute than is sometimes portrayed. First, as noted, the body of the risk assessment document contains analysis of the range of data sets, it is only in the "bottom line" conclusions that a single choice to represent the quantitative potency estimate is made. (Of course, everything except such conclusions tends to get stripped away as the risk assessment is used in practice. In a sense, then, the issue is more one of risk characterization and communication than of dose-response assessment *per se*.) Second, considerations such as the anticipated route of exposure among humans often lead to the choice of a data set obtained via the same route (or a more appropriate alternative, such as drinking water over oil gavage for an anticipated human inhalation exposure), even if this is not the most sensitive. Third, studies well designed for use in risk assessment (i.e., including adequate numbers of dose groups, numbers of animals per dose group, appropriately chosen dose levels, and so on) are chosen over less well designed studies, even when the latter show the highest apparent "sensitivity." Finally, when several data sets give roughly similar outcomes, a final potency is often calculated as the average (typically, the geometric mean) of the several possibilities.

The biological relevance of the data set is brought into consideration when there is information available. Appearance of tumors in animals that are considered similar to those appearing in human studies of the same compound adds weight to their presumed relevance. (The converse does not hold, however; the general default presumption of the relevance of animal tumors to potential human risk is not contradicted by lack of human responses at the same tumor sites. Indeed, animal response in organs entirely lacking in humans [e.g., Zymbal gland, forestomach] is not by itself taken as evidence of irrelevance. Comparison of positive cancer responses among animal species and from animals to humans shows that carcinogens often attack different organs or systems in different species, and the ability to generate malignancy in one setting is an indicator of the potential ability to do so at another site in another species.) It is not considered toxicologically tenable to argue that carcinogenic responses in male animals are more relevant to human male risk projections and in female animals to human females.

Instead, the relevance of animal data sets is judged by reference to knowledge of the mechanism of carcinogenic action and whether than mechanism corresponds to one that might be expected to operate in humans. This generally is taken to require a fair amount of biological insight into questions that are seldom well understood. The question of how much certainty should be sufficient in a mechanistic explanation of tumors in animals (and the irrelevancy of such a mechanism to humans) has been contentious. EPA has established a stance on this by way of a particular example, as documented in a Risk Assessment Forum publication, *Alpha-2u-Globulin: Association with Chemically Induced Renal Toxicity and Neoplasia in the Male Rat* (EPA, 1991). Based on a significant body of data, the EPA declared in this document that certain male rat kidney tumors are not to be considered as relevant to the assessment of human cancer risks. These tumors arise as a result of accumulation of droplets of a protein unique to male rats in renal tubules, leading to nephropathy and cell death. Kidney tumors are secondary to the compensatory hyperplasia induced by the toxicity and would not be expected to occur otherwise. The inhibition of degradation of the unique protein by

certain chemicals is responsible for the accumulation. Humans have no such protein subject to this effect.

It is frequently the case that, within a particular bioassay, animals show increases in more than one type or site of tumor. If it appears that these responses represent independent and multiple effects of the agent in that species and sex, a question of interpretation is posed for the use of the data in risk assessment: Are the independent responses to be considered alternative bases for potency estimation, each representing an observation of the tumorigenicity of the compound that can be projected to humans, or should one project the total rodent risk (consisting of the risk of developing any one of the several tumor types) to humans?

The EPA guidelines take the position that both types of analyses should be conducted, but that "to obtain a total estimate of carcinogenic risk, animals with one or more tumor sites of types showing significantly elevated tumor incidence should be pooled and used for extrapolation." That is, total risk should be projected, but this is to be done by combining the data into an incidence rate for bearing any of the tumors shown to be elevated by dosing before analysis (rather than analyzing each tumor type separately and projecting the summed risk, as is done at CPSC, for example [p.47]). In practice, this provision is not always followed; it is sometimes the case that a particular tumor response (usually the strongest one) among those seen in the data set becomes the basis for extrapolation.

A final issue about choice of data sets for quantitative analysis, and one not dealt with in the EPA guidelines, is that of correction of tumor incidence data for intercurrent mortality. Bioassays are nearly full rodent lifetime experiments. Cancers in animals appear mostly at older ages, just as they do in humans. It is often the case that consequential numbers of animals die (or go missing) before the full 104 weeks of the experimental exposure are complete, some of them quite early on. Sources of this loss vary, but they include escapes, mortality due to handling errors (e.g., injury during gavage), diseases, natural causes, death due to tumors of other types (whether induced by the test compound or spontaneous) and (in some cases) toxicity of the test compound. An animal that dies early has not had its capacity to develop a tumor of the type of interest over a full lifetime fully tested, a problem that gets worse the earlier the death. As noted in the next section, one way to handle this issue is by use of time-to-tumor statistical models of the dose-response relationship. Frequently, however, data for such an approach are not available, and an approximate correction must be applied to adjust the denominators (i.e., the numbers of animals in each dose group considered "at risk" for developing a tumor). There are several methods—for example, to eliminate from consideration all animals dying before the appearance of the first tumor of the type of interest—but the adequacy of the alternatives depends on the particular mortality pattern. In the interviews conducted for this study, it became apparent that this matter was treated largely *ad hoc*, without even unwritten policies in many cases. Although the alternatives are unlikely to result in large numerical differences, this is a source of potential inconsistency in the analysis by different groups of the same tumor data.

Choice of Mathematical Extrapolation Model:

The early history of quantitative risk assessment for carcinogens at EPA is recounted by Anderson et al. (1983). In many ways, it parallels the history of this process at FDA (as described in the section on that agency). In the early 1970's, the EPA's pesticides office moved to cancel most uses of three important pesticides (DDT, aldrin/dieldrin, and chlordane/heptachlor). Following the requirements of the pesticides law for balancing costs and benefits, the agency ran into controversy regarding assessment of benefits, raising fears that all carcinogenic agents for which thresholds could not be assumed would be treated as equally likely to cause cancer. In order to provide some means for ranking agents of high and low carcinogenic potency, and for making some estimate of the risks avoided by controls on use, the agency turned to early versions of quantitative risk assessment, publishing its first set of principles for such assessments in 1976 (41 FR 21402).

The initial EPA analyses evaluated potency by fitting the so-called one-hit model to tumor incidence data at different doses. This model assumes that tumors are initiated by a single rare event, and that the probability that this rare event occurs in a given individual (causing it to develop a tumor) rises in direct proportion to the dose of carcinogenic agent received. The model describes a dose-response curve that is linear up until substantial risk levels, and then gradually diminishes in slope, becoming asymptotic to 100% response at very high doses.

It is interesting that EPA chose a different modeling approach than did FDA at this early stage [see p.21]. The one-hit model is not only different from the FDA's initial use of the Mantel-Bryan procedure, it represents a different class of modeling approaches, embodying different ideas about why the incidence of disease differs at different dose levels. The Mantel-Bryan procedure is an outgrowth of the probit model, one of the so-called tolerance distribution models that rest on the idea that individuals vary among one another in the dose levels they tolerate without ill effect. When a population is exposed, those with their individual tolerances exceeded develop the effect and others do not. As dose rises, the proportion of the population whose tolerance is exceeded drops, leading to a dose-response curve that traces out the cumulative distribution of tolerances (in the case of the probit model, a lognormal distribution).

The one-hit model, in contrast, is the simplest of a class of models based on the idea that the chance accumulation of rare, discrete events is responsible for initiating the response. All individuals are presumed identical (rather than varying) in their sensitivity, that is, in the random chance that the rare events will occur. The increasing response to higher doses is modeled as the increasing chance that the requisite rare events (with probabilities increasing with dose) happen to occur in any particular individual. (Although in truth individuals vary in sensitivity, these models presume that the happenstance accumulation of chance events is the dominant factor in defining the shape of the dose-response curve observed in a population of exposed individuals.)

This approach to dose-response modeling was used in the radiation literature and is appropriate for describing the effects of mutagens. It also allows description of the

emerging pattern of response over time when exposure is continuous, explaining the convex age-incidence curves for cancer in epidemiological studies. During the 1970's it also seemed in accord with the growing understanding of cancer as a disease resulting in the rare loss of growth control in specific clones of cells (among their many unaffected neighbors), proceeding according to a series of discrete stages or transformations. These transformations could plausibly be thought of as mutations, an interpretation backed by the strong correlation over compounds of mutagenic and carcinogenic activity.

For these reasons, then, the EPA started down a different dose-response modeling path than did FDA, albeit in response to similar needs and with an approach that gave similar answers. Also like FDA, there was a change in procedure after a few years of experience and further thought. In response to public comments on the promulgation of water quality criteria in 1980 (44 FR 15926, 44 FR 43660, 44 FR 56628), EPA adopted a slightly more complex model in the same family of approaches, the linearized multistage model (Crump, 1981). This model has been used regularly by the agency since 1980 (Anderson, et al., 1983). Its use is endorsed in the 1986 guidelines, which state that various modeling approaches should be considered, but that unless biological or statistical evidence suggest a particular model, the linearized multistage model should be employed (51 FR 33997).

The multistage model has the form

$$P(d) = 1 - \exp - (q_0 + q_1 d + q_2 d^2 + \cdots + q_n d^n),$$

where *P(d)* is the probability of response at dose *d* and the *q*'s are fitted parameters that are constrained to be non-negative, fitted simultaneously by the method of maximum likelihood (Crump, 1981). This equation describes a curve that can be suitably S-shaped to fit observed patterns of tumor incidence, yet (if q_1 is positive) can decline in direct proportion to dose at low dose levels. Risk engendered by the dose over and above the background risk of spontaneous tumors is usually expressed as so-called "extra" risk, defined as

$$R(d) = \frac{P(d) - P(0)}{1 - P(0)},$$

where *P(0)* is the modeled risk at zero dose. With this definition of risk above background, at low dose levels (and providing that q_1 is positive), the above formula for added tumor risk from the compound at dose *d* reduces approximately to

$$R(d) \approx q_1 d,$$

that is, added risk is approximately proportional to dose times the agent's "potency," given by the fitted q_1 term in the multistage model equation and representing the low-dose slope of the fitted curve. That is, q_1 has the units of risk per unit of dose.

The fitted value of q_1 is critical; if it is estimated as zero, added risk is proportional to a higher power of dose, the low-dose region of the fitted curve is non-linear, and low-dose risk estimates are orders of magnitude lower than predicted by any positive q_1 value. Unfortunately (and unavoidably) the statistical power to fix the value of q_1 is limited by the very fact that it determines the critically important low-dose shape of the curve; its influence on curve shape is felt in the region of the curve where observations are lacking. Small changes in the high-dose data on tumor incidence (of the kind and magnitude expected by statistical fluctuation in bioassays limited to about 50 animals per dose group) may result in a positive q_1 becoming zero or a q_1 of zero becoming positive.

For this reason, EPA takes what amounts to a statistical upper bound on the value of the q1 term, designated q_1*, to yield a low-dose slope of the fitted curve that will be at the top end of the range of slopes that are among reasonable fits to the tumor incidence data. That is, the curve is chosen that has the curvature to fit the observations in the high dose range and has the largest reasonable linear slope in the unobservable low-dose range. If the unknown true curve is actually linear, this procedure overestimates the slope by a relatively small factor, usually two- to three-fold. If, however, the true curve is nonlinear at low doses, it is well below any linear extrapolation, and any such extrapolation (be it from the multistage model or elsewhere) will overestimate true risks by a ratio that grows as the dose level decreases.

The rationale that EPA employed in choosing the multistage model was that a curve was sought that had the flexibility to fit observations at high doses and yet produced a linear low-dose extrapolation. The EPA guidelines (51 FR 33997) cite the OSTP principles (50 FR 10372) recommendation that "When data and information [about expected low-dose curve shape] are limited,...and when much uncertainty exists regarding the mechanism of carcinogenic action, models or procedures which incorporate low-dose linearity are preferred when compatible with the limited information." In addition, the guidelines mention the argument of low-dose linearity due to additivity to background processes (Crump et al., 1976), stating that "If a carcinogenic agent acts by accelerating the same carcinogenic process that leads to the background occurrence of cancer, the added effect of the carcinogen at low doses is expected to be virtually linear."

That is, the use of the linearized multistage model is EPA's way to define a low-dose linear extrapolation, based on the general precept that, as a policy matter, such an extrapolation is appropriate when there is little information on the true low-dose shape of the curve. It is not applied with the understanding that most curves will indeed be linear, but only that a linear extrapolation is unlikely to underestimate risks for which it is very difficult meaningfully to characterize a "best" low-dose extrapolation The guidelines state "It should be emphasized that the linearized multistage procedure leads to a plausible upper limit to the risk that is consistent with some proposed mechanisms of carcinogenesis. Such an estimate, however, does not necessarily give a realistic prediction of the risk. The true value of the risk is unknown, and may be as low as zero" (51 FR 33997). Of course, the usefulness of such "one-sided" statements about low-dose

risk—that risk is likely to be no greater than a certain number without any indication as to how likely it is to be substantially lower—has been called into question, especially when the risk estimates are to be compared to one another or to costs. There is also controversy as to whether the policy stance that (as the guidelines put it) "An established procedure does not yet exist for making 'most likely' or 'best' estimates of risk within the range of uncertainty defined by the upper and lower limit estimates" (51 FR 33998) is still as true today as it was seen to be in 1986.

Within this framework of the multistage model, some details of application vary among EPA offices (and among agencies), as detailed in the sections on each office. The points of difference include how to choose the degree of the fitted equation (the number of q's and the highest power of dose included in the polynomial), the definition of risk over background, whether the full equation or the low-dose linear approximation is used, and whether risk calculations using the maximum likelihood estimate of the fitted curve are presented along with the upper bound. The effect of these variations is minor, however; the main uncertainty issue is the reality of a linear low-dose extrapolation. Variants in the way the linear extrapolation is defined (including ones not based on the multistage model, such as FDA's use of the modified Gaylor-Kodell procedure) all give about the same answer (generally within a factor of two, all else being equal) since they are all based on the same principles and aims.

The above methods deal with end-of-life tumor incidence data. As noted earlier, early mortality among animals from causes other than the tumor type of interest can lead to underestimates of the lifetime risk. If data on the time of death of each animal in the study are available, it may be possible to conduct a time-to-tumor analysis, in which one fits an analogue of the multistage model that also includes an estimate of the time-course of the appearance of tumors. Thus, end-of-life risks as they would be were there no competing intercurrent mortality can be estimated. Time-to-tumor analyses require that the cause of each death be classified as either random with respect to the presence of the tumor of interest (a so-called "incidental tumor" analysis) or as caused by the presence of the tumor (a so-called "fatal tumor" analysis). The distinction is necessary because fatal tumors prompt the time of their own discovery, whereas incidental tumors are found during necropsy at a time dictated by factors unrelated to the tumor's presence. Although standard reporting of data from lifetime bioassays now routinely includes time-of-death data, the problem of classifying causes of death often remains, and in practice, both incidental and fatal tumor analyses are conducted, which rarely differ appreciably.

Equivalent Exposure Units Among Species:

In the frequent case that human risks must be projected from the results of animal studies, the issue of what doses are of equivalent risk in rodents and humans must be addressed. The general toxicological precept is that, owing to the broad and fundamental similarity of much of mammalian anatomy, physiology, and biochemistry, toxic reactions in animals can serve as a useful insight into reactions that may occur in humans. Species-specific toxicological phenomena clearly exist, and the examination of underlying biological mechanisms during the hazard identification phase of risk assessment is intended to identify those cases where animal responses may be qualitatively misleading regarding potential for human risk.

Once it has been decided, either by default presumption or by biological insight, that results from an experimental animal bioassay qualitatively apply to potential human risk, it remains to account for the quantitative differences in scale between small, short-lived, quickly metabolizing rodents and large, long-lived, slowly metabolizing humans. This has traditionally been done by defining dose units that are presumed to lead to equivalent risk across species. That is, for the purposes of cross-species extrapolation, the rodents are treated as "scale model" humans, and it is assumed that if the magnitude of the challenge posed by the chemical can be adjusted to the magnitude of the recipient, a fundamental, scale-free similarity is preserved and the magnitudes of the responses can be treated as equivalent, at least as a first-cut approximation.

How to accomplish this adjustment for scale differences has been problematic, and different agencies have adopted somewhat different approaches, at least in the realm of assessment of carcinogens. Indeed, differences in assumptions about cross-species dose equivalency stand as the largest source of quantitative difference among carcinogen potency estimates by different elements of the Federal government. As noted below, there have been recent attempts to harmonize these methods.

For carcinogens, the EPA employs the assumption that end-of-life cancer risks will be equivalent across mammalian species when lifetime dosing in milligrams administered is proportional to each species' body surface area. (The principal alternative, used by FDA and OSHA, is to assume equal toxicity when daily doses are in proportion to each species' body weight.) The EPA cancer guidelines note the variety of scaling methods that have been used and then state "In the absence of comparative toxicological, physiological, metabolic, and pharmacokinetic data for a given suspect carcinogen, the Agency takes the position that the extrapolation on the basis of surface area is considered appropriate because certain pharmacologic effects commonly scale according to surface area" (51 FR 33998). Although the only mention of units in the guidelines is to "mg per m^2 body surface area," in practice the EPA uses the much more practical and essentially equivalent units of daily mg scaled by the 2/3-power of the species' body weight.

It should be noted that there is an implicit scaling for length of life, since the equivalent doses are expressed as amounts suffered daily for a full lifetime. The cumulative dose to a human over 70 years of daily dosing compared to only 2 years in rodents will be proportionally much higher than the ratio of daily doses presumed equivalent in toxicity.

The practice of surface-area scaling of carcinogen doses goes back to the origins of quantitative risk assessment at the EPA. It is not entirely clear why a different method was chosen than used by FDA. Such scaling has a history of use in certain clinical applications and early EPA documents cite its utility in projecting acute toxicity of antineoplastic drugs from animals to humans. The rationale behind its application to chronic effects and to cancer in particular has not been strongly articulated in EPA guidance, however. (By the same token, FDA's use of body weight scaling seems more readily attributed to the familiarity of this procedure from non-cancer safety assessment than to a well worked-out argument for how carcinogens should be treated.)

This is not to say that such a defense cannot be made. In 1992 a proposal for harmonizing the default methodology for scaling carcinogen doses across species was published jointly by EPA, FDA, and CPSC (*A Cross-Species Scaling Factor for Carcinogen Risk Assessment Based on Equivalence of mg/kg$^{3/4}$/day*; 57 FR 24152). This proposal reviewed the available empirical evidence on the success of various scaling methods in predicting cancer potency across species and also considered the rationale for scaling in terms of the underlying biological basis for potency differences in the patterns of scaling of physiological rates. The conclusion was that empirical tests, to the limited extent they could address the issue, support a scaling factor somewhere in the range between the FDA body weight method and the EPA surface area method, with neither existing method being clearly contradicted. The report found a great deal of chemical-to-chemical variability in relative potency, however, and departure from any scaling scheme can be substantial for particular agents. This report recommended a unified and harmonized default position for Federal agencies conducting risk assessment, namely that carcinogen doses should be presumed to be equivalently potent when daily doses are scaled by the 3/4-power of body weight. This position is in-between the current ones, and corresponds to the idea that dosing is best done in proportion to the general rates of metabolic and physiological processes in the dosed species. It is anticipated that the recommendation of this report will be incorporated into EPA's forthcoming revision of its cancer guidelines.

QUANTITATIVE ANALYSIS FOR EFFECTS OTHER THAN CANCER

For assessment of chronic toxic effects other than cancer, the methodology used by EPA is the approach of defining a "no observed adverse effect level" (NOAEL) from experimental data and then applying uncertainty factors. This is the basic approach to non-cancer assessment that is widely used. It was developed in the 1950's by FDA scientists as a way of defining daily intakes that humans could experience for a lifetime without appreciable probability of ill effect. (A brief discussion of this development is included in the section on FDA.) The rationale for this method is that non-cancer effects are expected to show a dose threshold below which no adverse reactions are to be found. The aim of the method is to identify the top of the range of dose levels without apparent ill effect in animals, and then to determine a dose to humans that will be similarly without effect, allowing for the possibility of greater sensitivity among at least some humans than in the experimental animals through the use of "safety factors."

Briefly, the existing data sets on chronic toxic effects are examined, and for each study a NOAEL is defined as the highest tested dose at which no statistically significant elevation over background in the incidence of the adverse effect was observed. (Doses are typically expressed in units of mg administered per kg of body weight per day.) Among all data sets, that producing the lowest NOAEL is chosen, and this NOAEL represents the highest daily administered dose level that was without apparent adverse effect in all available studies. This dose level is divided by a set of safety factors, typically factors of 10, to account for the possibility that the general human population might need a lower exposure to assure safety for each of the major extrapolations that must be made. These may include (1) the variation in sensitivity among members of the human population (over and above the variation seen in the experimental animals), (2) the uncertainty in extrapolating effects in animals to those in humans, (3) the possibility that full lifetime exposure entails greater sensitivity than the partial lifetime exposure tested, and (4) the occasional use of a study in which no dose was observed to be without effect. Which factors are used depends on which extrapolations are necessary in the particular case, but typically two or three factors, for a combined safety allowance of 100- to 1000-fold, is used.

This methodology presumes that daily doses scaled by species body weight are equally toxicologically effective across species (at least on average, with a 10-fold safety factor allowing for agent-by-agent variation). This assumption, which EPA shares with other Federal agencies, stands in contrast to the surface area scaling of doses presumed by EPA to be equivalent in carcinogenic potency.

EPA has a few features of its use of the "NOAEL/Safety Factor" methodology that are particular to this agency. The following points are drawn from the practices of the RfD/RfC workgroup, referred to earlier. The first is a matter of terminology. Rather than characterize the result of the procedure as an "acceptable daily intake" (as do FDA and CPSC, for example) the EPA prefers the more neutral term "reference dose"

(abbreviated RfD), since the question of acceptability is a value laden judgment not properly part of the risk assessment process. For similar reasons, EPA refers to "safety factors" as "uncertainty factors" so as not to imply that true safety has surely been achieved through their application.

Second, in addition to the usual uncertainty factors, EPA considers the application of further "modifying factors" on the basis of case-by-case professional judgment. These are additional factors (sometimes less than 10) that are applied to account for some particular uncertainty such as a data base for which only a few endpoints have been evaluated, a particularly troublesome route extrapolation, the possible role of development of tolerance, and so on. (In principle, but rarely in practice, a modifying factor can be less than one, i.e., reducing the total size of the uncertainty allowance, if data addressing the issue warrant.)

Third, EPA has made specific provisions for inclusion of pharmacokinetic data into its non-cancer assessments. The uncertainty factors for animal-to-human extrapolation and for variability among humans can be reduced (typically from 10 to 3, 3 being chosen as an approximate halving of the uncertainty factor on a multiplicative scale, since $\sqrt{10} \approx 3$) if pharmacokinetic data on each extrapolation are deemed informative. That is, the uncertainty in these extrapolations implicitly includes some unknown contribution of pharmacokinetic differences, and if these unknown differences can be characterized and numerically entered into the calculation of a reference dose, the corresponding uncertainty about that contribution can be reduced.

Fourth, EPA reference doses entered into the IRIS data base are accompanied by a "degree of confidence" statement (which may be either high, medium, or low). This statement, which is not reflected in the application of any numerical adjustment, is primarily intended to convey a judgment about the likelihood that the RfD might change upon the provision of further testing data. For instance, a less thoroughly tested compound has a higher probability than a well tested one that a new study (on a previously untested toxic endpoint) might yield a NOAEL lower than noted in previous studies.

Fifth, there are some particular variants of the methodology employed by EPA to account for special considerations about dosimetry. RfD's for developmental effects, for example, take into account the fact that there may be a critical time-window during fetal development for the induction of defects, and dose-rate during pregnancy, rather than chronic dosing over a lifetime, may be the appropriate metric to compare. By far the most important dosimetry consideration, however, is EPA's development of a variation on the RfD specifically for exposures by inhalation, known as the RfC (for reference concentration).

A full description of the RfC methodology is beyond the scope of this report. The methods (for there are several methods depending on whether the agent comprises particles or a gas and whether it has its effect at the site of deposition in the respiratory tract or systemically) are derived and documented in the EPA publication *Interim*

Methods for Development of Inhalation Reference Concentrations (EPA, 1990). (A revision of this document was released very recently, in October of 1995.) A good summary is presented in Rees and Hattis (1994).

Essentially, the RfC recognizes that for inhalation exposures, it makes more sense to express the encounter with the agent in terms of air concentrations rather than in terms of dose (i.e., amount taken in per unit of body weight). The air concentrations that will be toxicologically equivalent, however, will differ across species as a consequence of the different inhalation rates per unit of body weight in larger and smaller species. In addition, the relative rates and locations of deposition of the inhaled material across species will be affected by anatomical and geometric differences in the respiratory tract. Finally, the toxic effect from loading of the target tissue as a result of this deposition will depend on whether a respiratory tract surface is the target or whether the concentration in a volume of tissue is at issue (as, say, with systemic effects) since surface:volume ratios differ from experimental animals to humans. An adjustment for duration of exposure is also needed. The methods appropriate to addressing these issues differ if the agent is particulate or a gas, and if a gas, on its reactivity, water solubility, and readiness of absorption across the respiratory tract walls, as detailed in the aforementioned RfC document.

The RfC methods are based on the familiar idea of a NOAEL and uncertainty factors, but the NOAEL is expressed in terms of a human-equivalent concentration (HEC), adjusted according to the considerations listed above. That is, instead of presuming that doses in mg/kg/day are toxicologically equivalent across species, as the RfD methodology does, the RfC presumes that toxicologically equivalent air concentrations can be defined based on the idea of the concentration needed to produce an equivalent rate of loading of the target tissue with deposited or absorbed agent (in terms of amount per time per unit of surface area [or tissue volume] for specific target tissues or respiratory tract regions). Default methods and provisions for their replacement with agent- and species-specific dosimetric data are set out.

In general, the RfC methods will give different results than if inhalation exposures are simply rendered into estimated doses (by multiplying the volume of air breathed by the concentration of the agent in air). Thus, safe inhalation exposures as calculated by EPA's RfC methods will differ from those calculated by a dose-based ADI method at other agencies. Especially when deposition of the agent on respiratory surfaces is at issue, the RfC dosimetry adjustments bear some similarity to the surface area scaling methodology that is used by EPA for carcinogen risk assessment.

A final aspect of EPA's non-cancer risk assessment methodology that should be noted is this agency's interest in the developing methodology known as the "benchmark dose" approach (Crump, 1984). A recent publication of EPA's Risk Assessment Forum (EPA, 1995) reviews the agency's consideration of the issues involved in adopting this method. The benchmark dose (BMD) approach is an alternative to the NOAEL as a way to identify a dose without appreciable effect in an experimental study. Instead of considering only dose levels tested in the study, the BMD approach is to fit a dose-

response curve to the data in the observed experimental range. A lower bound on the dose causing some specified level of risk above background (say, 10% or 1%) is calculated, and this dose point is used as a point of departure for the application of uncertainty factors in place of the experiment's NOAEL. That is, it is taken as a standardized measure of a dose level near that at which the experimental response "disappears." Unlike the NOAEL, the BMD considers the entire set of data on doses and responses; it can bring to bear information on the overall pattern of response, including the steepness of the dose-response relationship, and is less sensitive to the specific placement of dose groups. Although the EPA has not widely used the BMD method, it has been employed in some RfC assessments. The agency is more actively working on developing it for use in non-cancer risk assessment than are other Federal regulatory groups.

It should be noted that benchmark doses have been extensively used since 1987 in ranking carcinogens under the Superfund and air toxics programs. The underlying paradigm is a ranking of relative potency, however, not the replacement of a NOAEL in a safety-factor based method, as is being considered for non-cancer risk assessment application.

EXPOSURE

As noted previously, the conduct of exposure assessment is typically the province of the individual regulatory programs as each one analyzes the exposure situations of interest to its regulatory responsibility. The main discussion of EPA exposure assessment in this report is accordingly in the sections on individual regulatory programs. Nonetheless, there are some overarching institutions to be mentioned that aim at keeping the methods and principles of exposure assessment in step across the EPA.

Among the EPA guidelines are the 1986 *Guidelines for Estimating Exposures* [51 FR 34042], which were updated in 1992 as *Guidelines for Exposure Assessment* [57 FR 22888-22938]. These guidelines provide general principles and practices for conducting exposure assessments, including considerations of measurement and modeling, model validation, and data quality assurance. They provide a series of definitions of terms such as "internal dose" and "toxicologically equivalent dose" that have in the past had imprecise meaning, leading to confusion and inconsistency. More importantly, they strive to provide a consistent framework for discussion distributions of exposure in populations. Exposure assessment tasks in various regulatory programs differ widely, as do the types of information available, their quality, depth, detail, and comprehensiveness. The aim of the guidelines is to provide a common terminology so that appropriate descriptions of each assessment, and appropriate comparisons, can be made.

The exposure guidelines urge that exposure assessors strive to provide both a central estimate of exposure and an estimate of "high end" exposure. The guidelines state "The primary objective when developing an estimate of high-end exposure or dose is to arrive at an estimate that will fall within the actual distribution, rather than above it"

[57 FR 22901]. That is, the guidelines seek to avoid generating a "worst case" exposure, especially a hypothetical exposure derived from worst-case assumptions on all elements of an exposure scenario. As a rule of thumb, they suggest that an exposure in the range of the 90th to 95th percentile should be the analyst's target for defining the "high-end."

A related document is the EPA *Exposure Factors Handbook* (EPA, 1989c). (This handbook is currently in the process of being expanded and updated; a citable draft will appear shortly.) The *Exposure Factors Handbook* contains (quoting the document) "a summary of the available data on various factors used in assessing human exposure including drinking water consumption, consumption rates of broad classes of food..., inhalation rate, skin area, lifetime, activity patterns, and body weight....Default values are presented as ranges from typical to reasonable worst case and as frequency distributions where appropriate data were available. Finally, procedures for assessing the uncertainties in exposure assessment are presented...includ[ing]...Monte Carlo and sensitivity analyses." The aim is to provide a sourcebook of parameter values for use across the agency, and to encourage use of reasonable exposure estimates by providing appropriate data sources and suggested methods.

A recently implemented EPA policy directs all regulatory programs to consider in its risk assessments exposures to an agent from all sources, direct and indirect, and not just from the source that is subject to regulation by the office doing the analysis. For example, the air program must estimate exposures not only to a compound in ambient air, but also by indirect pathways (such as might happen when crops grown in the same contaminated air take up the chemical). Some regulatory programs (e.g., the Office of Water) have made approximate allowances for such indirect exposure pathways in the past while others have not. At the present time, this policy is being put into practice across the agency, and most of the EPA personnel interviewed for this report mentioned some activity to update their program's exposure analysis methods accordingly. The practical consequences of this policy are not apparent yet. It is possible that the increase in overlap and duplication of exposure interests among regulatory programs will lead to a greater interaction and calls for standardization and perhaps centralization of exposure analysis, comparable to the forces acting on hazard identification and dose-response analysis.

RISK CHARACTERIZATION

As with exposure assessment, the conduct of risk characterization is typically done by the regulatory programs. There is EPA-wide guidance on how such characterization is to be done, however.

On February 26, 1992, then EPA Deputy Administrator Henry Habicht issued a memorandum entitled "Guidance on Risk Characterization for Risk Managers and Risk Assessors." This memorandum called upon agency risk analyses to give full characterization of the uncertainties and assumptions contained in agency risk assessments, including the provision of estimates of typical or central exposures, not just

worst case estimates or upper bounds. On March 21, 1995, EPA Administrator Carol Browner issued a "Policy for Risk Characterization at the U.S. Environmental Protection Agency." This document confirms and reinforces the guidance set out in the Habicht memorandum. It sets out standards of expectation for all agency risk assessments in terms of clarity, comparability, and consistency. It concludes that "a balanced discussion of reasonable conclusions and related uncertainties enhances, rather than detracts, from the overall credibility of each assessment." Among its standards are these (quoted from the policy document):

"[R]isk characterizations should be clearly presented, and separate from any risk management considerations....[They] should include a statement of confidence in the assessment that identifies all major uncertainties along with comment on their influence on the assessment....Information should be presented on the range of exposures derived from exposure scenarios and on the use of multiple risk descriptors (e.g., central tendency, high end of individual risk, population risk, important subgroups...)....In decision-making, risk managers should use risk information appropriate to their program legislation."

The policy addresses several points in particular having to do with the distribution of exposures. It states "For the Agency's purposes, high end risk descriptors are plausible estimates of the individual risk for those persons at the upper end of the risk distribution....The intent...is to convey estimates of exposure in the upper range of the distribution, but to avoid estimates which are beyond the true distribution. Conceptually, high end exposure means exposure above about the 90th percentile of the population distribution, but not higher than the individual in the population who has the highest exposure. When large populations are assessed, a large number of individuals may be included within the 'high end'...and information on the range of exposures received by these individuals should be presented." Monte Carlo and other modeling techniques are permitted if data allow. The document notes, however, that "unless a great deal is known about exposures and doses at the high end of the distribution, these estimates will involve considerable uncertainty which the exposure assessor will need to describe....If only limited information on the distribution of the exposure...is available, the assessor should approach estimating the high end by identifying the most sensitive variables and using high end values for a subset of these..., leaving others at their central values." As to central estimates, the policy endorses using "either the arithmetic mean...or the median exposure."

The estimation of population risks is endorsed and guidance given for describing proportions of the population with exposures or risks above key levels. Information on risk and exposure levels among subgroups of the population, especially those that might be particularly sensitive, should be developed as possible.

The EPA risk characterization policy is new and is in the process of implementation. The Science Policy Council is directing a program of implementation that includes efforts by each regulatory program and the regional EPA offices.

RISK MANAGEMENT CONSIDERATIONS

Overarching not only EPA risk management, but that of all Federal regulatory agencies, are President Clinton's executive orders (Executive Order 12866 on Regulatory Planning and Review and Executive Order 12875 on Enhancing Intergovernmental Partnership). These revoke and replace executive orders from President Reagan, but include many provisions on similar matters. The EO 12866 directs agencies to conduct cost-benefit analysis for all "significant regulatory actions" and to promulgate regulations only when necessary due to "compelling public need." Regulatory approaches should be chosen to maximize net benefits, minimize the overall regulatory burden on society, and to be the most cost-effective means of achieving the desired end.

A full discussion of these matters is beyond the scope of this report, but a summary of relevant executive orders under the Reagan and Clinton administrations, their comparative requirements, and their relation to risk legislation has been prepared (Schierow, 1994).

RISK ASSESSMENT IN EPA REGULATORY PROGRAMS

The preceding overview stresses general points applying to all EPA programs, together with a discussion of the institutions and policies that aim to keep EPA risk assessments coordinated and compatible. The following sections discuss each regulatory office in turn, presenting an overview of the relevant environmental legislation, its mandates about risk, risk assessment, and its use of risk analysis in developing its regulatory options.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

PESTICIDES - EPA OFFICE OF PESTICIDE PROGRAMS

The regulation of pesticides is carried out by EPA's Office of Pesticide Programs (OPP), which is a part of the Office of Prevention, Pesticides, and Toxics (OPPT). Pesticides are different than other potentially toxic compounds in that they are intended to be poisonous, at least to the pests they are designed to control, and they are intentionally introduced into the environment for that purpose. This situation naturally calls for the consideration of both costs and benefits, and the statutes under which pesticides are regulated provide for such analysis.

Pesticide regulation falls into two parts, and each part is accomplished under a different statute. The *registration* of pesticides (i.e., licensing for sale and use in agriculture or extermination) is carried out under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). No chemical may be sold in the U.S. as a pesticide without such registration, which establishes the conditions of legal use. The question of *tolerances* for pesticide residues on foods as encountered by the consumer is regulated under the Federal Food, Drug, and Cosmetic Act (FFDCA). The FFDCA is the primary statute of the Food and Drug Administration (and was discussed in the section on that agency, beginning on p.15), but the sections of that act applying to pesticide tolerances in foods are administered by the EPA. (The FDA retains authority over enforcement of tolerances, however.) Both FIFRA and the FFDCA existed before the EPA, but pesticide regulation was moved from FDA upon the founding of EPA in 1970.

This report contains an extensive discussion of food tolerance regulation under the FFDCA in the section on the Food and Drug Administration. The coverage of these matters will be briefer in the present section, which should be read in conjunction with the section on FDA.

FIFRA AND ITS MANDATES

The **Federal Insecticide, Fungicide, and Rodenticide Act** (FIFRA, 7 U.S.C.A. §§136 to 136y) provides for the regulation of sale and distribution of pesticides in the U.S. A pesticide is defined as "any substance or mixture...intended for preventing, destroying, repelling, or mitigating any pest, [or]...intended for use as a plant regulator, defoliant, or desiccant" [FIFRA §2(u)]. No pesticide may be introduced into commerce without obtaining a registration from the EPA. Registration is obtained through petition to the agency, with the petitioner providing information on the intended use, data on efficacy of the pesticide and its toxicological properties. The agency is empowered to ask for the provision of additional data, including the requirement for more toxicological testing, if the information is deemed necessary for the registration decision.

The Administrator may approve the petition if the pesticide "will perform its intended function" [§3(c)(5)(C)], and "when used in accordance with widespread and commonly recognized practice it will not generally cause unreasonable adverse effects on the environment" [§3(c)(5)(D)], which are defined as "any unreasonable risk to man or the environment, taking into account the economic, social, and environmental costs and benefits of ...use" [§2(bb)]. Pesticides are registered either for general use or for "restricted use" [§3(d)], with the latter category specifying conditions of use such as application methods, amounts used, target pests, geographic restrictions and so on.

Once granted, registrations expire after 5 years, at which time the petitioner can apply for renewal of registration [§6(a)]. There are provisions for EPA to cancel a registration early [§6(b)] if the Administrator finds adverse effects could indeed be caused, but a decision to cancel must take into account "the impact...on the agricultural economy." Much of the modern registration framework was introduced into FIFRA by 1972 amendments (the Federal Environmental Pesticide Control Act, 7 U.S.C.A. §136), and a large number of previously registered pesticides had been "grandfathered in" under the lax pre-1972 procedures. Further amendments in 1988 required re-registration (or cancellation) of these within 9 years, a large burden on the agency's risk assessment apparatus.

In sum, the registration process under FIFRA amounts to the granting of a license for sale and distribution of a potentially dangerous chemical. The license is not unlimited; it specifies the conditions of use that are permitted, potentially including restrictions on the target pests, the amounts of pesticide used, the application method, frequency, and timing of use, training of applicators, the time that must elapse after application before workers can reenter a treated field, and the time that must elapse after application before the crop can be harvested. Importantly, the registration also includes restrictions on which specific crops may be treated. Once registration is granted, however, all uses that fall within the specified restrictions become legal and permissible. That is, the regulatory power of registration is over permissible uses, not over actual practice within the permissible range.

To be granted a registration, the petitioner must demonstrate that the pesticide, when used on the proposed crops at the proposed levels, is effective at controlling pests and that, when used according to the restrictions, it will not cause unreasonable risk to man or the environment. The definition of such adverse effects in FIFRA is very vague, but in practice it includes risk to the applicators and farmworkers, ecological risks, risks to homeowners from extermination procedures, and (through interaction with the tolerance setting process of the FFDCA, as discussed below) risks to consumers of treated foodstuffs. The mandate in FIFRA for balancing of costs and benefits is similarly vague, comprising only the statement that "economic, social, and environmental costs and benefits" are part of the definition of what adverse effects are to be deemed "unreasonable." (The FFDCA is at least somewhat more specific on matters of both costs and benefits in regard to tolerances for residues on food.)

THE FFDCA AND ITS MANDATES REGARDING PESTICIDES

The **Federal Food, Drug, and Cosmetic Act** (FFDCA, 21 U.S.C.A. §§321 to 394) provides for regulation of permissible contents of toxic substances in or on food, and pesticides are explicitly considered in its provisions. While primarily an FDA statute, the parts of the FFDCA applying to pesticides are administered by the EPA. The FFDCA is discussed in the section on FDA, but some key provisions are briefly reiterated here.

Tolerances are the concentrations (on a per weight basis) permitted to remain in or on food as it is available to the consumer. The process of setting tolerances is also by petition, with the petitioner submitting proposed tolerance levels along with toxicological information to demonstrate that such tolerances will be sufficiently protective. Tolerances of pesticides on raw, unprocessed agricultural commodities are regulated under FFDCA §408, which mandates that tolerances should be set "for pesticide chemicals which are not generally recognized, among experts qualified by scientific training and experience..., as safe for use, to the extent necessary to protect the public health." However, "appropriate consideration" must be given "to the necessity for the production of an adequate, wholesome, and economical food supply."

If the commodities are processed (by "canning, cooking, freezing, dehydrating, or milling") and the processing results in concentrations of the pesticide that exceed the raw-product tolerances, then §402 of the act stipulates that the pesticide residues be considered as "food additives," which are regulated under §409. In such case, tolerances are to be set so that the substance "may be safely used," taking into account cumulative exposure to the substance from all dietary sources. ("Safe use" is defined elsewhere as "a reasonable probability of no harm.") Considerations of cost are excluded from tolerance decisions on food additives. Moreover, §409 contains the well known "Delaney Clause" which declares that no additive be deemed safe "if it is found to induce cancer when ingested by man or animal." As a consequence of these provisions, a pesticide that does not sufficiently concentrate during processing is regulated under a cost and benefit balancing standard (even if it is a carcinogen), while one that does concentrate is regulated under a strictly health-based standard, and if it has a positive carcinogenicity bioassay, it is banned altogether regardless of quantitative risk level.

RISK MANDATE

Registrations must set restrictions on permissible use so as to avoid "unreasonable risk to man or the environment, taking into account the economic, social, and environmental costs and benefits" [FIFRA §2(bb)]. At the same time, tolerances must be set so that residues experienced by the consumer "protect the public health" while allowing "the production of an adequate, wholesome, and economical food supply" [FFDCA §408], unless the pesticide sufficiently concentrates upon processing, in which case the residue should present "the reasonable probability of no harm," making no allowances for costs and benefits, and allowing no residues at all of carcinogens.

IMPLEMENTATION

The processes of petitioning for registration and petitioning for tolerances are interconnected, and in practice they often occur concurrently. Although regulated under separate laws and following different procedures, the two processes have a practical linkage in that the conditions and limitations for use of the pesticide established during registration must clearly lead to residues experienced by the consumer that will be below tolerances that can be approved on health grounds. The approval of tolerances is based on exposure from the total diet, so each new approved use of a pesticide in the registration process leads to potential residues that "use up" part of the total allowable intake. Since each use of a pesticide must employ enough of the agent to be effective against pests, a registrant must carefully choose the particular crop and use restrictions for which registration is being sought to ensure that the sum of resulting residues will be below the level for which a tolerance can be approved on consumer health grounds.

Because registration is regulation of a prospective activity, much of the analysis of exposures, use levels, benefits, and costs must be based on professional judgment. In many cases, the rigorous analysis of costs and benefits, and the economic and agricultural effects of using various alternative pesticides and pest control practices, arises when a registration renewal is in question or when a cancellation of registration is being considered.

Registration of a pesticide is a process that is carried out over time, comprising several steps of data submission, review by the agency, possible requests for more testing or data, and intermediate decisions before final action. A good deal of submission of raw data is required, some of which is confidential business information. As a consequence, a risk assessment in the pesticides office rarely consists of a single comprehensive document. A pesticide's case is more typically an ongoing and growing file of submissions and analyses, petitions and actions. This poses a challenge for risk communication and proves to result in some difficulty when a pesticide's potential risk comes to be of interest to another part of EPA. In her interview for this project, OPP Deputy Director Penelope Fenner-Crisp noted that sometimes the office feels difficulty in bringing its assessments to the CRAVE workgroup for approval of entry onto the IRIS data base because there are confidential and detailed data that cannot become part of IRIS's public record and yet have great bearing on the case.

The pesticides office has long grappled with the "Delaney Paradox" (NRC, 1987), the incongruous actions that must be taken under the Delaney Clause when pesticide residues become defined as food additives and regulation is moved from §408 of the FFDCA to §409. In 1988, the agency tried to implement a *de minimis* interpretation of the Delaney Clause with respect to such pesticide residues [53 FR 41104]. Basically, the agency argued that risks numerically small enough to be outside regulatory concern even under conservative estimating methods would not be considered as triggering the Delaney prohibition. This position was overturned in 1992 by the 9th Circuit Court [Les

v. Reilly, 968 F.2d. 985 (9th Cir. 1992); Cert. denied, 113 S.Ct. 1361 (1993)], stating that the agency has no discretion to permit use of food additives once a finding of carcinogenicity is made regardless of the degree of risk that may be involved. Thus, the EPA has had no better success at implementing a *de minimis* interpretation than has the FDA. This failure has raised interest in a legislative approach, since it appears that a repeal of the Delaney Clauses in the FFDCA is the only possible avenue to extract the agency from the Delaney paradox.

Unlike most EPA regulatory programs, the Office of Pesticide Programs conducts its own risk assessments, including the steps of hazard identification and dose-response assessment. This practice is made necessary by the chemical-specific nature of OPP's mandate, by the large amount of data the office receives in support of registrations, and by the statutory and court-imposed deadlines that it faces in considering registrations.

The OPP follows the EPA risk assessment guidelines, and most of its practices are equivalent to those used elsewhere in the agency, as outlined in the section on general EPA methods. There are some particular features distinguishing risk assessment in the pesticides office, however, that are noted in the sections that follow.

HAZARD IDENTIFICATION

For hazard identification of carcinogens, there are two different considerations faced by the pesticides office. For agents that do not concentrate sufficiently to trigger regulation under FFDCA §409, the hazard identification scheme is based on the weight of evidence procedure from the EPA *Guidelines for Carcinogen Risk Assessment* [51 FR 33992], as described in the general section on EPA [p.62]. For agents that *do* fall under §409 (and hence under the Delaney Clause), the question of carcinogenicity hinges on the Delaney criterion, whether the agent has been shown to cause cancer when ingested (or otherwise "appropriately" tested) in man or animal. As discussed in the section on FDA, Delaney is triggered by a single study, and does not involve weighing all the evidence.

The pesticides office has considered whether the reference to "appropriate" studies for testing food additives in §409 gives some latitude to require that the positive cancer result comes from the same route as anticipated human exposure, but has never really put this notion to a legal test. FDA interprets the Delaney reference to "cancer" quite literally in that only malignant tumors count as a positive response. At EPA, the cancer guidelines' acceptance in other contexts of benign tumors (technically not "cancer") as evidence contributing to the weight of evidence for carcinogenicity makes overlooking them in the context of Delaney a bit awkward. Benign animal tumors alone constitute only 'limited' evidence of carcinogenicity, but if a bioassay with some malignant tumors shows significant elevations only when the benign tumors are counted in the total incidence (as is EPA practice in other situations), there is the potentially difficult circumstance that a bioassay is counted as positive for non-Delaney purposes, but negative in the context of the Delaney Clause.

If an agent that triggers §409 has a positive cancer bioassay, then the pesticide tolerance must be set at zero for processed foods. In this case, EPA sets the tolerance at zero for the non-processed agricultural commodity as well, since there is no means to prohibit the processing. That is, if a pesticide is not allowed in apple sauce, then it cannot be allowed on apples either.

For purposes other than the Delaney Clause, the pesticides office uses the standard EPA weight of evidence for carcinogenicity. The pesticides program has the power under FIFRA to ask for toxicological testing that it feels is needed to make registration decisions. Thus, the data base is often quite extensive, and there is the possibility of resolving situations of ambiguous or meager data (although delaying the classification while new studies are conducted) that another office might have no choice but to grapple with.

Even though the rules of EPA's current evidence classification scheme are widely seen as being rigid—e.g., two positive animal bioassays always trigger a B2 classification—in fact the application of the criteria involves a good deal of professional judgment, and there is the possibility that different risk assessing organizations may have somewhat varying institutional views on when a marginal study result is to be judged "positive" and on the appropriate disposition of borderline cases regarding the body of evidence as a whole. For example, a study with a response that is at best marginal (and arguably spurious) might be judged positive or negative depending on which particular statistical tests are used, the level of significance chosen and the method for adjusting it for multiple comparisons, the weight given to historically rare tumors in boosting the degree of "biological significance," the weight given to historical fluctuation in incidence among controls, and so on. The criteria for judging whether a maximum tolerated dose has been achieved in a particular bioassay involves toxicological judgment. Although adhering to the same nominal criteria, in comparison with the rest of EPA, the pesticides office is often seen as more likely to doubt marginally positive responses, to question the relevance of benign tumors, and to resolve a borderline case in favor of the lower level of classification (i.e., a C for a B2/C decision or a D for a C/D decision) [Richard Hill interview].

The practice of hazard identification for effects other than cancer is similar in the pesticides office to that elsewhere in the agency. Frequently, the toxicity at issue is neurotoxicity, since many agents exert their pesticidal effect through this mechanism. A great deal of recent attention has been placed in the ability of compounds to act as hormone disrupters. Again, the ability under FIFRA to call for data and testing helps the pesticides office to resolve difficult cases.

DOSE-RESPONSE ANALYSIS

Quantitative assessment of the potency of carcinogens largely follows the usual EPA methods. In extrapolating potencies across species, the pesticides office has begun

(since August of 1994) using the new interagency proposal [57 FR 24152] to base equitoxic doses on scaling by the 3/4-power of body weight. This method is in contrast to the 2/3-power scaling specified in the EPA guidelines and still used elsewhere in the agency, although the forthcoming revision of the EPA guidelines is expected to establish the 3/4-power scaling as a default (as discussed in the general section on EPA [p.76]). Scaling by the newer method results in human risk estimates that are lower by a factor of about 1.5 to 2.

The EPA guidelines specify that "Agents that are judged to be in Group C will generally be regarded as suitable for quantitative risk assessment, but judgments in this regard may be made case-by-case" [51 FR 33996]. The pesticides office has historically been more willing to declare that quantitative analysis of an agent is not appropriate in particular cases. Indeed, for practical purposes the category has essentially been divided into "quantified C's" and "non-quantified C's." In the latter case, risks are interpreted according to the NOAEL/Safety Factor approach. The aforementioned tendency to be skeptical of studies showing marginal responses is reflected here as well.

EXPOSURE

Especially in the case of petitions for registration of new agents, the evaluation of exposures is often prospective. A key factor in the analysis of pesticide exposures is the rate of degradation of the compound in the environment, both on crops and in the air, water, and soil that also receive loadings when pesticides are used. Data on such degradation are an important component of a registration petition.

Another key factor is the proposal for conditions of use. Use levels, frequencies, methods of application, and so on must clearly be sufficient to allow the agent to be effective in controlling pests yet not so high as to prompt health effects to applicators or consumers of the treated products. The petition specifies what limits and conditions of use are proposed for approval, and the task of registration is to determine whether such use is both efficacious and safe.

Exposures to applicators are usually evaluated through standard scenarios depending on the mode of application and the proposed use levels. Estimation of exposures to consumers has two components, the estimation of pesticide residues that remain on agricultural products when encountered by consumers and the level of consumption by the consumer population of different agricultural products. Many of the considerations about estimating such exposures are discussed in the treatment of exposure in the section on the Food and Drug Administration [p.24], which faces a similar task of combining residue information with consumption levels, and the following discussion should be read in conjunction with that treatment. (There are important differences in the analysis of consumer exposures in the pesticides office and at FDA, however, as noted below.)

In the case of food additives intentionally added as part of processing, residues in food products can be known with some confidence. Residues of pesticides on crops are often more difficult to estimate. The conditions of initial application are known because they are specified, but the agent may differentially concentrate in different plant parts, and its rate of degradation in the environment can depend on weather conditions and other factors. Importantly, the interval from harvest to consumption, and the handling that the harvested crop may receive in the interim, may be highly variable. Finally, the mode of preparation of the food by the consumer (including peeling, washing, and cooking) will profoundly affect residues that appear in the food as finally eaten.

The pesticide residues on a particular kind of food, averaged over all sources of that food, will clearly depend on what percentage of the crop was treated with the pesticide and at what levels. FIFRA mandates, however, that the uses approved during registration be set so as to lead to no "unreasonable risk." Once the use of a pesticide at certain levels of application on a certain crop becomes registered, the EPA has no regulatory control over how much of this permitted use will actually occur, from maximal application on the nation's entire crop to occasional local use at reduced application levels. In order to ensure that tolerances will be sufficiently protective in the face of uses that have been declared permissible, the agency must assume (at least initially) that the agent is indeed used up to its permissible level for all permitted uses. That is, in estimating residues, the assumption is made that all crops for which the pesticide is approved are indeed treated at the limit of what is permitted on the label, that the interval between last treatment of the crop and consumption of the agricultural product is that minimally allowable, and that no interim handling (such as washing) to reduce residues has been done.

This initial estimate of residues is intended as a screening tool, defining the limit to residue levels that could occur given the uses of the pesticide that are to be declared legal. Because the intent is that the foodstuffs as presented to the consumer be safe for all modes of consumption, no allowance is made for the diminution of residues by washing, peeling, or cooking. As discussed below, these maximal residues are then crossed with "average" food consumption rates to determine potential exposure to the consumer population. If such hypothetical maximal residues cause no health concerns, the tolerances can be set accordingly. If not, it may be possible to make some more realistic projections of residues. For instance, residue projections can be reduced according to estimates of the percent of the crop that will be treated (usually a high-end estimate from historical data). (This implicitly assumes homogeneous distribution of the food supply over the country; to the extent that particular people consume disproportionally from treated or untreated sources, this assumption biases some individual exposure estimates up and others down.) When it seems warranted, further information may be developed to estimate anticipated residues with less conservatism (based on usual pesticide use patterns and food handling, when these can be known). The aim, however, is still to construct a picture of the upper end of residue levels that might reasonably be anticipated given the uses of the pesticide that are being approved, rather than measures of the actual distribution of residues, as a guard against changes in use patterns within the legal limits, over which there is no further regulatory control.

The exposure to the consumer population depends not only on residue levels on each food type, but also the amount of each food type that is consumed. Like FDA, OPP gets data on food consumption from existing food use surveys. (Some issues regarding the interpretation of such surveys are discussed in the section on FDA on p.26.) The pesticides office examines consumption patterns for 22 standard sub-groups, defined by different combinations of attributes including 5 ethnic groups, 4 broad geographic regions of the country, 4 seasons, age, sex, pregnancy, as well as nursing and non-nursing infants. General population estimates are also made based on lumping all of these sub-groups. For carcinogen assessment, only the general population estimates are developed.

As noted in the section on FDA [p.26], food use surveys represent "snapshots" of consumption patterns over short time spans, usually 3 or 14 days. The variation from person to person in such data represents a combination of variation in food preferences among individuals (some people don't eat spinach) and of day-to-day variation in food choices for each individual (even those who do eat spinach don't do so every day, and may not have happened to do so during the survey period). It is difficult to separate these components, and the assumptions that are made in this regard are quite different at FDA and OPP.

OPP assumes that everyone is a consumer of all food types in the long run, and the variations seen in a survey are presumed to represent the frequencies of use of different food types in the general population. This means that food consumption frequencies are averaged over the whole population (or at least over the whole demographic sub-group), and there is no information on variability among people in the amounts eaten. As a result, an item rarely consumed in the surveys will be assumed to make up a low fraction of everyone's diet, when in fact it might make up a high fraction of the diet of those few people who actually eat it.

It is noteworthy that FDA makes the exact opposite assumption; it assumes that variation in the survey represents variation in individual food preferences, and that the diet a person ate during the survey is typical of his or her diet throughout life. This method makes an (exaggerated) estimate of the distribution of consumption levels across the population. (Although the mean is the same, it has a somewhat different interpretation.)

Both agencies recognize their assumptions as not reflecting the true mix of these two phenomena, day-to-day and person-to-person variation, and both are working on methods to better address the actual features of the distribution of food consumption. To the extent possible, both agencies try to address the problem by using the limited longitudinal data in 14-day studies and other sources, but it remains that they are working toward the middle from different ends of the spectrum.

In the estimation of chronic consumer exposures to pesticides, then, high-end estimates of residues on different food types (which may be very conservative in some

settings) are combined with population average food consumption rates. For agents with linear low-dose potencies (as is assumed for carcinogens) such average consumption rates can give estimates of population risk; to the extent that consumption is uneven, the risks will be higher and lower for high and low consumers, respectively, but these average out in the population impact. For agents with thresholds (as is assumed for most non-cancer toxicities) population average consumption may be anticonservative, since it fails to account for the fact that many people eat more than the average amount of a food type. In the case of foods ever eaten by only a few people, most of those who do use the food consume much more than the population "average," which averages over users and non-users alike. Since a person consuming more than a safe level cannot "borrow" safety from someone not consuming the food at all, failure to account for the distribution of exposures can falsely imply that an exposure has no impact.

Whether the potential conservatism of residue levels outweighs the potential anticonservatism of consumption estimates in the OPP analyses is difficult to say, and will vary from case to case. It is important to understand the difference between OPP's analysis of food consumption and that employed by the FDA, which stresses the top end of the consumption distribution, but has generally less conservative residue estimates that does the OPP approach.

Acute exposures to pesticides, examined for use in judging risks of acute toxic effects, is quite different from the above procedure for chronic exposure. For acute effects, one is concerned for the person who undergoes a peak daily exposure by happening to eat a combination of foods that happen to bear the maximally permitted residues. Tolerances are set so as to be safe to the top end of the single-day total dietary exposures seen in the food surveys, assuming each food type has residues at its tolerance. (Anticipated residue data are used if available, but no correction for percent of crop treated is used, since a given individual on a given day will probably eat food from a single provider, which will either be treated or not treated.) If the pesticide in question is on a single crop, a very high percentile (usually the 99.5th) of the daily consumption is used, with the intent of protecting substantially the whole population of consumers. If the agent appears on several crops, however, it is unlikely that a single person on a single day will be among the very top consumers of each one, and so in compensation a lower percentile (say, the 95th) of the distribution of consumption of each is thought substantially to protect the whole population. In analysis of acute exposures, only 4 demographic groups are considered: infants, children aged 1-6, and males and females aged over 13 years. The top acute exposure levels that are estimated are compared with the NOAELs for toxicities thought to be of concern for acute exposure (e.g., reproductive, developmental, neurotoxicity, and others if appropriate) with an uncertainty factor of 10 or 100 depending on whether the NOAEL is from human or animal studies

The exposure analyses described above, both acute and chronic, are carried out on individual pesticide chemicals, taking no account of residues of other chemicals on the same foodstuffs. It is not general practice in OPP assessments to measure the total exposure to pesticides of all chemical types in the diet. (This practice of assessing

exposures [and risks] attributable to one agent at a time is in common with most regulatory programs, the chief exception being Superfund assessments, as discussed beginning on p.144.) There are exceptions, however, especially at the level of special review (for possible revocation of a registration). For example, triazine pesticides, which are chemically similar and share a mechanism of action, have recently been considered as a group.

The FFDCA mandates, however, that the *cumulative* exposure to an agent from all its sources in the diet be examined when setting tolerances. (Section 409, which, strictly speaking, applies to additives but not pesticide residues on raw commodities, also specifies that this cumulative exposure take into account "pharmacologically related substances," but in practice, the assessments are for one agent at a time.) That is, the tolerance for each agent is designed to protect against the cumulative load of exposure to that pesticide from all food consumption sources. The calculation does not include non-dietary sources, however, which sources may include exposure through contaminated groundwater, dust and other windblown exposure from treated fields reaching surrounding populations, and so on.

RISK CHARACTERIZATION AND REGULATION

As with other Federal regulatory programs, the pesticide regulators have had to grapple with the issue of acceptable risk posed by agents (such as carcinogens) that are assumed to be without a threshold for adverse effects. The Office of Pesticide Programs deals with cancer risks for pesticides that do not fall under the Delaney Clause, that is , for those that do not concentrate beyond tolerances during processing. These are to pose no "unreasonable risk," with the definition of reasonable to include considerations of costs, benefits, and the effects on agriculture and the food supply.

OPP considers three categories of exposure: to consumers, to those occupationally exposed (which in practice focuses on applicators, but also includes farmworkers generally), and the general public exposed via non-dietary means (i.e., through environmental contamination). As with most regulatory programs, there is no written rule or policy regarding the level of risk that must be deemed acceptable, but (also as with most agencies) there is understood unwritten practice that is revealed in the examination of regulatory decisions taken by the agency.

OPP generally tries to ensure that individual risks in all three categories do not exceed 10^{-6} for lifetime exposure. Until recently, the goal for occupational exposures was somewhat higher, closer to 10^{-4}, but this was lowered to match the other categories during the tenure of Assistant Administrator Linda Fisher, and has remained so since. In the case of consumers, the 10^{-6} risk applies to cumulative exposure to the pesticide from all dietary sources, with these estimates usually being based on conservative residue estimates but population average rates of consumption of food types. As noted earlier, it is difficult to determine when this combination is conservative, especially vis-à-vis the high end of levels of consumption of particular foods. For pesticide applicators, the

exposure assumptions are not particularly conservative in terms of exposure per treatment, but there may be assumptions about maximum allowable use of the agent that are not met in reality.

These risk criteria are nominally for individual risk levels. However, the fact that consumer risks are calculated based on consumption levels averaged over the entire population makes these risk calculations apply to the whole population (at least on average, and bearing in mind the conservative residue assumptions). Thus, the criterion really hinges on a kind of population risk measure. High individual cancer risks that result because of high consumption of the affected food products is not captured because of the nature of the exposure analysis.

For non-cancer risks, many of the same considerations apply; high end individual exposures are not captured by the exposure assessment. However, differences in average exposure in each of 22 demographic subgroups are considered.

The consideration of costs and benefits is vaguely specified in the pesticides statutes, but registrations and tolerances are set bearing in mind the balancing of the risks engendered with the costs to agriculture and food prices. As registrants tailor their petitions for which crop treatments are to be approved, limitations on uses, and tolerances, they consider the economic and agricultural benefits to be gained by different combinations of uses that might be approvable. That is, since each approved use leads to projected residues that "use up" some of the total tolerance for cumulative consumption in the diet, those specific uses that are most efficacious and economically favorable to agriculture are more likely to be proposed by the registrant since they will lead to a better market for the pesticide once registered.

The more rigorous analysis of costs, benefits, alternative pesticides that are available and so on tends to be most focused on the cases when an existing registration has been called into question, either through special review or at re-registration.

It should be borne in mind that the setting of tolerances can consider costs and economic considerations only for those pesticides not falling under FFDCA §409, i.e., those agents that do not concentrate sufficiently during processing of the agricultural products. For those that do concentrate, §409 does not allow consideration of costs in tolerance setting. Moreover, this section's Delaney Clause mandates tolerances of zero for agents showing evidence of carcinogenicity, overruling cost and risk considerations altogether.

TOXICS - EPA OFFICE OF POLLUTION PREVENTION AND TOXICS

The EPA's Office of Pollution Prevention and Toxics (OPPT) is a relative newcomer among EPA regulatory programs, having been founded (under the original name of the Office of Toxic Substances) to implement the 1976 Toxic Substances Control Act (TSCA). In addition to the original role as implementer of TSCA, OPPT has been given responsibility for pollution prevention programs, regulation of certain abatement programs (such as that for asbestos), and the administration of the Toxics Release Inventory, mandated under amendments to the Superfund law. The focus of risk assessment in OPPT, however, is under TSCA, and implementation of this statute will be the focus in the present discussion as well.

TSCA AND ITS MANDATES

The **Toxic Substances Control Act** (TSCA, 15 U.S.C.A. §§2601 to 2692) was conceived of as a "gap-filling" statute; Congress recognized that the existing array of environmental legislation covered risk posed by chemicals only under those particular exposure conditions each program was mandated to regulate. Moreover, this regulation was often in reaction to existing pollution, and its efficacy was hampered by lack of information on the chemicals in question. TSCA was passed in 1976 as an attempt to take a comprehensive approach to regulation of toxic substances, stressing properties of the chemical rather than of particular exposures to the chemical, and encouraging the development of information regarding toxic properties and exposures. The aim was to prevent risks from toxic substances that might "fall through the cracks" between other environmental statutes. This cross-cutting role has meant that throughout its history there have been ongoing questions about TSCA's overlap with other environmental statutes.

The provisions of TSCA implement a set of policy statements set out at the beginning of the act [TSCA §2(b)]. First, "adequate data should be developed with respect to the effect of chemical substances and mixtures on health and the environment." Moreover "the development of such data should be the responsibility of those who manufacture and...process such chemical[s]." Second, the government should have adequate authority "to regulate chemical substances...which present an unreasonable risk of injury to health or the environment," including imminent hazards. Finally, exercise of this authority should "not...impede unduly or create unnecessary economic barriers to technological innovation while fulfilling the primary purpose...to assure that...such chemical substances...do not present an unreasonable risk." Section 2(c) goes on to require that "the Administrator [of EPA] shall consider the environmental, economic, and social impact of any action" taken under the act.

Section 4 of TSCA relates to testing and gathering of information on chemicals. It authorizes rulemaking requiring manufacturers to conduct toxicological testing for "carcinogenesis, mutagenesis, teratogenesis, behavioral disorders, cumulative or synergistic effects, and any other effects with may present an unreasonable risk of injury to health or the environment" [§4(b)(2)(A)]. The burden is on EPA to show that such testing is necessary, however. (This is unlike testing mandates under FIFRA or FFDCA, in which the agency can without rulemaking call for all information needed to grant or deny petitions.) The substance must present possibilities of unreasonable risk, "enter the environment in substantial quantities," or be likely to have "substantial human exposure" [§4(a)], all criteria that require the agency to do some preliminary risk assessment. An Interagency Testing Committee is established to set testing priorities. (Through this means, §4 is a vehicle for various Federal regulatory groups to obtain testing mandates, as long as their interests parallel those of EPA.) In practice, testing is done through enforceable negotiated consent agreements ever since a lawsuit challenged the earlier practice of negotiated voluntary testing [NRDC v. EPA, 595 F.Supp. 1255 S.D.N.Y.1984)].

Section 4(f) requires that, if this mandated testing (or any other source of information) indicates "a significant risk of serious or widespread harm to human beings from cancer, gene mutations, or birth defects," the Administrator must initiate rulemaking action under TSCA §§5,6, or 7 to reduce or prevent such risks, or instead to state in the *Federal Register* why such risks are not unreasonable. Section 4(f) was exercised a number of times in the early 1980's, but the result was almost invariably a finding that another agency had jurisdiction, and this provision is little used today.

TSCA makes a distinction between "new" and "existing" chemicals. The latter are those on "a list of each chemical substance which is manufactured or processed in the United States," which §8(b)(1) mandates EPA to "compile, keep current, and publish." Anyone proposing to manufacture a "new" chemical (i.e., one not yet listed) or to undertake a "significant new use" of an existing chemical (in amounts over a volume threshold), must give notice to EPA, along with test data and information bearing on its potential risk. (Much of this submission, especially data on projected uses and formulations, must be kept confidential by OPPT.) EPA then has 90 days to review the case and decide whether to permit manufacture and distribution (the default upon lack of agency action), to suspend manufacture and distribution or restrict use pending the provision of further data (on the grounds that existing data are insufficient to determine whether there will be an unreasonable risk), or (upon finding that an unreasonable risk exists) initiate rulemaking to regulate manufacture or distribution. Once a chemical enters commerce, it is listed on the list of "existing chemicals" and becomes one of them.

Section 6(a) sets out the authority to control hazardous chemical substances and mixtures (be they "new" or "existing"). A rulemaking may be triggered by a finding by the Administrator that "there is a reasonable basis to conclude that the manufacture, processing, distribution in commerce, use, or disposal of a chemical substance...presents or will present an unreasonable risk of injury to health or the environment." The rule

must "protect adequately against such risk using the least burdensome requirements." These may include "prohibiting" or "limiting the amount...which may be manufactured, processed, or distributed in commerce," including limits to particular uses, mandates for handling, labeling, record keeping, as well as "regulating any manner or method of disposal." Section 6 specifies rulemaking procedures, which include an opportunity to petition for an informal hearing at which witnesses can be cross-examined.

In other words, EPA is given rather general authority to seek out and regulate any "unreasonable risk" wherever it may be found, but what might otherwise be sweeping authority is reigned in by the requirement to consider economic and social impact. The act also offers a myriad small checks on this authority in addition to one major one—"If...a risk of injury to health or the environment could be eliminated or reduced to a sufficient extent by actions taken under another Federal law" [§6(c); §9(a)(1)] that other law must be deferred to unless it can be shown to be in the public interest to regulate under TSCA. In practice, this "hand-off" to another regulatory authority almost always happens, and most assessments of risk due to major "existing" chemicals (as opposed to "new" chemicals, as discussed above) are referred to the CPSC, OSHA, or another part of EPA.

Section 8 of TSCA mandates that any manufacturer or distributor "who obtains information which reasonably supports the conclusion that...[a chemical] presents a substantial risk of injury to health or the environment shall immediately inform the Administrator" [§8(e)]. That is, any adverse results seen in testing, occupational hygiene, or epidemiological studies, even those voluntarily and privately conducted, must be reported.

RISK MANDATE

In essence, the Toxic Substances Control Act aims at establishing a system of both public and private vigilance against health and environmental risks from chemicals in commerce that might not be noted or covered by other regulatory authorities. The mandate is to avoid "unreasonable risk of injury to health or the environment," while balancing the benefits of any controls against "unnecessary economic barriers" [§2(b)]. The onus is on EPA to show that unreasonable risk exists, but if it does so, controls are to "protect adequately" against the risk [§6(a)]. In promulgating any such rule, the Administrator must "consider and publish a statement with respect to...the effects...on health and the magnitude of the exposure of human beings,...the effect on the environment,...the benefits of such substance...for various uses and the availability of substitutes..., and...the reasonably ascertainable economic consequences of the rule, after consideration of the effects on the national economy, small business, technological innovation, the environment, and public health" [§6(c)(1)].

IMPLEMENTATION

OPPT sees its goal under TSCA as being to prevent and mitigate risks, and they see this as aided by pursuing voluntary exposure reduction when possible. The aim is to foster a sense of stewardship over potentially hazardous substances on the part of manufacturers, distributors, and users. Compared to other EPA programs, there is more emphasis on negotiation of voluntary solutions aimed primarily at exposure reduction, rather than on promulgation of regulations *per se*.

Risk assessment carried out in support of testing and regulation initiatives is largely as practiced elsewhere in the agency, with some exceptions noted below. Assessment of "new" chemicals under the §5 premanufacture notice (PMN) review is somewhat different, however, owing both to the brief time span and usual paucity of data for such assessments. PMN assessments comprise the bulk of risk assessment activity in OPPT, with over 2000 such assessments being conducted annually.

The very fact that "new" chemicals are newly appearing to manufacturing generally means that there are few toxicity data and virtually no exposure data. Frequently the data do not go beyond the determination of physico-chemical properties, some mutagenicity assays, and perhaps some rudimentary pharmacokinetic tests and limited acute toxicity studies. There are generally no sub-chronic or chronic animal bioassays. Faced with such data, the EPA must make a decision within 90 days whether to allow manufacturing and distribution to go ahead. Section 5 does not ask EPA to pass on the compound's ultimate safety, but only to decide whether its entry into commerce should be delayed until further testing is done.

With thousands of compounds considered annually, it is impossible to require full testing for all agents, and so assessment concentrates on identifying indications that further testing is warranted. A great deal of reliance is put on structure-activity relationships, in which an examination of features of the compound's molecular structure (which is knowledge that is always available for new compounds) and knowledge about the properties of structurally related compounds are used in forming judgments about potential toxic activity. Exposure estimates can be made based on typical use and processing scenarios, perhaps based on analogous compounds with similar properties and uses. Carcinogenicity assessments are sometimes made in which potencies are determined not for substance in question, but rather for chemical analogues that happen to have animal bioassays. (Generally, only agents with an A or B classification are used as such analogues.)

Such assessments are rough screens, designed to flag situations in which further testing should be required. There is no question of such assessments being applied in another context, since the identity of the compound, the existing data, and the assessment of those data are kept confidential, even within the confines of OPPT on a "need-to-know" basis.

In the case of compounds that are being evaluated for regulation under §6 (or for possible further testing under §4) a more usual and rigorous risk assessment process is followed, with public analysis and public documentation. It is this type of analysis that can be compared with procedures in other agencies and EPA programs, and it is such analysis that is discussed further below.

HAZARD IDENTIFICATION

In hazard identification for carcinogens, OPPT application of the EPA guidelines is straightforward. Because they are typically concerned with exposures through several routes of uptake (typically, inhalation, oral, and dermal), presumption of applicability of results across routes is commonplace. Methods for non-cancer hazards are also similar to those used elsewhere in EPA.

DOSE-RESPONSE ASSESSMENT

The quantitative analysis of carcinogen potency is similar to that used elsewhere at EPA, as outlined earlier, with a few small exceptions. Historically, OPPT tended to employ several different dose response models and present the results along with those from the linearized multistage model, which (following the guidelines) was preferred. This process was discontinued in the mid-1980's, but a remnant lives on in the frequent practice of presenting the maximum likelihood estimates (MLE) of the fitted parameters of the multistage model in addition to the usual presentation of the upper bound potency defined by q_1^*. In the case of the assessment of some occupational exposures, risk estimates are high enough that the usual low-dose approximation used elsewhere, *viz.*,

$$R(d) \approx q_1 d \ ,$$

is not sufficiently accurate. In such cases, risks are calculated with the full multistage equation (and parameter values for the upper bound curve), i.e.,

$$P(d) = 1 - \exp - (q_0 + q_1 d + q_2 d^2 + \cdots + q_n d^n) \ ,$$

as defined earlier, in the section on general EPA methodology [p.72]. In these practices, OPPT is similar to OSHA, which also uses the full equation and presents analyses using different models, including both maximum likelihood estimates and upper bounds. (An important difference is that OPPT features the upper bound estimate of the multistage, while OSHA features the MLE.) The similarities reflect similar concerns with the robustness of estimates of high occupational risks that are not far extrapolated from the exposure levels seen to cause tumors in animal studies. (This issue is discussed in the section on OSHA [p.34].)

OPPT uses a different measure of risk over background, namely so-called "additional risk," defined as

$$R'(d) = P(d) - P(0)\,,$$

where $P(0)$ is the modeled risk at zero dose (i.e., the estimated background level of tumors of the type in question), and $P(d)$ is the total risk at dose d. This stands in contrast the definition used elsewhere, namely the so-called "extra" risk above background, given by

$$R(d) = \frac{P(d) - P(0)}{1 - P(0)}\,,$$

as discussed earlier in the general section on EPA methods [p.72]. The reason for this difference in methods is not clear, but it rarely makes an appreciable numerical difference, amounting to only a 10% difference in calculated potency when the tumor background incidence is 10%, and becoming vanishingly small as the background incidence diminishes.

Use of the multistage model requires specification of the degree of the dose polynomial, i.e., the number of fitted q parameters and the highest power of dose allowed to influence the shape of the curve. Historically, the degree has been set at one minus the number of dose groups (including the control group). That is, in analysis of the typical cancer bioassay design of a control group and two different dosed groups, the degree is set to two, and d^2 is the highest power of dose used in the fitting. (The maximum steepness of the curve that can be estimated is a function of the highest power of dose in the model; thus, the historical practice restricts curves to those rising no more steeply than proportional to d^2. The principal departure from this approach among Federal agencies is at CPSC, which tests fits using higher powers of dose, as discussed in the section on that agency [p.47].)

In 1985, OPPT initiated a discussion within the agency questioning the usual approach to choosing the model degree. The issue was whether it was wise to "use up" all the available degrees of freedom with parameter estimates, affecting the utility of goodness-of-fit tests to judge the success of the best-fitting model in describing the data. (Owing to inherent limits on curve shape and the method for parameter optimization, a curve with the number of parameters equal to the degrees of freedom does not necessarily fit the data perfectly, and the issue of the meaning of goodness-of-fit tests remains somewhat statistically problematic.)

The developer of the multistage model, Dr. Kenny Crump, was called in as a consultant, and he suggested an entirely new procedure: that all models from degree one to six be tried, and among those sufficiently fitting the data, the one providing the *lowest* value of potency (q_1^*) be chosen. The rationale was to allow the equation as much flexibility as it could profitably use in defining a fitted curve, with as much steepness as appeared necessary. As revealed by his simulations, if the unknown true curve was

indeed linear at low doses, all of the upper bound estimates were very close to their respective MLEs, which in turn were all very similar to one another. On the other hand, if the true curve were non-linear, all of the upper bounds were marked overestimates, and one might as well choose the lowest one among these. That is, the choice of degree matters only if the resulting upper bounds are overestimates anyway.

It was resolved to adopt Crump's proposed procedure as a matter of EPA standard methodology, but in practice, it is not clear that it is really used with any regularity except by the analysts in OPPT. The use of the new method of degree choice does not make a large difference, but when OPPT updated its assessment of formaldehyde carcinogenicity, the switch to the new method resulted in a drop in estimated potency of about a factor of 2 from what it would have been under the old method. (The formaldehyde data are such as to make the switch more consequential than it would be for more typical data sets.)

In the realm of quantitative non-cancer risk assessment, OPPT uses the usual NOAEL/Safety Factor procedure, as described under the general section on EPA methods. The office has yet to use the benchmark dose approach in its analyses.

EXPOSURE

It is difficult to generalize about exposure assessment in OPPT because of the diversity of situations the office must consider. Exposure assessments for PMN chemical screening analyses are very rudimentary, owing to the lack of data and the fact that the exposures of interest lie in the future, presuming the production of the new chemical is to be allowed. Nonetheless, approximate assessments can be made based on physico-chemical properties of the compound (vapor pressure, for example) and scenarios describing typical handling and usage conditions and settings, sometimes employing data on compounds of similar chemical structure and use. An attempt is made to define "typical" and "upper bound" degrees of potential exposure.

Within the realm of full risk assessments, analyses are diverse because of the all-encompassing scope of TSCA. Unlike other programs that focus on exposure through one medium, TSCA must assess all potential exposures to a chemical that may lead to unreasonable risk, whether they be occupational or to the general population, in manufacturing, distribution, use, or disposal. The range of regulatory options for which post-regulation exposure must be estimated is similarly broad, covering bans, use restrictions, labeling, and handling requirements. Because of the concern for multiple types and routes of exposure, and because of the frequent paucity of data, OPPT carries out a lot of route extrapolation. Again, the aim is to provide central estimates and upper bound estimates of exposure.

Because of the diversity of exposure situations covered, OPPT shares overlapping concerns about a number of different kinds of exposure with other Federal regulatory groups. For example, many of the exposures of concern to OPPT are occupational, an area shared with OSHA. Some of the OPPT standard exposure assumptions are different from those used by OSHA, however. For example, OPPT assumes that a full working

lifetime comprises 40 years, rather than the 45 years assumed by OSHA, and there may be other assumption differences about body weights, inhalation rates, and the like.

The aim of TSCA is mitigation of risks, and so risks are assessed separately for each mode of exposure, even if they occur in the same setting. For example, an industrial use that results in simultaneous dermal and inhalation exposure will have these exposure amounts kept separate in the analysis so that opportunities for exposure reduction can more clearly be seen.

RISK CHARACTERIZATION AND REGULATION

In the analysis of new chemicals, OPPT generally seeks margins of exposure relative to NOAELs of 100. Cancer risks are generally ruled acceptable if they fall below 10^{-4} lifetime individual risks for occupational settings and below 10^{-5} for general population exposures. It should be borne in mind that these are rough criteria given the screening nature of new chemical assessments.

Population risks are considered as well as individual risks. Estimates of the number of people likely to be exposed at different levels are made, and more concern is leveled at widespread exposures with substantial population risk estimates in terms of the number of projected cases of disease.

TSCA is a cost-benefit balancing statute, but a rigorous analysis of costs and benefits is usually only possible for actions contemplated under §6. The much more frequent new chemical analyses and development of risk justifications for test rules employ a more qualitative consideration of costs and benefits.

EPA OFFICE OF AIR AND RADIATION

The regulation of air pollution in the United States has a long and varied history. Since the 1960's, the main vehicle for Federal air pollution control efforts has been the Clear Air Act (CAA), and in many ways the history of application and amendment of this act has traced the shifts in prevailing ideas about pollution, about regulation and the Federal government's role, about the balance of health and economic concerns, and about effective risk management strategies. Of necessity, the present report will focus on how things stand now, but an understanding of present uses of risk assessment requires a brief historical overview. (The discussion that follows draws on accounts by Hattis and Minkowitz [1995] and Findley and Farber [1992].)

Until about the 1950's, air pollution regulation was framed in terms of control of public nuisances; local and state laws aimed to control particular emissions sources that created visible and direct public annoyance. Growing awareness of the chronic health effects of air pollution, and a growing concept of unsullied air as a public resource held in common and in need of public protection, led to various control measures, including the passage of the Clean Air Act in 1963. Initially, the Federal role was largely limited to research, with primary responsibility for control left to the states. It became evident, however, that state control alone was insufficient to deal with cross-boundary movement of polluted air. Moreover, states varied widely in the vigor of their enforcement, prompting fears that states would vie to attract industry by providing lax regulatory environments. The inherent conflict is that the sources of air pollution are local, and hence properly in the realm of state and local regulatory control, but the effects are on the common resource, so that irresponsibility of the few despoils the air for all—a classic "commons" problem, the notion of which was widely coming into the awareness of a public becoming increasingly alarmed about pollution and its chronic health effects.

This initial, desultory phase of air pollution control ended in 1970 with the passage of amendments to the Clean Air Act that for the first time created a strong Federal role. The act was put under the authority of the newly created Environmental Protection Agency. Implementation of pollution control plans, issuance of emissions permits, and enforcement were still the province of the states (as they continue to be today), but these state activities had to accomplish the meeting of Federally mandated and uniform standards for air quality, with provisions to ensure that the states would rigorously enforce the standards.

In these earlier years, the primary focus was on so-called "criteria" pollutants, a few ubiquitous, widely produced and distributed substances that comprise the bulk of emissions and to which everyone is exposed. There are six criteria pollutants: particulate matter (PM), carbon monoxide (CO), sulfur oxides (SO_x), nitrogen oxides (NO_x), ozone (O_3), and lead (Pb). (A seventh, the general class of hydrocarbons, was dropped in 1982.) The concern for these pollutants is primarily (although not exclusively) for

chronic non-cancer health effects. The criteria pollutants are mostly the products of combustion. Although there may be major sources, much of the production is associated with general human activity; thus control requires a comprehensive plan that addresses the spectrum of sources and somehow allocates the permissible emissions among them so that their cumulative environmental loading does not result in exceeding of air quality standards in any particular area. The CAA [§§108-109] mandates the Federal government to set and periodically review health-based standards for air quality for each criteria pollutant (a so-called National Ambient Air Quality Standard, or NAAQS), and for the states to prepare implementation plans (State Implementation Plans, or SIPs) [§110], specifying actions and regulatory activity designed to ensure that all regions come into compliance, with the SIPs requiring EPA approval. Federally mandated uniform performance standards are required for emissions from any newly constructed major sources [§111].

Another, related thrust of the Clean Air Act is the regulation of emissions from mobile sources [§§202-250] principally motor vehicles. Combustion of fuel by motor vehicles is a major source of the criteria pollutants, and so the CAA and its amendments have required ever more stringent limitations on allowable emissions from new vehicles. There is also authority to specify requirements for formulation of gasoline. Motor vehicle emissions regulation is framed primarily in terms of performance standards for exhaust content, and risk considerations generally are somewhat indirect, coming through the contribution of exhaust to the levels of criteria pollutants. Increasingly, however, as questions of fuel formulation grow in prominence, risk analysis of compounds other than criteria pollutants are being undertaken. Nonetheless, mobile source risk assessment will be treated only briefly in this report.

A third major area of the Clean Air Act is the regulation of emission to the atmosphere of toxic chemicals other than the six criteria pollutants [§112]. Early on, this was seen as a secondary provision; control of such "air toxics" was added almost as an afterthought in the zeal that prompted the 1970 amendments. Over time, however, this aspect of the CAA has grown in importance and impact owing to increasing concerns about chemicals in the environment (and in particular, carcinogenic chemicals) and to progress on controlling the more "traditional" criteria pollutants. As noted below, the 1970 amendments called upon the EPA to identify and control hazardous air pollutants so as to achieve "an ample margin of safety". A problem arose with regard to potentially carcinogenic air pollutants; if such agents could not be presumed to have an exposure threshold for their carcinogenic effects, no level of exposure could be declared "safe," certainly not with "an ample margin of safety." The EPA answered this quandary by developing an interpretation of the CAA §112 mandate (borrowing logic from the Clean Water Act) that if no "safe" level could be named, regulation should be based on the best available technology for control of emissions. (Such best available technology—abbreviated BAT—was the most advanced control that most facilities could afford to install without prompting plant closure.) That is, an unachievable health-based standard was transmuted into one based on economic and technical feasibility.

This interpretation was taken during the 1980's in the face of the Supreme Court's 1980 "benzene decision" [Industrial Union Department v. American Petroleum Institute, 448 U.S. 607 (1980)], discussed in the present report's section on OSHA [p.31]. In this decision, OSHA's similar practice under an act that *mandated* consideration of feasibility was struck down on the grounds that significant risk had to be shown, even for carcinogens, before maximal feasible controls could be entertained. For a variety of reasons, including procedural encumbrances, pressure from the Office of Management and Budget to include cost-benefit analysis, and discomfort over the health/feasibility question, only eight chemicals received completed National Emissions Standards for Hazardous Air Pollutants (NESHAPs) between 1970 and 1990. For many compounds, it was difficult to make the risk case look compelling, but even more difficult to dismiss the existence of risk, the only criterion the law seemed to allow.

Matters came to a head when EPA was sued on its regulation of vinyl chloride, a known human carcinogen. The decision [NRDC v. EPA, 824 F.2d. 1146 (D.C. Cir.1987)], generally known as the "vinyl chloride decision," struck down the best available technology approach, stating that the statute was clear that a health-based standard was necessary. The court drew heavily on the Supreme Court benzene decision, noting that "safe" did not mean "risk free," that determinations of risk could be made based on good faith judgments about science even when uncertainty remained, and that conservative methods of risk estimation in the face of that uncertainty were acceptable. A two-phase process was mandated in which an initial determination must be made as to what exposure levels were to be considered to pose no more than "acceptable" levels of risk based solely on health criteria, and a second phase, in which a policy decision could be made as to what was an "ample margin of safety" taking into account the uncertainties and limitations of the risk assessment. Only then could a regulation set a level based on feasibility, diminishing the already determined "safe" level.

By 1989, EPA had developed a policy, based on the vinyl chloride decision's mandated interpretation of the Clean Air Act, for how it would regulate hazardous air pollutants under §112 [54 FR 38044]. This required specifying what would be considered "acceptable risk" (a matter the courts left to the agency). The standard settled upon was to protect "the greatest number of persons possible to an individual lifetime risk level no higher than approximately 1 in 1 million," and to limit "to no higher than approximately 1 in 10 thousand the estimated risk that a person living near a plant would have if he or she were exposed to the maximum pollutant concentration for 70 years" [54 FR 38044]. A second step then defined the exposure limit that would provide "an ample margin of safety," based on factors that could include feasibility, economic impact, and the assumptions and uncertainties inherent in the risk assessment. Specifically, there would be consideration of the number of persons within 50 km of a source exposed within each of a series of individual lifetime risk ranges [Hattis and Minkowitz, 1995]. The EPA set about applying this new policy to the regulation of benzene emissions.

In 1990, however, Congress passed another major set of amendments (the Clean Air Act Amendments of 1990, or CAAA), that, among other changes, completely revamped the air toxics provisions in §112 and sets the stage for the application of risk

assessment as it now stands. A list of 189 hazardous air pollutants to be regulated is written into the act [§112(b)]. All new or existing sources of these pollutants are to require the use of maximum available control technology (MACT), which is judged to be "the best of the best" for new sources and at the top end of current emissions control performance for existing sources. Section 112(f) requires the EPA to develop "methods of calculating the risk to public health remaining, or likely to remain,...after application of [MACT] standards." A report to Congress on the adequacy of MACT standards is to be prepared; if Congress does not act on these recommendations, and if the technological standards "do not reduce lifetime excess cancer risks to the individual most exposed to emissions from a source...to less than one in a million," standards are to be promulgated meeting the criteria of §112 regulation *before* the 1990 Clean Air Act Amendments, that is, "to provide an ample margin of safety to protect the public health." (Standards are also to prevent environmental effects, but this consideration, unlike the human health criterion, allows consideration of costs.)

In other words, the former health-based criterion for standards under the previous version of §112, to be pursued chemical by chemical, is replaced with a criterion that is primarily technology based, mandating the most effective control available for a specified list of chemicals. If this technological fix is found to lower risks insufficiently, further regulation may be pursued to control this so-called "residual risk" under the policy worked out following the vinyl chloride decision. For cancer as a health effect, the residual risk "trigger" for the consideration of this further regulation is when the maximally exposed individual (MEI) to a plant's emissions suffers a 10^{-6} lifetime cancer risk. It is important to note that this risk level is a trigger for further regulatory consideration, not a statement of the degree of protection such a standard should afford. That degree of protection is presumably given by the post-vinyl chloride policy: a 10^{-4} risk to the MEI and protection of "the greatest number of persons possible" to a level of 10^{-6}.

Although it is not completely clear from the language of the amended §112, the MEI that could trigger further action is presumably an estimate of the highest exposure to an actual person, in contrast to the previous policy's hypothetical individual sitting at the plant's fenceline for 70 years. Indeed, many of the specific mandates and implementations of provisions of the 1990 Clean Air Act Amendments, having not yet come due, have yet to be put into practice. It is therefore difficult to speak confidently about risk assessment methodology that will be used in these circumstances. Section 112(o) mandated the National Academy of Sciences critique of EPA risk assessment that resulted in the report *Science and Judgment in Risk Assessment* (NRC, 1994), the recommendations of which must be considered (but not necessarily adopted) in a revision of the agency's carcinogen risk assessment guidelines, which must take place before any residual risk determinations occur. As discussed in the general section on EPA, this revision is in progress, but a proposal for new guidelines has yet to be released for public comment.

THE CAA AND ITS MANDATES

Federal regulation of air pollution is accomplished largely under the **Clean Air Act** (CAA, 42 U.S.C.A. §§7401 to 7671q). The act has as its declared purpose "to protect and enhance the quality of the Nation's air resources so as to promote the public health and welfare and the productive capacity of its population" [§101(b)(1)]. The Clean Air Act is large and complex, and all of its provisions cannot be covered here. The various regulatory programs administered by the Office of Air and Radiation (including some under laws other than the CAA), their purposes and scopes, the key statutory language, and key elements of their risk assessment methodology, are summarized in Table 5. This table was produced by the Office of Air and Radiation itself, and is reproduced directly with permission. The discussion that follows will focus on the National Ambient Air Quality Standards Program (i.e., the "criteria pollutants") and the Hazardous Air Pollutant Program (i.e., the "air toxics").

The basic structure of regulation under the CAA is that of Federal standards-setting and state implementation and enforcement. The Federal standards are of two basic kinds: standards for air quality and standards for the performance of pollutant sources in terms of allowable emissions. Standards for air quality specify uniform national definitions of what constitutes acceptably clean air, and regulatory programs (much of which occur at the state level with EPA oversight) covering the spectrum of sources of the pollutant by a variety of means are then aimed at achieving air quality at least up to those standards. Performance standards for sources are aimed at establishing uniform national limits on the emissions from particular kinds of sources, including motor vehicles (mobile sources) and stationary sources. (For some purposes, the CAA distinguishes among "major" and "minor" sources based on amounts of emissions, and on "point" and "area" sources based on whether the emissions come from a specific, identifiable facility or from more general human activity no easily localized to a few geographic coordinates.)

Sections 108 and 109 of the CAA calls for the development of air quality criteria for the widespread "criteria pollutants." (The criteria pollutants are not named in the statute, but are those with "emissions which...may reasonably be anticipated to endanger the public health or welfare...[and] result...from numerous or diverse mobile or stationary sources" [§108(a)(1)]. Over time, lead has been added to the list and hydrocarbons dropped.) The criteria "shall accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health and welfare" [§108(2)]. So-called "primary" ambient air quality standards are to be standards which "allowing for an adequate margin of safety, are requisite to protect the public health" [§109(b)(1)]. (There are also "secondary" standards that consider non-health effects.) Legislative history has led the "ample margin of safety" mandate to be interpreted as requiring protection of most of the population, including sensitive population groups (e.g., asthmatics, the elderly) but not the most exposed individual or the most sensitive member of a sensitive group. These are to be purely health-based criteria, and are not dependent on costs or technical feasibility.

It is up to the states to provide plans for controlling pollution so as to attain these National Ambient Air Quality Standards (or NAAQSs); section 110 calls on each state to submit to the EPA for approval "a plan which provides for implementation, maintenance, and enforcement of such primary standard in each air quality control region (or portion thereof) within such State" [§110(a)(1)]. Such State Implementation Plans (SIPs) are to include "enforceable emission limitations and other control measures...(including economic incentives such as fees, marketable permits, and auctions of emissions rights)...as may be necessary" [§110(a)(2)(A)] and must provide for monitoring and enforcement. Section 111 provides for Federal standards of performance for new sources of criteria pollutants "which may reasonably be anticipated to endanger the public health or welfare." Sections 160-169B provide for the prevention of significant deterioration of air quality in regions that are already in attainment of the NAAQSs.

Mobile source emissions are addressed in §202; emissions standards for new motor vehicles may be set for "any air pollutant...which may reasonably be anticipated to endanger public health or welfare" [§202(a)(1)]. Although the main concern has been motor vehicles as a source of criteria pollutants, mobile source toxics are also addressed in §202(l), which calls for study of "emissions that pose the greatest risk to human health or about which significant uncertainties remain" and calls for standards for these, including explicit requirements for regulation of benzene and formaldehyde. Fuel formulation may be regulated under §211, and manufacturers of additives may be required to conduct "tests to determine potential public health effects...including...carcinogenic, teratogenic, and mutagenic effects." Such regulations must consider technical and economic feasibility.

Air toxics are regulated under CAA §112. The amendments of 1990 added a list of 189 compounds designated as hazardous air pollutants [§112(b)]. Chemicals may be added to this list by rule if found to "present...a threat of adverse human health effects." Compounds may be deleted from the list by petition if "adequate data" determine that "emissions, ambient concentrations, bioaccumulation or deposition of the substance may not reasonably be anticipated to cause any adverse effects to human health or adverse environmental effects" [§112(b)(3)(C)]. The EPA must build and maintain a list of the principal areas sources and of "major sources" of these pollutants (i.e., those emitting more than 10 tons/year of any one listed chemical or 25 tons/year of any combination).

A category of sources may be delisted if no source has sufficient emissions to "cause a lifetime risk of cancer greater than one in one million to the individual in the population who is most exposed," or in the case of effects other than cancer, does not "exceed a level which is adequate to protect the public health with an ample margin of safety" and presents no environmental effects [§112(c)(9)(B)]. (This is known as the provision for *de minimis* delisting.)

Otherwise, EPA must promulgate for the categories of "new or existing sources" regulations establishing emissions standards that "require the maximum degree of reduction in emissions of the hazardous air pollutant...(including prohibition of emissions, where achievable) that..., taking into consideration the cost...is achievable"

[§112(d)(2)]. This is the so-called "MACT standard," for maximally achievable control technology. (For new sources, MACT consists of the performance of the best controlled existing source in the same category of sources; for existing sources, less stringent requirements are allowed, but must be as good as controls at the top 12% of sources in the category.)

Thus, §112 mandates that emissions of compounds on its specified list be controlled to the extent feasible on technical and economic grounds, regardless of the risk they may pose (excepting the *de minimis* delisting). Section 112(f) calls for the examination of risks that may remain after such technical controls are in effect; EPA must develop methodology to estimate such "residual risk" and recommend legislation to address any such risk that may be found. If Congress does not act on this recommendation, the EPA must promulgate emissions standards "with an ample margin of safety to protect the public health." That is, if residual risks exist after MACT standards are in effect, there is a fallback to the pre-1990 basis for air toxics regulation. In particular, the promulgation of such standards is triggered if MACT controls "do not reduce lifetime excess cancer risks to the individual most exposed to emissions from a source...to less than one in one million" [§112(f)(2)(A)]. As noted earlier, the standards adopted need not protect this maximally exposed individual to the 10^{-6} level—the criterion is the "acceptable risk" policy developed after the vinyl chloride decision [54 FR 38044]—but the existence of a 10^{-6} risk triggers the consideration of residual risk regulation.

Physical changes or changes in the method of operation at an existing facility that increase emissions can make it become a "new" source, and hence subject to stricter MACT standards. Section 112(g), however, exempts increases that "...will be offset by an equal or greater decrease in the quantity of emissions of another hazardous air pollutant...which is deemed more hazardous." (How practically to interpret this "offset" criterion has proved somewhat difficult).

Thus, despite the fact that the 1990 amendment of §112 was designed to reduce the role of risk assessment in air toxics regulation (and the consequent questions and delays as uncertainties in those assessments are debated), there are several places where risk assessment is called for in evaluating the technology-based controls. These include (1) the listing and delisting of hazardous air pollutants, which depends on whether a chemical may "present...a threat of adverse human health effects;" (2) the *de minimis* delisting of source categories, which requires less than a 10^{-6} risk to the MEI; (3) the triggering of post-MACT standards to address residual risk, which also requires less than a 10^{-6} risk to the MEI; and (4) the offset trading of one pollutant for another based on whether the increased emission is "more hazardous." Of these, the third (the residual risk determinations) is the one based primarily on EPA initiative, but it is one that will require extensive analysis, since each source of each hazardous air pollutant should in principle be evaluated at a level of detail such that the individual near each source at highest risk can be characterized. Two additional uses of risk assessment under §112 also allow consideration of economic and technical feasibility: the agency's analyses underlying "area source findings" for setting of MACT or GACT (Generally Available

Control Technology) standards for *area* sources under §112(d)(5), and those supporting an EPA decision to require a MACT standard stricter than the best 12% of existing sources under §112(d)(3).

IMPLEMENTATION FOR CRITERIA POLLUTANTS

The National Ambient Air Quality Standards (NAAQSs) must be reviewed every five years, an extensive process given the amount of research and data collection that has been done on the ubiquitous criteria air pollutants. The process involves many steps, with several opportunities for review and public comment (Padgett and Richmond, 1983). The statute specifies that standards be set on health criteria alone. (This interpretation has been challenged, but upheld, e.g., Lead Industries, Inc. v. EPA 647 F.2d. 1130 (DC Cir., 1980), cert. den. 101 S.Ct. 631 (1980).)

"Criteria Documents" containing compilations and reviews of the scientific evidence on health effects are prepared by the EPA Office of Research and Development. These documents are not themselves assessments, but rather critical examinations of the entire relevant scientific database on health effects associated with the pollutant. Criteria documents are reviewed by the Clean Air Scientific Advisory Committee (CASAC), a subcommittee of EPA's Science Advisory Board. When such a document is substantially complete, a "staff paper" is prepared by the air office's Office of Air Quality Planning and Standards. A staff paper develops interpretations of the key studies and identifies critical elements in the standard-setting process, bridging the gap between the compilation and review of scientific studies in the criteria document and the judgments required by the EPA Administrator in setting the ambient air standards (Padgett and Richmond, 1983). The staff paper summarizes the important findings of the criteria document and other documentation, and develops possible rationales for the choices among different standards that could be set, evaluating the evidence for judging the performance of candidate standards regarding health protection. (In essence, this is the non-cancer risk assessment of NAAQS development; differences in the assessment from the usual non-cancer assessment practices are discussed in the dose-response section below.) Staff papers also undergo CASAC review and public comment. Recommendations for possible criteria specifications are developed, along with a regulatory impact analysis, estimating the costs and of meeting alternative proposals for the standard. (The costs cannot be used in setting the standard, but they are useful in gauging the impact, and are necessary under current Executive Orders.) Standards are proposed, subjected to public comment, and then the rule specifying the final standards is promulgated.

NAAQS standards are framed in terms of air concentrations and averaging times over which those concentrations are determined. That is, they specify limits to concentration-time combinations that should not be exceeded when ambient air that is in compliance with the standard is monitored. The averaging time varies among pollutants in a way that depends on the nature of the health effects critically at issue. For example,

the current NAAQS for ozone is 0.12 ppm averaged over one hour, while that for nitrogen dioxide is 0.053 ppm averaged over one year.

State Implementation Plans comprise strategies for how each state proposes to limit, control, or otherwise regulate the various emissions so as to accomplish compliance with the ambient air standards. (Compliance is judged by a defined low frequency with which monitoring shows the standards to be exceeded.) The States have some flexibility in how they may propose to execute their responsibility to ensure compliance with the NAAQSs; there are many sources of emissions, and the States can decide how to apply and focus their regulatory efforts, in effect allocating among emitters the shares of total emissions that are permissible given the need to limit the cumulative effect on air quality. Of course, there are Federal uniform source performance standards that each source must obey.

There is no requirement for States to conduct risk assessment or exposure modeling in the course of developing their implementation plans; compliance is framed entirely in terms of ambient air quality, not on exposure or risk. Nonetheless, many States may do risk assessments to help gauge the relative health impacts on their populations of various strategies for achieving compliance.

IMPLEMENTATION FOR AIR TOXICS

Unlike the long, established history of regulation of criteria pollutants, the regulation of air toxics has been markedly recast by the Clean Air Act Amendments of 1990 and their changes to §112. As a result, much of the actual implementation is in the future. The amendments called for a rethinking of EPA's risk assessment methodology before the residual risk determinations are made. Section 112(o) specified a critique of EPA's methods by the National Academy of Sciences, which resulted in the report, *Science and Judgment in Risk Assessment* (NRC, 1994). The EPA must revise its *Guidelines for Carcinogen Risk Assessment* [51 FR 33992], taking into consideration the NAS recommendations, before residual risk determination can take place [§112(o)(7)]. This revision is in progress, but no proposal has yet appeared, as discussed in the general section on EPA.

There are other methods to be worked out as well. As noted above, several air toxics provisions require estimating risks to the maximally exposed individual near a source. How such a person's exposure is to be estimated remains to be determined, and is part of the development of methodology that EPA must undertake under §112(f)(1). A lifetime risk projection requires a lifetime exposure analysis. Presumably, the exposures nearby residents suffered *before* the MACT-based regulation was in place (the adequacy of which standard is being tested) are not counted. (Otherwise, the MEI is likely to be someone who lived near the plant for a long time before MACT regulation lowered emissions; such an MEI would be irrelevant to the intended key question of the adequacy of the MACT standards.) Thus, the exposures in question are primarily future exposures, and they must be estimated using projections of demographics, behavior patterns, and the

meteorological data that drive fate and transport models indicating the patterns of environmental dispersion of emissions. (That is, in an important sense, the exposure estimates still must be hypothetical.) To what degree these projections have to be site-specific is not clear at present, nor is it evident how the most-exposed individual among these projections of future exposures will be defined.

The "offsets" provision has also proved difficult to implement. Under §112(g), an increase in an emission can be traded for an equal or greater decrease in a "more hazardous" pollutant. Efforts to define practical criteria for such trading have run into controversies such as whether agents presumed to be without effect thresholds (basically, carcinogens) should be tradable against those with thresholds (currently, they are not), whether the trade should be on a weigh-for-weight basis or on a risk-for-risk basis (i.e., allowing a small decrease in a very potent agent to offset a large increase in a less potent one, as long as total risk does not increase, which is the current option), and whether the degree of hazard should be modified by the persistence of the compounds in the environment once emitted (currently, they are not).

As noted previously, the language of §112(f) seems to require that residual risk levels be estimated for every source for every hazardous air pollutant to a degree of detail sufficient to characterize the risk to the maximally exposed individual, a task that would seem to require a good deal of case-specific analysis and detailed local data on populations, their movements and habits, as well as local meteorological data and physiographic data for fate and transport modeling. (This question is discussed further under the exposure assessment section.) More generally, the practical questions of how to implement the residual risk provisions are being worked on, and are reported on in an appendix to *Science and Judgment in Risk Assessment* (NRC, 1994).

HAZARD IDENTIFICATION

The question of hazard identification applies mainly to air toxics. Before the 1990 amendments, hazard identification was similar in the air office as elsewhere at EPA. The 1990 amendments, however, have defined hazardous air pollutants as the compounds explicitly named in a 189-member list written into the law [§112(b)]. This list was compiled in large part by combining existing lists of "toxic" compounds maintained by various regulatory agencies, so ultimately most of the §112(b) list members gained membership through the more usual hazard identification procedures that led them to be included on previous lists. Nonetheless, the §112(b) list is heterogeneous, containing compounds that are known or suspect carcinogens as well as agents known to cause only effects other than cancer.

Several air toxics provisions in the CAA apply specifically to carcinogens, and so it is still important to determine which agents listed as hazardous air pollutants should be treated as carcinogens. These provisions include the 10^{-6} lifetime cancer risk criteria for *de minimis* delisting emissions sources, the similar cancer risk level that triggers post-MACT standards, the provisions for offset trading of pollutants, and others. The

statutory language on residual risk refers specifically to "pollutants...classified as known, probable or possible human carcinogens" [§112(f)(2)(A)], presumably a reference to the designations applied to the weight-of-evidence categories in the current *EPA Guidelines for Carcinogen Risk Assessment* [51 FR 33992], although they could also be taken to apply to the categories in the evidence evaluation scheme of IARC.

Section 112(o) requires EPA to revise its cancer guidelines and prohibits residual risk determination until this is done. Presumably, then, residual risk determination is to be done using the revised guidelines. It is perhaps noteworthy then that, at least according to current indications, the forthcoming EPA proposal for guidelines revision will not continue the "known, probable, and possible" categories of weight of evidence. Many agents now characterized as "possible" human carcinogens (i.e., in current EPA group C) will likely be lumped in with lesser agents in a new category for which human carcinogenicity "cannot be determined." How such agents (many of which will still be called "possibly carcinogenic to humans" by IARC) will be treated by the carcinogen-specific provisions of the CAA is not at present clear.

For the criteria pollutants, the "hazard identification" part of the analysis is essentially the air quality criteria document, mentioned above in the description of the development of standards. Criteria pollutants are unique in the kinds and amounts of data generally available for this analysis, a result of their importance, ubiquity, and the long history of study focused on the small number of agents . Generally, there are abundant human data, both epidemiological and experimental, for responses at directly relevant exposure levels. Animal data are also used, especially for very long-term chronic effects.

The criteria document examines all of the toxicity endpoints of concern. Unlike hazard identification elsewhere, no one toxic endpoint is defined as the "critical" one, in the subsequent quantitative analysis, all endpoints with adequate data are examined. Moreover, the definition of what responses comprise "adverse" outcomes is left to the risk manager; potential responses and the exposures that prompt them are described, but no decision is made about adversity or acceptability of particular responses.

DOSE-RESPONSE ANALYSIS

Dose-response analysis for air toxics has in the past been done largely through Health Assessment Documents produced by the Office of Research and Development for the air office, according to the methods discussed in the earlier general section on EPA. Carcinogen potency calculations for *de minimis* delisting and residual risk determination will in the future have to be done under the revised carcinogen assessment guidelines. As with hazard identification, there are CAA implementation issues that will have to be clarified if the new EPA guidelines contain anticipated revisions of dose-response analysis. Under the new guidelines as currently anticipated, there will be hazardous air pollutants with hazard classifications of "known/likely" human carcinogens and yet with quantitative analysis consisting solely of the so-called "non-linear" method—a

determination of an ED_{10} and uncertainty factors (where the ED_{10} is the estimated dose leading to a 10% tumor incidence over background). That is, there will be agents designated as carcinogens which have no means for calculating low-dose risks, even as upper bounds, according to the principles of the revised guidelines. How residual risk questions under the CAA will be handled for such cases is not clear at present.

Non-cancer risk questions for hazardous air pollutants are treated with the usual NOAEL/Safety Factor methodology. Although this method usually avoids route extrapolation, the air office may under specific circumstances (and with consultation with the Office of Research and Development) use RfDs from non-inhalation exposures when RfCs are not available.

In the realm of the criteria pollutants the principal concern is for non-cancer effects. These effects receive quantitative analysis that is very different from the usual NOAEL/Safety Factor approach used for non-cancer effects in most other settings. This is due to the relative abundance of data usually available, much of it on humans at directly relevant exposure levels. The principal issue is characterizing exposure-response patterns among variable humans in the range of observation. There is relatively little need for extrapolation across species or to low doses; the main extrapolation is from study populations of limited size to the national population as a whole (including its sensitive sub-populations).

Everyone is exposed to the criteria pollutants to some degree. Therefore, there are no real "control" populations, only lower and higher exposures. Moreover, the effects of concern (mostly but not exclusively respiratory effects) have an important background rate in the general population even among those with low exposure levels. That is, even though the health effects may in principle have a threshold, the tolerable exposure will vary greatly among people, and it appears that at least some members of the general population may have their thresholds exceeded at or near the lowest exposure levels. Usually, then, exposure-response curves are modeled as though they have no threshold, a practice quite different from that applied to non-cancer effects elsewhere.

Generally, then, exposure-response relationships are characterized without any "upper bound" methods. There is little or no extrapolation, and no use of thresholds or safety factors. As a result, in contrast to other risk assessment situations, there is little conservatism in quantitative analysis of the effects of criteria pollutants. Several varieties of Bayesian and probabilistic methods are used to characterize uncertainty in the fitted exposure-response relationships. A good deal of attention is given to characterization of exposure-time relationships, given the potential importance of temporal variation in criteria pollutant concentrations to the engendered effects.

EXPOSURE

For the criteria pollutants, the standards are *set* based on considerations of health protection, but they are *specified* in terms of allowable ambient air concentrations over

specified averaging times. That is, the criteria are set such that, if the ambient concentrations are adhered to, human exposures will be such that an acceptable degree of health protection will be achieved. Protection is sought for the substantial part of the national population, including sensitive subpopulations, and a single ambient air standard must prevail throughout.

For an area in compliance with the ambient air standard, the actual concentrations will generally be less than the standard (which defines a maximum not to be exceeded) and will vary in time (including daily, seasonally, and on other time scales) and in space. The human population is also spread unevenly geographically and will vary in habits and activities, including both indoor and outdoor activities. The exposure models that are used for the criteria pollutants attempt to characterize these variations to arrive at an overall population distribution of exposure levels that would be expected under a particular ambient air standard, presuming an area is marginally in compliance. (For comparison, distributions are also developed for non-complaint exposures as they may currently exist.) Concentrations may be estimated from ambient air monitors. Monte Carlo methods are used to characterize the temporal and spatial variation, including variation in activities and locations of the exposed people. Characterizations of the variation in exposure are made for the general population and for sensitive or at-risk subpopulations. Because of the long history of exposure analysis of criteria pollutants, such exposure modeling has been continually improved and expanded, and the models are now quite sophisticated and rich in data, with capabilities well beyond models used in other situations that do not have the benefit of decades of experience and application. The exposure estimates are designed to be unbiased estimates without any built-in conservatism.

The aim is that such distributions of exposure can be combined with the dose-response analysis to project an expected number of incidents of the relevant health endpoints in the actual population that might occur if compliance to a particular ambient standard is just achieved. The standard can then be chosen such that the health effect levels are acceptable, i.e., that they protect sensitive populations, but not necessarily the most sensitive or exposed individuals.

Monitoring for compliance with NAAQSs is the responsibility of the states. Clearly, the placement and operation of monitoring stations is key to determining whether a large area is in overall compliance; a monitoring program that misses local hot-spots, for instance, will underestimate the top end of the exposure distribution. Standards are set on the assumption that monitors will detect important lack of overall compliance, and there are standards for monitoring programs to ensure that this is the case.

In the realm of air toxics, the historical interest during exposure assessment has been to characterize the high end of exposures, both as currently experienced (in judging whether regulation is warranted) and after emissions standards have been set (in judging how to set those standards). The high end focus arose from the mandate to protect the public health with an ample margin of safety. Under the post-vinyl chloride policy on

the meaning of this mandate, there was a focus on both a hypothetical maximally exposed individual (who should have a risk less than 10^{-4}) and on the numbers of people exposed to various levels of risk between 10^{-4} and 10^{-6}. Thus, distributions of exposures in the population became increasingly important. The maximally exposed individual under this policy is hypothetical, consisting of a person experiencing the maximum fenceline exposure for a full 70 year lifetime. (The idea of the policy is not that this person surely exists, but that it should be possible for someone who wishes to have this exposure to do so if the emissions were to be declared "safe" under the meaning of the policy.)

Under the new §112, the exposure question has changed somewhat. The focus is on residual risk after maximal technical controls have been applied. The initial determination to be made is for the maximally exposed individual, whose lifetime cancer risk when exceeding 10^{-6} triggers further regulatory consideration. The intent here seems to be to define the actual most exposed person in the population, rather than a hypothetical fenceline sitter. Given the number of hazardous air pollutants (189) and the number of source categories (currently 174), and the many sources within each category, determining the MEI for each pollutant for each source is not a practical undertaking. The air office is developing a tiered approach in which screening analyses with conservative default values for key determinants of exposure are used for initial screening assessments, which will be sufficient in cases where no concern for a high MEI risk is evident. As needed, more site-specific data and realistic modeling assumptions are to be used, so that any critical determinations of MEI exposure can be based on more sophisticated modeling. Even so, there are questions to be answered regarding the ultimate level of detail and need to use site-specific data on physiography, meteorology.

As discussed above under implementation, even though the focus is on "actual" rather than "hypothetical" exposures, the analysis must be of future lifetime exposures, which must be hypothetical in an important sense. The approach being developed is based on Monte Carlo simulation expressing the variability to be expected in the key determinants of exposure as a way of characterizing a projected population distribution of exposure after MACT standards.

Exposures in the air toxics program have generally been estimated using a general purpose model largely based on fate and transport considerations for stack emissions. This Human Exposure Model (HEM) has historically been used with fixed values of the parameters, often at conservatively fixed levels. The model is being modified to facilitate the incorporation of variability in parameters and for the characterization of uncertainty in exposure projections. The projections of the amount and pattern of dispersion of emissions around sources must be coupled with data on local distribution of population. Geographical information systems are being investigated as a tool for compiling the necessary local data on population, meteorology, and other site-specific factors.

RISK CHARACTERIZATION AND REGULATION

For criteria pollutants, standards are set by using a complex characterization of the distribution of exposure levels in the population that would be expected under a specified air quality criterion. When combined with the exposure-response relationships, this gives a projection of the number of health effects incidents to be expected in the exposed population. Both the exposure and dose-response components are estimated based on extensive data; they require little extrapolation and few default assumptions, and the estimates of health impact are thus characterized as unbiased estimates without added conservatism. Point estimates rather than "upper bounds" are used. Ranges of risk are estimated corresponding to the experience of sensitive groups.

The risk mandate for protection of public health with an adequate margin of safety is accomplished by setting air quality criteria such that most of the population is protected, including sensitive sub-groups and highly exposed individuals, but not necessarily the most sensitive or most exposed person. There is no fixed level of acceptable risk, which depends on the nature of the health effect in question, the size of the group potentially affected, and the degree of uncertainty about effects and exposure. These decisions are prohibited from considering costs and feasibility.

Although the effects in question are non-cancer health effects, they are generally held not to display a practical threshold exposure for effects. The methods recognize that even quite protective standards do not banish the possibility of some few people being affected. In this way, the situation is similar to that of carcinogens, where "safety" cannot be absolute, and so a reasonable degree of protection must be defined. For criteria pollutants, the risk characterization focuses on population risk, that is on the health impact on the population as a whole, recognizing that that impact is most likely to appear among the most sensitive and most exposed. There is no real individual risk criterion.

In the analysis leading up to the development of a proposed ambient air quality standard, an analysis may be done of the effects that would be expected if the whole population were exposed to air just at the limit of the standard. Although this is not the

primary decision criterion, such an analysis provides an idea of potential impact if all the air were indeed as polluted as is being allowed. This situation is unlikely to occur in practice in a compliant area, since the air quality criteria represent the allowable maximum in what is always in reality a variable level of air quality. (It is interesting to compare the minor role this analysis plays for criteria pollutants to the major role that a similar analysis plays in the regulation of pesticide residues, as discussed in the section on the pesticides office. In that case, the regulatory decision is made on an analysis presuming that all foods contain their maximally allowed residues, even though a distribution with mostly lesser values is likely to be true. The chief difference, of course, is that pesticide residues are more readily manipulated up to their allowable level than is ambient air quality.)

In the case of air toxics, the application of analysis as now being formulated to regulatory decisions is still in the future, and so it is difficult to characterize with confidence. The presumption is that for most sources of most hazardous air pollutants, the maximally available control technology will be sufficient, and further regulation not needed. The residual risk trigger of a 10^{-6} lifetime cancer risk to the most exposed actual individual near a source is a quite stringent criterion, however. Actual regulations of residual risk will be made under the criteria prevailing before the 1990 amendments, that is, the criteria mandated by the D.C. District Court decision on vinyl chloride. These criteria have an individual risk component, that an individual exposed to the maximum fenceline concentration for 70 years should not have a risk exceeding 10^{-4}. They also have a population risk component, that as few people as possible should have a risk greater than 10^{-6}. The 10^{-4} level is the policy definition of "safe," fulfilling the mandate for a regulation that "protects the public health." It is intended that this level of safety be guaranteed even to someone who chooses to fulfill the fenceline exposure scenario, whether or not someone actually does so. The aim to protect as many people as possible from the 10^{-6} risk level is interpreted as the provision of an "ample margin of safety" as provided for in the CAA. In the case of non-cancer effects, it is presumed that exposures below the reference concentration (RfC) fulfill both the mandate for safety and for an ample margin of safety. Given the amount and site-specific detail of exposure analysis required to trigger post-MACT regulation, it is likely that the exposure assessments for such regulations will be much less conservative and "worst-case" than may have been the case prior to 1990. Even though the regulatory criteria are nominally the same, the risk outcome and the stringency of regulation may end up being somewhat different.

EPA OFFICE OF WATER

Regulation of water pollutants is carried out by EPA's Office of Water (OW). The Office of Water administers two major statutes, the Federal Water Pollution Control Act (better known as the Clean Water Act or CWA) and the Public Health Service Act (better known as the Safe Drinking Water Act, or SDWA). The Clean Water Act has as its goal to maintain and improve the cleanliness and biological integrity of the nation's waters, including lakes, rivers, and navigable waters. The aim is to make these waters "fishable and swimable." In many ways, the nature of the pollution problem and the nature of the statutory approach parallel that of the Clean Air Act, discussed in an earlier section [p.105]; the nation's waters constitute a broadly distributed common resource the quality of which is impinged upon by the activities of many local sources of contamination. Each source of effluent is not solely responsible for the resulting water quality, but the collective burden of discharges may result in unacceptable deterioration of the resource as a whole. The regulatory approach is the promulgation of nationwide uniform criteria defining the degree of water quality that is compatible with intended uses and states of different water bodies. (The criteria are health-based, but they are not rules, and are themselves unenforceable.) These water quality criteria are coupled with enforceable technology-based standards for allowable discharges from point sources, which (also like the Clean Air Act) are implemented through permitting regulations by the states. It is the responsibility each state to conduct regulation of discharges such that the applicable water quality criteria are met for the state's waters.

The CWA distinguishes "conventional" pollutants from "toxic" pollutants. The former are largely those associated with discharge of sewage and nutrients, such as fecal coliform bacteria, suspended solids, and sources of biological oxygen demand. In some ways, they are analogous to the criteria air pollutants, the inevitable, widespread products of human activity that are dangerous by virtue of their overproduction if uncontrolled. The present report will concentrate on the "toxic" water pollutants, analogous to the air toxics, that are treated and analyzed as exposures to toxic chemicals.

As enacted in 1972, the CWA required implementation of standards for toxic pollutants providing an "ample margin of safety;" that is, feasibility considerations were not allowed. For reasons similar to the difficulties seen in regulating air toxics under a similar standard, the CWA was amended in 1977 to include a named list of chemicals [§307(a)(1)] to be regulated within three year with regulation to be based on "best available technology" (abbreviated BAT, a feasible technology approach similar to the 1990 revision of the Clean Air Act). A more stringent "ample margin of safety" standard may be set if necessary [§307(a)(4)].

The second major statute administered by the water office is the Safe Drinking Water Act, which sets contamination level standards for "finished" (i.e., end-tap) drinking water provided by all but the smallest public water systems. The act took its

current form after 1986 amendments that followed a report from the Office of Technology Assessment documenting widespread serious incidents of contaminated drinking water (Findley and Farber, 1992). The standards are set on a health basis alone, but the requirement is to come as close to meeting them as is technologically feasible. Primary enforcement authority is with the states, which can opt for more stringent standards.

THE CWA AND ITS MANDATES

The **Clean Water Act** (CWA, technically the Federal Water Pollution Control Act, 33 U.S.C.A. §§1251 to 1387) opens with a "Congressional declaration of goals and policy" [CWA§101] that sets ambitious goals for the nation, declaring "it is the national goal that the discharge of pollutants into navigable waters be eliminated by 1985" and that "the discharge of toxic pollutants in toxic amounts be prohibited." The history of amendment of the CWA has been in part the history of rescheduling and delaying the milestones and timelines for achievement of the mandated complete solution to the nation's water pollution problems, as issues of feasibility and practical impediments are encountered. Nonetheless, the act has provisions for citizen lawsuits that has led to the agenda of water regulation being driven largely by court orders and consent agreements.

A "toxic pollutant" under the CWA is one that "after discharge and upon exposure, ingestion, inhalation, or assimilation into any organism, either directly from the environment or indirectly by ingestion through food chains, will, on the basis of information available to the Administrator, cause death, disease, behavioral abnormalities, cancer, genetic mutations, physiological malfunctions (including malfunctions in reproduction) or physical deformities in such organisms of their offspring" [§502(13)]. That is, chronic health effects on humans and other species are included, and exposure may be indirect.

Section 304 of the CWA calls on EPA to establish "criteria for water quality accurately reflecting the latest scientific knowledge...on the kind and extent of all identifiable effects on health and welfare," including ecological effects. That is, the criteria are to be entirely health- and effect-based. States must develop surface water quality standards consistent with the Federal criteria that are "such as to protect the public health and welfare, [and] enhance the quality of water" [§303(c)(2)(A)]. States have discretion to designate the intended uses of particular water bodies, but the criteria applicable for each intended use reflect the Federal criteria for that use. States must identify local spots where existing effluent limitations will not suffice to guarantee adequate water quality, and act to ameliorate them.

The Federal water quality criteria are unenforceable (but the EPA can require states to develop enforceable standards based on them). The second prong of Federal water regulation is the promulgation of enforceable performance standards for sources of effluent. As noted above, since 1977 establishment of effluent standards for toxic pollutants has been based on the "best available technology economically achievable"

(BAT) for sources in each particular category [§307(a)(2)]. The compounds to be so-regulated are specified in a list [§307(a)], and there are provisions for additions and deletions to the list.

There is, however, a "residual risk-like" provision to this BAT approach; standards must "be at that level which the Administrator determines provides an ample margin of safety" [§307(a)(4)], so that standards more stringent than BAT may be named at EPA discretion. (The courts have upheld that actions under §307(a)(4) need not consider feasibility [Findley and Farber, 1992]). (In practice, the Office of Water has had difficulty keeping up with the very ambitious schedule for promulgation of standards imposed first by the statute and then by the courts.)

Enforcement of standards promulgated under §307 is by requirement of a permit to discharge any substance for which standards exist. Section 402 establishes a permit system, the National Pollutant Discharge Elimination System, under which discharge permits can be granted by EPA or by the states under EPA-approved programs.

THE SDWA AND ITS MANDATES

The Safe Drinking Water Act (SDWA, Title XIV of the Public Health Service Act, 42 U.S.C.A. §§300f to 300j-26) regulates the contamination of drinking water provided by public water systems. (It also contains some provisions protecting drinking water sources.) As with the Clean Water Act, there are a number of statutory timelines for promulgation of regulations that set a very ambitious schedule, one that has been difficult to meet in practice. Regulation is based on the permissible levels of contamination of finished water, that is, as it appears to consumers at the end of the tap. These standards, called national primary drinking water regulations, are promulgated by EPA [§1412(b)(3)] and enforced by the states, which can opt to set more stringent standards [§1413]. The standards apply to all public water supplies serving at least 25 people. Section 1412(b)(3)(A) calls on the EPA Administrator to "promulgate national primary drinking water regulations for each contaminant...which...may have any adverse effect on health of persons and which is known or anticipated to occur in public water systems."

A standard specifies two levels of contamination of drinking water by the compound in question: a "maximum contaminant level goal" is set "at a level at which no known or anticipated adverse effects on the health of persons occur and which allows an adequate margin of safety." For each standard with such a goal there is also specified "a maximum contaminant level which is as close to the maximum contaminant level goal as is feasible" [§1412(b)(4)], where "feasible" means "feasible with the use of the best technology, treatment techniques and other means which...are available (taking cost into consideration)." (In practice, achievability is judged by affordability of control technology to larger public water suppliers; smaller suppliers may have economic difficulty complying. If contaminant levels cannot be measured, a standard can specify a treatment technique to be used.)

In other words, maximum contaminant level *goals* (known as MCLGs) are to be set solely on health grounds to protect with an adequate margin of safety. Maximum contaminant *levels* (known as MCLs) are levels that are practically achievable. It is the technically feasible MCL, and not the health-protective MCLG, that is the enforceable standard. The level set for the MCL depends on available technology, and the appropriate level can change with technological advance. Section 1412(b)(9) provides for periodic revision of MCLs to address this.

The main reason for the MCLG/MCL distinction is that carcinogens, being presumed to be without a threshold, have no safe level. (Clearly, it is also possible that an agent with a threshold has that threshold level lower than is technically achievable.) That is, the common problem faced under all statutes requiring "safety" (especially with an "adequate margin") when dealing with non-threshold toxicants is addressed under the SDWA by controlling contamination to as low a level as technically and reasonably possible without particular regard for how much risk is estimated to remain. This is similar to the "carcinogen policy" at OSHA as it existed before the Supreme Court benzene decision [see p.31] and practice at the EPA Office of Air and Radiation before the vinyl chloride decision [see p.107], both of which policies were overturned by those decisions, as discussed in the sections on those groups. The chief difference is that the SDWA explicitly decouples the risk and the feasibility issues.

It is important to remember that MCLs are set on a technical feasibility criterion, with the feasibility issue being affordability of controls by public water providers. In some cases, other regulatory programs (notably Superfund and Solid Waste) use the water office's MCLs as though they were health-based criteria, for example as standards to be attained for cleanup of or release into water. The entirely reasonable rationale is that requiring concentrations to be lower than allowable in tap water seems to be unwarranted, but the inappropriate implication is sometimes made that attainment of the MCL is a standard of health protection.

RISK MANDATE

The Clean Water Act calls on EPA to "protect the public health and welfare" by issuing Federal water quality criteria, specifying standards that surface waters should achieve according to the various intended uses of the water body. The criteria are health- and effect-based and not directly enforceable, but states must implement standards based on them. In addition, standards for performance of effluent sources are set to reflect best available technology.

The Safe Drinking Water Act calls for standards for contamination of tap water "at which no...adverse effects on health" occur, allowing "an adequate margin of safety." However, if these levels cannot be technically achieved, the lowest technically achievable level is required.

IMPLEMENTATION

Four years ago, the Office of Water underwent an internal reorganization, and to the extent feasible, it tries to combine its activities under the CWA and SDWA into one implementation scheme. Guidelines have been prepared for the methods to be used in Health Effects Assessment documents used in developing water quality criteria, and these guidelines address both methods for such criteria and for the determination of MCLGs [45 FR 79347-79357]. Specific guidance on setting MCLGs is also published [54 FR 22068]. A revision of methods for the preparation of ambient water quality criteria is in preparation. Health Effects Assessment documents are largely produced by the Office of Research and Development, but the Water office also produces its own Drinking Water Criteria Documents.

Standards are set in terms of concentrations of the applicable contaminants in water (be it ambient water or drinking water). The determination of health effects is based on measures of dose of the toxicant. Thus, the concentration levels in the standards must be linked to possible health effects through an exposure scenario. As discussed further under the exposure section, a standardized exposure scenario is used based on lifetime consumption of 2 liters of water per day and (in the case of water quality criteria) consumption of 6.5 grams of fish per day on average.

Because of the statutory focus on standards referring to health and effluent controls relating to technology, the more detailed exposure questions of the fate and transport of pollutants between their release into water bodies and their eventual consumption by humans (including modeling or measuring of the actual levels of consumption and spatiotemporal variation in the degree of water contamination) are not really the province of the Federal level of regulation (except as they enter the calculation of bioconcentration) . These questions enter into the analyses the states carry out in determining how to issue discharge permits, how to control hotspots, what uses to designate for various bodies of water, and so on. Fate and transport enter into consideration of how various patterns of effluent controls and permitted discharges will end up affecting compliance with the water quality criteria, but they are not as germane to the setting of those criteria as currently mandated.

Not all ambient water must be drinkable. It is up to each state to designate specific bodies of water for intended uses, and the criteria that must be met depend on the uses so specified. States may choose from Federally specified use categories (public water supply, recreational use, propagation of fish and wildlife, and others) or designate their own, but the criteria adopted in state standards for a given use must be in accord with appropriate Federal criteria for the uses intended, including numeric criteria for human health, aquatic life, wildlife, biological properties of the water body, sediments, and so on.

HAZARD IDENTIFICATION

As with most EPA regulatory programs (except OPP and OPPT), much of the hazard identification and dose-response components of risk assessment are done for the water office by the Office of Research and Development in the form of Health Effects Assessment documents setting out the data and basis for these analyses. Accordingly, the discussion of these processes in the general section on EPA applies [see p.62 *et seq*.]. The Water office also produces its own Drinking Water Criteria Documents.

Compounds that receive a carcinogenicity weight-of-evidence categorization of C (i.e., possible human carcinogens) may be treated differently by regulation under the SDWA than other agents with higher weight-of-evidence classifications. Unlike under other programs, group C compounds may be assessed with the NOAEL/Safety Factor method using an extra uncertainty factor of 1 to 10 to allow for uncertainty about the agent's possible carcinogenicity.

The program of development of MCLGs has a sort of hazard classification scheme that operates in addition to the usual EPA cancer classification system [54 FR 22068], but applies particularly to ingestion in drinking water. Category I chemicals are those with "strong evidence of carcinogenicity from ingestion in drinking water," which is generally in accord with EPA categories A and B. MCLGs for Category I agents are set at zero. Category II agents have "limited evidence of carcinogenicity," which generally correspond to EPA C's, and MCLGs are usually set upon non-cancer effects data, as noted in the previous paragraph. (If no such data are available, a lifetime cancer risk may be used.) There is a further Category III for agents with "inadequate or no evidence of carcinogenicity," (generally EPA D's and E's) that invariably have MCLGs set on non-cancer effects.

As in the EPA air program, to some degree which compounds are to be considered toxic pollutants is a matter of definition by law. Section 307(a)(1) specifies a list of compounds that are to be regulated as toxics.

DOSE-RESPONSE ANALYSIS

Dose-response analysis follows the usual EPA practice [see p.68 *et seq*.]. Especially for carcinogenicity assessments, the analyses are usually carried out for the water office by EPA's Office of Research and Development. For non-cancer effects, RfDs are developed. As noted in the previous section, substances in the EPA carcinogenicity weight-of-evidence category C (and in the water office Category II) are assessed by the Office of Water on non-cancer effects using the NOAEL/Safety Factor approach with an extra factor of 1 to 10.

EXPOSURE

As noted above, the main issue for regulation development at the Federal level is the setting of standards for concentrations in drinking or ambient water. The question

during this process is hypothetical: What health effects might be expected if people consumed water contaminated at the level of a candidate standard? The main function of exposure assessment is to link candidate standard water concentrations to the doses of the toxic compounds (and through this, to the health effects that might be projected). (Clearly, exposure assessment is also involved in monitoring tap and ambient water for compliance with standards; this is a state responsibility, however.)

This linking of water concentration to dose is conducted through a standardized exposure scenario of lifetime consumption of 2 liters of water per day. This value is considered somewhat high for an average, but below the high end of the distribution among individuals. No particular account is made for the fact that some of the water may be heated (for cooking, making coffee or tea), which could drive off volatile compounds (decreasing exposure by ingestion, but potentially increasing it by inhalation).

To express the RfD (a dose in mg/kg/day) and the water concentration (a concentration in mg/liter) in equivalent units, a Drinking Water Equivalent Level (DWEL) is calculated, which is the water concentration that would lead to the RfD daily dose level under the 2 liter daily consumption scenario for a 70 kg person. The DWEL thus corresponds to the water concentration that would lead to attaining an RfD dose.

For assessing standards under the CWA, the same water consumption rate is used, but in addition, the dose resulting from the daily average consumption of 6.5 grams of fish is added. This amount of fish consumption is considered appropriate for an average, but is well below the high end of individual fish consumption habits. The fish is presumed to have come from the same water source (since the usability of a particular source is what is at issue), and is assumed to be contaminated according to the equilibrium partitioning of the compound between the water and the fish tissue. That is, the concentration in fish is not based on any empirical observations of actual levels in fish taken from water contaminated at the level of interest. It is based on projected bioconcentration, not bioaccumulation; that is, there is no allowance for the accumulation of body burden in fish through the food chain, only for absorption directly out of the water column. This will tend to underestimate actual levels in fish flesh, sometimes considerably.

Most regulatory programs examine exposures only from the particular source for which regulation is being considered. Recently, EPA has established the policy that exposures from other and indirect sources should be considered. This is particularly important for non-carcinogens, where the fact that several exposure sources might individually be below the RfD level does not mean that collectively the exposure is above this presumably safe level. (For carcinogens, the assumption of additivity of risks from multiple exposures can be made; one can argue that each program is only looking at the additive effect of exposures under its regulatory purview. Total individual and population risks will be underestimated in this way, however.)

The water office, however, has for some time had an established policy of considering indirect exposures through the correction termed the *relative source contribution*. This is a method to try to account for exposures from sources other than drinking water, including exposures in air and in food. The procedures for calculating the relative source contribution vary with the amount of data available and with the estimate of the likely proportion of the total exposure coming through water. If data are available on exposures through other sources, they are used. (These should be middle range estimates, remembering that a middle range water consumption is the basis for regulation). The MCLG for drinking water is then lowered to allow for the presumed presence of the is other exposure. If data indicate that drinking water accounts for between 80% and 100% of total exposure, a value of 80% is assumed on the grounds that other minor sources may have been overlooked by the available data. When no data are available on other exposures, an assumption is usually made that drinking water accounts for 20% of total exposure, and the DWEL is adjusted accordingly.

The relative source contribution correction tends to counter the potential problem that volatilization of certain compounds out of tap water (from showers, toilets, etc.) has in some cases been found to lead to large indirect inhalation exposures that are nonetheless tied directly to concentrations in tap water. Indeed, such volatilization may account for the majority of uptake from tap water for some agents. There is also the possibility of dermal uptake during bathing and dishwashing that may be important in some instances. The Office of Water is actively working on ways to better characterize and account for these sorts of exposures.

In the realm of ambient water criteria, the setting of criteria is also based on standard consumption as a way of determining what health effects might be expected from compliant water. (A relative source contribution of 100% is usually used in this context.) At the state level there is the added problem of planning a program of controls and permitting of discharges at particular times and places in the state's water bodies so as to avoid any spots where the ambient water quality criteria are violated. That is, in this area of planning how to achieve compliance with standards (as opposed to setting of standards) there is a great use of fate and transport modeling. Such analyses may be done at the Federal level as well for planning purposes and for the mandated Federal evaluation and oversight of the adequacy of state water quality control plans. Generally, fate and transport models with somewhat conservative assumptions is used to determine the total loading of pollutants that can be tolerated by a body of water, and then this loading can be allocated among potential dischargers.

RISK CHARACTERIZATION AND REGULATION

Federal ambient water quality standards are set on a health basis without cost considerations, but they are not themselves enforceable. Instead, they serve as a guide for judging the appropriateness and adequacy of state standards. For carcinogens, no level can be named that fulfills the designation of "safe," so the criteria are presented as water concentrations that would be expected to lead to lifetime cancer risk levels of 10^{-5},

10^{-6}, and 10^{-7} when consumed at the standard rate for a lifetime. For non-carcinogens, water quality criteria are developed that will not violate the RfD. (Cancer risks and RfDs are calculated by the standard methods.)

These calculations are based on individual risk, but the criteria are to apply nationwide, so it is presumed that any criterion will apply to a significant number of people. Actual exposures for many people will of course be less, but exposures will be higher for a significant number, because of the midrange nature of the consumption assumptions, the high variability in fish consumption, and because much surface water in the country is not in compliance with the water quality criteria which (despite the policy statements set out at the beginning of the CWA) remain goals to be striven for in many cases.

For drinking water standards, the basis is not health but technical feasibility. Nonetheless, the setting of standards is not blind to risk, and an attempt is made to ensure that the drinking water standards result in projected lifetime individual risks in the range of 10^{-4} to 10^{-6}. Again, these are individual risks, but they apply to large blocs of people (all those served by a given water supply). For non-cancer effects, standards are desired that do not violate the RfD. These determinations are made taking account (via correction for relative source contribution) for exposures to the same compound from other sources.

Like all regulatory programs save Superfund, there is no accounting for the fact that people are exposed to more than one chemical at a time and that certain effects may be dose additive.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY.

HAZARDOUS WASTE - EPA OFFICE OF SOLID WASTE

The regulation of hazardous solid waste is the responsibility of EPA's Office of Solid Waste (OSW). The office implements the 1976 Resource Conservation and Recovery Act (RCRA), which amended the Solid Waste Disposal Act. The purpose of the act is to develop mechanisms for ensuring stewardship over hazardous compounds from their generation to their proper disposal. The act's provisions set up an extensive set of requirements for reporting and record keeping in addition to standards for generators and transporters as well as treatment and disposal practices. That is, the aim is to ensure that hazardous wastes are kept track of—and that ownership and responsibility for those wastes are not lost or obscured—during storage, transportation, and disposal. The provisions can be seen as a means to avoid the processes leading to dangerous hazardous waste sites, especially those at which responsibilities for the wastes are no longer assignable.

Most standards regarding transportation, handling, and disposal under RCRA have involved specifications of methods and technology to be used. For instance, there are criteria for the design and maintenance of landfills. Until relatively recently, risk questions were secondary, the presumption being that proper handling and disposal would preclude significant exposure. More recently, however, risk analysis has begun to play a larger role in several areas. These include the criteria for defining hazardous waste, the process for delisting substances as hazardous wastes, evaluation of the hazard posed by various waste streams, support for permitting of disposal, including incineration permits, and evaluation of the need for corrective action at disposal sites.

To a large degree, the evaluation of toxicity information and the potencies of substances is drawn from other EPA sources outside of OSW, including information on the IRIS database [see p.59], reports produced by the EPA Office of Research and Development, maximum contaminant levels taken from the EPA Office of Water, and methods borrowed from the Office of Emergency and Remedial Response (Superfund). In fact, the analysis of hazards posed by inadequate waste disposal sites has much in common with the analysis conducted by Superfund for abandoned sites. OSW combines this information with its own exposure analyses and conducts risk characterization appropriate to its uses of risk analysis, as discussed below.

RCRA AND ITS MANDATES

The **Resource Conservation and Recovery Act** (RCRA, amending the Solid Waste Disposal Act, 42 U.S.C.A. §§6901 to 6992k) declares it to be "the national policy of the United States that, wherever feasible, the generation of hazardous waste is to be reduced or eliminated as expeditiously as possible. Waste that is nevertheless generated should be treated, stored, or disposed of so as to minimize the present and future threat to

human health and the environment" [RCRA§1003(b)]. Hazardous waste is defined as solid waste that may "cause or significantly contribute to an increase in mortality or an increase in serious irreversible, or incapacitating reversible, illness" or otherwise present a potential hazard to human health or the environment [§1004(5)].

Section 1003 requires EPA to identify hazardous wastes and to list those wastes which should be subject to RCRA's provisions. (There is a provision for delisting a waste as well [§3001(f)].) Listing is to take into account "toxicity, persistence, and degradability in nature, potential for accumulation in tissue" as well as factors such as corrosiveness and flammability. EPA is empowered to issue standards "as may be required to protect human health and the environment" in three broad areas: generation, transport, and disposal. Generation is covered by §3002, requiring standards for record-keeping, handling, labeling, and use of appropriate containers. Section 3002 sets up a manifest system to ensure that the waste is kept track of and responsibility for it assigned, from its generation to eventual disposal, even if this involves transactions and transfers of ownership of the waste. Transport standards are mandated in §3003, which also incorporates the manifest system, as does §3004, which governs storage and disposal. Disposal standards are largely framed in terms of technology that must be used. Land disposal is prohibited unless "to a reasonable degree of certainty,...there will be no migration of hazardous consituents from the disposal unit...for as long as the wastes remain hazardous." RCRA also provides for EPA regulation of cleanup of currently active industrial sites that hold RCRA permits and requires permits for waste incineration and other disposal methods in addition to land storage.

RISK MANDATE

Given the largely technical and procedural nature of its provisions, RCRA has relatively little to say about risks and risk assessment. It simply calls for EPA to act to ensure that hazardous waste management practices "are conducted in a manner which protects human health and the environment" [§1003(a)(4)]. Section 3019(b) states that when, in the Administrator's judgment, "a landfill or a surface impoundment poses a substantial potential risk to human health, due to the existence of releases of hazardous constituents, the magnitude of contamination,...or the magnitude of the population exposed" a request may be made for the Agency for Toxic Substances and Disease Registry (ATSDR) to conduct a health assessment of the site. Such a health assessment is not a risk assessment *per se*, but it contains many of the elements of one, including characterization of the exposures and potential exposures around the site, identification of potential exposure pathways, review of the known health effects of the hazardous constituents present, surveys of health complaints in the population in the vicinity of the site, and the review of applicable health-based exposure standards that may exist. (Since the ATSDR is not a regulatory agency, and since its health assessments are not used in support of regulation, the methodology for such assessments is not addressed in this report.)

RCRA also has little to say about costs, neither requiring nor prohibiting their consideration (Schierow, 1994). Presumably, the stated concerns for health need to be adequately addressed.

IMPLEMENTATION

A difficulty faced with any regulatory treatment of hazardous waste is that waste typically comprises a complex mixture of constituents, some of which may be hazardous compounds and some of which will not be. Waste is hazardous by virtue of its hazardous consituents, but the components are generally not separable, and so regulation to control the handling, transport, storage, and disposal must address the mixture. When hazardous waste is mixed with other, non-hazardous materials, there is a question as to whether the resulting mixture should become a hazardous waste as well. The whole intent of RCRA, to foster "cradle to grave" stewardship and custody over hazardous wastes, argues against allowing the hazardousness of a mixture to be "diluted away." The provisions of RCRA regulating handling of waste aim to ensure that the hazardous consituents are properly dealt with and disposed of; this requires procedures for treatment of the mixture as a whole. (When the issue is leakage or emissions of materials out of a disposal facility, however, the constituents can be separately dealt with, and the release of "hazardous constituents"—i.e., individual compounds—can be the focus of risk analysis and regulation.) The question of what constitutes a hazardous waste, and how mixing and separation affect the definition, is complex both technically and legally. There are a number of exceptions and special cases written into the law, and wrangling about whether a substance falls under RCRA's definition (and hence under its considerable regulatory scope) has been part of the process from the start. A detailed account of these provisions and issues is beyond the scope of this report.

Such considerations put great importance on the listing provisions in RCRA, which define what wastes are to be treated as hazardous wastes and hence subject to RCRA provisions. OSW has published guidance on these methods [55 FR 11798 (1990)], but development and improvement of methods is continuing. There are two routes to being declared a hazardous waste: by properties of the waste *per se* and by being part of a waste stream from a particular industrial process that has been declared hazardous.

In the first route, by properties, certain immediate physical dangers such as corrosivity, reactivity, and flammability suffice, but there is also the possibility to be declared hazardous by virtue of toxicity. Toxicity of waste is evaluated by determining whether land disposal would be likely to result in leachate that poses health hazards to anyone in the surrounding area. The evaluation must be of hypothetical disposal anywhere in the nation, since the purpose is to decide whether the material should be defined as hazardous and subjected to disposal and other controls. A waste is taken to possess the characteristic of "toxicity," and hence be definable as hazardous, if such disposal results in leachate concentrations of any of about 40 constituents that exceed 100 times the Office of Water's maximum contaminant limit (MCL) (or a different

appropriate, health-based limit) or if consumption of the contaminated groundwater generates a cancer risk greater than 10^{-5}. As discussed under the exposure section, earlier methods used groundwater models with conservative assumptions, but more recently Monte Carlo techniques are used to characterize a distribution of likely disposal and leaching scenarios.

The second route to definition as a hazardous waste is that the waste streams from certain industrial processes (e.g., electroplating wastes, spent solvents) can be declared hazardous and in need of RCRA-approved disposal. Earlier analyses tended to examine the waste stream concentrations of particular hazardous compounds and compared these to established health-based standards. But more recent analysis takes more account of the probable modes of treatment and disposal (including present and likely future methods). The aim again is to determine whether likely disposal modes without RCRA regulation would be likely to cause undue risk to the populace surrounding disposal sites. The methods use questionnaires to industry on likely practices, models of all potential exposure routes, national meteorological and hydrological data for use in characterizing distributions of groundwater contamination that might arise, and other methods to characterize the likely "high end" of exposure (as defined in the revised EPA exposure guidelines and risk characterization policy, which are discussed in the general section on EPA, p.80). Cancer risks in the range of 10^{-4} to 10^{-6} may trigger definition of the waste stream as inherently hazardous, with risks higher in the range being more likely to cause such a designation, and risks above this range creating a "strong presumption" for listing. (These criteria are currently under internal discussion.)

Criteria for delisting of a listed toxic waste have also been developed, but the risk provisions are somewhat asymmetrical for listing versus delisting (Sadowitz and Graham, 1995). Delisting analysis tries to focus on a more site- or case-specific view of the likely exposures; i.e., there is an opportunity to account for factors that are particular to the specific waste substance that might mitigate concern compared to the standard disposal scenarios used in listing decisions. Conservative groundwater modeling is used, and risks below 10^{-6} are grounds for delisting (compared to risks up to 10^{-4} as triggers for listing). (There is thus a range of risks that are too low to cause a substance to be listed, but if it were listed, they are too high to allow it to be delisted.)

Standards for emissions from hazardous waste incinerators are under development. (Such incinerators require RCRA permits.) Both RCRA and Clean Air Act provisions must be obeyed by such facilities, but the CAA standards are based on maximum available control technology, while the RCRA ones are health-based. These will be national standards, but site-specific permits will be issued that take into account local populations and local land-use activities (farming, fishing, etc.), with distributional assessments of likely exposures. In the meantime, a "combustion strategy" (which is not a formal regulation) seeks 10^{-5} lifetime cancer risks as a criterion for issuance of an incinerator permit, based on a site-specific risk assessment. Non-cancer effects are judged against 25% of the RfD (to allow for other exposure sources, much like the water office's Relative Source Contribution calculations, as discussed in the section on that office). Conservative screening level assessments are done, with more realistic detail

added if health questions remain. The authority for this permitting is from the broad, non-specific "omnibus authority" in RCRA for EPA to set standards "as may be required to protect human health and the environment."

Another major area of application of risk assessment in RCRA implementation is the decision whether to take corrective actions at active waste sites, and on the choice of appropriate remediation approaches. This process has much in common with the similar actions under Superfund, and indeed, a lot of Superfund guidance is used in carrying out these RCRA responsibilities. Certainly, consistency among the programs is striven for.

Exposure analyses at particular sites can of course be much more site-specific than the hypothetical disposal analyses underlying prospective analysis of a substance's disposal. In particular, the risk estimation can focus on the migration off site of particular toxic compounds that constitute the local problem. When data sufficient for Monte Carlo modeling of exposures is not at hand, models may be used with some conservative parameter values and others set at "central" estimates, a version of the Superfund concept of the RME, or Reasonable Maximum Exposure methodology, as discussed in the section on the Office of Emergency and Remedial Response [p.142]. Like Superfund, OSW selects an option for site remediation that keeps eventual risks in the range of 10^{-4} to 10^{-6}. But (also like Superfund) the decision is not primarily based on the risk level, but rather on eventual land use at the site, costs and feasibility of clean-up, and local restrictions on permissible clean-up activities.

OSW once made distinctions between risks that are permissible from different weight-of-evidence classes of carcinogens, treating EPA group C compounds differently than those classified in groups A and B [e.g., 55 FR 17862]. This is no longer the case.

HAZARD IDENTIFICATION AND DOSE-RESPONSE ANALYSIS

In the narrow sense, hazard identification and dose-response analysis at OSW follows general EPA practice. Hazardous properties of particular compounds, RfDs, and cancer potencies are all garnered from EPA risk assessment documents on the appropriate compounds, from the IRIS data base, and other such sources.

The larger sense of hazard identification, the listing of certain substances and waste streams as "hazardous waste" subject to RCRA provisions, has been discussed under Implementation.

The frequent use by OSW of the maximum contaminant limits (MCLs) calculated by the Office of Water bears some mention. As discussed in the section on the water office [p.123], MCLs represent concentration standards for specific contaminants in drinking water. For carcinogens (and in principle for non-carcinogens with unattainably low thresholds) the MCLs are set at levels dictated by technical feasibility, where such feasibility is determined regarding either the affordability of control technology to public tap water suppliers or (perhaps more frequently) to the technical ability to detect the

substance in drinking water. That is, they are not health-based levels and do not represent a consistent standard of health protection from one compound to another.

Clearly, it would be difficult for OSW to require groundwater near waste sites to be cleaner than drinking water from the tap must be under the Safe Drinking Water Act. Nonetheless, MCLs are not health-based standards, and the feasibility considerations that go into them are not those applying to RCRA applications.

In dose-response analysis, OSW must deal with estimates of risk from mixtures of toxic compounds, and the issue is whether the mixture in general has certain associated risks. Thus, analysis can be for exposure to several chemicals at once, and the guidance of the EPA *Guidelines for the Health Risk Assessment of Mixtures* [51 FR 34014] are particularly important. Provisions for this are essentially the same as used by Superfund [p.144]. Basically, risks from mixtures of carcinogens are assumed to be additive, while mixtures of agents causing non-cancer effects are assumed to be dose-additive when affecting the same endpoint, and independent otherwise.

EXPOSURE

The exposure analyses are essentially of two sorts: (1) hypothetical exposures following disposal of wastes in particular ways (mainly in landfills), to determine if such disposal might constitute a source of undue risk, an aid in defining hazardous wastes and evaluating disposal options, and (2) evaluation of particular actual sites, either for waste site remediation decisions or for permitting incineration or other disposal activities.

The main question for both of these cases is fate and transport of hazardous constituents away from their intended confinement. The principal concern is often leaching in groundwater, but for incineration the pathways include inhalation as well as a variety of indirect routes (e.g., emissions outfall onto local crops and into local waters, etc.). In the past, the fate and transport models were usually fitted out with conservative values of their parameters, i.e., tending to make higher rather than lower exposure estimates. More recently, much of this modeling is now being done using a Monte Carlo approach, with distributions of parameter values. For instance, the former assumption that a drinking water well occurs 500 feet down gradient from the disposal area may now be replaced with a distribution of distances of the nearest well to disposal sites based on national data on many disposal sites, coupled with a distribution on the placement of such a well vis-à-vis the local hydrological gradient, along with distributions of other relevant factors. (It should be noted that the Monte Carlo simulation is applied to the groundwater modeling, reflecting a distribution of hydrological situations and well placements amid which a land disposal could be situated; the human exposure of concern for any such situation is someone consuming water from the nearest, most contaminated well, and population variability in consumption of the contaminated groundwater is not included in the analysis.) A higher percentile (usually the 85th for a listing decision, but the 95th us used for delisting) of the resulting distribution of well water concentrations gives a high end estimate of the likely degree of local contamination by the waste disposed of at the

site without compounding the conservatism in each parameter of the exposure model. For hypothetical exposures representing the potential consequences of a disposal practice occurring across the nation, national distributions of parameters may be used. For assessments of local sites, more site-specific data can be used as available.

In some cases, a procedure like the Superfund's Reasonable Maximum Exposure is used, in which certain exposure parameters are set at "high end" values and others at "central" values to yield an exposure estimate that is hoped to be in the high end of actual exposure levels. This approach may be used when appropriate information of distributions of parameters is not available.

As noted, the main focus is on fate and transport of contaminants in the environment. There are also distribution questions about individual behaviors that lead to different levels of human exposure for a given contaminant level in the affected medium, including variation in residence times, use of groundwater, amount of water consumption, and so on. These have played less of a role, since the question being asked is whether undue individual risks might be caused, not to characterize the actual distribution of exposures across the affected population. The analyses are not being done to balance the effects of disposal on populations around waste sites with other factors; rather they are done to define disposal practices that are unlikely to cause anyone undue risk.

RISK CHARACTERIZATION AND REGULATION

As noted, the risk calculations represent individual risks under exposures that are calculated with conservatism tempered where possible by the use of distributional and Monte Carlo analysis. The question is whether such individual risks are such as to trigger specific actions, such as listing or delisting a substance as a hazardous waste, granting a permit for incineration, or remediating an active waste site.

Individual lifetime cancer risk levels of 10^{-5} or so from unregulated disposal trigger listing of a waste as a hazardous substance and hence subject to RCRA controls on handling and disposal. Newer methods are adopting a range of 10^{-4} to 10^{-6} as a range in which this decision can be made. Delisting a substance as a hazardous waste requires a risk estimate less than 10^{-6} for unregulated disposal. Other questions are not triggered by specific individual risk levels, but a range of 10^{-4} to 10^{-6} is usually used as a guide. Incinerator permits have usually been granted if risks are below 10^{-5}. Remediation of active waste sites depends on many non-risk technical and other factors, but a post-remediation risk level of 10^{-4} to 10^{-6} is aimed at.

The role of risk assessment in decisions under RCRA is relatively new. Use is growing, but methods are also being reconsidered as more sophisticated methods, especially distributional models of exposure, become part of the analysis. Virtually all of the methods discussed in this section are under some kind of review or reconsideration. At the same time, OSW has to reach many assessment and methodology milestones

under court orders, and practical questions conflict with scientific ones as the office attempts to fulfill legal obligations.

SUPERFUND - EPA OFFICE OF EMERGENCY AND REMEDIAL RESPONSE

The Superfund program was created by the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980 and its subsequent amendments to address the need for cleanup of the nation' s hazardous waste sites. The program is administered by the EPA Office of Emergency and Remedial Response (OERR). With no state unaffected by past hazardous waste disposal practices, the Superfund program has perhaps done more than other programs to make the use of risk assessment a local issue. At the same time, it has become a lightning rod for criticism of the U.S. EPA's use of risk assessment for regulatory decisionmaking in general.

The aim of the Superfund program is to achieve permanent solutions to the hazards posed by waste sites. That is, the goal is not simply to manage exposure and access to sites, but to clean them up as near to the *status quo ante* as possible. Much of the program's focus is on abandoned sites or those for which responsibility for generation of the site (and hence, under the law, for the cleanup) are unclear, unknown, or in dispute. When responsible parties can be identified, they are responsible for cleanup costs, and a considerable portion of CERCLA concerns revolve around establishing and assigning liability, but there is a fund provided for covering unassignable costs.

Risk assessment is used under Superfund to define hazardous substances and the amounts of release that must be reported to EPA ("reportable quantities"), rank the risks posed by hazardous waste sites and identify the action priorities among them, including the addition of sites to a National Priorities List (NPL) of high-priority sites, and evaluating the effectiveness of options for remediation (which are chosen on various non-risk grounds in addition to considering risk reduction effectiveness).

CERCLA AND ITS MANDATES

The Superfund program is the implementation of the 1980 **Comprehensive Environmental Response, Compensation, and Liability Act** (CERCLA, 42 U.S.C.A. §§9601 to 9675). This act received major amendments in the 1986 Superfund Amendments and Reauthorization Act (SARA). Section 101(14) defines "hazardous substances" as those on lists of hazards and toxics maintained under other various other laws, while §102 requires the addition of any other substances that "may present substantial danger to public health or welfare or the environment." Any releases of these substances (in amounts determined by rough, screening level risk assessments defining the so-called "reportable quantities") must be reported to EPA.

The authority to undertake removal (a short-term emergency action) and remediation (a long-term, cleanup action) of sites is granted under §104 when "there is a release or substantial threat of release into the environment of any...contaminant which may present an imminent and substantial danger to the public health or welfare." (Responses may be carried out by Federal personnel, by contractors, or by cooperative agreement with the state.) In deciding which sites to act upon, a Hazard Ranking System is mandated by §105, which states "priorities...shall be based upon relative risk or danger to public health or welfare or the environment,...taking into account...the population at risk, the hazard potential of the hazardous substances" and the potential for contamination of air and drinking water, among other factors [§105(a)(8)(A)]. The system should "accurately assess the relative degree of risk to human health and the environment" [§105(c)(1)] and must be used in adding any site to the National Priorities List (NPL), the list of sites at highest priority for Federal action.

Cleanup standards are addressed in §121. As originally passed in 1980, CERCLA directed EPA to select remedies which are "to the extent practicable in accordance with the National Contingency Plan and provide a balance between the need for protection of public health and welfare and the environment" (Environmental Law Reporter, 1986). The law contained no detailed statutory requirements for cleanups; protection of public health remained vaguely defined. As a result risk assessment began to emerge as a tool to assist the Agency with the issues of when and how much remedial action was necessary to protect public health..

With passage of the only major amendments to the Superfund law, the Superfund Amendments and Reauthorization Act (SARA) of 1986, Congress moved to curb EPA's discretion substantially. This amended Superfund law has been the basis for the Superfund program for the last 9 years. SARA put in place a mixture of broad objectives and specific directives which were more proscriptive in nature than the language of the earlier statute. SARA strengthened the emphasis on public health; section 104(a) was amended to state that the "President shall give primary attention to those [sites] which the President deems may present a public health hazard." In §188(a), the Act also provided a stronger emphasis on the protection of groundwater. The section requires that priority be given to releases that may result in closing of drinking water wells or contamination of a principal drinking water supply. This language, combined with the EPA's regulations requiring that groundwater be returned to "beneficial uses, whenever practicable..." (according to the National Contingency Plan) favored remedial actions treating groundwater.

SARA introduced an oversight role for the Agency for Toxics Substances and Diseases Registry (ATSDR), requiring that the Agency perform a health assessment on every site on the NPL. The ATSDR was also charged with preparing toxicity profiles for several hundred contaminants commonly found on Superfund sites. (ATSDR health assessments are not risk assessments *per se*, but they contains many of the elements of one, including characterization of the exposures and potential exposures around a site, identification of potential exposure pathways, review of the known health effects of the hazardous constituents present, surveys of health complaints in the population in the

vicinity of the site, and the review of applicable health-based exposure standards that may exist. Toxicity Profiles present for particular hazardous compounds the base of human and animal data on toxic effects and the exposure levels known to cause them, presented in language appropriate to non-experts.)

SARA introduced more explicit language defining what remedies met the requirements to be protective of public health and the environment in §121 on cleanup standards. However, to focus only on the language specifically dealing with public health and environmental standards would be to miss the other powerful instructions on remedy selection which in many cases have come to dominate the remedy selection process.

Section 121 has three important sections. First is §121(a), which states that remedies selected must be cost-effective. Second, §121(b) specifies several criteria that must be considered in the selection of a remedial action including the preference for remedies "in which treatment permanently and significantly reduces the volume, toxicity or mobility of the...contaminants" and that transport of contaminants off-site is least favored. Third, §121(d) generally dictates the degree of cleanup that a remedy must meet. It specifies that the action must "at a minimum...assure... protection of human health and the environment" and must also satisfy any standards from other Federal and state environmental programs which constitute "legally applicable or relevant and appropriate requirements" (known as ARARs).

The implications of these sections and their impact on the role for risk assessment will be discussed below.

RISK MANDATE

Neither CERCLA nor SARA specifically mention risk assessment, when it is to be used, what procedures to follow, or what levels of risk warrant remedial action or (in the case of specific action) define what actions are to be deemed "protective." As discussed earlier, the statutes provide a broad mandate to pursue action on contaminated sites that "may present an imminent and substantial danger to the public health or welfare" [§102]. In the end, the action of risk significance is the remediation decision and its effectiveness. Thus, the mandates regarding cleanup levels in §121, discussed above, constitute the CERCLA risk mandate. As discussed below under implementation, the National Contingency Plan offers nine much more specific criteria for selection of remediation options, and these comprise the practical Superfund risk mandate.

IMPLEMENTATION

As noted above, risk assessment is used in a number of contexts in the Superfund program, ranging from use to define substances as hazardous for purposes of reporting spills and releases, through screening and priority setting among candidate sites for

listing, to more thorough evaluation of risks posed by particular sites and the evaluation of the effectiveness of alternative remediation options. The data available, and the rigor of the analysis, increase as one moves from identification to prioritization to evaluation of cleanup options. The analyses are rendered complex by the fact that many agents are typically present at a site, each with its own spatial dispersion, concentration profiles, and health hazards, and because site-specific exposure data are wanted for a large number of local assessments, each displaying a unique set of challenges.

Specific policies on risk assessment have been laid out in the National Contingency Plan (NCP, the body of regulations implementing CERCLA and its amendments) and in numerous guidance and policy directives issues pursuant to the NCP. The NCP, like the statutes themselves does not specifically define the use and form that risk assessment takes in the Superfund site assessment and remedy selection process. However, especially in the area of remedy selection, the NCP interpretation of SARA, sets the criteria which must be met and balanced in remedy selection and can profoundly affect the role that risk assessment plays in cleanup of hazardous waste sites.

Since the passage of CERCLA, EPA policy has been to define public health risk in the context of risk to an actual or hypothetically exposed individual. In the early years of the program, "worst case" assumptions were routinely used to assess individual exposure. Over time, specifically with the publication of *Risk Assessment Guidance for Superfund* (EPA, 1989d), the concept of "reasonable maximum exposure" (RME) assumptions evolved in order to represent an upper confidence limit on the mean exposure but without the badly battered image of worst case assumptions. (The RME concept is discussed further in the exposure section.) In reality, RME exposure have no consistent statistical meaning and EPA exposure assumptions have been the focus of heated debate. Some have argued that the compounded conservatism inherent in both the IRIS toxicity values and the RME assumptions lead to estimates of risk that are highly conservative (Hazardous Waste Action Project, 1993; Burmaster and Harris, 1993) Others argue that failings in the Superfund risk assessment process (e.g., exclusion of contaminants without toxicity values in IRIS from quantitative evaluation of overall risk) may understate risks (Finkel, 1989).

For carcinogens, risk is estimated as the excess individual lifetime risk of cancer for the individual. For noncarcinogens, exposure to individuals is assessed by comparison of estimated doses to the respective reference doses. These analyses follow quantitative methods as used elsewhere in the EPA; the chief aspect particular to Superfund is that simultaneous exposure to a number of agents is the norm and the aggregate risk is the object of the risk analysis. For carcinogens, risks are estimated from individual agents and the aggregate risk calculated as their sum. For non-cancer effects, exposures for each agent are compared to that agent's reference dose, the ratio forming what is known the Hazard Quotient, with values less than unity indicating exposures deemed safe if encountered in isolation. When several contaminants are present, however, the possibility that they add together in causing an adverse health effect arises, and levels that may be individually safe may be collectively hazardous. Agents that affect the same target organ with their toxic effects accordingly have their

Hazard Quotients added to form a Hazard Index, which should also be less than unity to ensure safety. That is, it is presumed that compounds affecting the same target are acting in a dose-additive manner.

Under EPA Superfund policy, population risks are not formally considered, so quantitative estimates of population risk rarely appear in risk assessments. However, some EPA Regional Project Managers (RPMs) have unofficially acknowledged that the magnitude of the potentially exposed population sometimes informally affects remedial decisions.

It is important to recognize, that although regulatory policy has given risk assessment a role in the evaluation and remediation of hazardous waste sites, it is one of many considerations in the selection of a final remedial alternatives. The NCP establishes nine criteria by which remedial alternatives must be evaluated:

- Overall protection of human health and the environment;
- Compliance with ARARs;
- Long-term effectiveness and permanence;
- Reduction of toxicity, mobility, or volume through treatment;
- Short-term effectiveness;
- Cost;
- Implementability;
- Cost;
- State acceptance; and
- Community acceptance.

The first two criteria are considered threshold criteria which must be met before a remedy can be evaluated fully by the other criteria. The "overall protection" includes consideration of risks that may be generated as a result of the remedial action (e.g., risks to remediation workers or to the public surrounding a site). However, the strong preference for permanent remedies voiced in SARA and codified in the NCP creates a more technology-based approach to remedy selection, which critics argue can override the implications of a risk assessment.

The most complete assessment done is of a site under the assumption that no action is taken at the site. It is in this phase that comparison of the elements of risk assessment with other regulatory programs is most appropriate. (In recent years, more attention has been given to quantitative assessment of the public health implications of implementing the remedial alternatives but no formal guidance has been issued on this subject.) With the publication of recent studies suggesting that the occupational health risk generated during cleanup of sites may exceed the initial risks posed at some sites the impact of remedial alternatives themselves will likely receive greater attention in the future.

HAZARD IDENTIFICATION AND DOSE RESPONSE ANALYSIS

The Superfund program does not have a unique approach to hazard identification or dose-response analysis. EPA risk assessment policy directs that toxicity values for carcinogens and non-carcinogens be obtained from the agency's Integrated Risk Information System. When the necessary data are not available from IRIS, EPA guidance lists other possible sources (ORD's Health Effects Assessment Summary Tables, for example) that may be consulted. In most cases, where toxicity values are not available for contaminants, qualitative assessments of toxicity are substituted.

EXPOSURE ASSESSMENT

Because the hazard identification and dose-response phases of the risk assessment are set by broader agency policy, the exposure assessment phase receives the greater emphasis in Superfund. Much of the guidance put forth by the Superfund program relates to assessment of exposure and dose (e.g., *Guidelines for Data Usability in Risk Assessment*[EPA, 1992c], *Guidelines for Exposure Assessment* [EPA, 1992b], *Exposure Factors Handbook* [EPA, 1989c], *Risk Assessment Guidance for Superfund* [EPA, 1989d].

Exposure, and ultimately dose (usually defined as intake or absorbed dose where absorption data are available) are estimated for individuals within the context of one or more exposure scenarios. An exposure scenarios is a collection of assumptions about an individual's activities, the frequency and duration of those activities, the possible pathways and route of contaminant exposure and rate of intake of contaminated media.

Whether implicitly or explicitly taken, the first step in identifying exposure scenarios for a site requires an assumption about the current and likely future land use for the site. The categories of land use considered typically include residential, industrial, recreational, or agricultural uses. Each may imply a different set of possible exposure scenarios. Risk assessments may be conducted for alternate land use assumptions if ultimate disposition of the land is uncertain. Historically, EPA policy has typically required that every site be evaluated under the assumption that the land might be used as residential property regardless of the current use of the property. A recent directive (OSWER, 1995) emphasizes the importance of considering reasonably anticipated future land use in remedy selection, however. Increasingly, there is interest in amending Superfund procedures so that industrial use may be made of cleaned up sites without the spectre of lingering liability for past contamination (so-called "brownfields"), freeing the remediation from the need to achieve the level of cleanup that would be needed for residential use.

Given land use assumptions and the nature and extent of contamination at the site, the next step is to develop the exposure scenarios which characterize the circumstances under which an individual may be exposed to contaminants at the site. EPA guidance largely dictates the types of scenarios that are considered under different land use

assumptions although site-specific scenarios may be developed. For example, under an assumption of residential land use the exposure scenarios typically evaluated include:

- ingestion of contaminated drinking water;
- inhalation of volatile contaminants released to indoor air from household use of contaminated water supply;
- ingestion of contaminated soil by children playing in the yard;
- ingestion of contaminants in or on home-grown vegetables;
- dermal exposure from contaminated soil;
- infiltration of volatile compounds from soil or groundwater into indoor air.

The general form of the equation used to estimate the dose resulting from a given exposure scenario is one common to risk assessments among regulatory programs at EPA:

$$AD = \frac{C \cdot IR \cdot F \cdot D \cdot AF}{BW \cdot AT},$$

where,
AD= dose in mg/kg/day
C= concentration in contaminated medium (e.g. mg/liter)
IR= intake rate of contaminated medium (e.g. liters/day of water)
F = frequency of exposure (e.g. days per year)
D= duration of exposure (years)
AF= absorption fraction (unitless)
BW = body weight (kg)
AT = averaging time (may vary for carcinogens, AT= D for noncarcinogens)

The exposure factors in this equation—IR, F, D, AF , BW, AT—are largely dictated by agency policy for the scenarios typically evaluated at sites (Exposure Factors Handbook, Exposure Assessment Guidance, Regional Guidance documents, etc.) although some may be varied on a site specific basis. For example, the frequency with which children play outside and potentially come into contact with contaminated soils may be assumed to differ between sites in northern and southern parts of the country. The concentrations of contaminants at the site are among the few exposure factors that may truly be considered to be site specific. Concentrations of contaminants in each medium are typically represented by arithmetic means and for the reasonable maximum exposure, by the upper 95% confidence limit on the mean (given sufficient data) or by the maximum value detected in the medium.

Given the limited number of samples typically collected at sites and other resource constraints, the mean and the 95% upper confidence limit on the mean of the concentration of contaminants in environmental samples are typically used to represent concentrations at the point of human exposure (particularly for exposures to soil and groundwater). These concentrations are typically assumed to be constant over the duration of exposure. Environmental modeling to predict concentrations of contaminants

over distance and time is not routine but is done for a small subset of sites (large, highly controversial, etc.) The widespread use of monitoring data to represent exposures reflects EPA efforts to minimize the cost of the remedial investigation and assessment and to expedite transition to the cleanup phase of the process.

Sensitivity analysis to these exposure assumptions, if conducted at all, tends to be based on a comparison of the impact of "average" exposure assumptions versus "reasonable maximum exposure" assumptions. For example, the average number of years for which an individual is assumed to live at a site is 9 years as opposed to 30 years for the reasonable maximum exposure assumption. Average concentrations of contaminants in environmental media, rather than 95% upper confidence limit on the average concentration might be used. Probabilistic exposure assessments using input distributions rather than point estimates are beginning to make their way into Agency risk assessments (see, for example EPA, 1994) but are still not widely used.

RISK CHARACTERIZATION

The size of the population exposed comes in as a practical criterion in ranking sites for cleanup consideration, but the nominal decisions about cleanup are influenced (to the degree they are based on risk at all) on individual risk levels. These risks are based on standard scenarios of exposure depending on the anticipated future land use, and on estimates (often upper end estimates) of the concentration of contaminants currently at the site. Exposures are often figured as RMEs, or reasonable maximum exposures, as discussed above.

Policies regarding the level of risk that constitutes a hazard have evolved in the Superfund Program. At the outset of the program, a one in one million (10^{-6}) lifetime risk of cancer was frequently the benchmark against which estimated risks for a site were judged. Under the current NCP and subsequent policy directives, estimated risks at a site are evaluated against a risk range of 10^{-4} to 10^{-6}. The NCP states: "For known or suspected carcinogens, acceptable exposure levels are generally concentration levels that represent an excess upper bound lifetime cancer risk to an individual of between 10^{-4} to 10^{-6} using information on the relationship between dose and response."

A 1991 directive of the Superfund office (OSWER, 1991), specified that remedial action should not be taken at sites where risks were less than 10^{-4} without adequate justification. Nonetheless, when setting remedial goals at a site the NCP specifies that: "The 10^{-6} risk level shall be used as the point of departure for determining remediation goals for alternatives when ARARs are not available or are not sufficiently protective because of the presence of multiple contaminants at a site or multiple pathways of exposure."

The NCP does not address the definition of "protective" in the context of exposure to non-carcinogens. In practice, however, exposures to contaminants resulting

in hazard quotients or hazard indices exceeding 1 are considered to carry an increased potential for adverse noncancer health impacts.

An important and unique feature of Superfund risk assessments is the consideration of exposure to many chemicals simultaneously. This is attributable to the need of risk assessment to evaluate waste sites as a whole (with all of their consituents) as health hazards, and not just risks from particular chemicals. Superfund does not consider the possible exposure of some people to multiple hazardous waste sites, however. There are ongoing efforts within Superfund to develop approaches for considering exposure to multiple sites, stemming from the Environmental Justice initiatives at the agency.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY.

CONCLUSIONS

It has been noted several times in the above report that risk assessment is a practical discipline. The need for risk assessment emerges because agents that may appear in the environment vary widely in the effects they have on living organisms, affecting different target organs in different ways in different circumstances. Compounds also vary extraordinarily widely in their potencies to cause these effects, with some agents capable of poisoning human beings at tiny, single doses, while others may have eventual effects if experienced for a lifetime at high level, and still others are quite benign even at these high exposures. Any program of action that seeks to protect people and the environment from chemically induced health effects must grapple with this diversity. If actions are to be effective and if efforts are to be sensibly deployed against a sea of agents of potential concern, then tools must be developed to make the differentiations of hazard and potency among agents that allow insight into the nature and magnitude of risks posed by different exposures.

The biological phenomena underlying chemical toxicity are complex and usually poorly understood. In many case, such as the potential for agents to cause effects at low doses, the phenomena of practical interest are not directly amenable to experimental elucidation, no matter how cleverly or carefully experiments are conducted, and conclusions about potential effects in these circumstances must be reached by analogy and by generalizing and extrapolating from observable circumstances to unobservable ones.

To a large degree, the body of environmental laws that seek to establish practices that will ensure safety (or at least mitigate risk) of chemical exposures were established before risk assessment was a well recognized and codified discipline. Most of the methodology of risk assessment has been invented in reaction to the calls by these laws to define limits on exposure that will "protect the public health" or lead to "a reasonable certainty of no harm." That is, in passing the laws, Congress called on the regulatory agencies to develop means to assess risks so as to define exposure levels that would achieve the stated qualitative goals of health protection. The presumption in this approach is that there will be relatively few such exposures in need of control and that controls that are clearly sufficient to achieve protection can be had at reasonable cost to those responsible and to society as a whole.

The history of the development of quantitative cancer risk assessment at FDA, recounted briefly in the section on that agency, is instructive. Conservative low-dose extrapolation was invented to determine whether minute exposures to a few compounds could conceivable lead to risks of any meaningful magnitude. The levels of exposure were not set according to risk assessment calculations—they were fixed by the intended uses of the agents. In this circumstance, a "one-way" conclusion was practically useful—at *worst*, risks were no bigger than some trivially small amount. How much less they

might be was not of practical concern, and the methodology developed for the particular question, despite the fact that it did not identify the "true" low-dose risk, was completely satisfactory from the point of view of the questions being asked by risk managers.

Once the precedent of calculating risk levels from different exposure levels was set, however, there was temptation to use the same methods for calculating how much one could *elevate* the exposure level without (unduly) affecting the risk level. That is, rather than just applying the method to demonstrate the practical safety of very low exposures, it came to be applied to defining ranges of "safe" exposures. Later, the methods began to be applied to estimating the degree of risks *above* the levels that might be deemed trivial, raising the issue of how much risk is "acceptable." Still later, these degrees of risk engendered by higher exposures were evaluated in comparison to the costs of control and the benefits derived from allowing higher levels in the environment. At present, risk assessment is asked to make much more precise and certain statements about the levels of risk at different exposures than it was ever intended to make and than it can reasonably be expected to produce. This has led to great criticism and controversy about the ability of risk assessment to make pronouncements on the risks it pretends to estimate. In sum, a tool that was developed as a practical, reasonable, and sufficient means to answer a particular practical question has struggled to keep up to the increasing demands put upon it. We now see the world of risk not as one of a few identifiable, easily controlled agents, but as a complex nexus of tradeoffs where every action (including action to avoid certain risks) engenders other risks.

The present report has attempted to examine the major environmental laws for their mandates on risk and for their calls for risk assessment to address these mandates. Since the laws largely precede risk assessment methodology, there is little call for specific analytical actions on the part of regulatory agencies. Nonetheless, the need for risk assessment is implicit in every call to define levels of exposure in regard to the potential health effects they may cause.

The different risk mandates are all rather vaguely worded, and it is not possible to discern calls for different methods of risk estimation from a mandate to assure "reasonable certainty of no harm" and one to "protect the public health with an adequate margin of safety." The chief difference among mandates is whether they call for balancing costs and benefits or whether they account for feasibility of controls, issues that affect the uses to which assessed risks are to be put in regulation but that do not affect the conduct of risk estimation itself. Only in the Consumer Product Safety Act are the criteria for balancing risks and benefits, and the particular findings in this regard that must be made to justify regulation, explicitly spelled out.

The environmental laws do not allow the regulatory agencies any action to control risks—they specify the nature of the regulatory actions to be undertaken, whether these be the issuance of permits or registrations, the definition of acceptable ambient concentrations, the limitations of discharges, and so on. The nature of the regulatory actions required vary more among laws than do the risk mandates, and the regulatory powers under each law are tailored to the nature of the regulated enterprise or activity,

hinging largely on practical questions regarding where regulatory control can be effectively administered to accomplish the ends and purposes intended.

From the point of view of risk assessment, this variation in regulatory powers tends to manifests itself in different exposure assessment methods. Consequently, there is more variation among regulatory agencies and programs in exposure assessment methods procedures than in assessment of toxic effects. In this report, an attempt has been made to relate the methods used in risk assessment (and in particular, exposure assessment) to the nature of the law's regulatory activities. Given these differences in the regulatory powers granted by the various laws, it is unreasonable to expect exposure and risk assessments to be equally realistic across regulatory groups. By their nature, laws acting through permits will define exposures above those usually seen in compliance since they regulate by specifying maxima; laws acting through ambient concentration standards that represent ambitions to control pollution will define exposures below those typically seen, since they regulate by specifying goals to be striven for; and laws acting through specification of difficult to achieve technical controls will define exposures (or at least emissions) close to that actually achieved, since they act by imposing uniformity in control.

When the express aim of a law is to *manage* risks to the population, the exposure assessment should attempt to characterize the full distribution of exposure levels in the population as accurately as possible, so that the distribution of risks can be examined (and changes or shifts in the burden of risk under different regulatory options noted). In this circumstance, it is important to attend not only the existence of high individual risks, but also to the total burden of risk on the population. Many current environmental laws, however, are written so as to require *protection* from risk. Permits are issued, standards are set, conditions of use are defined, or cleanups are mandated so as to set limits on exposure such that few if any of the population of concern will experience risk levels that are "unacceptable." In this setting, the focus is on setting regulations to protect those at the high end of the risk distribution. This focuses the attention of the assessment on defining the upper end of the range of exposure scenarios for which it is intended to furnish protection. Depending on the law, this may be the top end of the actual distribution of exposures near a source (as in the Clean Air Act §112), a person of somewhat above average consumption of a medium contaminated up to a limit deemed permissible (as in the Safe Drinking Water Act), or an especially frequent consumer of a foodstuff containing an additive (as in the Federal Food, Drug, and Cosmetic Act). Whether the protected exposure is actual or hypothetical (and whether a hypothetical exposure is high or low compared to the upper end of actual exposures) may have less to do with data availability or willingness to use different exposure estimation techniques than with the intent of the law. A key factor is which parts of the exposure equation are under regulatory control and which are not. For instance, in setting pesticide tolerances, the assumption is made that all foods on which the agent is permitted in fact bear it, and at the maximally permissible level, when conducting initial exposure assessments. This is done not simple to be "conservative," but because the law requires setting levels that will be safe for consumers of the foods, and this must include protection of someone who chooses to eat all the foods containing the agent, even though few people may actually do

so. Moreover, since permitting residues up to the tolerance level implies that such all such levels are acceptably safe, the tolerances have to be set such that they would be safe *if* they occur, irrespective of whether they in fact occur.

In other words, much of the attention to estimates of risk that are conservative in the face of uncertainty about potency and much of the focus on the upper end of exposures arise because these methods were invented to implement the calls from the statutes for defining regulatory actions that would ensure safety. As notions of effective risk management evolve, it is becoming clear that such methods are less well suited for estimating the actual burden of exposure and risk in populations. The discussions of each statute and regulatory program in this report attempts to examine how the methods that have evolved in each program reflect the tasks set for regulators, either explicitly or implicitly, by the various statutes as they set mandates about what is to be accomplished and by what regulatory actions.

The inconsistency of methods for dose-response assessment cannot be so easily explained in terms of response to different regulatory needs. The variety of methods seems to reflect the somewhat separate history of development of potency estimation in the different groups and the lack of a definitive scientific basis to guide these independent evolutions along exactly the same path. The variety of methods correctly reflects the uncertainty about the best or most appropriate procedures, but it results in the awkward result that different agencies can arrive at different characterizations of an agent's carcinogenic potency from the same set of data, based only on differences in preferred methods and precedents from earlier analyses. It would seem that harmonization of these methods to the extent achievable would be beneficial. At the same time, harmonization achieved through rigidity in rules for choice of methods would falsely imply that the mandated set of approaches is more correct than others and would stultify application of case-by-case judgment.

As with exposure assessment, the focus of much potency analysis is on defining levels of exposure that can be more or less assured of posing "acceptable" risk. The methods that are used in the face of uncertainty can usually be understood in this light. As the questions being asked by the risk management process move beyond such issues of assurance of safety, existing methodology and practices established in response to current environmental statutes become less appropriate.

Fundamentally, risk assessment methods are practical inventions put in place to address the kinds of questions asked of regulatory analysis by the mandates of the environmental laws. These laws and their mandates can be changed, and the methods for assessing risks will have to change with them, to respond to new needs.

REFERENCES

Anderson, E., and the Carcinogen Assessment Group of the U.S. Environmental Protection Agency (1983) Quantitative approaches in use to assess cancer risk. *Risk Anal*. 3:277-295.

Burmaster,, D.E. and R.H. Harris (1993) The magnitude of compounding conservatisms in Superfund risk assessments. *Risk Anal*. 13:131-134.

Crump, K.S. (1981) An improved procedure for low-dose carcinogenic risk assessment from animal data. *J. Environ. Pathol. Toxicol.* 5:339-348.

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

Crump, K.S., D.G. Hoel, C.H. Langley, and R. Peto (1976) Fundamental carcinogenic processes and their implications for low-dose risk assessment. *Cancer Res*. 36:2973-2979.

Environmental Law Reporter (1986) An annotated legislative history of the Superfund Amendments and Reauthorization Act of 1986 (SARA). *Environ. Law Reporter* 16:10360-10419.

EPA (1987) [U.S. Environmental Protection Agency] *Technical Analysis of New Methods and Data Regarding Dichloromethane Hazard Assessment*. Washington DC: Office of Health and Environmental Assessment, U.S. Environmental Protection Agency. EPA/600/8-87/029A.

EPA (1988) [U.S. Environmental Protection Agency] *Thyroid Follicular Cell Carcinogenesis: Mechanistic and Science Policy Considerations*. Washington DC: Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/625/3-88/014a.

EPA (1989a) [U.S. Environmental Protection Agency]*Workshop Report on EPA Guidelines for Carcinogen Risk Assessment*. Washington DC: Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/625/3-89/015.

EPA (1989b) [U.S. Environmental Protection Agency]*Workshop Report on EPA Guidelines for Carcinogen Risk Assessment: Use of Human Evidence*. Washington DC: Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/625/3-90/017.

EPA (1989c) [U.S. Environmental Protection Agency] *Exposure Factors Handbook*. Washington DC: Office of Health and Environmental Assessment, U.S. Environmental Protection Agency. EPA/600/8-89/043.

EPA (1989d) [U.S. Environmental Protection Agency] *Risk Assessment Guidance for Superfund. Vol. 1, Human Health Evaluation Manual (Part A), Interim Final*. Washington DC: Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency.

EPA (1989d) [U.S. Environmental Protection Agency] *Interim Methods for Development of Inhalation Reference Concentrations*. Research Triangle Park NC: Environmental Criteria and Assessment Office, U.S. Environmental Protection Agency. EPA/540/1-89/002.

EPA (1991) [U.S. Environmental Protection Agency] *Alpha-2u-globulin: Association with Chemically Induced Renal Toxicity and Neoplasia in the Male Rat*. Washington DC: Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/625/3-91/019F.

EPA (1992a) [U.S. Environmental Protection Agency] *Working Paper for Considering Draft Revisions to the U.S. EPA Guidelines for Cancer Risk: Review Draft*. Washington DC: Office of Research and Development, U.S. Environmental Protection Agency. EPA/600/AP-92/003.

EPA (1992b) [U.S. Environmental Protection Agency] *Guidelines for Exposure Assessment*. Washington DC: Office of Research and Development, U.S. Environmental Protection Agency.

EPA (1992c) [U.S. Environmental Protection Agency] *Guidance for Data Usability in Risk Assessment (Part B), Final*. Washington DC: Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency. Publication 9285.7-09B.

EPA (1994) [U.S. Environmental Protection Agency] *Soil Screening Guidance*. Washington DC: Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency. Publication 9355.4-14FS.

EPA (1995) [U.S. Environmental Protection Agency] *Use of the Benchmark Dose Approach in Health Risk Assessment*. Washington DC: Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/630/R-94/007.

Findley, R.W. and D.A. Farber (1992) *Environmental Law in a Nutshell*. St. Paul MN: West Publishing Co.

Finkel, A.M. (1989) Is risk assessment really too conservative? Revising the revisionists. *Columbia J. Environ. Law* 14:427-467.

Gaylor, D.W. and R.L. Kodell (1980) Linear interpolation algorithm for low-dose risk assessment of toxic substances. *J. Environ. Path. Toxicol*. 4:305-312.

Graham, J.D. (1993) The legacy of one in a million. *Risk in Perspective* 1(1):1-2. [Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, MA]

Graham, J.D., L.C. Green, and M.J. Roberts (1988) *In Search of Safety*. Cambridge MA: Harvard University Press.

Hattis, D. and W.S. Minkowitz (1995) Risk evaluation: Legal requirements, conceptual foundations, and practical experiences in the United States. In: *Risk Evaluation - A Survey of International Experiences*. O. Renn, ed. Amsterdam: Kluwer Academic Publishers.

Hazardous Waste Cleanup Project (1993) *Exaggerating Risk: How EPA's Risk Assessments Distort the Facts at Superfund Sites throughout the United States*. Washington DC: Morgan, Lewrs, and Bockius.

Hutt, P.B. (1985) Use of quantitative risk assessment in regulatory decisionmaking under Federal health and safety statutes. In: Risk *Quantification and Regulatory Policy*. D. Hoel, R. Merrill, and F. Perera (eds.) New York: Cold Spring Harbor Laboratory.

IARC (1987) [International Agency for Research on Cancer] *Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, Supplement 7:17-34. Lyon, France.

Lehman, A.J. and O.G. Fitzhugh (1954) 100-fold margin of safety. *Assoc. Food Drug Off. U.S. Q. Bull*. 18:33-35.

Mantel, N. and W.R. Bryan (1961) "Safety" testing of carcinogenic agents. *JNCI* 27(2):455-470.

NRC (1983) [National Research Council] *Risk Assessment in the Federal Government: Managing the Process*. Washington DC: National Academy Press.

NRC (1987) [National Research Council] *Pesticides in Foods: The Delaney Paradox*. Washington DC: National Academy Press.

NRC (1993a) [National Research Council] *Pesticides in the Diets of Infants and Children*. Washington DC: National Academy Press.

NRC (1993b) [National Research Council] *Issues in Risk Assessment*. Washington DC: National Academy Press.

NRC (1994) [National Research Council] *Science and Judgment in Risk Assessment*. Washington DC: National Academy Press.

Olin, S., W. Farland, C. Park, L. Rhomberg, R. Scheuplein, T. Starr, and J. Wilson (eds.) *Low-Dose Extrapolation of Cancer Risks: Issues and Perspectives*. Washington DC: ILSI Press.

OSWER (1991) [Office of Solid Waste and Emergency Response] Role of Baseline Risk Assessment in Superfund Remedy Selection Decisions. OSWER Directive No. 9355.0-30, April 22, 1991. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency.

OSWER (1995) [Office of Solid Waste and Emergency Response] Land use in the CERCLA Remedy Selection Process. OSWER Directive No. 9355.7-04, May 25, 1995. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency.

Padgett, J. and H. Richmond (1983) The process of establishing and revising national ambient air quality standards. *J. Air Pollut. Control Assoc*. 33:13-16.

Rees, D.C, and D. Hattis (1994) Developing quantitative strategies for animal to human extrapolation. In: *Principles and Methods of Toxicology, 3rd Edition*. A.W. Hayes (ed.) New York: Raven Press.

Rosenthal, A., G.M. Gray, and J.D. Graham (1992) Legislating acceptable cancer risk from exposure to toxic chemicals. *Ecology Law Quarterly* 19(2):269-362.

Sadowitz, M. and J.D. Graham (1995) A survey of residual cancer risks permitted by health, safety and environmental policy. *Risk: Health, Safety & Environment* 17:17-35.

Schierow, L.-J. (1994) *Risk Analysis and Cost-Benefit Analysis of Environmental Regulations*. CRS Report to Congress No. 94-961 ENR. Washington DC: Congressional Research Service, Library of Congress.

TABLES

Table 1: Organizational abbreviations; nesting indicates organizational hierarchy.

EPA	Environmental Protection Agency
OA	Office of the Administrator (of EPA)
OPPTS	Office of Prevention, Pesticides, and Toxic Substances
OPPT	Office of Pollution Prevention and Toxics ["Toxics"]
OPP	Office of Pesticide Programs ["Pesticides"]
OAR	Office of Air and Radiation
OAQPS	Office of Air Quality Planning and Standards
OMS	Office of Mobile Sources
OW	Office of Water
OSWER	Office of Solid Waste and Emergency Response
OERR	Office of Emergency and Remedial Response ["Superfund"]
OSW	Office of Solid Waste ["RCRA"]
ORD	Office of Research and Development
NCEA	National Center for Environmental Assessment
FDA	Food and Drug Administration
CFSAN	Center for Food Safety and Applied Nutrition
OSHA	Occupational Safety and Health Administration
CPSC	Consumer Product Safety Commission

Table 2: Environmental regulatory statutes addressed in this report.

Abbreviation/ Citation	Statute Title	Responsible Federal Office
CAA 42 U.S.C.A. §§ 7401 to 7671q	Clean Air Act	EPA, Office of Air and Radiation (OAR)
CWA 33 U.S.C.A. §§1251 to 1387	Clean Water Act (Federal Water Pollution Control Act)	EPA, Office of Water (OW)
SDWA 42 U.S.C.A. §§300f to 300j-26	Safe Drinking Water Act (Public Health Service Act)	EPA, Office of Water (OW)
RCRA 42 U.S.C.A. §§ 6910 to 6992k	Resource Conservation and Recovery Act (amending Solid Waste Disposal Act)	EPA, Office of Solid Waste and Emergency Response (OSWER), Office of Solid Waste (OSW)
CERCLA 42 U.S.C.A. §§ 9601 to 9675	Comprehensive Environmental Response, Compensation, and Liability Act	EPA, Office of Solid Waste and Emergency Response (OSWER), Office of Emergency and Remedial Response (OERR) ["Superfund"]
TSCA 15 U.S.C.A. §§2601 to 2692	Toxic Substances Control Act	EPA, Office of Prevention, Pesticides, and Toxic Substances (OPPTS), Office of Pollution Prevention and Toxics (OPPT)
FIFRA 7 U.S.C.A. §§ 136 to 136y	Federal Insecticide, Fungicide, and Rodenticide Act	EPA, Office of Prevention, Pesticides, and Toxic Substances (OPPTS), Office of Pesticide Programs (OPP)
FFDCA 21 U.S.C. §§ 321 to 394	Federal Food, Drug, and Cosmetic Act	Food and Drug Administration (FDA), Center for Food Safety and Applied Nutrition (CFSAN); and EPA, Office of Pesticide Programs
OSHAct 29 U.S.C.A. §§ 650 to 683	Occupational Safety and Health Act	Department of Labor (DOL), Occupational Safety and Health Administration (OSHA)
CPSA 15 U.S.C. §§ 2051n to 2084	Consumer Product Safety Act	Consumer Product Safety Commission (CPSC)
FHSA 15 U.S.C. §§ 1260 to 1278	Federal Health and Safety Act	Consumer Product Safety Commission (CPSC)
APA 5 U.S.C.A. §§ 551 to 559	Administrative Procedures Act	

Table 3: Federal regulatory agency personnel interviewed for this report.

Michael Bolger		FDA-CFSAN
Robert Cantilli	*	OW
James Cogliano		ORD
Nancy Crane		FDA-CFSAN
Pennelope Fenner-Crisp	*	OPP
Adam Finkel	*	OSHA
Richard Hill		OPPTS
Peter Infante	*	OSHA
James Kariya	*	OPP
Paul Kuznesof		FDA-CFSAN
Ronald Lorentzen	*	FDA- CFSAN
Elizabeth Margosches		OPPT
Carl Mazza		OAR
Alec McBride	*	OSWER
Bruce Means	*	OSWER
Bruce Mintz	*	OW
Jennifer Orme-Zavaleta	*	OW
George Pauli		FDA-CFSAN
Dorothy Patton		OA-Science Policy Council
William Perry	*	OSHA
Harvey Richmond	*	OAR-OAQPS
Kelly Rimer	*	OAR-OAQPS
Vanessa Vu	*	OPPT

* Interviews marked with an asterisk represent full interviews on all of the questions tabulated in Appendix A. Other interviews were on specific or selected topics.

Table 4: Summary overview of Federal regulation of potentially toxic chemicals, including risk mandates, key statutory language, and principal differences in risk assessment methodology among Federal regulatory programs, as detailed in the text.

Program Office	Statute/ Activity	Risk Mandate	Role of Carc Class.	Special Quant Methods	Individual Risks Considered	Population Risk Considered	Special Groups	Usual Acceptable Residual Risk	Practical Regul. Trigger or Criterion
OPPTS-OPPT "Toxics"	TSCA	avoid and mitigate "unreasonable risk" via risk-benefit balancing	no	"additional" cancer risk above background	yes, "reasonable worst case" for occup expos	yes, indirectly	workers, consumers, genl popn	unstated, but usually 10^{-5} to 10^{-6} for non-occup-ational, 10^{-4} to 10^{-5} for occup	
OPPTS-OPP "Pesticides"	FIFRA (registr.; use limits)	balance risks, benefits, social & economic costs; efficacious yet w/o "unreasonable risk to man or environment"	no QRA for *some* "C's"		yes, broadly, assume max permissible residues, but average food consumptions	yes		unstated, but usually 10^{-5} to 10^{-6} for non-occup-ational, 10^{-4} to 10^{-5} for occup	interplay of efficacy and tolerances for residues; registrant proposes use limits
	FFDCA (residue tolerances)	"Delaney Clause," no additives that are animal carcin.; "reasonable certainty of no harm" for residues	any pos cancer assay triggers Delaney		no for carcinogenic additives; yes for residue tolerances	yes for residue tolerances	demogr. sub-population diets considered	zero for additives; 10^{-6} for assumed max residues in average diet, 10^{-6} for non-dietary exposure	Delaney prohibition of carcinogenic additives
OW	SDWA (drinking water)	for carcinogens, unenforceable max contam limit goals (MCLG) of zero, but enforceable limits (MCL) set by technology if within adequate margin of safety	yes, "C's" may be treated as threshold	extra UF on NOAEL for "C's"	a standard exposure scenario in middle range	no	no	10^{-4} to 10^{-6} is range considered to be adequate	MCLG's primarily based on technical, cost feasibility if risk range hit.
	CWA (waterway water qual)	protect public health and welfare with non-enforceable, health-based water quality criteria and enforceable "best" technology based effluent standards	no	conserv. water transport models determine acceptable daily loading of water bodies	a standard exposure scenario in middle range	no	no	10^{-5} to 10^{-7}	standards set by states with EPA guidance; some consideration of residual risk after best avail tech effluent limits

Program Office	Statute/ Activity	Risk Mandate	Role of Carc Class.	Special Quant Methods	Individual Risks Considered	Population Risk Considered	Special Groups	Usual Acceptable Residual Risk	Practical Regul. Trigger or Criterion
OSWER	RCRA (haz waste handling, active disposal)	aim at "cradle-to-grave" stewardship; technology- and process-based, but also risk-triggered corrective action, to be protective of human health and the environment, excluding costs	in some haz waste ID criteria; C's may be treated specially	uses OW MCL's or its own QRA to list or delist as a haz waste	yes, a rather conservative estimate of hypothetical transport and exposure near a problem site, but uses some Monte Carlo modeling	no	hypothetical populations around haz waste facilities	listing: 10^{-5} corrective action: 10^{-4} to 10^{-6} incinerators: 10^{-5}	cleanup strategy chosen with site-use, feasibility considerations as long as within risk range of 10^{-4} to 10^{-6}
	CERCLA Superfund, abandoned and active haz waste site monitoring and cleanup	applicable other laws plus cleanup to be protective of human health and environment; risk-based but consider feasibility	no	consider cumulative risk of mixtures (but not exposure to multiple sites)	"reasonable maximum exposure" using mix of midrange and conservative assumptions	high population around site prompts listing on NPL	hypothetical populations around site, scenarios for special groups (real or hypothetical)	10^{-4} to 10^{-6}, depending partly on anticipated future use of site	site-specific "ranking" QRA for listing, prioritization of site; then more detailed risk assessment to choose actions reaching target risk range of 10^{-4} to 10^{-6}
OAR	CAA Criteria pollutants	adequate margin of safety to protect public health	non-cancer only	extensive data, including on humans	yes	yes			without harmful effects on most people
	CAA Hazardous Air Pollutants	Must apply Max Avail Control Technology; If residual risk to MEI $>10^{-6}$, further regulate to provide adequate margin of safety to protect public health, considering costs	no	Maximally Exposed Individual for each source can trigger residual risk provision	Only after MACT; MEI $>10^{-6}$ triggers further action; MEI $<10^{-6}$ before controls yields de minimis exemption	presumably yes, when assessing residual risk	populations around sources	$<10^{-6}$??	apply best controls as default, then consider further regulation if needed

Program Office	Statute/ Activity	Risk Mandate	Role of Carc Class.	Special Quant Methods	Individual Risks Considered	Population Risk Considered	Special Groups	Usual Acceptable Residual Risk	Practical Regul. Trigger or Criterion
FDA	FFDCA (food additives, colors & contam-inants; cosmetics)	“Delaney Clause,” no additives that are animal carcin.; "reasonable certainty of no harm" for residues, no cost considerations	any pos cancer assay triggers Delaney	"modified" Gaylor-Kodell procedure for carcinogens, body weight dose scaling	no for carcinogenic additives; yes for additives, contaminants	no	demogr. sub-population diets considered	zero for additives; 10^{-6} for assumed max residues in "high use" diet	Delaney prohibition of carcinogenic additives
OSHA	OSHAct (occup. exposures)	"no employee will suffer material impairment of health," considering feasibility of stds	no, frequent use of human data	MLE of multistage model, body weight dose scaling	yes, for full working life at permissible exposure limit	no	no	feasible controls	"significant" risk (in practice, 10^{-3})
CPSC	CPSA FHSA (consumer products)	"to protect...against unreasonable risk of injury" with "reasonably necessary" standards, considering cost/benefit	scheme similar to EPA's, focus on agents with "sufficient evidence"	MLE if linear, surface area dose scaling, combine tumor types	not explicitly	yes, in context of cost-benefit analysis	impact of regulation (not risk) on elderly, handicapped	unclear	"reasonably necessary," least burdensome standards with benefits "bearing a reasonable relationship" to costs

Table 5: Summary overview of regulatory programs of EPA's Office of Air and Radiation (OAR), including risk mandates, key statutory language, and principal differences in risk assessment methodology among OAR programs. This table was produced by the EPA Office of Air and Radiation, and was reproduced in the printed document with permission. A reproducible table has not yet been made available for electronic publication.

APPENDIX A

Interviews for
Commission on Risk Assessment and Management
Project on Risk Assessment Differences Among EPA Programs

I. What does the *enabling legislation* for your program mandate about the assessment of health risks? How do you see this mandate as *differing* from those in other environmental laws?

A. Primarily health-based, technology-based, or cost-benefit balancing?

B. Individual risks or population risks? Consideration of special populations?

C. How specific are the mandates? Do they mention risk assessment or its methods explicitly? Do they define "safe" levels of exposure?

D. Any distinctions among *parts* of the law?

E. Any insights into the *intent* of the law from legislative history, events preceding passage, etc.?

II. Through what process were specific risk assessment *practices* in your program defined and implemented? Are policies and practices formally codified? (Where?) How are these practices seen as fulfilling the legislative mandate?

A. To what extent do such policies reflect:

1. specific, explicitly promulgated EPA-wide policies? (on "acceptable risk," comparing costs and benefits, conservative estimation, and so on)
2. program-specific development of policy? (which may resemble other program's policies as a result of being designed to meet common issues)
3. "borrowings" from pre-existing methods for sake of consistency, weight of precedent, or reliance on established methods?

B. Have there been notable *judicial challenges* to the risk assessment methods? Were changes made as a result?

C. To what extent does your program's risk assessment practice rely on centralized or EPA-wide analyses (e.g., ORD documents, IRIS)? Do program-specific methods address:

1. Hazard identification (including listing, delisting)?
2. Potency and RfDs?
3. Exposure?

D. Are there any differences in risk assessment methods used for:

1. Setting regulatory levels?
2. Listing/Delisting or *de minimus* determinations?
3. Screening?
4. Priority setting?

III. Broadly speaking, what are the principle kinds of *regulatory options* available in your program? That is, what regulatory "control points" are available with which to achieve your legislation's aims? (e.g., granting or denying emissions permits, mandating restrictions on a compound's use, specifying technical abatement or control methods, etc.)

A. How are the risk consequences of such regulatory options projected? (e.g., fate of emissions are modeled, and exposure to a hypothetical, conventionally defined maximally exposed individual is projected; national average consumer exposures are estimated for a given use-restriction option.)

B. Is there an identifiable *practical criterion* for acceptable or unacceptable options? (e.g., maximally exposed individual's risk not to exceed a cut-off; population risk to consumers not to exceed a certain number of projected annual cancer cases.)

C. What is the source of policies on these issues?

IV. Are there any special considerations for *hazard identification* in your program?

A. Are "hazards" defined in your enabling legislation? Is there a legislated list of toxic compounds? A mandate to maintain a list?

B. Where are criteria for inclusion on such lists specified? Are they primarily qualitative (classification) or quantitative (potency)? (Are only IRIS listings used?)

C. Are there special, program-specific considerations regarding route of exposure? (e.g., consider only inhalation toxicity data for identifying hazardous air pollutants.)

D. Is there any special role for epidemiology vis-a-vis use of animal data in your program?

V. Are there any special considerations for *dose-response analysis* in your program?

A. Choice of data sets? (Potential issues: route-of-administration; benign tumors; background rates; relevance-to-humans)

B. Combining multiple tumor responses.

C. Corrections for intercurrent mortality.

D. Dose-response model;
1. Number of stages in LMS;
2. Additional or Extra Risk;
3. Best fits or upper bounds.

E. Equivalent dose units across species.

F. Presenting/combining results from different data sets.

G. Use of epidemiological data.

H. For non-cancer assessment, are there any special, program-specific policies about uncertainty factors? About the definition of "adverse" outcomes?

VI. For several reasons, *exposure assessment* is the area of most difference among EPA regulatory programs. Not the least of these is that the various programs are primarily differentiated by the sources of exposure to be controlled. But beyond this, there are differences among programs in *whose* exposure is at issue (the most exposed individuals, individuals with average exposure, the total numbers of people exposed, exposure of special subgroups, such as children, etc.) and in *how* that exposure is estimated (measurement, modeling, standardized scenarios, etc.).

A. Does your enabling legislation specify whose exposure is to be considered in determining safety? Are there such specifications in your program's policy documents?

B. Broadly speaking, in your regulatory program, whose exposure is key to determining regulatory actions? (e.g., a hypothetical person making life-long "average" use of water just downstream from the emissions source of concern; the most-exposed actual individual near a particular facility during its projected operational lifetime; the distribution of actual exposures among all people living within a certain distance of a contaminated site.)

C. How is the choice of such key exposures seen as reflecting the risk mandate of the regulatory program?

D. How would the assessment of these exposures be characterized along the following dimensions:
1. Hypothetical vs. Actual exposures;

2. National or overall patterns vs. Local or site-specific exposures;
3. Individual exposures vs. Characterization of the whole exposed population;
4. High end exposures vs. Typical, "average," or middling exposures;
5. Conservative estimates and assumptions vs. "Best" estimates *(NB, point 4. Is about exposure* variability *among individuals, while 5. is about* uncertainty *of the estimate--one can make a conservative estimate of an average persons exposure or a "best" estimate of a highly exposed person's exposure*);
6. Present (and past) exposures vs possible or likely future exposures;
7. Averaged over time and conditions vs. Varying with time and changing conditions, including peak exposures;
8. General population vs. Particular demographic or sensitive segments of the population.

E. For the following elements of exposure assessment, is the usual basis (a) national measurements, (b) local or site-specific measurements, (c) modeling, or (d) hypothetical standard scenarios? To what extent are estimates regarded as conservative?

1. Estimation of sources of emissions, uses of chemicals;
2. Fate, persistence, and transport of agents through the environment;
3. Patterns of exposure of humans, including numbers exposed, concentrations, frequencies, and durations;
4. Uptakes and doses, given exposure.

VII. Are there any special considerations for *risk characterization* in your program? Are they mandated in legislation or policy documents?

A. How does the scientific strength (or weakness) of the analysis figure in the risk analysis of regulatory options?

B. Are individual risks or population risks (number of cases) stressed in the risk analysis of regulatory options?

C. How is the perceived conservatism of quantitative risk estimates factored in when risks are compared to other factors?

ACKNOWLEDGMENTS

The author offers special thanks to all of the people listed in Table 3 who graciously gave their time and expertise to aid in developing information for this project. In addition, particular thanks are due to Katherine Walker, who drafted the section on Superfund, and Eric Ruder, who provided valuable comments on that chapter.